The evolutionary and functional characterisation of the ecdysteroid kinase-like (EcKL) gene family in insects

Jack Louis Scanlan
ORCID: 0000-0001-9163-2203

Submitted in total fulfilment of the requirements of
the degree of Doctor of Philosophy

August 2020

School of BioSciences
Faculty of Science
The University of Melbourne

Ask our esteemed panel,

'Why are we alive?'

And here's how they replied,

'You're what happens when two substances collide,
and by all accounts, you really should have died.'

Andrew Bird
excerpt from 'A Nervous Tic Motion of the Head to the Left'
in *Andrew Bird & the Mysterious Production of Eggs* (2005)

Abstract

Many thousands of gene families across the tree of life still lack robust functional characterisation, and thousands more may be under-characterised, with additional unknown functions not represented in official annotations. Here, I aim to characterise the evolution and functions of the poorly characterised ecdysteroid kinase-like (EcKL) gene family, which has a peculiar taxonomic distribution and is largely known for containing an ecdysteroid 22-kinase gene in the silkworm, *Bombyx mori*. I hypothesised that EcKLs may also be responsible for insect-specific 'detoxification-by-phosphorylation', as well as ecdysteroid hormone metabolism.

My first approach was to explore the evolution of the EcKLs in the genus *Drosophila* (Diptera: Drosophilidae), which contains the well-studied model insect *Drosophila melanogaster*. *Drosophila* EcKLs have evolutionary and transcriptional similarities to the cytochrome P450s, a classical detoxification family, and an integrative 'detoxification score', benchmarked against the known functions of P450 genes, predicted nearly half of *D. melanogaster* EcKLs are candidate detoxification genes. A targeted PheWAS approach in *D. melanogaster* also identified novel toxic stress phenotypes associated with genomic and transcriptomic variation in EcKL and P450 genes. These results suggest many *Drosophila* EcKLs function in detoxification, or at least have key functions in the metabolism of xenobiotics, and additionally identify a number of novel P450 detoxification candidate genes in *D. melanogaster*.

I then broadened the phylogenomic analysis of EcKLs to a manually annotated dataset containing an additional 128 insect genomes and three other arthropod genomes, as well as a number of transcriptome assemblies. Phylogenetic inference suggested insect EcKLs can be grouped into 13 subfamilies that are differentially conserved between insect lineages, and order-specific analyses for Diptera, Lepidoptera and Hymenoptera revealed both highly conserved and highly variable EcKL clades within these taxa. Using phylogenetic comparative methods, EcKL gene family size was found to vary with detoxification-related traits, such as the sizes of classical detoxification gene families, insect diet, and two estimations of 'detoxification breadth' (DB), one qualitative and one quantitative. Additionally, the rate of EcKL duplication was found to be low in lineages with small DB—bees and tsetse flies. These results suggest the EcKL gene family functions in detoxification across insects.

Building on my previous 'detoxification score' analysis, I used the powerful genetic toolkit in *D. melanogaster* and developmental toxicology assays to test the hypothesis that EcKL genes in the highly dynamic Dro5 clade are involved in the detoxification of selected plant and fungal toxins. Knockout or misexpression of Dro5 genes, particularly *CG13659* (Dro5-7), modulated susceptibility to the methylxanthine alkaloid caffeine, and Dro5 knockout also increased susceptibility to kojic acid, a fungal secondary metabolite. These results validate my evolutionary and integrative analyses, and provide the first experimental evidence for the involvement of EcKLs in detoxification processes.

Finally, I aimed to find genes encoding ecdysteroid kinases in *D. melanogaster*, focusing on *Wallflower* (*Wall*/*CG13813*) and *Pinkman* (*pkm*/*CG1561*), orthologs of a known ecdysteroid 22-kinase gene. *Wall* and *pkm* null mutant animals developed normally, but misexpression of *Wall* caused tissue-specific developmental defects, albeit not those consistent with inactivation of the main ecdysteroid hormones, ecdysone and 20-hydroxyecdysone. In addition, my hypothesis that *Wall* encodes an ecdysteroid 26-kinase was not supported by hypostasis experiments with a loss-of-function allele of the ecdysteroid 26-hydroxylase/carboxylase gene *Cyp18a1*. Combined with existing expression and regulatory data, these results suggest *Wall* encodes an ecdysteroid kinase with an unknown substrate, and hint at previously unknown complexity in ecdysteroid signalling and metabolism in *D. melanogaster*.

Overall, this thesis provides a detailed exploration of the functions of the EcKL gene family in insects, showing that these genes comprise a novel detoxification gene family in multiple taxa, and that they may also contribute to understudied aspects of ecdysteroid metabolism in a model insect. This work also demonstrates the power and potential of integrating evolutionary, genomic, transcriptomic and experimental data when characterising genes of unknown function.

Declaration

This is to certify that:

i. the thesis comprises only my original work towards the PhD, except where indicated in the preface;

ii. due acknowledgement has been made in the text to all other material used; and

iii. the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Jack Scanlan

August 2020

Preface

This thesis comprises six chapters: an introduction/literature review (Chapter 1), four results chapters (Chapters 2–5) and a general discussion (Chapter 6), as well as three appendices. The bulk of Chapter 2 has been published as an article in the journal *Insect Biochemistry and Molecular Biology* (Scanlan *et al.* 2020), but the thesis chapter contains additional introductory and discussion sections (Chapters 2.1 & 2.7).

Some experiments, analyses or data presented in this thesis were conducted or collected by myself during my MSc (Genetics) research project (2013–14): the annotation and some preliminary evolutionary analyses of the EcKL gene family in the *Drosophila* genus found in Chapter 2 (which were subsequently revised and re-analysed for this thesis and Scanlan *et al.* 2020), the RNAi knockdown experiments in Chapter 2 and Scanlan *et al.* (2020), the preliminary annotation of the EcKL gene family in 29 of the insect genomes found in Chapter 3 (which were subsequently revised for this thesis), and some data presented in Appendix 2 (as noted there).

In addition, other students and scientists materially contributed to this work. Paul Battlay conducted the PheWAS on EcKL and P450 genes (Chapter 2) and the *in silico* genotyping of *CG31370* in the DGRP (Chapter 4); Charles Robin and Paul Battlay assisted with writing and editing the published manuscript (Scanlan *et al.* 2020) that forms the bulk of Chapter 2; and Rebecca Gledhill-Smith and Pontus Leblanc assisted with the RNAi knockdown experiments in Chapter 2—I would also like to acknowledge Rebecca's complementary PhD project on the biochemical and structural characterisation of EcKL proteins. Philip Batterham, Kieran Harvey, Michael Murray, Michael O'Connor and Trent Perry provided fly stocks used in Chapters 2, 4 and 5, and Simon Bullock provided the pCFD6 plasmid vector used in Chapters 4 and 5.

Acknowledgements

Over the course of my time in his lab (over eight years, if you can believe it—from helping Llew Green out as an undergrad, to my MSc project, to this PhD project), Charlie Robin has been a fantastic supervisor—always enthusiastic, encouraging, thoughtful and willing to pin down exciting research questions with me during long meetings. I am grateful of the time and independence he has given me to explore this research topic so broadly, and his mentorship has made me a better scientist. Thanks for everything, Charlie.

All the members of the Robin Lab past and present have been wonderful to work alongside—special thanks to Paul Battlay, Rebecca Gledhill-Smith, Dave Clarke, Natalia Hernandes and Rob Good for advising or helping out with various lab or bioinformatic aspects of this PhD, as well as Caitlyn Perry and Rebecca again for many discussions and conversations about the areas where our PhD projects overlapped. The old Department of Genetics cohort (now a BioSciences oasis in Old Micro), along with the *Drosophila* people at Bio21, have been great to work with over the years too, especially students in the Murray, Batterham, Perry and Andrianopoulos groups, and in particular Shane Denecke and Danielle Christesen for discussions about *Drosophila* genetics, and Harshini Weerasinghe for molecular biology advice.

My co-supervisor, Phil Batterham, has always been so encouraging and helpful over the years, and together with my other committee members, Mike Murray and Alex Andrianopoulos, they have always given me great guidance throughout this project—I've been lucky to have them in my corner. I'd also like to thank Christen Mirth and Mike O'Connor for discussions about ecdysteroids, Lars Jermiin for discussions about phylogenetics, and Alex Fournier-Level for practical advice about pulling a PhD together. Of course, doing *Drosophila* work in Australia would be much more painful without the OzDros facility and the rest of the AusFly community, who are great people.

Doing a PhD is tough work and I definitely wouldn't have gotten through it without the support of my lovely family and friends—many thanks to you all.

And to my wonderful partner, Mel Stewart: you know what you did.

Table of Contents

Chapter 1: Introduction 1

1.1. The evolution and functional analysis of gene families 2

1.1.1. Uncharacterised gene families and domains 4
1.1.2. Models and mechanisms of gene family evolution 5
1.1.3. Genomic annotation of gene families 7
1.1.4. Phylogenetic analyses of gene families 7
1.1.5. Gene family nomenclature 9
1.1.6. Functional genetics of gene families 11
1.1.7. Comparative phylogenomics and gene family-phenotype associations 12

1.2. Xenobiotic detoxification in insects 15

1.2.1. Detoxification breadth 16
1.2.2. Detoxification mechanisms and phases 19
1.2.3. Classical detoxification gene families 22
1.2.4. The evolution of detoxification gene families 24
1.2.5. Tissues and organs involved in detoxification 26
1.2.6. Transcriptional regulation of detoxification genes 27
1.2.7. Identifying and characterising detoxification genes 29
1.2.8. Detoxification by phosphorylation 30

1.3 Ecdysteroid hormones 33

1.3.1. Biosynthesis of ecdysteroids 34
1.3.2. Ecdysteroid signalling 38
1.3.3. Transport of ecdysteroids 39
1.3.4. Inactivation of ecdysteroids 39
1.3.5. Recycling of ecdysteroids 44
1.3.6. Ecdysteroid-phosphate conjugates 47
1.3.7. Ecdysteroid kinases 53
1.3.8. Ecdysteroid-phosphate phosphatases 55

1.4. The ecdysteroid kinase-like (EcKL) gene family 58

1.4.1. EcKLs in *Drosophila melanogaster* 59
1.4.2. EcKLs in other insects 61
1.4.3. What are the functions of EcKLs? 62

1.5. Summary, thesis aims and hypotheses 65

Chapter 2: Genomic, transcriptomic and evolutionary analyses of EcKLs in *Drosophila* 67

2.1. Chapter Introduction 68

2.2. Paper Introduction 70

2.3. Materials and Methods 76

2.3.1. Gene family annotation 76
2.3.2. Phylogenetic analyses 76
2.3.3. DGRP PheWAS 77
2.3.4. Tissue-specific transcriptome data 78
2.3.5. Gene induction datasets 78
2.3.6. Detoxification scores 79
2.3.7. Review of published P450 functions 79
2.3.8. RNAi knockdown 80
2.3.9. Statistical analyses 80

2.4. Results 81

2.4.1. Evolution of the EcKL gene family in the *Drosophila* genus 81
2.4.2. EcKL and P450 genes are transcriptionally enriched in detoxification tissues in *D. melanogaster* 85
2.4.3. Associations between xenobiotic induction, CncC induction, evolutionary stability and detoxification tissue enrichment 88
2.4.4. The detoxification score method is concordant with known functions of *D. melanogaster* P450s 91
2.4.5. DS method suggests many *D. melanogaster* EcKLs are candidate detoxification genes 97
2.4.6. Genomic and transcriptomic variation in EcKL and P450 genes is associated with toxic stress phenotypes in *D. melanogaster* 97
2.4.7. Ubiquitous RNAi knockdown of EcKLs in *D. melanogaster* 101

2.5. Paper Discussion 102

2.5.1. EcKL evolution in *Drosophila* 102
2.5.2. Detoxification gene properties and tissue-specific enrichment 103
2.5.3. Suitability of the DS method for identifying candidate detoxification genes in insects 104
2.5.4. Integration of EcKL and P450 PheWAS associations and candidate detoxification genes 107
2.5.5. Functional characterisation of EcKLs 108

2.6. Paper Summary 111
2.7. Chapter Discussion 112
2.7.1. PheWAS methylmercury phenotypes 112
2.7.2. Detoxification candidate EcKLs 113
2.7.3. Developmental EcKLs 114
2.8. Supplementary Materials 115
2.8.1. Supplementary Figures 115
2.8.2. Supplementary Tables 123
Chapter 3: Evolution and comparative phylogenomics of the EcKL gene family in insects 124
3.1. Introduction 125
3.1.1. What is known about the evolution of EcKLs across insects? 125
3.1.2. Why conduct comparative phylogenomic analyses of the EcKL family? 127
3.1.3. Chapter Aim 130
3.2. Materials and Methods 131
3.2.1. Genome and transcriptome annotation 131
3.2.2. Multiple sequence alignment, phylogenetic inference and ancestral clade assignment 132
3.2.3. An estimated time tree for insects 133
3.2.4. Detoxification gene family sizes, dietary groups and DB groups 135
3.2.5. PGLS regression of gene family sizes in insects 144
3.2.6. PGLS regression of EcKL family size and host plant diversity in Lepidoptera 144
3.2.7. PGLS regression of gene family sizes and diet/detoxification breadth (DB) 145
3.3. Results 147
3.3.1. Annotation of the EcKL gene family in insect and arthropod genomes 147
3.3.2. EcKL subfamilies in insects 148
3.3.3. Evolutionary relationships between EcKLs in insects and other arthropods 150
3.3.4. Ancestral EcKL clades in Diptera 157
3.3.5. Ancestral EcKL clades in Lepidoptera 161
3.3.6. Ancestral EcKL clades in Hymenoptera 164
3.3.7. Non-standard sequence features of insect EcKL proteins 167
3.3.8. EcKL family size is positively associated with the size of known detoxification gene families across insects 168
3.3.9. EcKL family size is positively associated with host plant diversity in Lepidoptera 173
3.3.10. EcKL family size varies with diet and estimated detoxification breadth (DB) in insects 178
3.3.11. EcKLs have low rates of duplication in tsetse flies (Diptera: Glossinidae) 182
3.3.12. EcKLs have low rates of duplication and loss in bees and sphecoid wasps (Hymenoptera: Apoidea) 184
3.4. Discussion 186
3.4.1. Insect-wide EcKL conservation 186
3.4.2. Intra-order EcKL conservation 187
3.4.3. Co-evolution of detoxification gene family sizes 189
3.4.4. Associations between dietary phenotypes and detoxification gene family sizes 190
3.4.5. Is detoxification breadth (DB) associated with EcKL stability as well as size? 193
3.4.6. Candidate detoxification EcKLs in Diptera 194
3.4.7. Candidate detoxification EcKLs in Lepidoptera 195
3.4.8. Candidate detoxification EcKLs in Hymenoptera 197
3.4.9. Candidate detoxification EcKLs in other insect orders 199
Box 3.1. EcKL gene family nomenclature 200
Box 3.2. Limitations of annotation and phylogenetic methodologies 202
3.5. Conclusion 205
3.6. Supplementary Materials 206
3.6.1. Supplementary Figures 206
3.6.2. Supplementary Tables 210
Chapter 4: Functional genetic analyses of detoxification candidate EcKLs in *Drosophila melanogaster* 216
4.1. Introduction 217
4.1.1. Detoxification in *Drosophila* 218
4.1.2. The chemical ecology of *Drosophila* 219
4.1.3. Dro5 EcKLs in *Drosophila melanogaster* 220
4.1.4. Plant and fungal toxins selected for experiments 223
4.1.5. Chapter Aim 224
4.2. Materials and Methods 227
4.2.1. Fly lines and husbandry 227
4.2.2. Generation of UAS-ORF lines 228
4.2.3. Generation of CRISPR-Cas9 mutant lines 229
4.2.4. DGRP genotyping and statistical analyses 233

4.2.5. Lab media developmental viability assays 233
4.2.6. Single-dose developmental toxicology assays 234
4.2.7. Dose-response developmental toxicology assays 234
4.2.8. Citrus-based media developmental assays 236

4.3. Results 238

4.3.1. CRISPR-Cas9 mutagenesis of the Dro5 EcKL clade 238
4.3.2. Developmental viability of Dro5 knockout and misexpression 239
4.3.3. A *CG31370* loss-of-function allele is associated with adult caffeine susceptibility in the DGRP 244
4.3.4. Dro5 loss-of-function mutants are developmentally susceptible to caffeine 247
4.3.5. Misexpression of two Dro5 genes in detoxification tissues decreases developmental susceptibility to caffeine 253
4.3.6. *Dro5$^{A3\text{-}B7}$* homozygotes are developmentally susceptible to kojic acid 253
4.3.7. No evidence that *Dro5$^{A3\text{-}B7}$* homozygotes have increased developmental susceptibility to salicin, curcumin, escin, esculin or quercetin 256
4.3.8. *Dro5$^{A3\text{-}B7}$* homozygotes are developmentally viable on citrus-based media 259

4.4. Discussion 261

4.4.1. Which Dro5 EcKLs are most likely to be involved in caffeine detoxification? 261
4.4.2. Caffeine toxicology and metabolism in *Drosophila* 264
4.4.3. Is caffeine detoxification in *Drosophila* ecologically relevant? 270
4.4.4. Dro5 EcKLs and other plant toxins 271
4.4.5. Dro5 EcKLs and kojic acid 272
4.4.6. Are all Dro5 EcKLs detoxification genes? 273
4.4.7. Future research directions for Dro5 genes and detoxification 275

4.5. Conclusion 277

4.6. Supplementary Figures 278

Chapter 5: Functional genetic analyses of ecdysteroid kinase candidate EcKLs in *Drosophila melanogaster* 280

5.1. Introduction 281

5.1.1. Identifying and characterising ecdysteroid kinase-encoding genes 281
5.1.2. Ecdysteroid phosphorylation in *Drosophila melanogaster* 282
5.1.3. *Wallflower* (*Wall*/*CG13813*) and *Pinkman* (*pkm*/*CG1561*) are *D. melanogaster* orthologs of *BmEc22K* 283
5.1.4. Published data on *pkm* 283
5.1.5. Published data on *Wall* 286
5.1.6. Chapter Aim 287

5.2. Materials and Methods 289

5.2.1. Fly lines and husbandry 289
5.2.2. Generation of UAS-ORF lines 291
5.2.3. Generation of CRISPR-Cas9 mutant lines 291
5.2.4. Germline knockout, somatic knockout, RNAi knockdown and misexpression developmental viability assays 295
5.2.5. Delayed-onset misexpression assay 295
5.2.6. Food avoidance assay 295
5.2.7. Misexpression epistasis assays 296
5.2.8. *EPPase*/*Wall* co-misexpression assays 296
5.2.9. Starvation assays 297

5.3. Results 298

5.3.1. CRISPR-Cas9 mutagenesis of *Wall* and *pkm* 298
5.3.2. *Wall* and *pkm* single and double mutants have no obvious developmental defects 300
5.3.3. Somatic knockout of *pkm* produces an immobility phenotype in adult flies 302
5.3.4. Putative *Wall* and *pkm* RNAi knockdown causes developmental arrest, but may be due to off-target effects 304
5.3.5. Ubiquitous misexpression of *Wall*, but not *pkm*, *BmEc22K* or *EPPase*, causes developmental arrest 308
5.3.6. Ubiquitous misexpression of *Wall* causes developmental arrest until the middle of metamorphosis 311
5.3.7. Tissue-specific misexpression of *Wall* causes developmental arrest distinct from *Cyp18a1* 311
5.3.8. Larval tracheal misexpression of *Wall* causes tracheal defects and food aversion 316
5.3.9. *Wall* misexpression phenotypes are not hypostatic to *Cyp18a1* 318
5.3.10. Co-misexpression of *EPPase* does not rescue developmental arrest caused by misexpression of *Wall* 320
5.3.11. *pkm* mutants are not more susceptible to starvation than wild-type controls 322

5.4. Discussion 324

5.4.1. Discrepancies between gene disruption phenotypes 324
5.4.2. Lack of *pkm*, *BmEc22K* and *EPPase* misexpression phenotypes 325
5.4.3. Possible functions of *pkm* 327
5.4.4. Future experiments on *pkm* 328
5.4.5. Possible functions of *Wall* 329
5.4.6. What is Wall's substrate? 330
5.4.7. How does *Wall* misexpression produce developmental defects? 334
5.4.8. Future experiments on *Wall* 337

5.5. Conclusion 339

5.6. Supplementary Materials **340**
5.6.1. Fly crossing for genotype creation 340
5.6.2. Supplementary Figures 342
5.6.3. Supplementary Tables 345

Chapter 6: General Discussion **348**

6.1. Introduction **349**

6.2. EcKLs as ecdysteroid kinases **351**
6.2.1. How many EcKLs encode ecdysteroid kinases? 351
6.2.2. Conservation of ecdysteroid kinases in insects 352
6.2.3. Candidate ecdysteroid kinase EcKLs 354

6.3. EcKLs as detoxification enzymes **356**
6.3.1. Phosphorylation and the classical model of detoxification 356
6.3.2. Physiological considerations of detoxicative phosphorylation 357
6.3.3. Natural X-class EcKL substrates 358

6.4. Other EcKL functions **360**
6.4.1. Juvenile hormone and EcKLs 360
6.4.2. Non-ecdysteroidal E-class EcKL substrates 360
6.4.3. S-class EcKL functions 361

6.5. The evolution of EcKL functions **363**
6.5.1. Evolutionary transitions between EcKL functions 363
6.5.2. What was the ancestral EcKL function in Tetraconata? 364
6.5.3. What was the origin of the EcKL family? 365

6.6. Structural properties of EcKL proteins **367**

6.7. Future research directions **369**
6.7.1. Outstanding questions in *Drosophila melanogaster* 369
6.7.2. Functional genetics of EcKLs in other insects 370
6.7.3. EcKLs in non-insects 371
6.7.4. EcKL biochemistry 371
6.7.5. Practical applications of EcKL research 372

6.8. Conclusion **374**

References **375**

Appendices **427**

Appendix 1: Ubiquitous misexpression of EcKLs in *Drosophila melanogaster* **428**

Appendix 2: *CG31098* may be required for choriogenesis in *Drosophila melanogaster* **433**

Appendix 3: Supplementary Methods **440**

List of Figures

Figure 1.1. Biosynthetic chemodiversity, detoxification breadth and detoxification capacity 18
Figure 1.2. Generalised xenobiotic metabolism/detoxification pathways 21
Figure 1.3. Phosphate metabolites of xenobiotic compounds in insects 32
Figure 1.4. Generalised ecdysteroid biosynthesis, signalling and catabolism 37
Figure 1.5. Ecdysteroid catabolic reactions 43
Figure 1.6. Ecdysteroid-phosphate conjugates 46
Figure 1.7. The ovarian/embryonic ecdysteroid-phosphate (OEEP) recycling system of *Bombyx mori* 52

Figure 2.1. EcKLs in 12 species of *Drosophila* 82
Figure 2.2. Genes per species, duplications and losses for ancestral EcKL clades in *Drosophila* 83
Figure 2.3. Gene tree of ancestral EcKL clades in *Drosophila* 84
Figure 2.4. Summary of tissue enrichment of P450s and EcKLs in *Drosophila melanogaster* 87
Figure 2.5. Associations between detoxification characteristics in *Drosophila melanogaster* P450s and EcKLs 90
Figure 2.6. Detoxification score detail of P450s in *Drosophila melanogaster* 93
Figure 2.7. Detoxification score detail of EcKLs in *Drosophila melanogaster* 99
Figure 2.8. Detoxification score summaries of P450s and EcKLs in *Drosophila melanogaster* 100
Figure S2.1. Tissue expression of P450s in *Drosophila melanogaster* 115
Figure S2.2. Tissue expression of EcKLs in *Drosophila melanogaster* 116
Figure S2.3. Tissue enrichment of P450s in *Drosophila melanogaster* 117
Figure S2.4. Tissue enrichment of EcKLs in *Drosophila melanogaster* 118
Figure S2.5. Tissue enrichment differences between stable and unstable P450s and EcKLs in *Drosophila melanogaster* 119
Figure S2.6. Tissue enrichment differences between xenobiotically induced and uninduced P450s and EcKLs in *Drosophila melanogaster* 120
Figure S2.7. Tissue enrichment differences between CncC induced and uninduced P450s and EcKLs in *Drosophila melanogaster* 121
Figure S2.8. RNAi knockdown of EcKLs in *Drosophila melanogaster* 122

Figure 3.1. Published phylogenetic trees of the EcKL gene family 126
Figure 3.2. Representative phylogeny of Ecdysozoa, Arthropoda, Tetraconata and Insecta 129
Figure 3.3. Master phylogeny of genomically annotated insect species 134
Figure 3.4. Lepidopteran host plant data from the HOSTS Database 146
Figure 3.5. Gene tree of EcKL subfamilies in insects 151
Figure 3.6. EcKLs per subfamily per genome in insects 152
Figure 3.7. Summary of EcKLs per subfamily per insect order 153
Figure 3.8. EcKL subfamily retention per insect order for genomes 154
Figure 3.9. EcKL subfamily presence/absence per insect order for genomes and transcriptomes 155
Figure 3.10. Unrooted gene tree of Tetraconata EcKLs 156
Figure 3.11. Gene tree of ancestral EcKL clades in Diptera 158
Figure 3.12. EcKLs per ancestral clade per species in Diptera 159
Figure 3.13. Gene tree of ancestral EcKL clades in Lepidoptera 162
Figure 3.14. EcKLs per ancestral clade per species in Lepidoptera 163
Figure 3.15. Gene tree of ancestral EcKL clades in Hymenoptera 165
Figure 3.16. EcKLs per ancestral clade per species in Hymenoptera 166
Figure 3.17. Gene family sizes, diet and detoxification breadth for genomically annotated insects 169
Figure 3.18. Heatmap of EcKL, P450, GST and CCE family sizes across insects 170
Figure 3.19. PGLS regressions of EcKL, P450, GST and CCE family sizes 171
Figure 3.20. PGLS regressions of EcKL family size against host plant diversity in Lepidoptera 175
Figure 3.21. PGLS regressions of Lep1 and Lep8 clade size against host plant diversity in Lepidoptera 176
Figure 3.22. PGLS regression of EcKL, P450, GST and CCE family sizes against diet 179
Figure 3.23. PGLS regression of EcKL, P450, GST and CCE family sizes against detoxification breadth 180
Figure 3.24. Gene tree of EcKLs in *Glossina* 183
Figure 3.25. EcKLs per ancestral Hymenoptera clade per species for Apoidea genomes and transcriptomes 185
Figure S3.1. Relationships between the proportion of partial gene models and genome assembly N50 and L50 206
Figure S3.2. The number of partial-only EcKL subfamilies by the number of genomes and transcriptomes queried 207
Figure S3.3. Phylogenetic correlograms for autocorrelation of EcKL, P450, GST and CCE family sizes in insects 208
Figure S3.4. Gene trees for the Lep1 and Lep8 EcKL clades in Lepidoptera 209

Figure 4.1. The chromosome 3R cluster of 26 EcKLs in *Drosophila melanogaster* 222
Figure 4.2. Chemical structures of toxins used in toxicological experiments 225
Figure 4.3. Crossing schemes for CRISPR-Cas9 mutagenesis on chromosome 3 232
Figure 4.4. The Dro5 locus in *Drosophila melanogaster*, with mutant alleles 241
Figure 4.5. Developmental viability of Dro5 loss-of-function alleles on lab media 242
Figure 4.6. Developmental viability of Dro5 ubiquitous misexpression on lab media 243
Figure 4.7. The *CG31370^{del}* allele in the DGRP, with caffeine phenotypes and expression data 246
Figure 4.8. Caffeine dose-response assays with the *Dro5$^{A3\text{-}B7}$* allele 249

Figure 4.9. Caffeine dose-response assays with the *Dro5$^{A3\text{-}B7}$* allele and *Df(3R)BSC852* 250
Figure 4.10. Caffeine dose-response assays with the *Dro5$^{A3\text{-}B7}$* allele and single-gene disruption alleles 251
Figure 4.11. Caffeine dose-response assays with the *CG31370^{MI}* allele 252
Figure 4.12. Caffeine dose-response assays with *HR*>Dro5 misexpression 254
Figure 4.13. Kojic acid dose-response assays with the *Dro5$^{A3\text{-}B7}$* allele 255
Figure 4.14. Developmental viability assays with the *Dro5$^{A3\text{-}B7}$* allele and curcumin, escin, esculin and quercetin media 257
Figure 4.15. Salicin dose-response assays with the *Dro5$^{A3\text{-}B7}$* allele 258
Figure 4.16. Developmental viability assays with the *Dro5$^{A3\text{-}B7}$* allele on citrus juice media 260
Figure 4.17. Working hypothesis for caffeine detoxification via Dro5 enzymes 269
Figure S4.1. Gel electrophoresis of failed Dro5B CRISPR-Cas9 mutagenesis 278
Figure S4.2. Independent caffeine dose-response assays with the *Dro5$^{A3\text{-}B7}$* allele 279

Figure 5.1. Expression of *Wall* and *Cyp18a1* through *Drosophila melanogaster* development 288
Figure 5.2. Crossing scheme for CRISPR-Cas9 mutagenesis on the X chromosome 293
Figure 5.3. The *Wall* and *pkm* loci in *Drosophila melanogaster*, with mutant alleles 299
Figure 5.4. Developmental viability of *Wall* and *pkm* single and double mutants 301
Figure 5.5. Ubiquitous somatic CRISPR-Cas9 mutagenesis of *Wall* and *pkm* 303
Figure 5.6. Ubiquitous RNAi knockdown of *Wall* and *pkm* (egg-to-adult viability) 306
Figure 5.7. Ubiquitous RNAi knockdown of *Wall* and *pkm* with UAS-*Dcr2* at 29 ºC (larval-to-adult viability) 307
Figure 5.8. Ubiquitous misexpression of *Wall*, *pkm*, *BmEc22K*, *Cyp18a1* and *EPPase* 309
Figure 5.9. Visible phenotypes of *act*>*Wall^{EY}* larvae 310
Figure 5.10. Delayed-onset misexpression of *Wall* with *act-GAL4* 314
Figure 5.11. Tissue-specific misexpression of *Wall* and *Cyp18a1* 315
Figure 5.12. Hypoxia and tracheal and pigmentation defects in *btl*>*Wall* individuals 317
Figure 5.13. Tests of hypostasis between *Wall* misexpression and *Cyp18a1* 319
Figure 5.14. Tests of interaction between *Wall* misexpression and *EPPase* misexpression 321
Figure 5.15. Starvation experiments with *pkm* loss-of-function alleles 323
Figure S5.1. Gene tree of subfamily A EcKLs from Diptera and Lepidoptera 342
Figure S5.2. Developmental arrest phenotype of *HR*>*Wall^{EY}* animals at 25 ºC 343
Figure S5.3. Basal expression of EcKLs in the S2-DGRC and S2R+ cell lines 344

Figure A1.1. Ubiquitous misexpression of 13 EcKLs with the *tub-GAL4* driver 430
Figure A2.1. Phenotypes associated with transheterozygous deficiency genotypes around the *CG31098* locus 436

List of Tables

Table 2.1. Phosphorylated metabolites of xenobiotic compounds in insects 69
Table 2.2. Comparison of detoxification scores with known functions for *Drosophila melanogaster* P450s 94
Table 2.3. Novel detoxification candidate P450s in *Drosophila melanogaster* 96

Table 3.1. Gene family sizes, diet and detoxification breadth for genomically annotated insects 137
Table 3.2. Mapping between insect subfamilies, ancestral Diptera clades and ancestral *Drosophila* clades for the EcKLs 160
Table 3.3. Model outputs for PGLS regressions of EcKL, P450, GST and CCE family sizes 172
Table 3.4. Model outputs for PGLS regressions of EcKLs against host plant diversity in Lepidoptera 177
Table 3.5. Model outputs for PGLS regressions of EcKL, P450, GST and CCE family sizes against diet and detoxification breadth 181
Table S3.1. All arthropod genome assemblies annotated for the EcKL family 210

Table 4.1. Induction by caffeine and PheWAS associations for the Dro5 genes 226
Table 4.2. Sequences of primers used in Chapter 4 237
Table 4.3. Collated evidence for caffeine detoxification for individual Dro5 genes in *Drosophila melanogaster* 263

Table 5.1. GAL4 drivers used in Chapter 5 290
Table 5.2. Sequences of primers used in Chapter 5 294
Table S5.1. Phenotypic descriptions of tissue-specific *Wall* and *Cyp18a1* misexpression 345

Common Abbreviations

20E	20-hydroxyecdysone
aa	amino acid(s)
ABC	ATP-binding cassette transporter
AGRF	Australian Genome Research Facility
bp	base pair(s)
CCE	carboxylcholinesterase
CDS	coding sequence
chr	chromosome
CI	confidence interval
CNS	central nervous system
CNV	copy number variation
CRISPR	clustered regularly interspaced short palindromic repeats
DB	detoxification breadth
DGRP	Drosophila Genetic Reference Panel
DNA	deoxyribonucleic acid
E	ecdysone
EcK	ecdysteroid kinase
EcKL	ecdysteroid kinase-like
EPPase	ecdysteroid-phosphate phosphatase
EST	expressed sequence tag
EtOH	ethanol
FPKM	fragments per kilobase of transcript per million mapped reads
GFP	green fluorescent protein
GLM	generalised linear model
gRNA	guide RNA
GST	glutathione S-transferase
GWAS	genome-wide association study
HMM	hidden Markov model
kb	kilobase pairs
LB	lysogeny broth
LC_{50}	median (50%) lethal concentration
MSA	multiple sequence alignment
NEB	New England Biolabs
nt	nucleotide(s)
OEEP	ovarian/embryonic ecdysteroid-phosphate
ORF	open reading frame
P450	cytochrome P450
PCR	polymerase chain reaction
PheWAS	phenome-wide association study
RNA	ribonucleic acid
RNAi	RNA interference
SNP	single-nucleotide polymorphism
TE	transposable element
TWAS	transcriptome-wide association study
UAS	upstream activating sequence
UGT	UDP-glycosyltransferase
UTR	untranslated region

Chapter 1

Introduction

This introductory literature review encompasses three areas of biology—gene families (Chapter 1.1), xenobiotic metabolism in insects (Chapter 1.2) and ecdysteroid moulting hormones (Chapter 1.3)—and an overview of what is currently known about the ecdysteroid kinase-like (EcKL) gene family, the specific focus of this thesis (Chapter 1.4).

1.1. The evolution and functional analysis of gene families

Genes are the functional units of heredity in a genome and terminally encode protein or RNA molecules that carry out the biological functions of cells and organisms as a whole—in this thesis, 'function' refers to selected-effect function ala. Graur *et al.* (2015); 'biochemical function' will be used to refer to biochemical activity *per se* without regard for a gene's *raison d'être*. There are four major natural ways through which a gene (or part of a gene) can find itself in a genome: *de novo* origination from non-genic sequence (Van Oss & Carvunis 2019); duplication of an existing sequence in the genome (Innan & Kondrashov 2010); horizontal transmission from another genome (horizontal gene transfer, or HGT; Soucy *et al.* 2015); and inheritance from a parental genome. The last of these—inheritance—is clearly the most common, with the other three having various degrees of importance depending on the taxon (e.g. HGT is much more common in prokaryotes than in eukaryotes). The interaction of these four processes—as well as others, such as deletion, gene fusion and gene conversion—shape the evolution of genes within and between species. Genes that are evolutionarily related to one another and therefore can be traced back to a single ancestral gene are homologous (or 'homologs'); orthologs are homologs that diverged due to speciation, paralogs are homologs that diverged due to duplication, and xenologs are homologs that diverged due to HGT (Treangen & Rocha 2011).

That many genes between and within genomes are homologs has long suggested the natural classification of genes into 'families'. However, there are two competing definitions of a gene family, one that classifies single-copy orthologs as a gene family (definition A), and one that restricts the definition to gene lineages in which duplication has occurred (definition B; Demuth & Hahn 2009). For example, under definition A, a gene that was born *de novo* in the common ancestor of all mammals and has been retained as a single copy in all extant mammalian species would be classified as a gene

family, while definition B would call them single-copy orthologs and require some amount of duplication to have occurred, such that it would have become a gene family by 'expanding'.

An issue with definition A is that it necessitates the use of some other terms to describe larger gene families: 'multigene' families, gene families in which duplication has occurred; and gene 'superfamilies', very large multigene families where the members have become highly diverged at the sequence, expression and functional levels (Ohta 2001). The distinction between multigene families and superfamilies is unclear in practice, with many gene families classified as one or the other depending on their perceived size and complexity, with no objective consistency. Additionally, if *de novo* gene birth is truly rare, then a large proportion of all genes in nature are orthologous and paralogous over deep evolutionary time, in which case the gene family concept can be extended to include many more genes than may be intuitively appropriate. Also, as suggested by Ohta (2001), superfamilies may include multiple families and multigene families within themselves; this all makes delineation between terms impractical.

Complicating gene family definitions are the competing classification schemes of proteins, which tend to be organised by a combination of structural, functional and evolutionary information applied to individual domains. The Structural Classification of Proteins (SCOP) hierarchy (Murzin *et al.* 1995), for example, includes as its levels 'proteins' (high-identity sequences with essentially identical biochemical functions), 'families' (moderate-identity sequences with similar biochemical functions), 'superfamilies' (low-identity sequences with related biochemical functions), 'folds' (sequences with related tertiary structures; typically the last point at which homology is asserted) and 'classes' (sequences with related secondary structures and organisation). Alternatively, in the Pfam database, a domain family is defined as a curated collection of related protein regions that share a Hidden Markov Model (HMM) and may or may not share a biochemical function (El-Gebali *et al.* 2018; Punta *et al.* 2012). As genes can encode proteins with multiple domains, this can result in different parts of the same gene belonging to different classes or folds, even though over moderate amounts of evolutionary time they are evolving together within the same gene family.

For the purposes of this thesis, which deals mostly with single-domain enzymes, a

'gene family' will be defined as a collection of orthologs and paralogs that share a core conserved protein domain defined by the Pfam database (El-Gebali *et al.* 2018; Punta *et al.* 2012). Intra-family classification nomenclature will depend on the family in question, but the term 'gene family' will always refer to the family—as defined by the domain—as a whole, not a particular clade of orthologs or paralogs within the family.

1.1.1. Uncharacterised gene families and domains

A key promise of the genomics 'revolution' was that sequencing an organism's genome would allow inferences about the organism's molecular biology from sequence data alone (Galperin & Koonin 2010). Such inferences rely on experimentally determined data on homologous genes, but a large proportion of all known genes have unknown functions (Jaroszewski *et al.* 2009). In 2018, 22% (3,961) of all domain families in Pfam were 'domains of unknown function' (DUFs), the same fraction as in 2010 (Bateman *et al.* 2010; El-Gebali *et al.* 2018); DUF families do not possess members with characterised or suspected functions, and overall, 25% of all domains do not have functions that have been experimentally validated (El-Gebali *et al.* 2018). These data suggest that functional characterisation has barely kept pace with the increase in novel sequence discovery due to taxonomically broad-sampled genome sequencing, particularly in the field of microbial metagenomics (Bernard *et al.* 2018). An analysis in 2015 suggested that a majority of DUFs are taxonomically restricted, suggesting they may have lineage-specific functions (Mudgal *et al.* 2015), highlighting the need for the characterisation of genes and proteins outside of the core model organisms.

Improving the research community's understanding of the complete diversity of protein domains and gene families has implications not only for the fundamental challenge of linking genotype to phenotype, but also for the fields of biotechnology and synthetic biology (Ellens *et al.* 2017; Wright & Nemhauser 2019). Many DUF-encoding genes have possible biotechnological applications in plants (Yang *et al.* 2017a; 2017b; Zhang *et al.* 2016), and one-third of genes in the minimal synthetic *Mycoplasma mycoides* JCVI-syn3.0 genome are of unknown function, despite being essential for cell growth and survival (Danchin & Fang 2016), something also seen in naturally occurring bacteria (Goodacre *et al.* 2014).

1.1.2. Models and mechanisms of gene family evolution

Multiple models have been proposed to explain the large-scale patterns of evolution within gene families (Nei & Rooney 2005). Some families exhibit 'concerted evolution', whereby paralogs within a genome share higher sequence similarity than paralogs between genomes; this is thought to occur through recombination-based mechanisms such as gene conversion and unequal crossing-over (Chen *et al.* 2007; Holliday 1964; Roman & Ruzinski 1990). However, strict concerted evolution appears to explain the evolution of relatively few gene families, and so another model—'birth-and-death'—was developed, whereby the interplay between gene duplication ('birth') and pseudogenisation/deletion ('death') over time produces a stochastic pattern of gene gain and loss throughout the family, with some paralogs retained for long periods of time and others lost quickly (Nei & Rooney 2005). It is likely that these two models of evolution lie at the ends of a spectrum, with most gene families falling somewhere between the two extremes; gene conversion likely has an underappreciated role in the sequence evolution of paralogs in families that otherwise show broad birth-and-death patterns, and may be partially responsible for challenges in reliably detecting orthological relationships in large gene families (Hasić & Tannier 2019).

Many gene families have substantial differences in size (the number of genes in the family) between taxa (Demuth *et al.* 2006; Hahn *et al.* 2007; Morales-Cruz *et al.* 2015). Gene family size evolves through the interaction of multiple molecular mechanisms, including unequal crossing-over (resulting in local duplication or deletion), retrotransposition and whole genome duplication (Demuth & Hahn 2009), but there are disagreements between researchers about whether gene family size differences are predominantly explained by neutral (Hahn *et al.* 2005) or adaptive (Good *et al.* 2014; Low *et al.* 2007) evolutionary forces. However, it seems unlikely that very large gene families with substantial sequence divergence between paralogs would be retained in the genome without the action of directional selection in genomes where the DNA loss rate is high (Hawkins *et al.* 2009; Petrov & Hartl 1998; Petrov *et al.* 1996), although there may be a role for neutral forces in smaller-scale changes in gene family size, or in concerted evolution scenarios.

Gene duplicates have three main possible fates: neofunctionalisation, wherein one or both duplicates acquire a novel biochemical function through gain-of-function

mutations; subfunctionalisation, wherein multiple functions (or properties such as expression domains) of the ancestral gene are independently retained in each of the duplicates through complementary, modular loss-of-function mutations; and pseudogenisation, wherein one duplicate acquires total loss-of-function mutations and becomes a pseudogene (Prince & Pickett 2002). Neofunctionalisation is thought to be largely adaptive (e.g. Zimmer *et al.* 2018), but it is possible genes can acquire preadaptive biochemical functions—in the absence of a selective pressure—that later become adaptive in a novel environment (Francino 2005); gene duplicates can also be selected (non-adaptively) through selfish mechanisms such as meiotic drive (Brand *et al.* 2015). Subfunctionalised duplicates are not adaptive *per se* (although they can lead to future neofunctionalisation) but are retained through purifying selection if the ancestral gene was also retained by purifying selection. Gene loss can be adaptive (Borges *et al.* 2015; Sharma *et al.* 2018) but is usually understood to be due to the neutral accumulation of inactivating mutations under relaxed selection. If inactivating point mutations, small indels and/or large deletions occur more frequently than duplication events, functional gene duplicates are unlikely to be retained over long periods of time in the absence of directional or purifying selection, and thus will tend to be lost from the genome as pseudogenes (Katju & Bergthorsson 2013; Petrov *et al.* 1996; Petrov & Hartl 1997).

Gene duplicates may also be retained for relatively short periods of time if increased dosage of the ancestral gene product is adaptive; an extreme example of this is the extensive gene 'amplification' of insecticide resistance-related esterases in aphids (Field *et al.* 1999; 1988). However, so long as total gene dosage is at a level that is favoured by selection, such 'high-dosage' alleles can neutrally accumulate mutations that increase the expression of some duplicates and reduce the expression of others, which can lead to the pseudogenisation of low-expression duplicates (Demuth & Hahn 2009; Gout & Lynch 2015; Thompson *et al.* 2016); such a process makes the long-term retention of gene duplicates for the purposes of high gene dosage unlikely without the action of neofunctionalisation or subfunctionalisation, although there are examples of the retention of duplicated core 'housekeeping' genes such as histones and rRNAs via purifying selection and gene conversion, due to stoichiometric and protein-protein interaction requirements (Eickbush & Eickbush 2007; Scienski *et al.* 2015).

1.1.3. Genomic annotation of gene families

In the modern 'post-genomic' era, genes can be identified and analysed in genomes quickly and easily, and the rapid sequencing and freely available nature of genomes from across the tree of life allows for the 'mining' of genome sequence to analyse gene families in many taxa simultaneously. Determining which genes in a genome belong to a particular gene family can be done with direct searches against HMM databases such as Pfam (Punta *et al.* 2012) or using genome-wide tools such as HMMER (Finn *et al.* 2011).

A necessary precondition for accurate analyses of gene families, however, are accurate gene model annotations; unfortunately, some characteristics of large gene families and assemblies of eukaryotic genomes produce problems for automated annotation methods, such as large tandem arrays of highly similar genes that tend to produce implausible 'merged' gene models, and fragmented assemblies that split gene models across multiple non-contiguous scaffolds, interpreted by annotation pipelines as multiple partial genes or pseudogenes. Automated methods that primarily use homology to genes in other species—such as NCBI's Gnomon—can also fail to accurately annotate highly divergent members of a gene family. Incorporating RNA sequencing (RNA-seq) reads into gene prediction algorithms substantially improves accuracy, but can miss genes that have low read depth (Ambrosino & Chiusano 2017). The best automated methods use phylogenetic information, as well as RNA-seq data, to annotate multiple genomes simultaneously (Emms & Kelly 2019; König *et al.* 2018), but are likely not completely accurate.

Despite recent advances in genome annotation algorithms, manual annotation of gene models remains the 'gold standard' for producing and curating accurate data for gene family analyses. This is a time-consuming, low-throughput method that requires expertise in the gene family to identify possible problems with gene models, but produces high-confidence datasets; as such, many genome sequencing projects use community-based manual annotation methods to target gene families of interest (e.g. Johnson *et al.* 2018c; Kanost *et al.* 2016).

1.1.4. Phylogenetic analyses of gene families

Once all the members of a gene family in a collection of genomes have been found,

the evolution of the family can be studied using phylogenetic tree-building. Typically, for large gene families being studied over taxonomic divergences greater than a few million years, the predicted amino acid sequence of the protein is used over the nucleotide sequence, given that it evolves less rapidly and allows for a greater signal-to-noise ratio at deep divergence points. Phylogenetic trees are estimated by first creating a multiple sequence alignment (MSA) from the sequences of the genes in question, optionally 'polishing' and trimming the MSA to adjust or remove poorly aligned regions, then running the MSA through a tree-building program that attempts to determine the evolutionary relationships between sequences via one of various methods, including parsimony, nearest-neighbour-joining, maximum likelihood, or Bayesian approaches (Nascimento *et al.* 2017; Penny & Hendy 2004; Saitou & Nei 1987; Yang 1996). Once a gene tree is produced, a tree of the known relationships between the species from which the genes were derived (a 'species tree') can be used to infer ortholog and paralog relationships between genes, by reconciling the two trees (Emms & Kelly 2019).

Various factors complicate this approach. First, some gene families evolve so rapidly at the amino acid level that phylogenetic signal is poor for any nodes far from the tips of the tree, making deeper level relationships between genes hard to confidently determine (Clifton *et al.* 2016). Second, large numbers of sequences in an MSA can reduce the power of tree-building methods to accurately estimate the parameters needed to find a good tree, especially when the length of the MSA is short (recommended in Burnham & Anderson 2004). Third, if a well-resolved species tree is not available, accurately determining orthologs and paralogs may be difficult for large and complex gene families. Fourth, gene conversion and paralog exchange may act to reduce the diversity between paralogs in the same genome, confounding true phylogenetic signal (Brady & Richmond 1992; Robin *et al.* 2009)—the impact of these processes on gene family phylogenetics is poorly understood but is currently receiving some attention from researchers (Chan & Robin 2019; Hasić & Tannier 2019). Fifth, in some instances HGT can confound gene tree/species tree reconciliation, but its relative rarity in most eukaryotic genomes means it is typically not an issue when studying gene families in insect taxa. Sixth, incongruence between gene trees and species trees, due to incomplete lineage sorting or introgression on short branches of the species tree, can naively result in true 1:1 orthologs being designated as a complex collection of paralogs. Overall, these factors make the phylogenetic analysis of gene families a non-trivial exercise,

and it needs to be approached in a conscious manner.

1.1.5. Gene family nomenclature

Understanding the evolutionary relationships between genes in a gene family naturally highlights the importance of developing a family-wide naming system, or nomenclature. This serves multiple purposes: it helps to clarify ortholog and paralog relationships without resorting to consulting a phylogenetic tree, and it also allows for quickly summarising the sizes and properties of different parts of a gene family for comparative purposes, through the use of higher-level nomenclature designations. These are important for linking a gene family's functions with its evolution: for example, direct orthologs—which have a high likelihood of possessing similar or identical functions—should have similar or identical nomenclature designations in a good system. However, despite the importance of developing a robust and appropriate gene family nomenclature, there is no consensus on the best way to do so.

Individual genes can have numerous names, which can cause confusion among researchers and non-experts (Editors 2007; Mitchell *et al.* 2003). Sometimes this occurs through simultaneous discovery or mistakes in the scientific literature (Seringhaus *et al.* 2008; Ziemann *et al.* 2016), but can also happen in a valid manner by which a gene encoding an enzyme may be named before the substrate or catalytic mechanism of the enzyme is determined. Some genes have more than two names, such as the cytochrome P450-encoding Halloween genes of arthropods, which are known by a 'phenotypic group' name (Halloween genes), gene family nomenclature (e.g. *Cyp306a1*), gene name in a particular organism (e.g. *phantom* in *Drosophila melanogaster*) and an enzyme name (e.g. 2,22,25dE-ketodiol 25-hydroxylase). Sometimes to distinguish orthologous genes in different species, one of these names may be prefixed or suffixed by a short identifier (e.g. *DmCyp306a1* or *Cyp306a1-Dm* in *D. melanogaster*). In light of all these possible names, a robust nomenclature system for a gene family can serve as the backbone to make the identification of genes intelligible.

For relatively small gene families, nomenclature is often determined by simple and obvious groupings, commonly with single number or letter suffixes (e.g. Holmes *et al.* 2010), but for large gene families, it can become necessary to develop a more complex nomenclature system to highlight the various, nested relationships between particular

gene family members. The cytochrome P450 (P450) 'superfamily' has a nomenclature system based primarily on sequence similarity. P450 genes with >40% amino acid sequence identity belong to the same family, and genes with >55% identity belong to the same subfamily. P450s are formally named in the manner '*CypXYZ*', where X (a number) is the family, Y (one or more letters) is the subfamily, and Z (a number) denotes the particular gene in question—e.g. *Cyp12d1*. The highest level of P450 classification is the clan, which is determined by orthology and is not typically attached to the name of any particular P450. As the P450 gene family currently contains tens of thousands of annotated genes, this nomenclature has grown exceptionally complicated (Nelson 2006). The UDP-glycosyltransferase (UGT) nomenclature is similar to that of the P450s, in the form '*UGTXYZ*', with families (X, a number) sharing >45% identity, subfamilies (Y, a letter) sharing >60% identity and genes (Z, a number) sharing ~100% identity (Mackenzie *et al.* 2005). Notably, the UGT nomenclature does not incorporate phylogenetic information.

An obvious issue with nomenclature based solely on sequence similarity is that the clear grouping of 1:1 orthologs, and other functionally relevant ortholog and paralog groupings, becomes unlikely as both the divergence between species and the evolutionary rate increases; at the other end of the scale, higher level groupings of genes can become unnecessarily exclusionary, possibly leading researchers to assume, based on such a nomenclature, that paralogous genes in different groups may be functionally dissimilar even when they are relatively closely related. Another—more extreme—issue with this approach is that highly diverged genes may be excluded from being grouped with their orthologs, while related but clearly paralogous genes appear closer under the nomenclature, effectively producing paraphyletic groupings.

To deal with these problems, gene family nomenclatures need to be based on evolutionary information: the various levels of the nomenclature need to be cladistic. I call this approach an 'orthology-based' nomenclature. Theoretically, the 'perfect' orthology-based nomenclature of a given gene family is represented by the true phylogenetic relationships of all of its members (i.e. a tree), and groups would be defined at each node of the corresponding tree. However, there are practical problems with this: for many gene families, phylogenetic inferences are at least somewhat inaccurate (especially at deep nodes) due to low signal-to-noise ratios and/or inadequate sequence sampling; also, such a nomenclature would be essentially impossible to use in plain-

language scientific discourse. As such, orthology-based nomenclatures need to be a practical compromise between intelligibility, accuracy and sufficient useful detail on the evolutionary relationships between genes.

1.1.6. Functional genetics of gene families

Functional genetics—linking genotype to phenotype—is split into two general strategies: forward genetics, wherein phenotypes of interest are selected from natural or artificial pools of genomic or transcriptomic variation and the causal genes responsible are identified; and reverse genetics, wherein genes of interest are manipulated and the consequent phenotypes are analysed in order to explore the genes' biological roles (Hardy *et al.* 2010; Peters *et al.* 2003). In order to study the functions of an uncharacterised gene family, reverse genetics approaches—utilising targeted gene knockout, knockdown and misexpression—are typically the most appropriate, given that the members of the gene family will be known and there may be specific functional hypotheses to test. However, gene families present some challenges to such methods of genetic analysis and require careful thought before research is undertaken.

A general problem is working out which genes to study: in large gene families, there may be many members that could plausibly function in any particular biological process of interest. Narrowing down a list of candidates with transcriptomic information may be useful in this situation—for example, if a gene is transcriptionally induced under certain environmental conditions or enriched for expression in a particular set of tissues, this may point towards specific functions.

Paralogous genes can be functionally redundant, meaning that knockout or knockdown of one paralog may be compensated by another, maintaining a wild-type phenotype—for example, the juvenile hormone receptor paralogs *Met* and *gce* are partially redundant, meaning *Met* null alleles show wild-type phenotypes through compensation by *gce* (Abdou *et al.* 2011; Jindra *et al.* 2015). This may sometimes be due to genetic compensation via transcriptional adaptation in mutant organisms, wherein paralogous genes can be up-regulated through a poorly understood, sequence-dependent process that involves the nonsense-mediated decay pathway to restore wild-type function (El-Brolosy *et al.* 2019). Conversely, in situations where paralogs are not redundant and are functionally distinct, RNAi—especially that which uses long

double-stranded RNA molecules to produce siRNA fragments *in vivo*—can risk unintentionally knocking down paralogs with high sequence similarity (Qiu *et al.* 2005). Likewise, CRISPR approaches use guide RNA sequences that can have similar paralog off-target problems if not carefully designed (Fortin *et al.* 2019).

In some cases, the genomic organisation of gene families can facilitate a way to skirt the problem of paralog redundancy—since recently duplicated paralogs are often clustered into tandem arrays, gene deletion strategies that aim to delete multiple paralogs simultaneously can increase the chances of seeing mutant phenotypes (e.g. Wang *et al.* 2018). However, care must be taken to demonstrate that such phenotypes are not due to the disruption of any one particular gene, and so additional single gene knockouts should probably also be performed. Advances in multiplexed gRNA delivery methods are also opening the door for the generation of double, triple, quadruple etc. mutants at unlinked genetic loci (Port & Bullock 2016); such would be required in the case of close paralogs derived from retrotransposition or that have become unlinked due to genomic rearrangements.

1.1.7. Comparative phylogenomics and gene family-phenotype associations

Genetic or molecular biology experiments are not the only way to discover functions of uncharacterised gene families or their members—phylogenetic information can also be used to provide independent evidence of gene family functions, as well as generate hypotheses for other methods to test. Comparative phylogenomics (CPG) is the use of whole genome assemblies to explore the evolution of one or more gene families in relation to trait variation across a particular group of taxa, using phylogenetic comparative methods—phylogenetically informed statistical techniques to detect relationships between traits across species (Cornwell & Nakagawa 2017). For CPG, these traits of interest are genomic traits, typically related to one or more gene families (e.g. gene family sizes, the size of particular clades in a family, or the binary presence or absence of clades, individual genes or even whole gene families), and phenotypic traits related to organismal biology. CPG can be thought of as an extension of 'genome-wide association' methods that aim to associate genetic variation with phenotypic variation within populations, except applied to deeper evolutionary timescales (Dutilh *et al.* 2013).

There are two approaches to CPG, analogous to those in functional genetics: 'forward' CPG and 'reverse' CPG. Forward CPG analyses aim to explain phenotypic differences between taxa by genomic changes—for example, which gene families might explain trophic niche differences between beetles (Seppey *et al.* 2019) or the genomic impacts of the loss of plant-microbe symbioses (Delaux *et al.* 2014). On the other hand, reverse CPG analyses aim to explain genomic differences between taxa by phenotypic changes—for example, the evolutionary trajectories of P450 clans in lepidopterans with differing dietary breadths (Calla *et al.* 2017) or the size of the vertebrate bitter taste receptor family in relation to diet (Li & Zhang 2014). These approaches appear similar in principle but are different in focus: forward analyses are usually genome-wide and 'unbiased', while reverse analyses focus on a defined set of genes, typically just one gene family.

Some CPG analyses are simply descriptive and aim to generate hypotheses, rather than test them, such as a recent analysis of gene gain and loss events in metazoan evolution by Fernańdez & Gabaldón (2020). However, while most forward CPG analyses are framed as ways of discovering the causal genotypic changes behind the evolution of dramatic and/or unique phenotypic transitions, to do this they need to be conducted in a phylogenetically informed way (i.e. through phylogenetic comparative methods). Without the proper application of these methods, CPG is particularly susceptible to treating 'Darwin's scenario'—the single, unreplicated co-occurrence of two traits on a phylogeny—as evidence of association between phenotype and genotype (Maddison & FitzJohn 2015; Uyeda *et al.* 2018). For example, Thomas *et al.* (2020) recently note gene family changes at various transition points in arthropod evolution (such as the evolution of wings or eusociality) but they lack power to determine if these changes are causative or merely correlative. Smaller studies sometimes attempt to use CPG-like framing to explore the biology of a particular organism—this is common in 'genome papers' for single organisms, where unique or notable phenotypic traits are suggested to be explained by large expansions or contractions in particular gene families—for example, the genome papers of the housefly *Musca domestica* (Scott *et al.* 2014), the Chagas vector *Rhodnius proxilus* (Mesquita *et al.* 2015) and the invasive beetle *Anoplophora glabripennis* (McKenna *et al.* 2016) show these species have large numbers of genes in classical detoxification families (typically P450s, GSTs and/or CCEs; see Chapter 1.2.3)—compared to a relatively small number of other

organisms—which is claimed as evidence of adaptation to their ecological niches; but without a robust statistical framework, these are just hypotheses.

Of course, the above comments are not meant to be critiques of these sorts of studies *per se*, just a note about their suitability for answering questions about 'genotype-phenotype maps' and how the results of the analyses should be interpreted. The papers above may lead researchers to functionally test hypotheses about the genes highlighted. Even without a robust CPG approach, there are a number of observations that can be made about gene function in relation to gene family evolution that can lead to testable hypotheses: direct orthologs or closely related paralogs are likely to share similar or identical functions; highly conserved genes across many taxa are likely to have essential functions; and rapidly expanding parts of a gene family are likely to be involved in functions related to environmental, niche-dependent processes (Kawashima & Satta 2014; Tan & Low 2018; Thomas 2007). Ultimately, a combination of CPG and functional genetics is a powerful way to hypothesise about, and test, the functions of gene families, particularly those that are poorly characterised.

1.2. Xenobiotic detoxification in insects

Organisms are exposed to a wide variety of organic and inorganic environmental molecules. Organic compounds of biological origin that adversely affect survival, development or general fitness at physiologically plausible doses are considered toxins. Exposure to toxins is often linked to the ecological niche of an organism (Ibanez *et al.* 2012), and so niche adaptation often requires insects to evolve strategies to deal with toxins, including behavioural changes, excretion of toxins, molecular or physiological changes to the targets of toxins, and metabolic detoxification.

Detoxification (herein also referred to as xenobiotic metabolism, although endogenously produced compounds can also be detoxified) is well studied in humans and other mammals, wherein it is a vital aspect of pharmacology and medicinal chemistry. However, mammals and insects have been diverging for over 600 million years (Reis *et al.* 2015), suggesting the organ systems and genes involved in detoxification may be substantially different between these taxa. Historically, detoxification in insects has been studied to understand two main phenomena: the development of insecticide resistance (Feyereisen 1995), and the chemical ecology of host plant/pest insect interactions (Ibanez *et al.* 2012); these speak to the deep impacts of pest insects on agriculture, as well as the importance of insect vectors in the prevalence of diseases such as malaria. As such, detoxification research has often focused on either agricultural pest species (such as polyphagous caterpillars) or disease vectors (such as mosquitos), often in the context of metabolic resistance to synthetic insecticides (Li *et al.* 2007). Understanding toxin metabolism generally can shed light on insecticide resistance by pest species (Alyokhin & Chen 2017; Li *et al.* 2007; Rane *et al.* 2016), as well as understanding the range of adverse effects they might have on non-target species (Arena & Sgolastra 2014; Claudianos *et al.* 2006; Mao *et al.* 2017). But fundamental biological questions centre on detoxification mechanisms in insects, such as the chemical ecology of plant and insect interactions (Brattsten 1988; Simon *et al.* 2015), how insects deal physiologically with toxic stress (Chahine & O'Donnell 2011), how insects chemically interact with symbionts (De Fine Licht *et al.* 2013) and the role of naturally occurring genetic variation in response to toxic compounds (Green *et al.* 2019; Highfill *et al.* 2017; Najarro *et al.* 2015).

Toxins can be absorbed from the environment passively, introduced into the insect by a competing organism, and/or ingested from food sources. Passive exposure to toxins through the cuticle is a common mode of action for synthetic insecticides, but may be of only limited importance in exposure to natural toxins (Balabanidou *et al.* 2018). The detoxification of dietary toxins is thought to be important for adaptation of many insects to ecological niches that have fuelled bursts of diversification, such as feeding on plants; the evolution of plant secondary metabolism is thought to be partially driven by co-evolutionary 'arms races' between insect herbivores and plant hosts (de Castro *et al.* 2017; Edger *et al.* 2015). An area of currently limited research is the detoxification of venoms from parasitoids and predators, which tend to be introduced by injection, and primarily contain protein and peptide toxins (Arbuckle *et al.* 2017; Danneels *et al.* 2010; Santos-Pinto *et al.* 2018) that are chemically distinct from most insecticidal plant secondary metabolites and are therefore likely to be detoxified by distinct mechanisms.

1.2.1. Detoxification breadth

In order to discuss the concept of detoxification in the broader context of whole-organism biology and ecology, I would like to introduce a novel concept: 'detoxification breadth' (DB). I define DB as an organism's ability to detoxify greater or fewer unique toxins: taxa with a large DB can detoxify many compounds, while species with a small DB can detoxify comparatively few compounds (Fig. 1.1B). This is in contrast to a related, but distinct, concept called 'detoxification capacity' (DC; Fig. 1.1C), which has been used to refer to how quickly a particular quantity of toxin(s) can be detoxified by an organism (e.g. Nobler *et al.* 2018). DB is arguably more useful when discussing the large-scale evolution of detoxification, as it relates to how organisms adapt to new toxins, while DC mainly relates to the optimisation of tolerance towards a particular set of toxins—however, DC may be important for toxicological adaptation in specialists, who can often efficiently metabolise far greater quantities of specific toxins than generalists (Ali & Agrawal 2012; Orr *et al.* 2020).

It may be useful to conceptualise DB as a catabolic counterpart to the anabolic 'chemodiversity' of plant secondary metabolism (Weng *et al.* 2012)—just as some plants can produce a broad range of secondary metabolites (high chemodiversity), some organisms can detoxify a broad range of environmental compounds (large DB), and vice

versa (Fig. 1.1A). Both biosynthetic chemodiversity and DB are derived from complex complements of enzymes, many of which are catalytically promiscuous (Atkins 2020; Weng *et al.* 2012); as such, selection for a larger DB may result in a larger number of detoxification enzymes, or at least a broadening of their overall substrate specificity.

While useful as a concept, estimating DB for a particular species in practice would likely require measuring proxy traits that correlate with the chemodiversity of toxin exposure. While detoxification is not always related to oral exposure to toxins, it is likely the most relevant route for toxin exposure in nature, and therefore an insect's diet strongly impacts its exposure to toxins—as such, a useful proxy for DB may be the number of different diets on which an insect can survive and/or develop, although this may be experimentally confounded by the nutritional value of different plants. For herbivorous insects, this could be measured by the number of plant species they have been observed viably feeding on, either in the wild or in laboratory settings. As closely related plants likely produce similar secondary metabolites, the higher-level taxonomy of a herbivore's hosts may be important to consider—for example, an insect that feeds on 100 species from just one genus in one family likely has a smaller DB than an insect that feeds on 100 species from 50 genera in 10 families. For other consumer groups, estimating DB may be less quantitative. Saprophages and detritivores have relatively unpredictable diets, depending on the microbial species present during feeding and the secondary metabolites they produce. For carnivores and parasitoids, the chemical diversity of their diet, and therefore their DB, depends on the *de novo* defensive, sequestered or otherwise ingested secondary metabolites present in their prey/hosts—for example, lacewing larvae have the capacity to detoxify plant-derived glucosinolates that accumulate in their prey, diamondback moth larvae (Sun *et al.* 2019). In general, one might expect predators of large-DB insects to have a larger DB than predators of small-DB insects, e.g. larval parasitoids of generalist caterpillars might have a larger DB than larval parasitoids of specialist caterpillars. Like herbivores, the prey/host taxonomic breadth of a carnivore or parasitoid may also contribute to DB. DB requirements might also change throughout an insect's lifecycle due to different feeding strategies or diets—for example, mosquito larvae feed on aquatic detritus that may be largely chemically unpredictable, while adult mosquitos feed on blood, which is chemically predictable and lacking in insecticidal secondary metabolites.

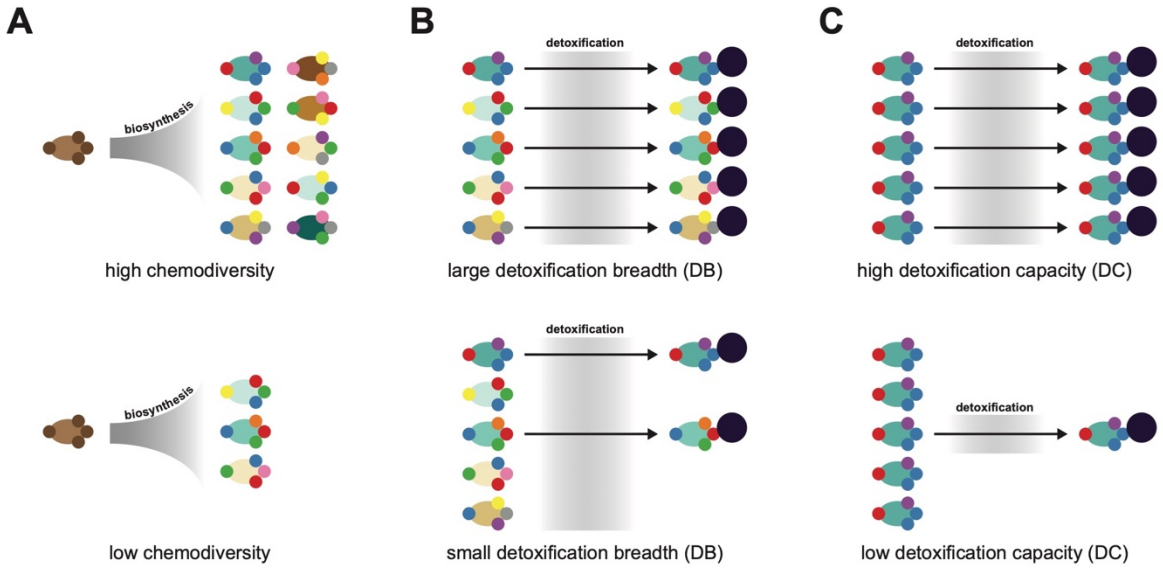


Figure 1.1. A comparison between different levels of (A) biosynthetic chemodiversity, (B) detoxification breadth (DB), and (C) detoxification capacity (DC). Unique chemical species and moieties are represented by different colours. (A) Taxa with high chemodiversity (top) can synthesise a greater number of secondary metabolites from precursor molecules (brown) than taxa with low chemodiversity (bottom). (B) Taxa with large DB (top) can detoxify a greater number of unique xenobiotic compounds than taxa with small DB (bottom). (C) Taxa with high DC for a particular xenobiotic compound (top) can detoxify greater quantities of that compound than taxa with low DC (bottom).

1.2.2. Detoxification mechanisms and phases

A compound is detoxified if it is chemically modified such that its presence no longer harms the organism and/or it can be excreted so rapidly that its effective toxicity is reduced. In order to meet these goals, detoxification is classically divided into two or three 'phases' (Fig. 1.2A; Omiecinski *et al.* 2011; Williams 1951). These phases involve the action of enzymes or transporter proteins that are usually (but not always—see below) encoded by the organism's genome.

Phase I—modification—involves the addition of functional groups to the xenobiotic compound, or cleavage of part of the compound that reveals functional groups, that facilitate the addition of further moieties. Common enzyme families that act in this phase are the cytochrome P450s (P450s), which typically act as monooxygenases, and the carboxylcholinesterases (CCEs), which hydrolyse ester linkages; modification by both enzyme families typically result in the addition or exposure of hydroxyl groups (Oakeshott *et al.* 2005; Omiecinski *et al.* 2011).

Phase II—conjugation—involves the attachment of a bulky, typically highly hydrophilic, molecule to the xenobiotic compound, at the site of phase I modification. The addition of such a moiety can substantially increase the hydrophilicity of many toxins, aiding in their excretion, but can also reduce or prevent the biochemical action of the toxin. Common enzyme families that act in this phase are the glutathione S-transferases (GSTs), which typically catalyse the addition of the tripeptide glutathione, and UDP-glycosyltransferases (UGTs), which catalyse the addition of sugars such as glucose (Wilkinson 1986).

Phase III—excretion (not a metabolic step *per se*)—involves the removal of conjugated xenobiotic compounds from target cells and tissues by transmembrane protein transporters. The most highly studied of these belong to the ABC transporter family, which can transport a wide variety of compounds across plasma membranes (Bretschneider *et al.* 2016; Denecke *et al.* 2017a).

There are some problems with this three-phase model of detoxification: not all compounds need to be modified and/or conjugated before being excreted from target cells and tissues (Fig. 1.2B–C; Chahine & O'Donnell 2011; Torrie *et al.* 2004); direct phase II

conjugation can occur without prior phase I modification (Fig. 1.2D; Ahn *et al.* 2011); some compounds are sequentially conjugated, such that the first conjugation reaction allows for the second (Fig. 1.2E; Boeckler *et al.* 2016); and non-catalytic enzymes can sometimes bind and sequester compounds, reducing their effective toxicity without metabolism occurring *per se* (Fig. 1.2F; Hopkins *et al.* 2017). It is also important to note that xenobiotic metabolism can sometimes result, not in detoxification, but 'toxification' (also called bioactivation or toxication), wherein a compound is converted to a more toxic form (Pirmohamed *et al.* 1994), such in the case of organophosphate insecticides being converted from thionates to oxons by P450s *in vivo*, which markedly increases their ability to inhibit acetylcholinesterase (Siegfried & Scharf 2001). Toxification is mainly studied in insects in the context of exposure to synthetic insecticides, although there are examples involving natural toxins (Zeng *et al.* 2013).

There is an increasing body of evidence that detoxification can be carried out by an insect's gut microbiome before toxins cross the body wall, or by bacterial endosymbionts (Ceja-Navarro *et al.* 2015; Fusetto *et al.* 2017; Hammer & Bowers 2015; Kikuchi *et al.* 2012; Scates *et al.* 2019; van den Bosch & Welte 2017). The relative importance of microbial detoxification compared to the insect's own metabolism is currently unclear, but it likely varies between taxa. Indeed, caterpillars appear to lack a substantial gut microbiome, and elimination of gut microbes in *Manduca sexta*, *Danaus chrysippus* and *Ariadne merione* does not affect caterpillar growth or development (Hammer *et al.* 2017; Phalnikar *et al.* 2019), suggesting microbiome detoxification may play an insignificant role in these species.

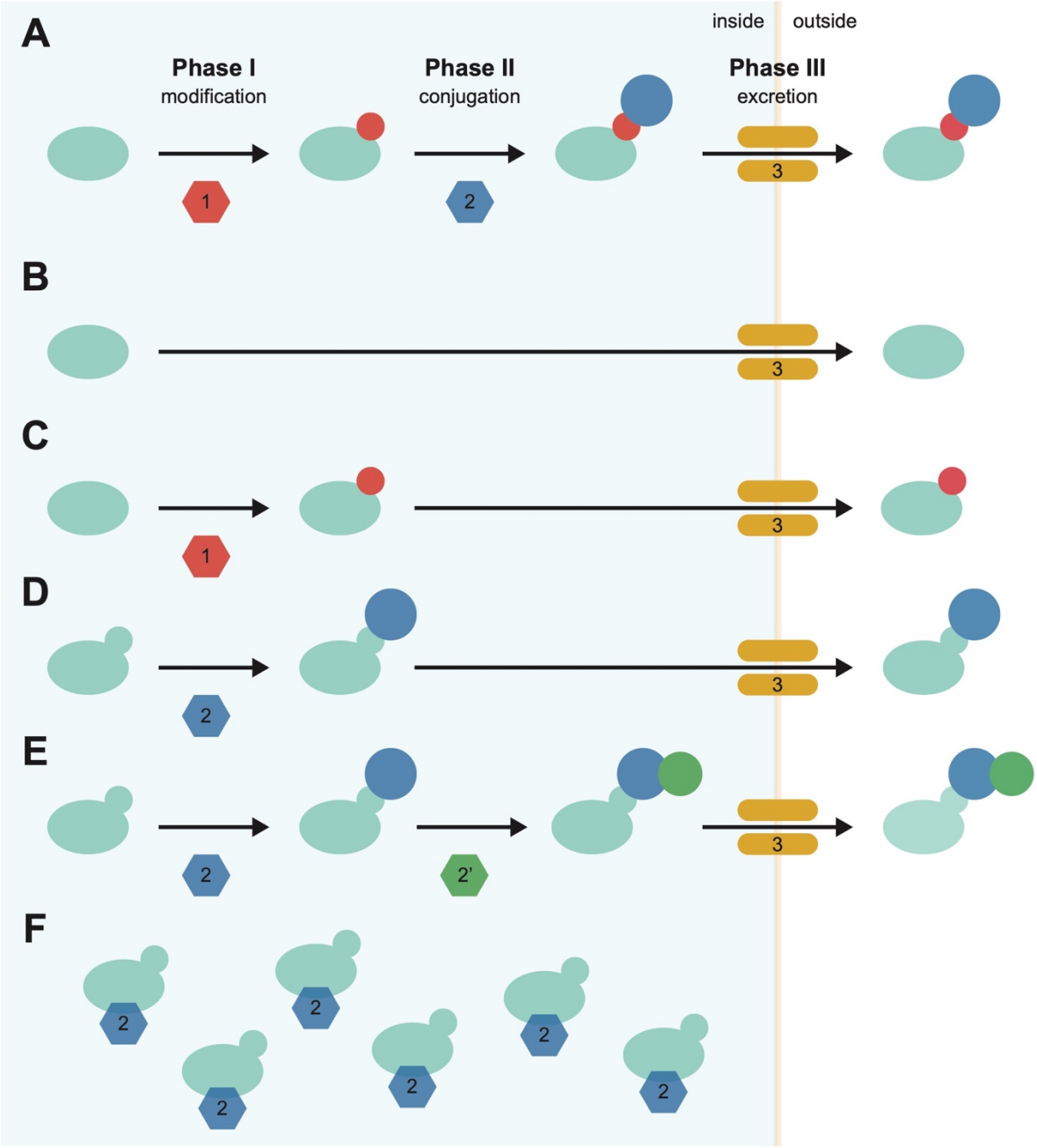


Figure 1.2. Generalised xenobiotic metabolism/detoxification pathways. (A) The classical 'three-phase' model of detoxification, from Williams (1959), involving the sequential action of a modifying enzyme (1; Phase I), a conjugating enzyme (2; Phase II), and an efflux transporter (3; Phase III). (B–E) Alternative pathways for particular xenobiotic compounds, not found in the classical model: (B) excretion without metabolism, (C) modification and excretion without conjugation, (D) conjugation and excretion without modification, (E) multiple conjugation (with an additional enzyme, 2'), and (F) sequestration without metabolism.

1.2.3. Classical detoxification gene families

Despite the overall complexity of detoxification metabolism, it is thought that most genes encoding detoxification enzymes belong to a relatively small group of large gene families (Omiecinski *et al.* 2011). Many such 'classical' detoxification gene families are known in insects, all of which have well-characterised roles in mammalian detoxification; however, not all members of these gene families are involved in detoxification, and many have essential roles in housekeeping (endogenous) metabolism. It is also important to note that not all detoxification enzymes belong to large gene families—prominent examples are the alcohol dehydrogenases, which convert alcohols to aldehydes and ketones (Winberg & McKinley-McKee 1998), nitrile-specifier proteins, which prevent the activation of glucosinolates to toxic isothiocyanates (Fischer *et al.* 2008), and flavin-dependent monooxygenases, which can metabolise some insecticides (Mallott *et al.* 2019; Tian *et al.* 2018).

Probably the most well-known detoxification gene family, the cytochrome P450s (henceforth P450s; Pfam family PF00067) are a family of prototypical phase I enzymes that primarily function as monooxygenases, although some are known to catalyse other reactions, including dealkylation, epoxidation and reduction (Bernhardt 2006). P450s have been linked to the metabolism of a wide variety of natural and synthetic toxins and are thought to dominate 'first-pass' metabolism in insects (Daborn *et al.* 2002; 2007; Denecke *et al.* 2017b; Schuler 2010). However, many insect P450s have essential housekeeping functions, the most notable being the 'Halloween' P450s comprising the biosynthetic pathway of ecdysteroid moulting hormones (Gilbert 2004); other insect P450s may have roles in secondary metabolism and pheromone degradation (Beran *et al.* 2019; Wojtasek & Leal 1999).

Carboxylcholinesterases (CCEs, also known as carboxylesterases; Pfam family PF00135) hydrolyse the ester linkage between carboxylic acids and alcohols and are generally considered phase I enzymes. They are most well studied for their role in catabolism of organophosphate and pyrethroid insecticides (see references in Oakeshott *et al.* 2005), but it is unclear if CCEs act on natural toxins. CCEs with housekeeping functions include acetylcholinesterase, which catabolises the neurotransmitter acetylcholine to allow proper muscular contraction (and is the target of organophosphate insecticides; Fournier *et al.* 1995), and juvenile hormone esterases, which

catabolise the sesquiterpenoid juvenile hormones that are essential for insect development and physiology (Kamita & Hammock 2010).

Glutathione S-transferases (GSTs; Pfam families PF00043 and PF02798) are typically phase II enzymes that conjugate toxins with the tripeptide glutathione (Armstrong 2002), but some catalyse other reactions, such as dechlorination (Low *et al.* 2010). GSTs are involved in the detoxification of insecticides (Enayati *et al.* 2005) as well as natural toxins (Shabab *et al.* 2014), and also contribute to the antioxidant response to oxidative stress (Enayati *et al.* 2005; Hu *et al.* 2019a). Housekeeping functions of GSTs include ecdysteroid synthesis (Chanut-Delalande *et al.* 2014; Enya *et al.* 2014), pigment biosynthesis (Kim *et al.* 2006) and the regulation of energy metabolism (Kim *et al.* 2012).

UDP-glycosyltransferases (UGTs; Pfam family PF00201) are phase II enzymes well studied in mammalian detoxification, where they conjugate glucuronic acid to a wide variety of xenobiotic compounds, but in insects, UGTs typically transfer glucose groups. Insect UGTs are involved in the conjugation of gossypol (Krempl *et al.* 2016b) and capsaicin (Ahn *et al.* 2011) in lepidopteran species, and appear to be able to affect susceptibility to nicotine in *D. melanogaster* through an unknown biochemical reaction (Highfill *et al.* 2017). In additional, UGTs have been associated with resistance to various insecticides (neonicotinoids, ryanoids and pyrethroids) in many different insects, although the enzymatic reactions they catalyse are typically not known (Kaplanoglu *et al.* 2017; Li *et al.* 2017b; Pan *et al.* 2018; Tian *et al.* 2019; Zhao *et al.* 2019; Zhou *et al.* 2019). Housekeeping functions for UGTs in insects include the conjugation of ecdysteroids and tyrosine (Thompson *et al.* 1987a). Curiously, an ecdysteroid UGT encoded in the genome of baculoviruses allows them to control the development and behaviour of its lepidopteran hosts (Hoover *et al.* 2011), and may have been acquired by horizontal gene transfer from an insect host (Hughes 2013).

ATP-binding cassette (ABC) transporters (Pfam family PF00005) are transmembrane proteins that actively transport small molecules through the plasma membrane, and are the dominant gene family involved in phase III detoxification (Dermauw & Van Leeuwen 2014; Wu *et al.* 2019a). ABC transporters can transport both modified and unmodified xenobiotic compounds out of target cells where they exert their toxic effects (e.g. neurons in the CNS for many synthetic insecticides) and into cells where xenobiotic metabolism takes place (e.g. in the Malpighian tubules; Denecke *et al.*

2017a; O'Donnell 2009). However, while the interaction between ABC transporters and conjugation reactions is thought to be important, it is understudied (Chahine & O'Donnell 2011).

1.2.4. The evolution of detoxification gene families

Individual detoxification enzymes lie somewhere on a spectrum between two theoretical endpoints: highly promiscuous (generalist; able to catabolise virtually any substrate with the correct functional group/s) and highly specialised (only able to catabolise one distinct molecular species). Most enzymes probably lie in the middle between these extremes and are limited to groups of substrates with similar structures—highly generalist detoxification enzymes exist in some insects, such as Cyp6g1 in *Drosophila melanogaster*, which can metabolise multiple chemically distinct insecticides (Daborn *et al.* 2007; Fusetto *et al.* 2017; Joußen *et al.* 2008; Schmidt *et al.* 2010), but many other detoxification enzymes have much more limited substrate specificities (see Daborn *et al.* 2007; Krempl *et al.* 2016a; Wang *et al.* 2018).

If most detoxification enzymes are limited in their substrate specificity, and if most belong to a relatively small number of large gene families, it follows that evolving an increase in DB typically requires increasing the size of these detoxification gene families. This has only recently begun to be studied empirically using the wide variety of sequenced insect genomes, and although most studies have not conducted this work within a robust phylogenetic comparative framework, there is some evidence suggesting an association between the sizes of classical detoxification gene families and dietary complexity (Calla *et al.* 2017; Edger *et al.* 2015; Rane *et al.* 2016; 2019). The co-evolution of size between detoxification gene families has not been robustly investigated either, and although Rane *et al.* (2019) note that the number of P450s, CCEs and GSTs in their dataset of 160 insect species have significant correlations (r) between 0.23 and 0.25, this may be due to the phylogenetic non-independence of the underlying data.

While increasing DB can occur by changes in gene expression (Daborn *et al.* 2007; Denecke *et al.* 2017b), it is mostly thought to take place by changes in the amino acid sequence of detoxification enzymes. There are three main scenarios for this: modification of an existing enzyme (or transporter) away from its ancestral function (or in addition to its ancestral function) and towards a novel detoxification function;

neofunctionalisation after gene duplication (which produces a new gene with the ability to detoxify the toxin; Francino 2005; Zimmer *et al.* 2018); and gene fusion (which generates a chimeric gene that contains sequence from two different genes; Battlay *et al.* 2018; Joußen *et al.* 2012; Rogers & Hartl 2011). Modification without duplication creates enzymatic novelty but can remove the ancestral function from the genome (except where alleles are functionally diverged; see below), and so is not always the most likely scenario when DB needs to increase overall, not simply rerouted to a different set of toxins. Neofunctionalisation after duplication resolves this problem by allowing the ancestral function of the gene to persist in the genome; rarely, however, both genes can evolve towards a novel function simultaneously in a process called 'concerted neofunctionalization' (Heidel-Fischer *et al.* 2019). Gene fusion can also create enzymatic novelty by creating a completely new active site that is able to metabolise the toxin, but will result in the loss of the ancestral functions of both progenitor genes unless both progenitors are new (redundant) duplicates themselves and/or the chimeric gene has overlapping substrate specificity with its progenitors.

The evolution of new detoxification functions requires innovation; in many models of neofunctionalisation, this is derived from the slight promiscuity of an ancestral enzyme: minor functions (an inefficient side-reaction or substrate) are reinforced and optimised over time by step-by-step changes to the amino acid sequence, eventually resulting in a 'fully functional' new enzyme. The 'escape from adaptive conflict' model of neofunctionalisation can also apply in these situations, where multiple beneficial alleles of a gene are present at the same time due to different innovative modifications, resulting in duplication to fix both copies of the gene in the genome at the same time (Marais & Rausher 2008), which has also been called 'permanent heterozygosity' (Remnant *et al.* 2013; Spofford 1969); in nature, blowflies have become resistant to two different organophosphate insecticides by fixing two alleles of the same esterase via duplication (Newcomb *et al.* 2005; Smyth *et al.* 2000). Consistent with a key role for duplication and neofunctionalisation in the evolution of detoxification gene families, standing structural variation is common in detoxification gene families in *Drosophila melanogaster* (Chakraborty *et al.* 2018; Good *et al.* 2014), and alleles derived by duplication are commonly associated with insecticide resistance (Battlay *et al.* 2018; Chakraborty *et al.* 2018; 2019; Schmidt *et al.* 2010; Schmidt & Robin 2011; Zimmer *et al.* 2018), although this may often be due to initial selection for increased gene dosage.

The evolution of P450s, GSTs, CCEs, UGTs and ABC transporters has been studied in various insect taxa (Ahn *et al.* 2012; e.g. Good *et al.* 2014; Labbé *et al.* 2010; Low *et al.* 2007; Robin *et al.* 2009), in general showing that these families evolve rapidly, with frequent gene gain and loss throughout parts of the family, consistent with a birth-and-death model of evolution (Nei & Rooney 2005). P450 family evolution has received particular attention, especially in the *Drosophila* genus, where the functions of many genes in the family are known (Chung *et al.* 2009; Good *et al.* 2014). *Drosophila* P450s belong to 77 ancestral clades, over half of which are 'unstable', in that they have experienced gene gain or loss during the evolution of the genus, while the rest are 'stable' and have 1:1 orthologs in all 12 species; this appears roughly consistent with known functions of the genes—stable genes tend to have essential functions, while unstable genes tend to be involved in detoxification (Good *et al.* 2014).

1.2.5. Tissues and organs involved in detoxification

Like in other animals, xenobiotic metabolism in insects is localised to specific tissues and organs, typically thought to be the midgut, the Malpighian tubules and the fat body (Yang *et al.* 2007). The midgut—the primary site of digestion in the insect—is also considered an important detoxification tissue, and transcriptomic surveys of insect midguts consistently detect many genes from classical detoxification families (Gazara *et al.* 2017; Harrop *et al.* 2014; Pauchet *et al.* 2009b; 2010). The Malpighian tubules (MTs) are blind-ended tubes attached to the pyloric valve separating the midgut and hindgut, and control osmoregulation and the conversion of nitrogenous waste to uric acid. Like the midgut, MTs express many detoxification genes (Dow *et al.* 2018; Yang *et al.* 2007), and their detoxification functions are linked to their additional role in excretion: toxins and their metabolites move from the hemolymph to the MT lumen, then to the hindgut lumen for expulsion in the faeces (Chahine & O'Donnell 2011; 2009; Maddrell & Gardiner 1976). The fat body is a lipid storage organ that is also thought to play a role in detoxification, although it has more clearly defined roles in development, energy metabolism, lipid homeostasis and reproduction (Li *et al.* 2019). A number of P450s with known roles in detoxification are expressed the *Drosophila melanogaster* fat body (Chung *et al.* 2009), and P450-mediated metabolism of furanocoumarin may occur in the fat body of the caterpillar *Papilio polyxenes* (Petersen *et al.* 2001). However, little to no work appears to have been done unequivocally demonstrating that the fat body is an important site of xenobiotic metabolism in insects.

An under-appreciated detoxification tissue in insects is the blood-brain barrier (BBB)—a layer of glia that encapsulates the CNS and protects neurons from toxins present in the haemolymph (Schirmeier & Klämbt 2015). The *D. melanogaster* BBB is enriched for the expression of a large number of genes from classical detoxification gene families (DeSalvo *et al.* 2014; Hindle & Bainton 2014), and xenobiotic metabolism and efflux has been observed in the BBB of the locust *Schistocerca gregaria* (Hellman *et al.* 2016), suggesting the BBB's role in detoxification is conserved across insects. However, the contribution of BBB-specific detoxification to whole-body xenobiotic metabolism is unknown, and it likely plays only a minor role towards the total detoxification capacity of individual insects.

As insects have an open circulatory system, it is possible that many other tissues may metabolise toxins to some degree, to protect themselves against toxins that fail to be fully detoxified after passing from the midgut into the haemolymph. Odorant-degrading enzymes may also be related to detoxification genes, as indicated by the expression of detoxification gene family members in the sensory organs of some insects (Gu *et al.* 2015; Younus *et al.* 2014). Insecticide-metabolising P450s are present in the nervous systems of some pyrethroid-resistant insects (Korytko & Scott 1998; Zhu *et al.* 2010), but it is unclear if similar enzymes—outside of the BBB—protect the brain against natural toxins.

1.2.6. Transcriptional regulation of detoxification genes

Given that detoxification genes are involved with responding to toxins, it makes sense that control of their expression may be regulated by the presence or absence of toxins (Terriere 1984). Dietary shifts can result in large-scale transcriptomic changes in insects, some of which are linked to plastic responses to new dietary toxins (Birnbaum *et al.* 2017; Hoang *et al.* 2015; Koenig *et al.* 2015; Yu *et al.* 2016). This is also true for exposure to individual toxins, but many toxins induce far more genes than might be plausibly involved in their detoxification (Coelho *et al.* 2015; Willoughby *et al.* 2006), suggesting the transcriptional response to these toxins involves general pathways that co-regulate many detoxification genes, as well as genes that function in repairing physiological damage caused by toxin exposure.

The most well-characterised general detoxification-response pathway in insects signals through the transcription factor cap-n-collar isoform-C (CncC), an ortholog of mammalian Nrf2. CncC, along with its binding partner Maf, activates detoxification gene expression in detoxification tissues by binding to their promoters and driving transcription (Deng & Kerppola 2014; Misra *et al.* 2011). CncC is negatively regulated by Keap1, overexpression of which is sufficient to cause loss of CncC signalling, even in the presence of activating toxins; suppression of Keap1 by RNAi in the midgut, Malpighian tubules or fat body is sufficient to reduce susceptibility to the organophosphate insecticide malathion (Misra *et al.* 2011). Constitutive expression of CncC has been associated with resistance to a number of other synthetic insecticides (Green *et al.* 2019; Hu *et al.* 2019b; Misra *et al.* 2013; Wilding 2018). Importantly, CncC does not appear to regulate all detoxification genes (Misra *et al.* 2011) and the genes it does regulate are not restricted to detoxification, as it also has roles in development (Deng 2014; Deng & Kerppola 2013). This is explained by the fact that CncC is proximally responsive to the presence of reactive oxygen species (ROS), which are produced by the damage caused by some toxins and result in oxidative stress, and ROS also plays a role in developmental signalling in insects (Andrew Pitoniak 2015; Brock *et al.* 2017; Hu *et al.* 2019a; Sykiotis & Bohmann 2008). As such, CncC is not responsive to specific toxins *per se*, but rather their biochemical consequences; CncC is clearly a 'big tent' response pathway that up-regulates a large number of genes in a scattershot way, to deal with both the consequences of high ROS levels and also the cause of the ROS: the toxins. Causal links between toxins that induce the CncC pathway and specific genes induced by CncC are therefore hard to demonstrate without more direct experimental evidence (e.g. transgenic manipulation).

The nuclear receptor DHR96, an ortholog of the mammalian aryl hydrocarbon receptor (AHR), is also thought to be a detoxification-response transcription factor (Afschar *et al.* 2016; King-Jones *et al.* 2006). Like CncC, DHR96 also regulates developmentally important genes, and is involved in cholesterol and lipid homeostasis (Bujold *et al.* 2010; Horner *et al.* 2009; Sieber & Thummel 2012). The function of DHR96 in detoxification has only been studied in *Drosophila melanogaster*, and so its importance in the regulation of detoxification genes in insects more broadly is unknown.

There are likely more pathways—as yet undiscovered—that respond to dietary toxins in insects; these are likely induced by a narrower set of xenobiotic compounds, which

would explain the transcriptional plasticity seen in generalist herbivores, which differentially express detoxification genes when feeding on different host plants. It is unclear how such pathways would operate in specialist insects, which exhibit much more limited transcriptional plasticity (reviewed in Birnbaum & Abbot 2020), and it is unknown if specialists have evolved novel toxin-response transcriptional pathways or if they have co-opted ancestral pathways to focus on their niche-defining toxins.

1.2.7. Identifying and characterising detoxification genes

The identification of detoxification genes is key to understanding how detoxification breadth (DB; Fig. 1.1B) differs between insect taxa. For many decades, detoxification reactions have been identified by the presence of specific metabolites, with specific enzyme families implicated by the use of chemical inhibitors, such as piperonyl butoxide (PBO) for P450s, S,S,S-tributyl phosphorotrithioate (DEF) for CCEs and diethyl maleate (DEM) for GSTs (Casida 1970; Pasay *et al.* 2009; Sanchez-Arroyo *et al.* 2001). In some cases, isozymes can be distinguished from one another by native gel electrophoresis, and specific genes cloned if the particular proteins involved are identified (e.g. Field *et al.* 1993). Gene expression data can also be used to find candidate genes that may underpin a particular detoxification phenotype (Daborn *et al.* 2002; Puinean *et al.* 2010). This approach could be considered a 'forward' approach in the style of 'forward genetics', whereby a detoxification reaction leads to a specific gene/enzyme of interest being identified and characterised.

In the modern, post-genomic research landscape, where many detoxification gene families are known, there are a great deal many genes in any particular insect's genome that could be involved in detoxification processes yet have not been associated with any particular detoxification reaction or toxicological phenomenon. In addition, given the fact that not all members of detoxification families are involved in detoxification, assumptions that any gene in, for example, the P450 gene family is a detoxification gene may not be valid. Regardless, there now exist an abundance of 'orphan' detoxification genes that have not been associated with a particular toxin or set of toxins. This makes the identification of true detoxification genes in general, as well as specifically (tying them to specific detoxification reactions), an important part of toxicological genomics.

Detoxification genes can be identified through direct or indirect lines of evidence. Indirect evidence of detoxification function comes from a variety of sources, all of which relate to general properties that have been associated with detoxification genes: significant expression in the 'detoxification tissues'—midgut, Malpighian tubules and fat body (Yang *et al.* 2007); transcriptional regulation by detoxification-response pathways such as CncC/Keap1 (Misra *et al.* 2011); transcriptional induction by toxins (Willoughby *et al.* 2006); and evolutionary instability—a demonstrated signal of detoxification function in large gene families such as the P450s and GSTs in insects and mammals (Kawashima & Satta 2014; Tan & Low 2018; Thomas 2007). However, each type of indirect detoxification evidence is unlikely to be completely diagnostic—for example, not every unstable gene is a detoxification gene, and not every stable gene is *not* a detoxification gene—and there have yet to be systematic attempts at combining these lines of evidence together; this will be a focus of Chapter 2 of this thesis.

Ultimately, the best way to determine whether a gene is involved in detoxification is through direct evidence from a combination of genetic and biochemical experiments. Toxin susceptibility upon gene knockout and/or toxin resistance upon gene overexpression, combined with substrate kinetics assays on recombinant enzymes, is powerful evidence for detoxification functions (see Denecke *et al.* 2017b; Fusetto *et al.* 2017; Joußen *et al.* 2008). However, these experiments can be difficult to perform, especially in non-model insects. Advances in transgenic manipulation of a wide variety of insects is likely to improve the tractability of robustly identifying detoxification genes in many different insects, not just *Drosophila melanogaster*.

1.2.8. Detoxification by phosphorylation

Phosphorylation—the addition of a phosphate moiety, typically to a hydroxyl group—is a crucial biochemical reaction in nature that plays major roles in primary metabolism and intracellular signalling. Due to the high polarity of phosphate moieties, phosphorylation also has the potential to be a phase II-style detoxification reaction—despite this, it is rarely a component of xenobiotic metabolism in animals (Mitchell 2015). The exception to this is in insects, where a body of evidence for detoxicative phosphorylation has been slowly growing since the 1960s. Phosphorylated metabolites of exogenous phenolic, steroidal and glycosidic compounds (e.g. Boeckler *et al.* 2016; Ngah & Smith 1983; Rharrabe *et al.* 2007; Smith & Turbert 1964), as well as of

hydroxylated and glycosylated drug metabolites (Olsen *et al.* 2014; 2015), have been identified in a broad range of insect species (Fig. 1.3; for more detail, see Chapter 2.1), suggesting phosphorylation is a general detoxification reaction across all insects. To date, however, no detoxicative kinases have been cloned or otherwise identified at the genetic level, and so the gene family (or families) responsible is not known. Attempting to solve this mystery is one of the aims of this thesis.

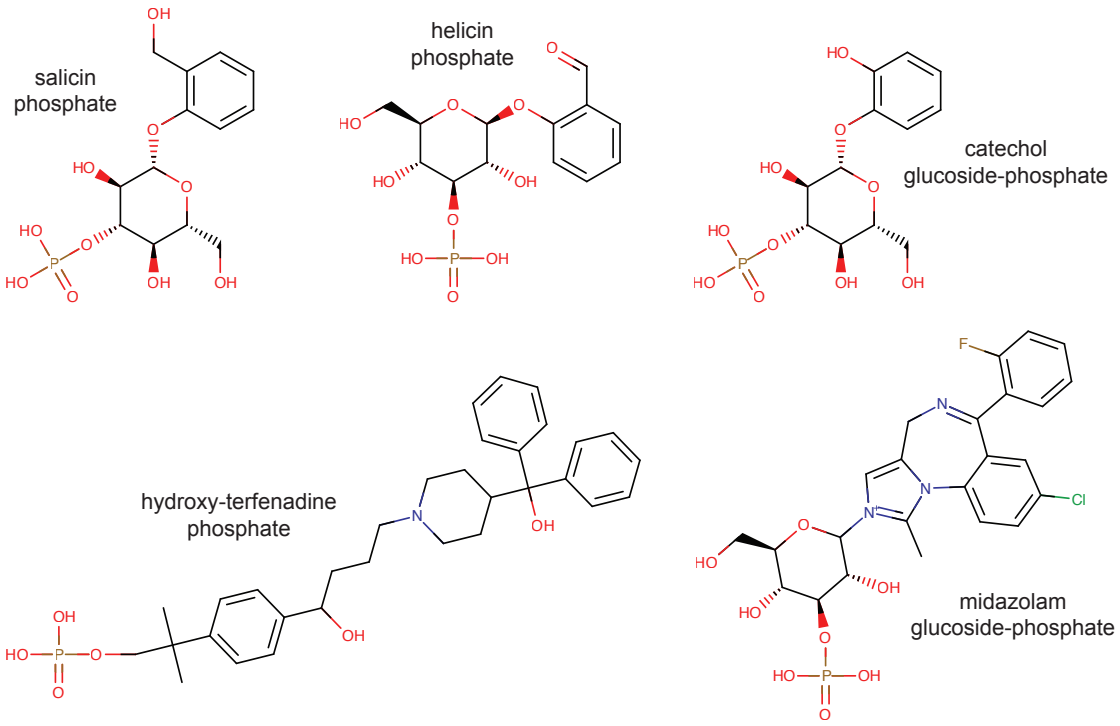


Figure 1.3. Examples of phosphate metabolites of xenobiotic compounds isolated from insects: salicin phosphate, helicin phosphate and catechol glucoside-phosphate (top) from *Lymantria dispar* (Lepidoptera: Erebidae; Boeckler *et al.* 2016), and hydroxy-terfenadine phosphate and midazolam glucoside-phosphate (bottom) from *Schistocerca gregaria* (Orthoptera: Acrididae; Olsen *et al.* 2015; 2014). The exact phosphate position on midazolam glucoside-phosphate is unknown; shown is a possible structure.

1.3 Ecdysteroid hormones

Ecdysteroids are polyhydroxylated steroid hormones that control various aspects of arthropod and insect biology, and are synthesised from dietary sterol precursors (Carvalho *et al.* 2010; Feldlaufer *et al.* 1995). 20-hydroxyecdysone (20E) is typically considered the dominant active ecdysteroid in most insects, while ecdysone (E) is usually considered a weakly active prohormone (e.g. Parvy *et al.* 2014). Confusingly, 20E is frequently referred to simply as 'ecdysone' in research focusing on higher-level aspects of ecdysteroid-controlled biology (e.g. Gautam *et al.* 2015; Tan *et al.* 2014). In addition, while many aspects of ecdysteroid biochemistry remain to be discovered, it is clear that 20E is not the only main ecdysteroid in most insects—other 20-hydroxylated ecdysteroids, such as makisterone A and 24,28-dehydromakisterone A, can be synthesised depending on available dietary sterols and appear to fulfil the same molecular and physiological functions in insect biology (Baker *et al.* 2000; Harmatha & Dinan 1997; Lavrynenko *et al.* 2015). Likewise, the 20-deoxy precursors to these hormones—24-methyl-E and 24-28-dehydromethyl-E—may act in the same manner as E (Lavrynenko *et al.* 2015). As such, in the remainder of this document I will refer to the E-like precursors as 'Es', and the 20E-like 'main' hormones as '20Es'—usage of the singular 'E' and '20E' will refer to the specific compounds ecdysone and 20-hydroxyecdysone, respectively.

Ecdysteroids are essential for a large number of important processes in insects, the most well studied of which is development, from which they derive their moniker of 'moulting hormones' (Carlisle & Jenkin 1959; Yamanaka *et al.* 2013). As implied, pulses of ecdysteroids are required for the moulting process that transitions insects between developmental stages (Truman 2019), but also for other developmental checkpoints, such as minimum viable weight and critical weight (Mirth *et al.* 2009; Ono 2014). Ecdysteroids are also required for morphogenesis and cuticle deposition during embryonic development (Kozlova & Thummel 2003), a role that led to the discovery of the ecdysteroid biosynthetic pathway through the characterisation of the 'Halloween' mutants, which fail to produce embryonic cuticle and are rendered 'ghostly' (reviewed in Gilbert 2004). The action of ecdysteroids during metamorphosis in holometabolous insects is also well studied, where they direct the timing and initiation of larval wandering, pupariation and pupation (Danielsen *et al.* 2016; Rewitz *et al.*

2010; Zirin *et al.* 2013). The seasonal polyphenism observed in some butterflies is also controlled by the ecdysteroid titre during metamorphosis (Oostra *et al.* 2015; Rountree & Nijhout 1995). Ecdysteroids are also crucially important for reproduction, where they control multiple processes in oogenesis, including stem cell maintenance and proliferation, cyst differentiation, follicle formation, chorion formation and ovulation (Buszczak *et al.* 1999; Domanitskaya *et al.* 2014; Gaziova *et al.* 2004; Ge *et al.* 2015; Knapp & Sun 2017; Swevers 2019). Ecdysteroids also mediate the reproductive response to starvation in females, with high levels of 20Es leading to the apoptosis of mid-stage egg chambers and a cessation of fertility (Terashima *et al.* 2005); however, the reproductive roles of ecdysteroids in male insects are currently less clear (Hentze *et al.* 2013). Physiological processes controlled by ecdysteroids include the functioning of the Malpighian tubules (Gautam *et al.* 2015; Gautam & Tapadia 2010) as well as regular circadian rhythm (Oostra *et al.* 2015). Insect immunity also appears to be linked to ecdysteroids (Regan *et al.* 2013; Rus *et al.* 2013; Tan *et al.* 2014; Verma & Tapadia 2015), as are various aspects of behaviour and neurobiology, including sleep, feeding and memory (Hindle *et al.* 2017; Ishimoto *et al.* 2009; Wang *et al.* 2010).

1.3.1. Biosynthesis of ecdysteroids

The biosynthesis of ecdysteroids (ecdysteroidogenesis) is an essential and highly conserved process in insects that converts sterols into active ecdysteroids (Fig. 1.4). Insects are sterol auxotrophs and therefore must obtain sterols from their diet (Carvalho *et al.* 2010; Hobson 1935a; 1935b); the specific sterols ingested determine which 20-hydroxylated ecdysteroids (20Es) end up synthesised, the most common of which are 20E (from cholesterol, an animal C_{27} sterol) and makisterone A (from plant and yeast C_{28} sterols), although in the model dipteran, *Drosophila melanogaster*, cholesterol appears to be preferentially utilised over other sterols, even at trace concentrations in experimental media (Lavrynenko *et al.* 2015), likely resulting in its dominance in most published ecdysteroid isolation experiments.

In most insects, ecdysteroidogenesis is divided into two parts: the biosynthesis of Es from sterols (which I will call 'early' ecdysteroidogenesis; EE) and the activation of Es into 20Es by an ecdysteroid 20-monooxygenase ('late' ecdysteroidogenesis; LE). Curiously, in some insects, and most lepidopterans, the primary end products of EE are 3-oxoecdysteroids (e.g. 3-oxo-E), which are converted to Es in the haemolymph by 3-

oxoecdysteroid 3β-reductases (Kiriishi *et al.* 1990). EE takes place in different tissues at different life stages. The most well-studied EE tissue is the prothoracic gland (PG), which in the dipteran suborder Brachycera is fused with other endocrine tissues (the corpora allatum and corpora cardiacum) into the ring gland (RG); the PG/RG is the dominant site of ecdysteroidogenesis in larval and prepupal holometabolous insects (Redfern 1983). In *D. melanogaster*, the RG degrades throughout metamorphosis, suggesting there is another source of ecdysteroids during the pupal-adult transition (Dai & Gilbert 1991). In embryos, ecdysteroids may be produced by the hydrolysis of inactive conjugates in some species (Sonobe & Ito 2009), but EE occurs in the developing PG/RG (as in larvae and prepupae), while LE takes place in the epidermis (Petryk *et al.* 2003). Larval and prepupal EE occurs in the PG/RG (Yamanaka *et al.* 2013), while adult EE occurs in the ovary in females and in the accessory glands in males (Ameku *et al.* 2017; Hentze *et al.* 2013). LE (20-monooxygenation) occurs in peripheral target tissues of 20Es, including the midgut and fat body, in pre-adult stages (Akagi *et al.* 2016; Petryk *et al.* 2003), as well as in the adult ovary and testis (Ameku *et al.* 2017; Hentze *et al.* 2013).

The enzymes that catalyse the biosynthesis of ecdysteroids were first identified in the Heidelberg Screen for larval patterning mutants in *D. melanogaster* (Jurgens *et al.* 1984; Nüsslein-Volhard *et al.* 1984; Wieschaus *et al.* 1984), which identified a group of 'Halloween' mutants that lacked cuticle and failed to complete embryogenesis (Wieschaus & Nüsslein-Volhard 2016). Many of these mutants have been mapped to genes that encode ecdysteroidogenic enzymes, which have subsequently been deemed the 'Halloween genes' (Gilbert 2004; Rewitz *et al.* 2006). While the classic Halloween phenotype involves defects in embryogenesis, disruption of Halloween gene function also affects ecdysteroidogenesis at other life stages (with one exception), strongly suggesting that the same biochemical pathway is present in all ecdysteroidogenic tissues. The ecdysteroidogenic genes/enzymes currently identified are *shroud* (*sro*), *spook*/*spookier* (*spo*/*spok*), *phantom* (*phm*), *disembodied* (*dib*), *shadow* (*sad*) and *shade* (*shd*), along with a two additional genes—*neverland* (*nvd*) and *noppera-bo* (*nobo*)—that are known to have roles in ecdysteroidogenesis but were not identified in the Heidelberg Screen (Enya *et al.* 2014; 2015; Škerlová *et al.* 2020; Yoshiyama *et al.* 2006); most of these, except *nvd*, *sro* and *nobo*, are P450s that catalyse hydroxylation reactions. *spo*, *spok* and *sro* act in the 'Black Box', a collection of unknown reactions that convert 7-dehydrogenated sterols to 5β-ketodiol and have yet to be fully characterised, although they may involve 3-

oxo-ecdysteroids (Ono *et al.* 2012; Saito *et al.* 2016). *spo* and *spok* are paralogs in *D. melanogaster* that are stage-specific—*spo* is expressed in the yolk nuclei and amnioserosa of early embryos, as well as in oocytes in the adult ovary, while *spok* is expressed in the RG of late embryos, larvae and prepupae—although they likely catalyse the same reaction in the ecdysteroidogenic pathway (Ono *et al.* 2006; Sztal *et al.* 2007).

As ecdysteroids need to be secreted in discrete pulses, their biosynthesis is tightly temporally regulated, mostly through negative feedback loops (Moeller *et al.* 2013; Parvy *et al.* 2014; Sakurai & Williams 1989). Many external signalling pathways also control ecdysteroidogenesis, including the insulin, prothoracicotropic hormone (PTTH), nitric oxide (NO), Activin, TOR and EGFR pathways, which integrate nutritional and environmental information to control developmental progression and timing through the production of ecdysteroids (Cruz *et al.* 2020; Yamanaka *et al.* 2013).

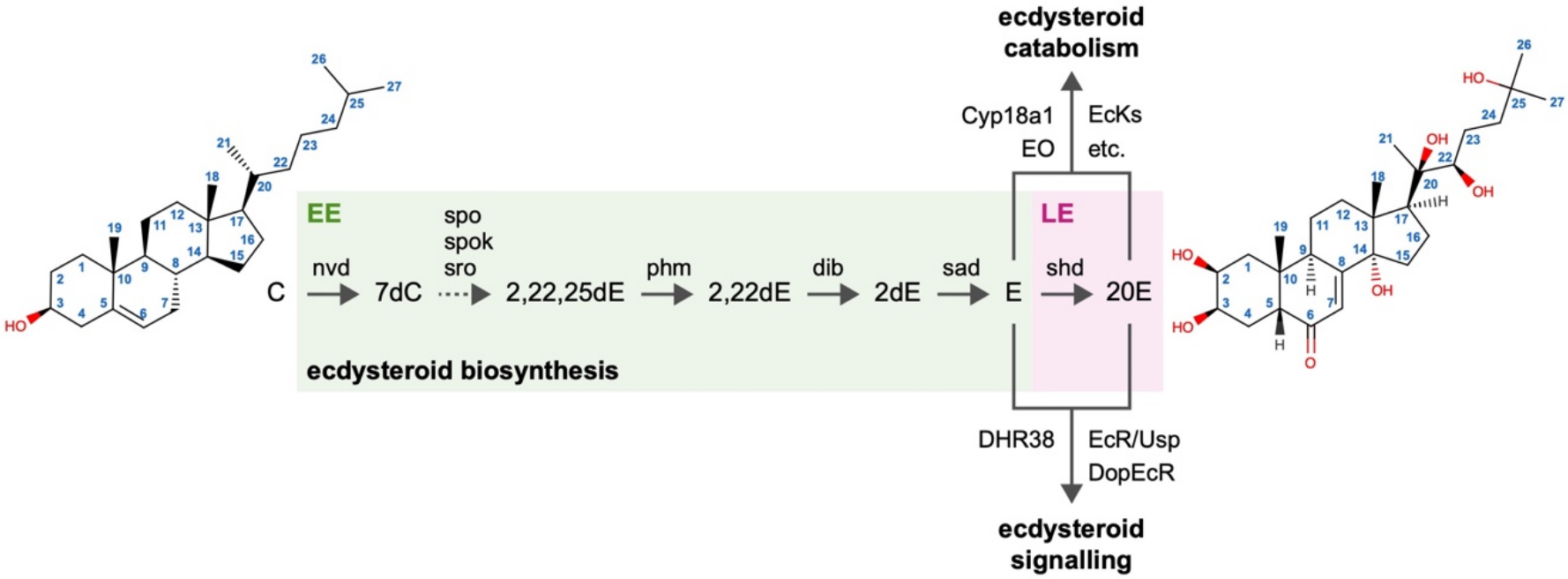


Figure 1.4. Diagram of generalised ecdysteroid biosynthesis from the C_{27} animal sterol cholesterol (C; left) to 20-hydroxyecdysone (20E; right) via the Halloween gene products, in *Drosophila melanogaster*, a species where early ecdysteroidogenesis (EE; green) in the prothoracic gland produces ecdysone (E), not 3-oxo-E as in some insects; late ecdysteroidogenesis (LE; pink) occurs in various target tissues of ecdysteroids. Blue numbers are the carbon locator numbers (C-1, C-2, C-3, etc.) of the sterol backbone. 7dC, 7-dehydrocholesterol; 2,22,25dE, ketodiol; 2,22dE, ketotriol; 2dE, 2-deoxyecdysone; EO, ecdysone oxidase; EcKs, ecdysteroid kinases. The 'Black Box' (dashed arrow), in which various unknown reactions take place, is denoted by a dashed arrow. The Halloween gene product nobo, which may act as a steroid isomerase (Škerlová *et al.* 2020), is not shown, as its precise function in the biosynthetic pathway is not clear.

1.3.2. Ecdysteroid signalling

Ecdysteroids exert their biological effects through a number of receptors, the best studied of which is the heterodimer of the nuclear receptor Ecdysone Receptor (EcR) and its partner Ultraspiracle (Usp), which is primarily activated by binding 20Es (Baker *et al.* 2000; Wang *et al.* 2000). EcR/Usp binds to the promoters of 20E-responsive genes and activates them in the presence of an activating ligand, otherwise repressing their expression (Cherbas *et al.* 2003; Hu *et al.* 2003; Uyehara & McKay 2019). The nuclear receptor DHR38 also appears to be an ecdysteroid-sensing receptor, although it does not have a ligand-binding pocket and likely mediates its effects indirectly through an unknown binding partner with ecdysteroid-binding capacity (Baker *et al.* 2003). Like vertebrate steroids (Wilkenfeld *et al.* 2018), ecdysteroids can also activate rapid, non-genomic signalling; this occurs through the G-protein coupled receptor DopEcR, which is responsible for many of the connections between ecdysteroids and insect behaviour (Evans *et al.* 2014; Ishimoto *et al.* 2012; Petruccelli *et al.* 2020; 2016).

20Es are typically considered the dominant active ecdysteroids for the reason that most ecdysteroid-dependent developmental events can be traced back to their action via EcR/Usp. The canonical direct target genes of the EcR/Usp/20E complex are the 'early genes', which are located within polytene chromosome puff loci inducible by 20E (Ashburner 1990) and are primarily transcription factors—such as DHR3, E74 and E75 (Bialecki *et al.* 2002; Carney *et al.* 1997; Fletcher & Thummel 1995)—that coordinate the downstream ('late gene') response to 20E by integrating information from other signalling pathways. However, there are thousands of genes in the insect genome that respond to 20Es, many of which are uncharacterised (Beckstead *et al.* 2005; Stoiber *et al.* 2016; Uyehara & McKay 2019).

Es are typically considered prohormones, but may have signalling roles distinct from 20Es (Baker *et al.* 2003; Beckstead *et al.* 2007; Ono 2014), and while E is a very poor activator of *D. melanogaster* EcR/Usp (Baker *et al.* 2000; Wang *et al.* 2000), it is a better activator of DHR38 than 20E (Baker *et al.* 2003). RNAi against the cholesterol 7,8-dehydrogenase gene *nvd*, which acts at the very start of the ecdysteroid biosynthetic pathway, causes developmental arrest that can be rescued completely by feeding on 7-dehydrocholesterol (7dC; nvd's catalytic product) but not 20E (Yoshiyama *et al.* 2006), consistent with the hypothesis that there are other ecdysteroids or ecdysteroid

precursors downstream of 7dC that have roles in development. Other ecdysteroids largely thought to be inactivation products, such as 3-oxoecdysteroids and 3-epiecdysteroids, may have specific functions in insect development as well (Baker *et al.* 2003; Sommé-Martin *et al.* 1990), although their physiological functions are very poorly characterised in comparison to Es and 20Es.

1.3.3. Transport of ecdysteroids

It is likely that ecdysteroids—which are polyhydroxylated lipids—have a very limited capacity to diffuse passively across plasma membranes, and therefore require membrane proteins to assist in their transport into and out of cells (Neuman & Bashirullah 2018; Schweizer *et al.* 2019). While little attention has been paid to this area compared with the biosynthesis of ecdysteroids, a small number of ecdysteroid transporters have been identified in *D. melanogaster*. Ecdysone Importer (EcI) is an organic anion transporting polypeptide (OATP) required for the uptake of E and 20E. *EcI* null mutant animals fail to progress past the 1st larval instar but have no obvious defects during embryogenesis, suggesting multiple ecdysteroid importers act redundantly during this developmental stage (Okamoto *et al.* 2018); *EcI* is also required for the passage of 20E across the blood-brain barrier (Okamoto & Yamanaka 2020). In the *D. melanogaster* RG, biosynthesised E is secreted into the haemolymph in a vesicle-mediated manner that requires the ABC transporter Atet, which loads E into intracellular vesicles (Yamanaka *et al.* 2015), although the function of Atet in other tissues has not been characterised. *Early gene at 23* (*E23*) is an 20E-inducible gene that encodes an ABC transporter, ectopic expression of which results in suppression of EcR/Usp/20E signalling, consistent with a hypothesis that the protein transports ecdysteroids out of cells (Hock *et al.* 2000), although E23 appears not to transport E and therefore may be selective for 20-hydroxylated ecdysteroids (Yamanaka *et al.* 2015). Another ABC transporter, Mdr65, can transport 20E across the blood-brain barrier (Hindle *et al.* 2017), even though it is primarily known as a xenobiotic efflux transporter (Denecke *et al.* 2017a).

1.3.4. Inactivation of ecdysteroids

Given the highly pulse-like specificity of the developmental ecdysteroid titre in insects, ecdysteroid inactivation is likely as important for development (and other ecdysteroid-related processes) as ecdysteroid biosynthesis, but comparatively little is

known about ecdysteroid inactivation through catabolism. Ecdysteroid catabolites are often considered physiologically inactive because they cannot activate EcR/Usp (Baker *et al.* 2000; Makka *et al.* 2002), but given that there are ecdysteroid signalling pathways unconnected to EcR/Usp (Baker *et al.* 2003; Ishimoto *et al.* 2012; Sommé-Martin *et al.* 1990), this is not completely convincing evidence.

Many types of ecdysteroid catabolites have been identified in insects but they are produced by only two general reaction types, analogous to those found in xenobiotic metabolism: modification through the addition and subtraction of hydroxyl groups and oxygen atoms (Fig. 1.5A); and conjugation of various polar and non-polar moieties (Fig. 1.5B). Virtually all of these reactions take place at either one of four crucial positions on the ecdysteroid backbone: C-2, C-3, C-22 and C-26. Ecdysteroid inactivation is not just important for the regulation of endogenous ecdysteroids but also exogenous ecdysteroids—some plants produce phytoecdysteroids and ecdysteroid-like compounds as insect antifeedant secondary metabolites, which must be detoxified by insect herbivores (Dinan 2001; Rharrabe *et al.* 2007; 2009). Insect parasitoids of other insects are also exposed to their host's endogenous ecdysteroids, which may require inactivation by the parasitoid to maintain developmentally appropriate ecdysteroid titres (Brown & Reed-Larsen 1991; Buergin & Connat 1989; Melk & Govind 1999).

Two common ecdysteroid modification reactions are oxidation (the formation of oxoecdysteroids, also called dehydroecdysteroids) and epimerisation at the C-3 position (3-epiecdysteroids have a 3α hydroxyl group, while non-epimerised ecdysteroids have a 3β hydroxyl group; Fig. 1.5A). 3-oxo- and 3-epiecdysteroids are major catabolites in many insect species (Beydon *et al.* 1981; 1987; Sommé-Martin *et al.* 1988a; Weirich *et al.* 1991) although, as previously mentioned, some insect species secrete 3-oxoecdysteroids from early ecdysteroidogenic tissues, so they cannot be considered true catabolites in every context. In fact, C-3 oxidation is fully reversible—3-oxoecdysteroids are converted to 3-epiecdysteroids by 3-oxoecdysteroid 3α-reductases, but they can also be converted back to 'active' ecdysteroids by 3-oxoecdysteroid 3β-reductases (Chen *et al.* 1996; Sakurai *et al.* 1989); C-3 epimerisation, however, appears irreversible. The enzyme that converts ecdysteroids to 3-oxoecdysteroids—ecdysone oxidase (EO)—has been identified in multiple species (Sun *et al.* 2012; Takeuchi *et al.* 2001; 2005), and inactivation of EO in *Bombyx mori* results in an extended final larval instar, suggesting the enzyme is important for development (Li *et al.* 2015); *B. mori* 3-

oxoecdysteroid 3β-reductase has also been shown to increase active ecdysteroid titres during immune challenge (Sun *et al.* 2016). EO and 3α-reductases are thought to negatively regulate active ecdysteroid titres in lepidopteran species, and may do the same in *D. melanogaster* (Sun *et al.* 2012; Takeuchi *et al.* 2005; 2001). C-3 reductases have not been functionally characterised in *D. melanogaster*, but *spidey*, a gene encoding a putative ecdysteroid reductase that catalyses an unknown reaction, has essential functions in developmental and physiological processes connected to ecdysteroid signalling (Chiang *et al.* 2016; Cinnamon *et al.* 2016).

Another common modification reaction is carboxylation, which exclusively occurs at C-26 after 26-hydroxylation (Fig. 1.5A); ecdysonoic acids irreversibly formed by this process are major catabolites in most insects, particularly during metamorphosis (Beydon *et al.* 1981; Li *et al.* 2014; Moribayashi *et al.* 1985; Rewitz *et al.* 2010). Ecdysteroid 26-hydroxylation and carboxylation are carried out sequentially by Cyp18a1, expression of which is induced by high ecdysteroid titres (Bassett *et al.* 1997; Chen *et al.* 1994; Kayser *et al.* 1997) and is essential during metamorphosis in *D. melanogaster*, where it reduces the prepupal and pupal ecdysteroid titres, inducing the key metamorphosis factor βFTZ-F1 (Guittard *et al.* 2011; Rewitz *et al.* 2010). Given that ecdysonoic acids are taxonomically widespread and *Cyp18a1* is broadly conserved across insects (Guittard *et al.* 2011), carboxylation may be one of the dominant ecdysteroid inactivation pathways in insects. However, *Cyp18a1* has been lost in the *Anopheles gambiae* species complex (Neafsey *et al.* 2015), suggesting carboxylation is not a broadly essential catabolic pathway in every insect species, or another enzyme has evolved to replace Cyp18a1 in this taxon.

Aside from producing a pathway to carboxylation, the function of 26-hydroxylation *per se* is unclear—while 26-hydroxyecdysteroids are physiologically inactive in dipterans (Baker *et al.* 2000; Thompson *et al.* 1967), they are the predominant ecdysteroids in *Manduca sexta* eggs and embryos (Feldlaufer *et al.* 1987; Kaplanis *et al.* 1973; 1980) and undergo dynamic catabolism during embryogenesis (Thompson *et al.* 1987a; 1987b; 1985), suggesting they coordinate development at this stage. This is in contrast with *M. sexta* metamorphosis, during which 20-hydroxyecdysteroids are also drivers of development (Lozano *et al.* 1989; Warren & Gilbert 1986), as they are in other species (Garen *et al.* 1977; Mirth 2005). It remains to be seen if 26-hydroxyecdysteroids have important signalling roles in other insects.

The most common ecdysteroid conjugation reaction in insects is phosphorylation, which can occur at the C-2, C-3, C-22 and C-26 positions (Figs. 1.5B & 1.6) and is found in many orders of insects (Sonobe & Ito 2009). Ecdysteroid-phosphate conjugates are thought to play important roles in the reproduction and early development of some insects, particularly lepidopterans and orthopterans (Isaac *et al.* 1983b; Ito & Sonobe 2009; Thompson *et al.* 1987a); however, their importance to other developmental and physiological processes is unknown. Other conjugation reactions appear to be comparatively less common (Fig. 1.5B). The 22-glucosylation of 26-hydroxy-E has been observed in the developing embryos of *Manduca sexta* (Thompson *et al.* 1987a) and ingested 20E is 22-glucosylated in *Myzus persicae* (Rharrabe *et al.* 2007); ecdysteroid glucosylation is also a crucial step in the developmental and behavioural control of caterpillars by baculoviruses (Hoover *et al.* 2011) and the UGT gene responsible—*egt*—is thought to have been horizontally acquired from a host lepidopteran genome (Hughes 2013). Fatty acid conjugation—acylation—occurs in orthopterans, lepidopterans and dipterans (Grau & Lafont 1994; Hoffmann *et al.* 1985; Kubo *et al.* 1987; 1994; Rharrabe *et al.* 2007; Robinson *et al.* 1987), while acetylation occurs in orthopterans and dipterans (Maróy *et al.* 1988; Modde *et al.* 1984). Some ecdysteroid catabolites undergo multiple modification and/or conjugation reactions, sometimes at the same carbon position: various 2- and 3-acetylecdysteroid 2- and 22-phosphates are found in orthopterans (Isaac & Rees 1984; Lagueux *et al.* 1984; Modde *et al.* 1984; Tsoupras *et al.* 1982), while 3-epiecdysteroid 2-, 3- and 22-phosphates are present in lepidopterans (Kamba *et al.* 1995; 2000; Mamiya *et al.* 1995), and 3-epiecdysonoic acids are found in both dipterans (Moribayashi *et al.* 1985) and lepidopterans (Thompson *et al.* 1988).

With few exceptions, such as 26-carboxylation, carbon-specific ecdysteroid inactivation pathways appear poorly conserved between insect taxa (Rharrabe *et al.* 2007; Williams *et al.* 2002), suggesting that many of these catabolic pathways are evolutionarily labile and perhaps largely functionally interchangeable. This further suggests that genes encoding ecdysteroid catabolism enzymes may be much more evolutionarily unstable than ecdysteroid biosynthesis genes, which are (largely) highly conserved across insects (Ono *et al.* 2006; Rewitz *et al.* 2007; Sztal *et al.* 2007).

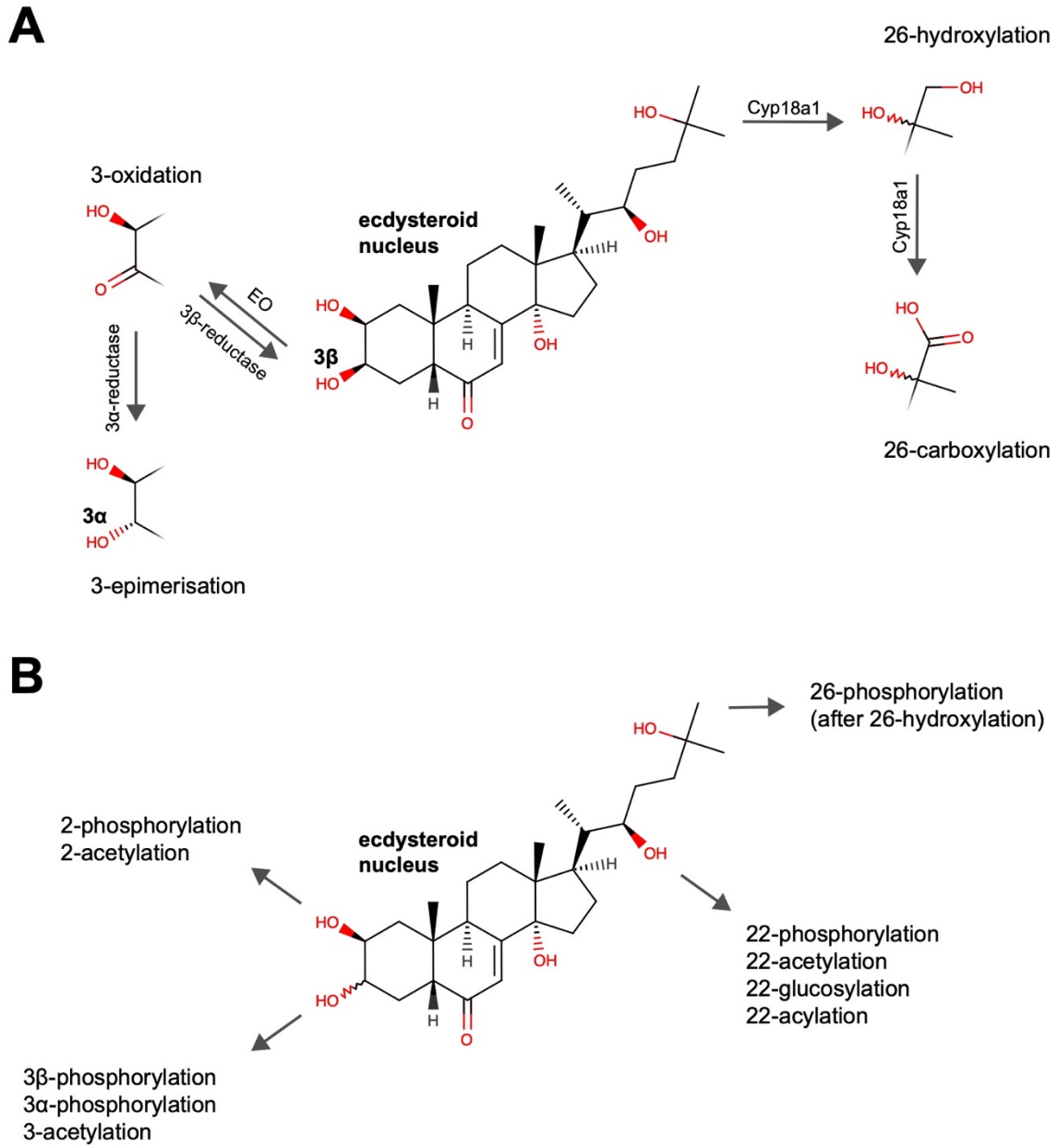


Figure 1.5. Ecdysteroid catabolism in insects via (A) modification reactions (oxidation, epimerisation, hydroxylation and carboxylation) and (B) conjugation reactions (phosphorylation, acetylation, glucosylation and acylation). EO, ecdysteroid oxidase.

1.3.5. Recycling of ecdysteroids

An important endocrinological distinction between the modification and conjugation of ecdysteroids is that the former tends to be irreversible, while the latter tends to be reversible; as such, conjugates are thought to play fundamental roles in the recycling of ecdysteroids (Delbecque *et al.* 1990; Sonobe & Ito 2009). Ecdysteroid recycling is the reuse of ecdysteroids by conjugation, followed by storage and then deconjugation, via a reciprocal enzyme system; sites of storage and release are called 'secondary' sources of ecdysteroids by Delbecque *et al.* (1990), in contrast to ecdysteroidogenic 'primary' sources.

The most well-known example of recycling is the reproductive ecdysteroid kinase/phosphatase system in *B. mori*, where oocytes are loaded with ecdysteroid-phosphates, which are then hydrolysed during early embryogenesis to supply part of the early ecdysteroid titre independent of biosynthesis (Fig. 1.7A; Sonobe & Ito 2009)—this will be discussed in more detail later in this chapter. Recycling may also be important for metamorphosis in some insects. OA, an uncharacterised conjugate of 20E, is formed in the larvae and early pupae of the fleshfly *Sarcophaga peregrina* and then deconjugated during the pupal-adult transition (Moribayashi *et al.* 1985; Moribayashi & Ohtaki 1980); ecdysteroid-phosphate conjugation and deconjugation also occurs during early/mid-metamorphosis in *Manduca sexta* pupae (Lozano *et al.* 1989). It has also been hypothesised that the very high pupal-adult ecdysteroid titre in *D. melanogaster* must come from one or more sources other than the RG, which degenerates from the start of metamorphosis and secretes relatively little E as a consequence (Dai & Gilbert 1991; Redfern 1983); the hydrolysis of stored conjugates from a secondary source may be one mechanism by which the pupal-adult ecdysteroid titre is produced, apart from the possibility of non-RG primary sources during this period (Delbecque *et al.* 1990). In addition, *D. melanogaster* females temporarily unable to synthesise ecdysteroids (due to the temperature-sensitive loss-of-function ecd^1 mutation) mated to wild-type males produce embryos with high rates of developmental defects (Kozlova & Thummel 2003), suggesting maternally derived ecdysteroids—likely ecdysteroid-acyl conjugates that are hydrolysed during early embryogenesis (Bownes *et al.* 1988)—are important for embryonic development in this species.

Explanations for why ecdysteroid recycling occurs at all are poorly theorised. One

benefit of recycling may be a reduction in energy use due to forgoing *de novo* biosynthesis, but it may also be a way of quickly increasing the ecdysteroid titre in a tissue or life stage independent of regulating the expression of the biosynthetic pathway. Recycling may also allow the selective retention of particular ecdysteroid species and therefore control specific types of ecdysteroid signalling in a temporal and/or spatial manner.

There may also be an unexplored functional relationship between ecdysteroid conjugation and transport in the context of recycling: as insect development requires precise pulses of ecdysteroids, the recycling of conjugates likely requires secretion of free ecdysteroids from specific storage sites in a tightly controlled manner. If so, conjugates likely cannot move freely between tissues or cell types, and ecdysteroid efflux transporters must either be unable to transport the conjugates across cell membranes or not expressed in these tissues at the relevant points in development. Given that ecdysteroid conjugates are often chemically distinct from free ecdysteroids (i.e. far bulkier and of substantially different polarity), it is unlikely that the free ecdysteroid transporters already identified (see Chapter 1.3.3; Hindle *et al.* 2017; Okamoto *et al.* 2018; Yamanaka *et al.* 2015) can transport conjugated ecdysteroids. Alternatively, if the sites of ecdysteroid conjugation, storage and/or deconjugation are spatially separated, there may be conjugate-specific transporters that move conjugates between these sites—however, such separation has yet to be demonstrated.

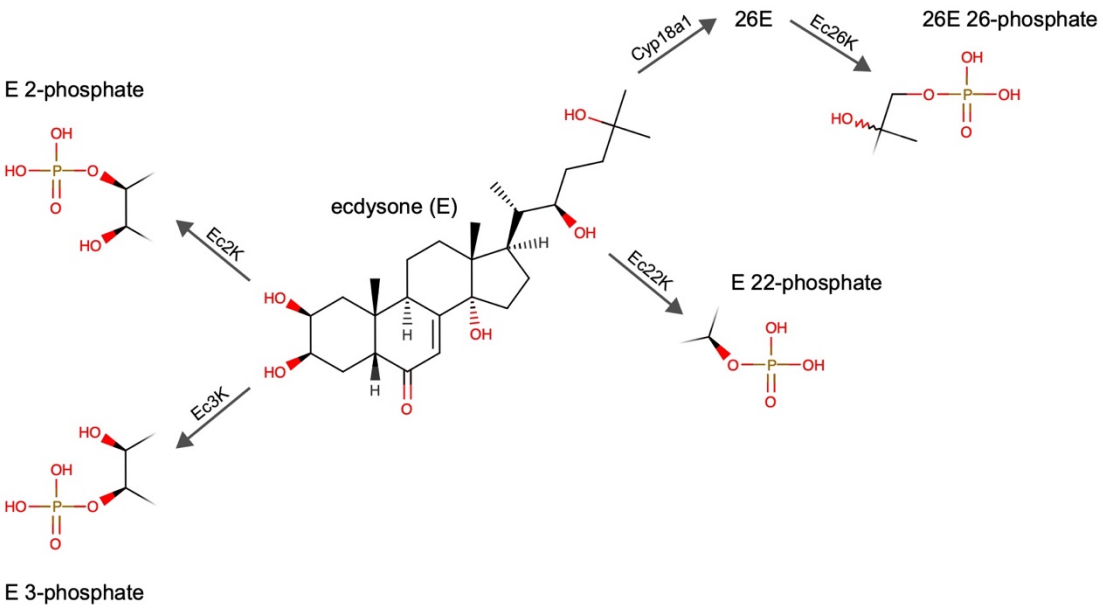


Figure 1.6. The four types of ecdysteroid-phosphate conjugates—2-phosphates, 3-phosphates, 22-phosphates and 26-phosphates—illustrated on the ecdysone (E) backbone. Each phosphorylation reaction is catalysed by a corresponding ecdysteroid kinase (Ec2K, Ec3K, Ec22K or Ec26K); it is currently unclear if any ecdysteroid kinases can efficiently phosphorylate multiple hydroxyl positions, meaning these could be distinct enzymes. 3-phosphorylation can take place at 3β hydroxyl groups (shown) or 3α hydroxyl groups (not shown)—it is unclear if different enzymes are required for these two different hydroxyl orientations. 26E, 26-hydroxyecdysone.

1.3.6. Ecdysteroid-phosphate conjugates

The phosphorylation of ecdysteroids, as the most widespread ecdysteroid conjugation reaction, received considerable attention in the insect endocrinological community in the 1980s through 2000s, although work in the area has greatly slowed in recent years—ecdysteroid-phosphate conjugates are produced by ecdysteroid kinases (called 'ecdysteroid phosphotransferases' in some papers), but only one such kinase has been genetically identified (Chapter 1.3.7; Sonobe *et al.* 2006). Due to their bulk and polarity, phosphate conjugates have very low affinity for EcR/Usp (Makka *et al.* 2002) and are likely physiologically inactive *in vivo*—and they can also often be readily hydrolysed back to free ecdysteroids by ecdysteroid-phosphate phosphatases (EP-Pases; Chapter 1.3.8), lending them a reputation as a storage form of hormone (Rees & Isaac 1984; Sonobe & Ito 2009). However, it is not clear that all ecdysteroid-phosphate conjugate species can be hydrolysed and therefore some may function as terminal catabolites.

Ecdysteroid-phosphate conjugates have been best studied in lepidopterans and orthopterans, where four species (two in each order) are known to possess ovarian/embryonic ecdysteroid-phosphate (OEEP) recycling systems, which involve maternal deposition of ecdysteroid-phosphate conjugates into oocytes that are subsequently hydrolysed to free ecdysteroids during early embryogenesis (Rees & Isaac 1984; Sonobe & Ito 2009).

The desert locust, *Schistocerca gregaria*, stores 22-phosphate conjugates of maternally derived E, 2-deoxy-E, 20E and 2-deoxy-20E in newly laid eggs that are hydrolysed and modified to active ecdysteroids during embryogenesis (Isaac *et al.* 1983a), but 2-phosphate conjugates predominate by the end of embryogenesis, suggesting they are terminal catabolites (Isaac & Rees 1985). An ecdysteroid 22-phosphate phosphatase in *S. gregaria* embryos has been biochemically characterised but not cloned (Isaac *et al.* 1983b), as has a 2-deoxy-E 22-kinase from the follicle cells of the ovary in the same species (Kabbouh & Rees 1991).

The migratory locust, *Locusta migratoria*, stores 22-phosphate conjugates of maternally derived 2-deoxy-E and E in its eggs, which are hydrolysed and modified to active ecdysteroids throughout embryogenesis; of note, these are bound to vitellin in yolk

granules before being hydrolysed (Lagueux *et al.* 1984). 3- and 22-phosphate conjugates of free and acetylated ecdysteroids, respectively, slowly accumulate throughout embryogenesis, suggesting they may be terminal catabolites during this developmental stage (Lagueux *et al.* 1984).

Uniquely, the tobacco hornworm, *Manduca sexta*, stores predominantly maternally derived 26-hydroxy-E 26-phosphate in its eggs, along with small amounts of 26-hydroxy-E 2-phosphate (Thompson *et al.* 1988; 1987b; 1985). The 26-phosphate is hydrolysed to free 26-hydroxy-E throughout embryogenesis, some of which is 20-hydroxylated, and all of which is eventually converted into 2-phosphates, 22-glucosides or 20-hydroxyecdysonic acids (Thompson *et al.* 1988; 1987a; Warren *et al.* 1986). The kinase(s) and phosphatase(s) involved in this system have yet to be characterised.

The silkworm, *Bombyx mori*—which has the only genetically characterised OEEP system—stores 22-phosphate conjugates of 2-deoxy-E, E and 20E. These conjugates are synthesised in the oocyte by an ecdysteroid 22-kinase, BmEc22K, and bind to vitellin in yolk granules (Fig. 1.7B; Sonobe & Ito 2009). It is important to note that two kinds of eggs are laid in this species: non-diapause eggs (NDEs), from which larvae hatch 10–11 days after oviposition; and diapause eggs (DEs), which suspend embryonic development at the late gastrula stage until triggered by favourable environmental conditions. Diapause can be broken, or prevented in otherwise fated eggs, by the injection or application of 20E, suggesting a low active hormone titre is required for diapause initiation and maintenance (Gharib *et al.* 1981; Makka *et al.* 2002). In NDEs, 22-phosphate conjugates remain bound to vitellin only until the early gastrula stage, at which point the yolk granules are acidified, releasing the conjugates, which are hydrolysed to free ecdysteroids by an ecdysteroid-phosphate phosphatase (BmEPPase; Fig. 1.7C); this process of hydrolysis is greatly suppressed in DEs, keeping active ecdysteroid titres low and maintaining diapause (Yamada *et al.* 2005). 2- and 3-phosphate conjugates are also present in *B. mori* eggs; although their functions are not well understood, it is suspected that the 2-phosphate conjugates at least may be terminal catabolites (Sonobe & Yamada 2004). Both *BmEc22K* and *BmEPPase* have been cloned and belong to the ecdysteroid kinase-like (EcKL) and histidine phosphatase families, respectively (Sonobe *et al.* 2006; Yamada & Sonobe 2003).

Phosphate conjugates are also present at other life stages in lepidopterans and

orthopterans. Final-instar *Pieris brassicae* larvae convert injected 3-epi-20E to 3-epi-20E 3-phosphate (Beydon *et al.* 1987), while *L. migratoria* and *S. gregaria* nymphs of various ages metabolise E to 3-acetate 2-phosphate double conjugates (Gibson *et al.* 1984; Modde *et al.* 1984). Phosphate conjugates of E, 3-epi-E, 3-epi-20E, 26-hydroxy-E, 20,26-dihydroxy-E and 3-epi-20,26-dihydroxy-E (the phosphate positions of which are unknown, except 26-hydroxy-E 26-phosphate) are formed during metamorphosis in *M. sexta* (Lozano *et al.* 1989).

Outside of Lepidoptera and Orthoptera, ecdysteroid-phosphate conjugates are poorly characterised. *In vivo* 3-oxo-E 2-phosphorylation is induced by ingestion of the non-steroidal ecdysteroid agonist RH-5849 in larvae of the housefly *Musca domestica* (Williams *et al.* 2002), and up to four ecdysteroid-phosphate conjugates have been detected in *D. melanogaster*:

- a possible 3-oxo-E 2-phosphate is formed *in vitro* by 3rd-instar larval homogenate after oral induction by 20E (Hilton 2004),
- a possible 3-epi-20E 3-phosphate is formed *in vivo* after injection of 3rd-instar larvae with 20E (Sommé-Martin *et al.* 1988a),
- E 22-phosphate is present in adult ovaries (Grau *et al.* 1995; Pis *et al.* 1995); and
- the S2 cell line constitutively converts 20,26-dihydroxy-E to a 26-phosphate conjugate (Guittard *et al.* 2011).

Both the 2- and 3-phosphate conjugates in *D. melanogaster* are only tentatively identified, and the functions of any of the conjugates are not known. 26-phosphorylation competes with the Cyp18a1-mediated carboxylation of 26-hydroxyecdysteroids in S2 cells (Guittard *et al.* 2011) and so may play a role in regulating the irreversible carboxylation of ecdysteroids, or it could conceivably be part of a biochemical mechanism to produce and retain 26-hydroxyecdysteroids without the subsequent (rapid) conversion to ecdysonoic acids. Ovarian E 22-phosphate is bound to an unknown 50 kDa protein that is not a yolk protein (Grau *et al.* 1995; Pis *et al.* 1995); Grau *et al.* (1995) speculate that this ecdysteroid-phosphate/protein complex is the same ecdysteroid-related antigen detected in oocyte follicle cells (Grau & Gutzeit 1990), but further experiments on these phenomena have not been published. Ecdysteroid-phosphates have not been detected in the eggs and embryos of *D. melanogaster*, which seem to instead use apolar conjugates—possibly ecdysteroid-acyl esters—as a source of

ecdysteroids during early embryogenesis (Bownes *et al.* 1988; Grau *et al.* 1995). This is consistent with the major yolk proteins of *D. melanogaster* and other dipterans being orthologs of vertebrate lipases, which suggests they have the capacity to bind to ecdysteroid-acyl conjugates (Bownes 1992), in contrast to true vitellins, which bind ecdysteroid-phosphates in other species (see above).

It is currently unclear if OEEP systems (or other ecdysteroid recycling systems involving non-phosphate conjugates) are truly essential for insect embryogenesis. In *B. mori*, free ecdysteroids are synthesised from sterols during non-diapause embryogenesis, even as 22-phosphate conjugates supply much of the active ecdysteroid titre (Sonobe & Yamada 2004). However—as noted by Niwa & Niwa (2014)—*B. mori* animals homozygous for loss-of-function alleles of *sro* (also called *non-moulting glossy* in this species) or *nobo*—genes essential for ecdysteroidogenesis—appear to progress through embryogenesis normally and only arrest development during larval stages (Enya *et al.* 2015; Nagata *et al.* 1987; Niwa *et al.* 2010; Tanaka 1998); this is in contrast to *sro*, *nobo* and other Halloween gene loss-of-function mutants in *D. melanogaster*, which uniformly fail to complete embryogenesis (Chanut-Delalande *et al.* 2014; Chavez *et al.* 2000; Enya *et al.* 2014; Niwa *et al.* 2010; Ono *et al.* 2006; Petryk *et al.* 2003; Warren *et al.* 2002; 2004). In addition, it appears that ecdysteroidogenesis is only responsible for—at most—a very small proportion of the ecdysteroid titre during *M. sexta* embryogenesis, as the vast majority of ecdysteroids present in embryos are derived from 26-hydroxy-E 26-phosphate (Thompson *et al.* 1988) and the low levels of E and 20E can be explained by the presence of small quantities of maternally derived E and 20E conjugates in newly laid eggs (Warren *et al.* 1986). Overall, these data suggest that the embryonic active ecdysteroid titre can be satisfactorily supplied by hydrolysing ecdysteroid-phosphates in species with OEEP systems and that these conjugates may be essential for early development.

Outside of insects, ecdysteroid-phosphate conjugates have also been found in the ovaries and embryos of some crustaceans (Subramoniam 2000). *Macrobrachium rosenbergii*, the giant river prawn (order Decapoda), produces high levels of 2,3-diacetyl-E 22-phosphate during late embryogenesis, suggesting the compound is a terminal catabolite (Young *et al.* 1991). *Penaeus monodon*, the giant tiger prawn (order Decapoda), possesses 20E 22-phosphate in the very early stages of vitellogenesis in the ovary, but only a relatively minor quantity (Young *et al.* 1993). *Armadillidium vulgare*, the common

pill-bug (order Isopoda), contains conjugated E and 20E in its ovaries, which can be hydrolysed by alkaline phosphatase and are therefore likely phosphate conjugates; these conjugates accumulate as the oocytes in the ovary mature, consistent with an OEEP system (Suzuki *et al.* 1996). These data suggest that some crustaceans may have OEEP systems similar to those in some insects, although it is unclear how widespread this may be.

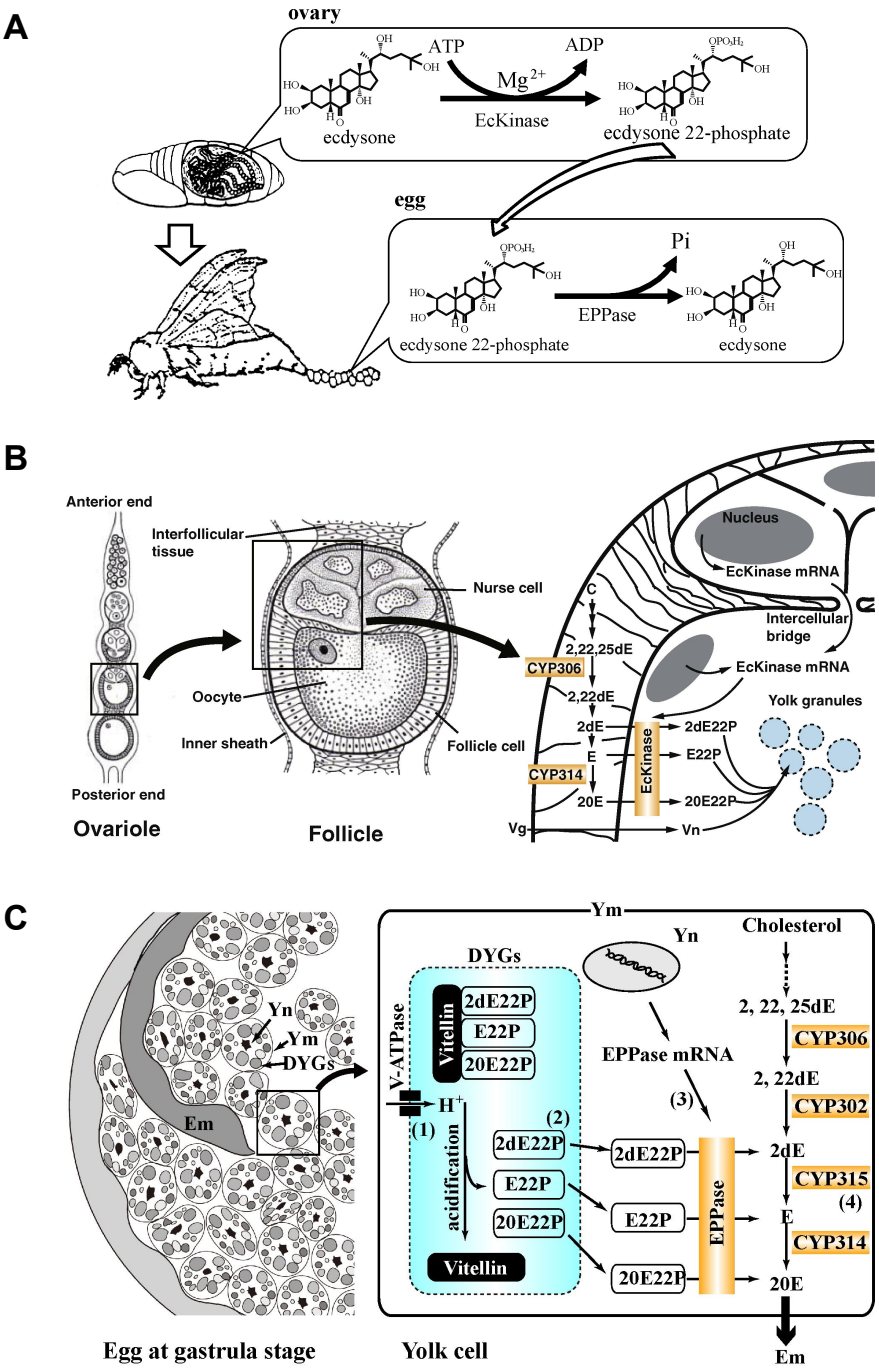


Figure 1.7. The ovarian/embryonic ecdysteroid-phosphate (OEEP) recycling system of *Bombyx mori*. (A) Overview of conversion of ecdysteroids (here ecdysone) into ecdysteroid 22-phosphate conjugates via the ecdysteroid 22-kinase BmEc22K in the ovary (top), with the reciprocal reaction occurring via the ecdysteroid-phosphate phosphatase BmEPPase in the egg/embryo (bottom). (B) Detail of the ecdysteroid phosphorylation process in the follicles of *B. mori* ovaries, whereby ecdysteroids synthesised in follicle cells are phosphorylated by BmEc22K in the oocyte and stored in yolk granules after binding to vitellin. (C) Detail of the ecdysteroid dephosphorylation process in the yolk cells of *B. mori* embryos, whereby ecdysteroid-phosphate conjugates are released from vitellin under acidic conditions and hydrolysed by BmEPPase in the yolk cell cytoplasm. Vg, vitellogenin; Vn, vitellin; Em, embryo; Yg, yolk granule; Ym, yolk cell membrane; Yn, yolk cell nucleus; DYGs, dense yolk granules; 2dE22P, 2-deoxyecdysone 22-phosphate; E22P, ecdysone 22-phosphate; 20E22P, 20-hydroxyecdysone 22-phosphate. Adapted from Sonobe & Ito (2009).

1.3.7. Ecdysteroid kinases

Ecdysteroid kinases convert ecdysteroids into ecdysteroid-phosphate conjugates by phosphorylating one of four hydroxyl groups at C-2, C-3, C-22 or C-26 (which is only present on 26-hydroxyecdysteroids; Fig. 1.6). While the existence of ecdysteroid kinases can be inferred from the presence of ecdysteroid-phosphate conjugates, very few of these enzymes have been characterised, either biochemically or genetically. As such, it is unclear whether all ecdysteroid kinases belong to the same gene/protein family, or if multiple families have ecdysteroid kinase activity in insects.

Cytosolic extracts from fat body and ovaries of *Bombyx mori* female pupae possess ATP- and Mg^{2+}-dependent ecdysteroid kinase activities sufficient to produce 2-deoxy-E 22-phosphate, E 22-phosphate and 2,22-dideoxy-20E 3-phosphate conjugates (Takahashi *et al.* 1992). This discovery led to the eventual characterisation of the best characterised ecdysteroid kinase, BmEc22K, the ovarian ecdysteroid 22-kinase in *B. mori*. BmEc22K is cytosolic, uses ATP and Mg^{2+} as cofactors and is inhibited by Ca^{2+}. Known substrates of BmEc22K are 2-deoxy-E, 2-deoxy-20E, E and 20E, but not 22-deoxyecdysteroids, which has been interpreted as evidence that this enzyme cannot phosphorylate C-3 hydroxyl groups and is not responsible for 3-phosphate conjugates present in ovaries (Sonobe *et al.* 2006); if this interpretation is correct, a hypothetical BmEc3K enzyme remains to be identified. The *BmEc22K* gene has been cloned from a cDNA library (and subsequently annotated in the *B. mori* genome), making it the only ecdysteroid kinase to be identified genetically. The BmEc22K polypeptide sequence contains the ecdysteroid kinase-like (EcKL) domain, formerly known as 'domain of unknown function 227' (DUF227; see Chapter 1.4). The molecular weight of purified BmEc22K is 44 kDa, which is consistent with the *in silico* estimation of the polypeptide weight encoded by *BmEc22K* (44.2 kDa), suggesting it is not drastically post-translationally modified (Sonobe *et al.* 2006). *BmEc22K* mRNA is transcribed in the oocyte and nurse cells, but appears to only be translated in the oocyte; BmEc22K protein is localised to the external region of the oocyte—this is likely where ecdysteroid 22-phosphate conjugates are synthesised before binding to vitellin and being imported into yolk granules for storage (Ito *et al.* 2008).

Other species of Lepidoptera have ecdysteroid kinase activity in their larval midguts. Cytosolic midgut homogenates from *M. sexta* 5^{th}-instar larvae convert E and 3-epi-E

to four and two unknown phosphate conjugates, respectively, with ATP and Mg^{2+} as cofactors (Weirich *et al.* 1986). Injections of the non-steroidal ecdysteroid agonist RH-5849 (but not 20E) induce 3-epi-E 2- and 3-kinase activity in the same tissue, suggesting 3-epi-E 2- and 3-phosphates are some of the unknown conjugates formed by *M. sexta* midguts; induction of these kinase activities is independently inhibited by actinomycin D and cycloheximide, suggesting the genes encoding the enzymes are actively transcribed and translated in response to RH-5849 (Williams *et al.* 1997). Cytosolic midgut homogenates from *Spodoptera littoralis* (Lepidoptera: Noctuidae) 6th-instar larvae contain E 2- and 22-kinase activities, both of which require ATP and Mg^{2+} as cofactors; notably, these activities are greatest during feeding periods when the endogenous active ecdysteroid titre is low, suggesting that these enzymes might be involved in phytoecdysteroid inactivation rather than developmental processes (Webb *et al.* 1996; 1995). Curiously, in the process of trying to clone the *S. littoralis* 2- and 22-kinases, Hilton (2004) discovered that the presence of reducing agents or non-ionic detergents severely reduced ecdysteroid kinase activity in midgut cytosol, suggesting these enzymes may need to adopt a specific, perhaps homomeric, quaternary structure before they are catalytically active.

Schistocerca gregaria (Orthoptera: Acrididae) has two biochemically characterised ecdysteroid kinases: a 2-deoxy-E 22-kinase (Sg2dEc22K) that is present in the follicle cells of the ovary (Kabbouh & Rees 1991) and a 3-acetyl-E 2-kinase (Sg3aEc2K) that is present in the fat body of 5th-instar nymphs (Kabbouh & Rees 1993); both enzymes are cytosolic, use ATP and Mg^{2+} as cofactors and are inhibited by Ca^{2+}. Sg2dEc22K activity increases throughout ovarian development, peaking at the same time as the titre of conjugated ecdysteroids in the ovary (Kabbouh & Rees 1991). Sg3aEc2K appears specific for 3-acetyl-E, as neither 2-acetyl-E nor E can serve as substrates; the enzyme also has an approximate molecular weight of 45 kDa (Kabbouh & Rees 1993), which is very close to the molecular weight of BmEc22K (44 kDa).

That all ecdysteroid kinase activities are cytosolic and use ATP and Mg^{2+} as cofactors—and that the molecular weights of BmEc22K and Sg3aEc2K are very similar—is consistent with a hypothesis that all ecdysteroid kinases in insects are encoded by the same gene family as *BmEc22K*, the EcKLs. However, since no EcKL enzymes outside of BmEc22K have been heterologously expressed and biochemically characterised, nor any EcKL genes genetically characterised for functions in ecdysteroid-related

processes, this hypothesis remains tentative—testing this hypothesis in *Drosophila melanogaster* is an aim of this thesis.

1.3.8. Ecdysteroid-phosphate phosphatases

Just as ecdysteroid-phosphate conjugates are produced by ecdysteroid kinases, they are hydrolysed by ecdysteroid-phosphate phosphatases (EPPases). While slightly more is known about EPPases genetically than ecdysteroid kinases, EPPase biochemical activities have been characterised in fewer organisms and little is known about their functions outside of OEEP systems.

The best characterised EPPase is BmEPPase in *B. mori*, which is responsible for the hydrolysis of maternal 22-phosphate conjugates in non-diapause eggs (NDEs) and is a cytosolic enzyme with a molecular weight of 42 kDa. The *BmEPPase* gene has been cloned from cDNA and the polypeptide sequence belongs to the histidine phosphatase family (Yamada & Sonobe 2003). However, the functional BmEPPase enzyme purified and cloned by Yamada and Sonobe (2003) was a short isoform (herein $EPPase_S$; also called $EPPase_{PGM}$ in the literature) containing only a phosphatase domain—a longer isoform ($EPPase_L$) is also present in *B. mori* and other insects, which contains at least two additional N-terminal domains: an SH3 domain and a UBA domain (discussed later), and a possible 2H phosphoesterase domain, which is not currently recognised by the Pfam and Interpro databases (Davies *et al.* 2007). $BmEPPase_S$ activity is extremely low in diapause eggs (DEs) but very high in NDEs, peaking around early organogenesis (just before the peak of free ecdysteroids) and slowly decreasing to DE levels by the time of hatching (Yamada & Sonobe 2003). $BmEPPase_S$ hydrolyses 22-phosphate conjugates with an approximately 200-fold greater specificity than 3β-phosphate conjugates, suggesting it sparingly hydrolyses 3β-phosphate conjugates during embryogenesis. $BmEPPase_S$ mRNA is not detectably expressed in the pupal ovary or testis, nor in the Malpighian tubules, fat body or midgut of 5th-instar larvae (Yamada & Sonobe 2003). Recent work has demonstrated that RNAi knockdown of *BmEPPase* (using a dsRNA molecule that targets both isoforms) in early NDEs reduces the expression of ecdysteroid-response genes compared to controls, which is—to my knowledge—the first piece of functional genetic evidence that BmEPPase activity controls the active ecdysteroid titre during *B. mori* embryogenesis (Matsushima *et al.* 2019).

EPPase$_L$ from *D. melanogaster* (DmEPPase$_L$; CG13604) has also been characterised *in vitro*. DmEPPase$_L$ can use E 22-phosphate, 3-epi-E 22-phosphate and 3-epi-E 2-phosphate as substrates, in decreasing order of specificity, however it cannot use 3-epi-E 3-phosphate, possibly due to the 3α orientation of the phosphate moiety (Davies *et al.* 2007); it is unknown if either DmEPPase isoform can hydrolyse 3β-phosphate conjugates. The *in vivo* functions of either DmEPPase isoform have yet to be explored with genetic experiments.

The enzyme responsible for EPPase-like activity detected in the embryos of *S. gregaria* (Isaac *et al.* 1983b) has not been purified or cloned, but there are reasons to believe it may not be an EPPase ortholog but instead a non-specific acid phosphatase: first, its optimal pH is 4.0, while BmEPPase and DmEPPase activity is highest at pH 7.4–7.5; and second, its subcellular distribution is largely lysosomal, while BmEPPase and DmEPPase are cytosolic enzymes (Davies *et al.* 2007; Yamada & Sonobe 2003). Further work characterising this EPPase-like activity or cloning the associated gene in *S. gregaria* has not been published.

While EcKLs have no mammalian orthologs (Mitchell *et al.* 2014), EPPases are orthologs of the *Homo sapiens* protein Sts-1 (Suppressor of T-cell receptor signalling 1; also known as UBASH3B and TULA-2), which modulates the GPVI and T-cell receptor pathways through protein phosphatase activity (Tsygankov 2019). Curiously, insect EPPases can dephosphorylate not just ecdysteroid-phosphate conjugates, but synthetic vertebrate steroid-phosphate conjugates and phosphorylated peptides and proteins—a trait shared with *H. sapiens* Sts-1 and a nematode ortholog in *Caenorhabditis elegans* (Chen *et al.* 2008; Davies *et al.* 2007). While insect EPPases are very unlikely to encounter vertebrate steroid-phosphates *in vivo* and therefore their activity towards them is likely coincidental, it is unclear if their protein phosphatase activities are of physiological importance. In addition, the SH3 and UBA domains of Sts-1 are involved in binding to ubiquitin ligases and ubiquitin, respectively, which indirectly complexes them with the membrane-bound receptors whose activity they modulate (Kowanetz *et al.* 2004). This raises the possibility that insect EPPase$_L$ indirectly interacts with the G-protein coupled dopamine/ecdysteroid receptor DopEcR through its own SH3 and UBA domains, as suggested by Davies *et al.* (2007); consistent with this hypothesis, both *DopEcR* and *EPPase$_L$* (but not *EPPase$_S$*) transcripts appear specifically

expressed in the *D. melanogaster* nervous system (Leader *et al.* 2018). If a functional interaction exists between these two proteins, it would suggest that ecdysteroid-phosphate conjugates are involved in the regulation of insect behaviours ala. DopEcR (Petruccelli *et al.* 2020).

Of note, the *EPPase* ortholog in the crustacean *Daphnia magna* (Cladocera: Daphniidae) is strongly expressed in early non-diapause embryos (but not diapause embryos) and putative RNAi knockdown of the gene results in arrested embryonic development (Asada *et al.* 2014), suggestive of an OEEP-like system in this species. However, polar ecdysteroid conjugates (which may or may not be ecdysteroid-phosphates) are present in relatively small quantities—compared to free ecdysteroids and apolar conjugates—in newly deposited *D. magna* eggs (Martin-Creuzburg *et al.* 2007). It is possible that the hypothetical phosphate conjugates in *D. magna* eggs are a storage form of a developmentally essential yet scarce ecdysteroid species that is not produced by ecdysteroidogenesis in the embryo; alternatively, this EPPase may have some non-ecdysteroid-related function during this period, or the RNAi knockdown phenotype observed by Asada *et al.* (2014) is the result of an off-target effect.

Thus far, only ecdysteroid 2-, 3- and 22-phosphate conjugates have been shown to be substrates—to varying degrees—of purified insect EPPases. As such, it is unclear if the 26-phosphate deconjugation in *M. sexta* embryogenesis and metamorphosis (Lozano *et al.* 1989; Thompson *et al.* 1987a) occurs via an EPPase ortholog or if another enzyme is responsible.

1.4. The ecdysteroid kinase-like (EcKL) gene family

The ecdysteroid kinase-like (EcKL) gene family (Pfam family PF02958) is a poorly characterised group of genes encoding proteins containing the EcKL domain, with a peculiar taxonomic distribution. Based on data in the Pfam database, 85% of EcKL genes are in eukaryotic genomes, with the remaining 15% in bacteria; of the eukaryotic genes, over 85% (>72% of all EcKLs) are in the arthropod clade Tetraconata (insects, hexapods and crustaceans; Schwentner *et al.* 2018), with the remainder in fungi, oomycetes and a very small handful of sequences in the Chordata, including teleost fish and the cephalochordate *Branchiostoma floridae*; EcKLs appear completely absent from mammalian genomes (El-Gebali *et al.* 2018). It is important to note that there may be some uncertainty around the precise classification of some proteins annotated as EcKLs, particularly in vertebrates and bacteria (Sun *et al.* 2015), where the sequence identity with arthropod EcKLs is very low.

The EcKL domain used to be called 'domain of unknown function 227' (DUF227) until the ecdysteroid 22-kinase function of the member gene *BmEc22K* (Sonobe *et al.* 2006) was recognised by Pfam (El-Gebali *et al.* 2018). The domain family belongs to the PKinase clan (CL0016), a group of serine/threonine- and tyrosine-protein kinases and non-protein kinases that contains 38 Pfam families; the domains with the highest similarity to the EcKLs are the choline/ethanolamine kinase (PF01633), aminoglycoside phosphotransferase (PF01636) and DUF1679 (PF07914) domains, which are all known or predicted small-molecule kinases (El-Gebali *et al.* 2018). The EcKL domain is defined by the HMM in Pfam, which is 294 aa long, but the mean length of the domain in actual proteins is 243.4 aa. The EcKL HMM contains a number of characteristic motifs, the most striking of which is the Brenner kinase motif (HxDxxxxN), which binds ATP (Brenner 1987). However, the sequence conservation of the domain throughout the family is overall quite low, with the average pairwise identity of sequences in the 'full' Pfam dataset at just 19%. The vast majority (87%) of EcKL-containing proteins contain a lone EcKL domain, while 8% have two EcKL domains; the remainder possess more than two EcKL domains or a combination of other domains (El-Gebali *et al.* 2018), but it is unclear which of these domain models are true or drawn from incorrectly annotated gene models, which is a pervasive problem in protein databases (Nagy *et al.* 2008).

As noted, the DUF1679 gene family, found primarily in nematodes, is a close relative of the EcKLs, and members in *Caenorhabiditis elegans* were previously annotated as DUF227 genes (see McElwee *et al.* 2004; Patel *et al.* 2008; Taubert *et al.* 2008; Thomas 2006). Curiously, some *C. elegans* DUF1679 genes appear to be co-regulated with genes from classical detoxification families (McElwee *et al.* 2004; Taubert *et al.* 2008), suggesting they might have a role in detoxification in nematodes. However, to the best of my knowledge, phosphate conjugates of xenobiotic compounds have yet to be detected in nematodes.

1.4.1. EcKLs in *Drosophila melanogaster*

Drosophila melanogaster is one of the best-studied model organisms in biology and has had a published genome sequence for two decades (Adams *et al.* 2000), but a large fraction of the genes in its genome remain uncharacterised. This extends to the EcKL gene family (of which there are 51 members in *D. melanogaster*), where only a few genes have been studied in any detail.

Juvenile hormone-inducible protein 26 (*JhI-26*) was cloned by Dubrovsky *et al.* (2000) in a study that aimed to discover genes that were transcriptionally induced by the juvenile hormone (JH) agonist methoprene in the S2 cell line. *JhI-26* is rapidly induced by methoprene in S2 cells and *in vivo*, and its expression throughout the lifecycle of *D. melanogaster* matches the JH titre, suggesting it may be a JH primary-response gene; consistent with this, the promoter of *JhI-26* is sufficient to confer responsiveness to JH in S2 cells (Dubrovsky *et al.* 2000). Later work discovered that *JhI-26* is strongly induced in *D. melanogaster* larval testes by infection with the maternally inherited endosymbiotic bacteria *Wolbachia* (Zheng *et al.* 2011) and that JhI-26 is present in *D. melanogaster* sperm (Wasbrough *et al.* 2009); this led to the hypothesis that *Wolbachia*-induced cytoplasmic incompatibility (CI) involves the JH pathway and is mediated by *JhI-26* (Liu *et al.* 2014). CI is a consequence of *Wolbachia* infection that prevents infected males from successively reproducing with uninfected females, while infected females are unaffected by the infection status of their partner—this reduces the fitness of uninfected females and spreads *Wolbachia* throughout insect populations (Werren *et al.* 2008). *Wolbachia* infection in *D. melanogaster* increases the expression of JH biosynthesis and signalling genes in testes, which is sufficient to induce *JhI-26*. Transgenic testis-

specific or ubiquitous overexpression of *JhI-26* results in CI-like egg hatching failure in offspring due to mitotic defects, unless the flies are mated with *Wolbachia*-infected females; this is due to the induction of the accessory gland protein CG10433 by *JhI-26*, which—when overexpressed itself—causes a similar CI-like phenotype. Overexpressing *JhI-26* in the testes of male flies also reduces the remating receptivity of their female mates, which is partially phenocopied by overexpression of *CG10433*, and is also seen in *Wolbachia*-infected males (Liu *et al.* 2014). Overall, these data strongly suggest that *JhI-26* is involved in regulating the sperm proteome and therefore reproductive outcomes and post-mating female behaviour—however, the molecular mechanisms by which this takes place, including JhI-26's enzymatic substrates, are unknown.

CHKov1 is another *D. melanogaster* EcKL that was first identified by Aminetzach *et al.* (2005) as possessing an unusually high-frequency insertion of a *Doc1420* transposable element into its second exon. This $CHKov1^{Doc}$ allele has low levels of nucleotide diversity and is found at high frequencies in non-African populations, suggesting it has undergone a soft selective sweep sometime in the past 25–240 years, even though the original *Doc1420* insertion likely arose many thousands of years prior (Garud & Petrov 2016; Magwire *et al.* 2011). Aminetzach *et al.* (2005) suggested—based on the distant homology between EcKLs and choline/ethanolamine kinases—that *CHKov1* might be involved in choline metabolism and therefore $CHKov1^{Doc}$ might affect resistance to organophosphate (OP) insecticides, which target acetylcholinesterase; a backcross experiment comparing $CHKov1^{Doc}$ flies to wild-type flies suggested that $CHKov1^{Doc}$ homozygosity increases adult resistance to the OP azinphos-methyl (Aminetzach *et al.* 2005), however, $CHKov1^{Doc}$ is not associated with developmental resistance to azinphos-methyl in the DGRP (Battlay *et al.* 2016) and did not associate with adult resistance to the OPs malathion and parathion in other DGRP GWAS (Battlay *et al.* 2018; Duneau *et al.* 2018).

An alternative—and better supported—cause for the selective sweep around $CHKov1^{Doc}$ is resistance to the vertically transmitted sigma virus, which infects *D. melanogaster* at low frequencies in wild populations and causes permanent paralysis upon exposure to CO_2 gas. Multiple viral resistance alleles have been isolated in *D. melanogaster* (Gay 1978), one of which—*ref(3)D*—was mapped to the *CHKov1* locus by Magwire *et al.* (Magwire *et al.* 2011) and found to be $CHKov^{Doc\text{-}Dup}$—a derived allele of $CHKov1^{Doc}$ that contains two partial 5′ duplications of $CHKov1^{Doc}$ and a whole and

partial duplication of the neighbouring EcKL gene *CHKov2*. *CHKov1*^{Doc} is strongly associated with sigma virus resistance in the DGRP, while *CHKov*$^{Doc\text{-}Dup}$—which confers even greater resistance—is not found in the DGRP (Magwire *et al.* 2011; 2012). The mechanism of viral resistance for these alleles is currently unknown, although it appears to depend on the presence of novel *CHKov1*^{Doc} transcripts and not an absence of ancestral *CHKov1* functionality (F. Jiggins, personal communication), suggesting the *CHKov1*^{Doc} allele possesses a novel antiviral function; this is consistent with the increased resistance of *CHKov*$^{Doc\text{-}Dup}$ (compared to *CHKov1*^{Doc}), which increases the copy number of 5′ *CHKov1*^{Doc} transcripts. However, as the polypeptides putatively translated from *CHKov1*^{Doc} mRNA isoforms likely do not possess kinase functionality, since they largely exclude the EcKL domain and are missing the ATP-binding Brenner motif, this implies that the viral resistance function of *CHKov1*^{Doc} may be completely unrelated to wild-type EcKL functions. Curiously, the open reading frame of the 5′ *CHKov1*^{Doc} transcript appears to have accumulated seven hydrophilic amino acid substitutions at its C-terminus, suggestive that these may be associated with its novel function (Aminetzach *et al.* 2005). Overall, it seems likely that sigma virus resistance, not OP resistance, was responsible for the selective sweep around *CHKov1*, but that the ancestral function of *CHKov1* may be unrelated to either insecticide resistance or viral resistance.

1.4.2. EcKLs in other insects

Outside of *D. melanogaster*, EcKLs are equally poorly characterised, apart from a few notable examples. The most obvious is *BmEc22K* (Chapter 1.3), which encodes an ecdysteroid 22-kinase and is responsible for the maternal deposition of ecdysteroid-phosphate conjugates in oocytes (Sonobe *et al.* 2006; Sonobe & Ito 2009). *BmEc22K* has yet to be studied with gene disruption techniques (i.e. CRISPR knockout or RNAi knockdown) and so it is unknown if it has important functions outside of reproduction and early embryogenesis. Another characterised EcKL is in the pine engraver beetle, *Ips pini* (Coleoptera: Scolytidae): *Ipi10G08*—encoding a 410 aa polypeptide (47.4 kDa)—was cloned by Bearfield *et al.* (2008) and is rapidly inducible by the juvenile hormone JH III in male and female adults. *Ipi10G08* is basally expressed in a variety of tissues, predominantly the fat body and appendages in males and the hindgut in females, but is expressed strongly in the gut after JH III exposure. Immunoblotting revealed an expected ~48 kDa band, but also a ~100 kDa band and a band larger than

250 kDa (which was also observed upon heterologous expression of the protein in bacteria), suggesting Ipi10G08 may form homomeric complexes. Immunoblotting also revealed that despite strong mRNA induction upon JHIII treatment, Ipi10G08 protein levels are much lower in females, suggesting post-transcriptional regulation may sex-bias the translation of the polypeptide. Bearfield *et al.* (2008) hypothesised that *Ipi10G08* may be involved in regulating the biosynthesis of male pheromones, given this occurs in the gut, is induced by JH III and involves sex-biased translation of biosynthetic gene transcripts (Keeling *et al.* 2006; Seybold & Tittiger 2003).

Not much is known about EcKLs in other insects, but members of the gene family have appeared in various published datasets. EcKL proteins are expressed in the columnar cell microvilli of the *M. sexta* 4th-instar larval midgut (Pauchet *et al.* 2009a) and 20 EcKL transcripts were identified in an EST library derived from midguts dissected from multiple larval instars of *M. sexta* (Pauchet *et al.* 2010); 21 EcKL transcripts were also identified in a similar EST library derived from 3rd-instar larval midguts of the poplar leaf beetle, *Chrysomela tremulae* (Pauchet *et al.* 2009b), and 16 EcKLs were identified in a 5th-instar larval midgut transcriptome of the rice stem borer *Chilo suppressalis* (Ma *et al.* 2012). A genomic locus containing four EcKLs appears to be under positive selection in the cotton bollworm, *Helicoverpa armigera*, subspecies *H. armigera conferta* (Anderson *et al.* 2018). EcKL family size has expanded in the broadly omnivorous german cockroach, *Blattella germanica*, compared to termites, a related wood-specialist group, and some genes in the family are differentially expressed between termite castes, raising the possibility they may be involved in regulating caste differences (Harrison *et al.* 2018).

1.4.3. What are the functions of EcKLs?

A fundamental question about any gene family is: what are the functions of its members? For enzymes such as the EcKLs, this inevitably relates to the types of reactions the enzyme can catalyse, as well as the substrates upon which they act. Due to their sequence similarity to choline/ethanolamine kinases and aminoglycoside phosphotransferases, as well as the confirmed enzymatic activity of one member (Sonobe *et al.* 2006), it is reasonable to speculate that all EcKL proteins are small-molecule kinases. However, it is theoretically possible that some EcKLs have protein kinase activity, given their relationship to protein kinases in the PKinase clan (El-Gebali *et al.*

2018).

The one experimentally known function of an EcKL is ecdysteroid kinase activity, as seen in BmEc22K (Sonobe *et al.* 2006), and it is likely that other members of the family are responsible for the ecdysteroid-phosphate conjugates and ecdysteroid kinase activities observed in other species (Chapter 1.3.6–1.3.7). In addition, there are likely undiscovered ecdysteroid-phosphate species and associated kinases that contribute to endocrinologically important processes such as ecdysteroid inactivation and recycling. As such, functionally characterising the EcKL gene family may shed light on poorly understood aspects of ecdysteroid biology and insect development and reproduction—this will be a major focus of this thesis.

In addition to this, preliminary searches of protein domain databases suggest that insects possess many multiple members of the EcKL family in their genomes, with 51 in *Drosophila melanogaster* alone—it seems unlikely that so many genes would be required for ecdysteroid kinase activity, given that there are only four positions on the ecdysteroid nucleus that can be phosphorylated (Fig. 1.6) and likely a relatively small number of ecdysteroid species to regulate in any particular organism (Lafont *et al.* 2012). The question then becomes: what other substrates might EcKL proteins have?

A possible answer to this question is that some EcKLs may phosphorylate xenobiotic compounds and toxins. As explored earlier (Chapter 1.2.8), insects phosphorylate a variety of xenobiotic compounds (Fig. 1.3), yet no detoxicative kinases have been genetically identified in any species. *D. melanogaster* EcKLs have been noted by various researchers to be transcriptionally induced after ingestion of phenobarbital (King-Jones *et al.* 2006), piperonyl butoxide (Willoughby *et al.* 2007), caffeine (Zhuo 2014) and secondary metabolites produced by the filamentous fungus *Aspergillus nidulans* (Trienens *et al.* 2017), and their taxonomic distribution (i.e. in insects but not mammals) is consistent with that of detoxicative phosphorylation (Mitchell 2015). The possible homology between aminoglycoside phosphotransferases—known detoxicative kinases in bacteria—and EcKLs provides further justification for taking this hypothesis seriously. Indeed, the idea that EcKL genes are involved in detoxification may have been first suggested by Lespinet *et al.* (2002), who note, likely referring to the EcKL family,

> *The...independent [expansion] of predicted small-molecule kinases related to ethanolamine and*

> *aminoglycoside kinases...in* D. melanogaster *[is] particularly enigmatic. Given the role of the related bacterial kinases...in xenobiotic resistance...these enzymes might be used to modify a range of xenobiotics encountered by the animals in their specific environments.*

Until now, there have been no follow-up studies to test their suspicions—another major focus of this thesis will be addressing this long-standing question in insect genetics and toxicology.

1.5. Summary, thesis aims and hypotheses

This first chapter has highlighted the current state of knowledge on, as well as outstanding questions in, three topics: gene families, xenobiotic detoxification and ecdysteroid hormones. As just discussed (Chapter 1.4), these converge on the focus of this thesis, the EcKL gene family, whose evolution, biochemistry and higher-level functions are poorly understood, despite these genes being mentioned in many dozens of studies over the past two decades.

The aim of this thesis is to further characterise the biology of the EcKL gene family in insects by testing two broad hypotheses regarding their function:

1. Some EcKLs mediate the catabolism and recycling of ecdysteroid hormones, and therefore play important roles in developmental endocrinology, and
2. Some EcKLs are responsible for the phosphorylation of dietary toxins seen in insects.

It is hoped that the pursuit of these hypotheses will shed light on not just a single gene family, but also the genomics, toxicology and development of insects.

In Chapter 2, largely published as a paper (Scanlan *et al.* 2020) in the journal *Insect Biochemistry and Molecular Biology*, I annotate the EcKL gene family in 12 species of the genus *Drosophila* and perform phylogenetic analyses to explore their evolution and classify individual genes into ancestral *Drosophila* clades. The evolutionary stability of these clades is integrated with transcriptomic data mined from publicly available databases and publications to develop a 'detoxification score' (DS) method for predicting the possible detoxification functions of genes in *D. melanogaster*. The DS method is validated on the known functions of a classical detoxification gene family, the cytochrome P450s, and predicts how many *D. melanogaster* EcKLs may be involved in detoxification in this species. In addition, a phenome-wide association study (PheWAS) in *D. melanogaster* identifies some novel detoxification-related phenotypes associated with uncharacterised P450s and EcKLs, laying the groundwork for future studies of these candidate detoxification genes.

In Chapter 3, I further test the detoxification hypothesis by annotating the EcKL gene

family in a wide selection of publicly available insect and arthropod genomes and perform detailed phylogenetic analyses to explore the evolution of EcKLs between and within insect taxa. Insect EcKLs are classified into subfamilies, and three orders—Diptera, Lepidoptera and Hymenoptera—have ancestral clades defined. These data are used to test associations between the size of the EcKL gene family and multiple detoxification-related traits in insects, including the sizes of known detoxification gene families, diet, and an estimation of detoxification breadth (DB). As a further test of the detoxification hypothesis, associations are also explored in herbivorous Lepidoptera between host plant diversity, a quantitative proxy for DB, and total EcKL gene family size and the sizes of specific lepidopteran ancestral EcKL clades. The stability of EcKL genes in taxa with small DB (tsetse flies and bees) is also contrasted with EcKLs in the large DB genus *Drosophila*.

In Chapter 4, I experimentally test the detoxification hypothesis by utilising the powerful reverse genetic toolkit available to *D. melanogaster*. Using CRISPR-Cas9 mutagenesis, I create deletion and loss-of-function alleles in a candidate clade of EcKLs identified in Chapter 2 and test the susceptibility of mutant animals to a number of plant and fungal secondary metabolites with plausible connections to *D. melanogaster* ecology. I also use UAS/GAL4 gene overexpression to test if increasing individual EcKL dosage correspondingly increases resistance to these toxins.

In Chapter 5, I experimentally test the ecdysteroid inactivation hypothesis by focusing on two paralogous candidate ecdysteroid kinase genes in *D. melanogaster*: *CG13813* (named here *Wallflower*/*Wall*) and *CG1561* (recently renamed *Pinkman*/*pkm*; Santana *et al.* 2020). Using CRISPR-Cas9 mutagenesis and UAS/GAL4 gene overexpression, I explore their possible roles in development and ecdysteroid biology. The hypothesis that *Wallflower* is an ecdysteroid 26-kinase is tested by epistasis experiments with *Cyp18a1*, and its ecdysteroid kinase activity is tested with co-misexpression of the ecdysteroid-phosphate phosphatase *CG13604*.

In Chapter 6, I integrate the phylogenomic and experimental results from previous chapters and discuss how they have shed light on the roles of the EcKL gene family in insects and other arthropods. I also propose new research questions that will hopefully be answered by future studies.

Chapter 2

Genomic, transcriptomic and evolutionary analyses of EcKLs in *Drosophila*

2.1. Chapter Introduction

The use of phosphorylation as a detoxification reaction has been observed in a taxonomically wide selection of insects, from Holometabola to Polyneoptera (Table 2.1; see Fig. 3.2 for insect taxonomic groups), suggesting this biochemical trait may be inherent to all insects. These detoxicative phosphorylation reactions are 'orphaned' (Lespinet & Labedan 2005), as they have yet to be associated with particular genes or enzymes. However, there are good reasons to suspect that the EcKL gene family is responsible for at least some of these reactions (Chapters 1.4.3 & 2.2).

In this chapter, I exploit the rich collection of genomic, transcriptomic and phenotypic data available in the dipteran genus *Drosophila*, and particularly *Drosophila melanogaster*, to test the hypothesis that some EcKLs are involved in detoxification, by analysing their evolutionary and transcriptomic characteristics and comparing them with the cytochrome P450s, a well-established detoxification gene family.

The bulk of this chapter (Chapters 2.2–2.6 & 2.8) has been published in the journal *Insect Biochemistry and Molecular Biology* as 'Genomic and transcriptomic analyses in *Drosophila* suggest that the ecdysteroid kinase-like (EcKL) gene family encodes the "detoxification-by-phosphorylation" enzymes of insects' (Scanlan *et al.* 2020). The content of that publication is presented here with minor modifications to match the style and format of this thesis, with additional introductory (Chapter 2.1) and discussion (Chapter 2.7) sections.

Table 2.1. Phosphorylated metabolites of xenobiotic compounds in insects reported in the literature.

Phosphorylated compounds	Species	Order	References
p-nitrophenol	*Gromphadorhina portentosa*	Blattodea	Yang & Wilkinson 1973; Gil *et al.* 1974
p-nitrophenol, p-nitrophenol glucoside	*Periplaneta americana*	Blattodea	Ngah & Smith 1983
1-napthol, 2-napthol, p-nitrophenol, 1-napthol glucoside, p-nitrophenol glucoside; possibly phenolphthalein, m-aminophenol, 8-quinolinol and 4-methylumbelliferone	*Costelytra zealandica*	Coleoptera	Darby *et al.* 1966; Binning *et al.* 1967; Heenan & Smith 1974
p-nitrophenol, p-nitrophenol glucoside	*Tenebrio molitor*	Coleoptera	Ngah & Smith 1983
p-nitrophenol, p-nitrophenol glucoside	*Anisolabis littorea*	Dermaptera	Ngah & Smith 1983
p-nitrophenol, p-nitrophenol glucoside	*Calliphora vicina*	Diptera	Ngah & Smith 1983
harmol	*Drosophila melanogaster*	Diptera	Baars *et al.* 1980
1-napthol, 2-napthol, p-nitrophenol, 1-napthol glucoside, p-nitrophenol glucoside; possibly phenolphthalein, m-aminophenol, 8-quinolinol and 4-methylumbelliferone	*Lucilia sericata*	Diptera	Darby *et al.* 1966; Binning *et al.* 1967; Heenan & Smith 1974; Ngah & Smith 1983
1-napthol, 2-napthol, p-nitrophenol, 1-napthol glucoside, p-nitrophenol glucoside; possibly phenolphthalein, m-aminophenol, 8-quinolinol and 4-methylumbelliferone	*Musca domestica*	Diptera	Darby *et al.* 1966; Binning *et al.* 1967; Heenan & Smith 1974; Ngah & Smith 1983
p-nitrophenol, p-nitrophenol glucoside	*Galleria mellonella*	Lepidoptera	Ngah & Smith 1983
salicin, arbutin, helicin, phenol glycoside, catechol glucoside	*Lymantria dispar*	Lepidoptera	Boeckler *et al.* 2016
p-nitrophenol	*Manduca sexta*	Lepidoptera	Yang & Wilkinson 1973
salicin, catechol glucoside	*Orgyia antiqua*	Lepidoptera	Boeckler *et al.* 2016
p-nitrophenol	*Wiseana cervinata*	Lepidoptera	Ngah & Smith 1983
20-hydroxyecdysone 3-acetate	*Locusta migratoria*	Orthoptera	Modde *et al.* 1984
terfenadine, oxidised terfenadine metabolites, midazolam glucoside	*Schistocerca gregaria*	Orthoptera	Olsen *et al.* 2014; Olsen *et al.* 2015
p-nitrophenol	*Acanthoxyla intermedia*	Phasmatodea	Ngah & Smith 1983
p-nitrophenol	*Clitarchus* sp.	Phasmatodea	Ngah & Smith 1983

2.2. Paper Introduction

Toxins play central roles in competition and trophic interactions between species—this is especially true for insects, where they often define ecological niches and can limit the food sources species can exploit. A well-known example is the wide diversity of toxins produced by plants, which aim to kill or otherwise dissuade herbivorous insects feeding on their tissues; the capacity of an insect to tolerate these toxins partially defines which plants it can consume (Mithöfer & Boland 2012) and contributes to plant-insect co-evolution (Edger *et al.* 2015; Ehrlich & Raven 1964). Toxin tolerance has many components, one of the most well studied of which is metabolic detoxification. In animals, detoxification is carried out by collections of enzyme and transporter systems present in a number of organs and tissues throughout the body, which were originally conceptualised as a series of 'phases' that result in the sequential modification (phase I), conjugation (phase II) and excretion (phase III) of toxins (Williams 1959). After decades of pharmacological and biomedical research, the specific phase I and II reactions and enzymes present in mammals and other vertebrates are well known, many of which have also been identified in insects (Berenbaum & Johnson 2015; Chahine & O'Donnell 2011; Yu 2008). However, given that the lineages leading to arthropods and vertebrates diverged approximately 600 m.y.a. (Reis *et al.* 2015) and many gene families are not conserved over deep evolutionary time (Danchin *et al.* 2006; Lespinet *et al.* 2002), it is likely that those involved in detoxification may differ between these taxa.

One such difference is xenobiotic phosphorylation, which is rare in mammals but common in bacteria and insects (Mitchell 2015; Ramirez & Tolmasky 2010; Wilkinson 1986). Phosphorylation has the potential to be a phase II detoxification reaction, as phosphate groups are highly polar and can be conjugated to hydroxyl moieties, and the formation of phosphorylated metabolites of xenobiotic compounds has been observed in at least 18 insect species across seven insect orders (Table 2.1). Many of these metabolites were discovered by John Smith and colleagues in the mid 1960s to early 1970s (Binning *et al.* 1967; Darby *et al.* 1966; Heenan & Smith 1974; Smith & Turbert 1964) before the widespread adoption of modern analytical methods like LC-MS, but more recent papers have characterised phosphate conjugates in some detail (Olsen *et al.* 2015; 2014). An example can be found in the caterpillars of the gypsy moth,

Lymantria dispar, which phosphorylate the glycoside moiety of salicinoids found the leaves of a host plant, *Populus tremula* x *tremuloides*, a hybrid poplar tree. These phosphate conjugates—formed in the gut and perhaps also the Malpighian tubules—comprise a substantial proportion of excreted salicinoid-like compounds, especially when caterpillars are previously fed poplar leaves, suggesting this detoxification process is induced by poplar secondary metabolites (Boeckler *et al.* 2016). Likewise, the detoxicative kinase activity identified in *Gromphadorhina portentosa*, the Madagascar cockroach, is present in the midgut, fat body and Malpighian tubules of the insect and is inducible by *in vivo* exposure to phenobarbital (Gil *et al.* 1974; Yang & Wilkinson 1973), a compound commonly used to induce detoxification gene expression. Phytoecdysteroid detoxification may also involve phosphorylation (Rharrabe *et al.* 2007), but the phosphorylation of ingested ecdysteroids has been studied in only one species (Modde *et al.* 1984), even though ecdysteroids can be phosphorylated *in vitro* with midgut tissue homogenates from a handful of other species (Webb *et al.* 1996; 1995; Weirich *et al.* 1986). Overall, many insects appear able to phosphorylate xenobiotic phenols, glycosides and/or steroids directly, and it is possible other xenobiotics may be metabolised in a similarly direct manner, or after hydroxylation, hydrolysis or glycosylation. To date, however, no detoxicative phosphotransferase enzymes have been cloned or otherwise identified at the genetic level. In this paper, we wish to highlight this outstanding question in insect toxicology and raise a hypothesis about the identity of these unknown enzymes.

The identification of detoxification genes and enzymes is an important part of bridging the gap between toxicology, chemical ecology and functional genomics in insects. In the past, the main method of finding detoxification enzymes involved cloning and biochemical characterisation based on a known detoxification reaction. However, detoxification genes tend to have other characteristic properties, including transcriptional induction by xenobiotics (Willoughby *et al.* 2006) and enriched expression in tissues with known detoxification roles, like the midgut, Malpighian tubules and fat body (Yang *et al.* 2007). Phenotypes discovered during genetic experiments can also indicate detoxification functions, including susceptibility via gene disruption (Wang *et al.* 2018) or tolerance via overexpression (Daborn *et al.* 2007); genome- and transcriptome-wide association studies can also identify candidate genes for detoxification-related phenotypes using naturally occurring variation (Robin *et al.* 2019). Additionally, genes encoding detoxification enzymes are thought to undergo gene duplication and

loss at a faster rate than those encoding enzymes with important housekeeping functions (Kawashima & Satta 2014; Thomas 2007). The broad availability of '-omic' data in some insect taxa, particularly the *Drosophila* genus, raises the possibility that candidate detoxification genes could be identified by integrating evolutionary and transcriptomic data without any prior knowledge of biochemical functions. However, to our knowledge, this has yet to be attempted in a systematic way.

To validate this approach, we examine the cytochrome P450s (henceforth P450s), which are a large multigene family of enzymes that largely function as monooxygenases and catalyse phase I detoxification reactions such as hydroxylation, although a subset also catalyse a wider variety of reactions, including dealkylation, epoxidation and reduction (Bernhardt 2006). P450s are an established detoxification family in virtually all animals, including insects (Heidel-Fischer & Vogel 2015; Yu 2008), and the number of P450 genes per genome in insects varies dramatically, from 38 in the fig wasp *Ceratosolen solmsi* to 222 in the little fire ant *Wasmannia auropunctata* (Rane *et al.* 2019). It has been suggested that diversity in the size of the P450 gene family may be linked to differences in detoxification capacity between taxa (Calla *et al.* 2017; Rane *et al.* 2019; 2016), although this has not been rigorously studied.

Multiple attempts have been made to classify P450s according to their biological functions. Thomas (2007) suggests a split between 'endogenous-substrate' and 'xenobiotic-substrate' enzymes, while Kawashima & Satta (2014) suggest a similar split with 'biosynthesis-type' and 'detoxification-type'. For the purposes of this paper, we adopt the classification system proposed by Gotoh (2012) of xenobiotic (X-class), secondary (S-class) and endogenous (E-class) functions. E-class enzymes synthesise or degrade compounds that are important for developmental or physiological processes, such as hormones and cuticular hydrocarbons; S-class enzymes synthesise or degrade secondary metabolites, such as defensive compounds or pigments; and X-class enzymes detoxify xenobiotic compounds found in the diet or otherwise derived from the environment. E- and X-class P450s have been studied for many decades. A classic example of E-class enzymes are the Halloween P450s, which synthesise ecdysteroid moulting hormones from dietary sterols (Rewitz *et al.* 2007), but other E-class P450s belong to the biosynthetic pathways of juvenile hormones (Christesen *et al.* 2017; Helvig *et al.* 2004a) and cuticular hydrocarbons (Qiu *et al.* 2012). Many X-class P450s have also been characterised and implicated in the detoxification of both natural and synthetic toxins (Li

et al. 2007). S-class P450s are the least understood of the three classes, although some are involved in the biosynthesis of cyanogenic glycosides (Beran *et al.* 2019) and the degradation of pheromones (Wojtasek & Leal 1999).

The evolutionary dynamics of P450s have been well studied in insects and the *Drosophila* genus specifically (Feyereisen 2011; 2006; Good *et al.* 2014). Within *Drosophila*, 30 ancestral P450 clades are stable—that is, they contain 1:1 orthologs in all studied species—while 30 clades have gene gain and gene loss in the genus, and 17 have only gene loss (Good *et al.* 2014). The genome of *Drosophila melanogaster* specifically contains 87 P450 genes, some of which have been studied in great detail (Chung *et al.* 2009). While evolutionary stability and developmentally essential (E-class) functions are thought to be linked in *D. melanogaster* (Chung *et al.* 2009), the link between evolutionary instability and X-class or S-class functions has yet to be rigorously established in this species.

The ecdysteroid kinase-like (EcKL; Interpro entry IPR004119) gene family is taxonomically restricted, being predominantly present in insect and crustacean genomes (Mitchell *et al.* 2014). EcKL enzymes are predicted to conjugate phosphate to secondary alcohols, using ATP as a phosphodonor (EC 2.7.1.-) and contain the EcKinase domain (Pfam accession PF02958), formerly known as DUF227 (domain of unknown function 227), a member of the Protein Kinase superfamily (CDD accession cl21453; (El-Gebali *et al.* 2018). This superfamily contains protein kinases, as well as kinases with small molecular substrates, such as choline/ethanolamine kinases, aminoglycoside 3′-phosphotransferases and phosphoinositide 3-kinases (Marchler-Bauer *et al.* 2015). The EcKLs are currently named after a single member, *BmEc22K*, which encodes an ecdysteroid 22-kinase in the silkworm, *Bombyx mori*. BmEc22K phosphorylates the C-22 hydroxyl group of ecdysteroids, producing physiologically inactive ecdysteroid 22-phosphate conjugates (Sonobe *et al.* 2006). These conjugates are stored in the oocyte where they bind to vitellin in the yolk, and are hydrolysed to their active free form by an ecdysteroid-phosphate phosphatase (EPPase) after fertilisation to supply the ecdysteroid titre required for embryonic development (Sonobe & Yamada 2004; Yamada *et al.* 2005; Yamada & Sonobe 2003). This reciprocal conversion process may also occur in other insects and crustaceans, as orthopteran and other lepidopteran species also store ecdysteroid-phosphate conjugates in their eggs (Feldlaufer *et al.* 1987; Isaac *et al.* 1983a; Isaac & Rees 1984; Sonobe & Ito 2009), as may some crustaceans (Subramoniam

2000; Young *et al.* 1991). However, no ecdysteroid kinase/EPPase system has been molecularly characterised in any species besides *B. mori*, although EPPase orthologs exist in many insect genomes (Sonobe & Ito 2009) and at least one crustacean genome (Asada *et al.* 2014).

No EcKLs besides BmEc22K have had their substrates identified, and very few other EcKLs have been functionally characterised. *Juvenile hormone-inducible protein 26* (*JhI-26*) is an EcKL found in *D. melanogaster* that has been implicated in *Wolbachia*-mediated cytoplasmic incompatibility (Liu *et al.* 2014). *CHKov1* and *CHKov2* are also EcKLs in *D. melanogaster*; a *CHKov1* allele containing a transposable element (TE) insertion confers resistance to the vertically-transmitted sigma virus, and a derived allele containing complete and partial duplications of both the *CHKov1*-TE allele and its neighbouring gene *CHKov2* confers an even greater level of resistance (Magwire *et al.* 2011). The mechanism underlying the resistance conferred by *CHKov1*-TE is currently unknown, although it is important to note that the TE insertion likely destroys the kinase function of the encoded polypeptides.

We raise the hypothesis that members of the EcKL gene family encode kinases responsible for the detoxicative phosphorylation seen in insects, based on four observations: first, the apparent taxonomic distribution of EcKLs is consistent with the limited taxonomic distribution of detoxicative phosphorylation in animals (Mitchell 2015; Chapter 3); second, ecdysteroid kinase activity has been linked to detoxicative phosphorylation of phytoecdysteroids in some insects (Rharrabe *et al.* 2007); third, the size of the EcKL family appears to vary considerably between insect taxa (Mitchell *et al.* 2014; Chapter 3), suggesting not all members encode E-class enzymes; and fourth, EcKLs are at least distantly related to the aminoglycoside 3'-phosphotransferases, known detoxicative phosphotransferase enzymes (Marchler-Bauer *et al.* 2015).

Here we conduct the first evolutionary analysis of the EcKL gene family, focused on the dipteran genus *Drosophila*, demonstrating there is wide variability in the stability of EcKL orthologs within this taxon. We then show that integrating evolutionary, xenobiotic induction, transcriptional regulation and tissue expression datasets can be used to predict which members of the P450 gene family are involved in xenobiotic detoxification, and apply this method to the EcKLs, suggesting they also contribute to the insect detoxification system. We also show, using a targeted phenome-wide

association study (PheWAS) approach in the Drosophila Genetic Reference Panel (DGRP), that EcKL and P450 genomic and transcriptomic variation is associated with toxic stress phenotypes, providing candidate detoxification functions for members of these two gene families. Finally, we perform RNAi knockdown on a subset of EcKLs in *D. melanogaster* to find developmental lethality phenotypes and identify candidate E-class members of this gene family.

2.3. Materials and Methods

2.3.1. Gene family annotation

Eleven *Drosophila* genomes were accessed from NCBI (Coordinators 2016; Drosophila 12 Genomes Consortium 2007)—*D. simulans* (GCA_000259055.1), *D. sechellia* (GCA_000005215.1), *D. erecta* (GCA_000005135.1), *D. yakuba* (GCA_000005975.1), *D. ananassae* (GCA_000005115.1), *D. pseudoobscura* (GCA_000149495.1), *D. persimilis* (GCA_000005195.1), *D. willistoni* (GCA_000005925.1), *D. mojavensis* (GCA_000005175.1), *D. virilis* (GCA_000005245.1) and *D. grimshawi* (GCA_000005155.1)—while *D. melanogaster* (Release 6) was accessed from FlyBase (Thurmond *et al.* 2019). Genomic scaffolds containing EcKLs were identified with TBLASTN searches (Altschul *et al.* 1990) using *D. melanogaster* EcKL protein sequences as queries and were annotated in Artemis (Carver *et al.* 2012). Gene model annotation was performed by reciprocally performing BLASTP and TBLASTN searches (Altschul *et al.* 1990) between putative gene models and *D. melanogaster* proteins with the highest sequence similarity. Annotated EcKL gene models in FlyBase in non-*D. melanogaster* species were used as starting points but were not assumed to be correct. EcKL domain completeness in translated gene models was checked with searches to the Pfam database (Punta *et al.* 2012). Gene models were considered pseudogenes if more than one inactivating mutation (frameshifts or premature stop codons) were found; gene models with only one inactivating mutation were considered null alleles, the stop codons of which were ignored for phylogenetic analyses or reverted to the most likely previous codon with a single nucleotide change. Gene models with missing exons due to incomplete genome assembly were considered 'partial' genes and were assumed to be functional and not pseudogenes. Gene models were mapped to Muller elements (chromosome arms in *Drosophila*) using data from Schaeffer *et al.* (2008).

2.3.2. Phylogenetic analyses

Predicted protein sequences of EcKLs were aligned with MAFFT 7.402 (Katoh & Standley 2013); proteins containing two domains were split into N- and C-terminal halves. MSAs were manually trimmed at the N- and C-termini to remove poorly-aligned regions using AliView (Larsson 2014), and columns containing non-gap characters from only one sequence were also removed. Gene trees were generated with

IQ-TREE v1.6.10 using ModelFinder to automatically pick the most appropriate model for the data (Kalyaanamoorthy *et al.* 2017; Nguyen *et al.* 2015), with the following command:

```
iqtree_1.6.10_comet -s infile.txt -bb 10000 -bnni -st AA -nm 50000 -msub nuclear -nt AUTO -pre output -m TESTNEW
```

IQ-TREE was run five times on each alignment and the tree with the highest log-likelihood was used. MAFFT and IQ-TREE runs were conducted through the CIPRES Science Gateway (Miller *et al.* 2010).

A gene family clade was defined as a group of genes that all descended from a single gene inferred to exist in the most recent common ancestor of all 12 *Drosophila* species ($MRCA_D$) based on parsimony, while a subclade was defined as a group of genes that all descended from a single gene inferred to be the product of a duplication event after the $MRCA_D$, if that duplication event occurred in a common ancestor of at least two of the 12 *Drosophila* species. Clade and subclade nomenclature was developed to help refer to groups of orthologous genes, and individual genes have a clade ID in the format 'DroC-S', where C is the clade number and S is the subclade number. If a gene does not belong to a subclade, its subclade number is 0. This nomenclature is cladistic and based purely on inferred evolutionary relationships within the *Drosophila* genus, and clade IDs are not intended to be official gene names.

Ancestral clades were demarcated into four stability categories—stable (1:1 orthologs in every genome), blooming (at least one duplication, no complete losses), wilting (at least one complete loss, no duplications) and mixed (at least one duplication, at least one complete loss). For some downstream analyses, blooming, wilting and mixed clades were categorised as 'unstable' clades. A gene loss event was defined as the loss (absence or pseudogenisation) of a gene lineage that existed in a species lineage after the $MRCA_D$. A complete clade loss event was defined as the loss any of that clade's genes in a particular species or lineage. A duplication event was defined as any inferred duplication of a gene that occurred after the $MRCA_D$.

2.3.3. DGRP PheWAS

One hundred and forty six DGRP phenotypes were collected from 32 publications

(Table S2.2) and aligned by DGRP line using the merge function in R (R Core Team 2019). Abbott's correction for control mortality was applied to methylmercury and caffeine phenotypes from Montgomery *et al.* (2014). DGRP genotypes, transcriptomes and annotations were recovered from the DGRP website (http://dgrp2.gnets.ncsu.edu/data.html). DGRP genotypes were first filtered for variants with a minor allele frequency > 0.05, and then for whether variants had been annotated to within—or within 1 kb of—EcKL genes or cytochrome P450 genes, resulting in 2,472 and 5,938 variants, respectively. Male and female transcriptomes were filtered for transcripts from EcKL or cytochrome P450 genes, including both copies of *Cyp12d1* found in the DGRP (*Cyp12d1-p* and *Cyp12d1-d*). PheWAS were performed by fitting a linear model in R between each variant or transcript and each DGRP phenotype. Genomic and transcriptomic associations were filtered using significance thresholds of $p < 1\times10^{-5}$ and $p < 1\times10^{-3}$ respectively.

2.3.4. Tissue-specific transcriptome data

Tissue-specific RNA-seq data in *D. melanogaster* were mined for the EcKL and P450 families from FlyAtlas 2 (Leader *et al.* 2018). Tissue 'enrichment' was defined as (tissue FPKM + 1)/(whole body FPKM + 1) for each gene—this is roughly the ratio of tissue expression to whole body expression, but avoids undefined values and moves enrichment scores towards 1 for genes where the absolute difference between tissue and whole body FPKMs is small, effectively deweighting enrichment when expression measurement uncertainty is high and/or expression is less likely to be biologically relevant. Genes with enrichment values greater than or equal to 2 (expression ~≥ 2-fold higher compared to whole body) were considered 'enriched' in a tissue, while genes with enrichment values less than 2 were considered 'not enriched'.

2.3.5. Gene induction datasets

Seven xenobiotic differential gene expression datasets were mined for EcKL and P450 genes: phenobarbital in 3rd instar larvae (Sun *et al.* 2006); phenobarbital in adult flies (King-Jones *et al.* 2006; Misra *et al.* 2011); piperonyl butoxide in adult flies (Willoughby *et al.* 2007); methamphetamine in adult flies (Sun *et al.* 2011); tunicamycin (8 hr timepoint) in adult flies (Chow *et al.* 2013); and fungal toxins (6 hr timepoint, wild-type vs. *ΔlaeA Aspergillus nidulans*) in 1st-instar larvae (Trienens *et al.* 2017). We chose these datasets because they contained a 4-8 hr post-exposure timepoint (Willoughby

et al. 2006), involved ingested single toxins or ecologically appropriate mixtures of toxins, were on larval or adult *D. melanogaster*, and had a genome-wide focus. Genes were considered induced in a dataset if they were up-regulated ≥ 1.5-fold with a reported p-value < 0.05, except in the case of Chow *et al.* (2013), where their average up-regulation across all 20 lines needed to be ≥ 1.5-fold. Genes not meeting these criteria were considered not induced. Genes were considered induced (positively regulated) by CncC if they were up-regulated ≥ 1.5-fold ($p < 0.05$) upon ectopic expression of CncC in adult male flies, as reported by Misra *et al.* (2011), otherwise they were considered not induced; we did not consider repression in *CncC* mutant animals (Deng & Kerppola 2013) evidence for detoxification function due to CncC's role in developmental gene expression (Deng & Kerppola 2013).

2.3.6. Detoxification scores

D. melanogaster genes in the EcKL and P450 gene families were each given 'detoxification scores' from 0 to 4 based on four criteria (1 point was given for each criterion met): membership in a gene clade that is unstable in the *Drosophila* genus; induction in at least one xenobiotic induction dataset; induction by ectopic expression of CncC; and enrichment in at least one detoxification tissue (midgut, Malpighian tubules and/or fat body) at one or more life stages (3rd-instar larva, adult female and/or adult male). Data on the stability of P450 genes in *Drosophila* were taken from Good *et al.* (2014).

2.3.7. Review of published P450 functions

Published data on *D. melanogaster* P450 gene functions were collated from both FlyBase release FB2019_05 (Thurmond *et al.* 2019) and direct searches for each gene name in the literature. Asserted gene functions were only recorded if they were supported with functional genetic or biochemical evidence (gene induction alone was not considered sufficient), and this evidence was rated as 'weak' (RNAi-only characterisation), 'moderate' (gene disruption and/or transgenic overexpression) or 'strong' (direct biochemical evidence, such as enzyme assays) depending on the experimental methods used. Unvalidated GWAS or TWAS associations with toxic stress phenotypes were not considered functional evidence for these purposes.

2.3.8. RNAi knockdown

tubulin-GAL4/TM3, *act-GFP*, *Ser*^{1} females, which strongly express GAL4 in all tissues, were crossed to *UAS-dsRNA* males and the offspring were phenotypically scored for the presence or absence of the TM3, *act-GFP*, *Ser*^{1} balancer chromosome. Significant deviation in genotypic ratios towards balancer-containing individuals was considered evidence for partial or complete developmental lethality associated with the inheritance of the *tubulin-GAL4* construct and the expression of the *UAS-dsRNA* construct. All fly crosses were conducted on yeast-cornmeal media at 25 ºC. dsRNA responder lines were obtained from the Vienna Drosophila Resource Center GD and KK libraries (www.vdrc.at; Table S2.5); KK lines were PCR genotyped for their hairpin landing site as per Green *et al.* (2014)—lines with annotated site insertions upstream of *tiptop* (which can produce dominant lethal effects) have been noted (Fig. S2.8). The *tubulin-GAL4*/*actGFP*, *Ser*^{1} line was a gift from Philip Batterham (The University of Melbourne). The 40D line was a gift from Kieran Harvey (Peter MacCallum Cancer Centre).

2.3.9. Statistical analyses

Interactions between xenobiotic induction, CncC induction and clade instability were modelled as a homogeneous association loglinear model with *glm* in R. Differential tissue enrichment between xenobiotic-induced/-uninduced, CncC-induced/-uninduced and stable/unstable genes was determined by comparing the median $\log_2$(enrichment) between groups of genes using bootstrap estimation with the *dabestr* package in R (Ho *et al.* 2019). Effect sizes with 95% confidence intervals that did not include 0 were considered significant. The diagnostic test for the sensitivity and specificity of the DS method was performed with the 'diagnostic' function in the *ThresholdROC* package (v2.7) in R. RNAi knockdown developmental lethality was assessed by comparing offspring genotypic ratios with the 'binom.test' function in R, with a Bonferroni correction for multiple tests.

2.4. Results

2.4.1. Evolution of the EcKL gene family in the *Drosophila* genus

We explored the evolution of the EcKLs in the *Drosophila* genus by collating a dataset of gene models in the genomes of 12 species (Table S2.1). We manually annotated 564 EcKL gene models in the eleven non-*D. melanogaster* species in the *Drosophila* genus, to produce a dataset of 618 total EcKL gene models across the 12 species—of these, 605 are likely full gene models (of which 21 were considered likely pseudogenes), while 13 were partial gene models with clear missing sequence due to genome assembly incompleteness. The total number of putatively functional EcKLs (full and partial genes) in the genome of each species varies considerably; *D. willistoni* has the most with 61, while *D. mojavensis* has the least with 41 (Fig. 2.1). There are also differences between closely related species, the most striking of which is the nine-gene difference between *D. simulans* (56 genes) and *D. sechellia* (47 genes). Two of these nine genes are missing from the assembly, while the remaining seven appear to have been pseudogenised in the *D. sechellia* lineage.

By examining gene trees and using the known species phylogeny, we inferred the existence of 46 ancestral EcKL clades, which each represent a single EcKL that was present in the genome of the $MRCA_D$. The vast majority of these clades have very strong support, however the reconstruction of the Dro26 clade is currently tentative and it may need to be split into smaller clades in the future. 18 clades (39%) are stable across the *Drosophila* genus, while 28 (61%) are unstable; of the unstable clades, 14 are blooming (at least one duplication after the $MRCA_D$), 10 are wilting (at least one complete clade loss after the $MRCA_D$) and four are mixed (at least one duplication and one complete clade loss after $MRCA_D$; Figs. 2.2 and 2.3). The unstable clades are not all equally labile; while the mean number of inferred duplications and losses per unstable clade is 2.39 and 2.54 respectively, there are substantial outliers (Fig. 2.2). The Dro5 clade has experienced 20 duplications (and 10 losses) across the genus, and also has the highest number of genes in a single genome, with nine in *D. simulans*, while the Dro1 clade has the largest number of losses with 11.

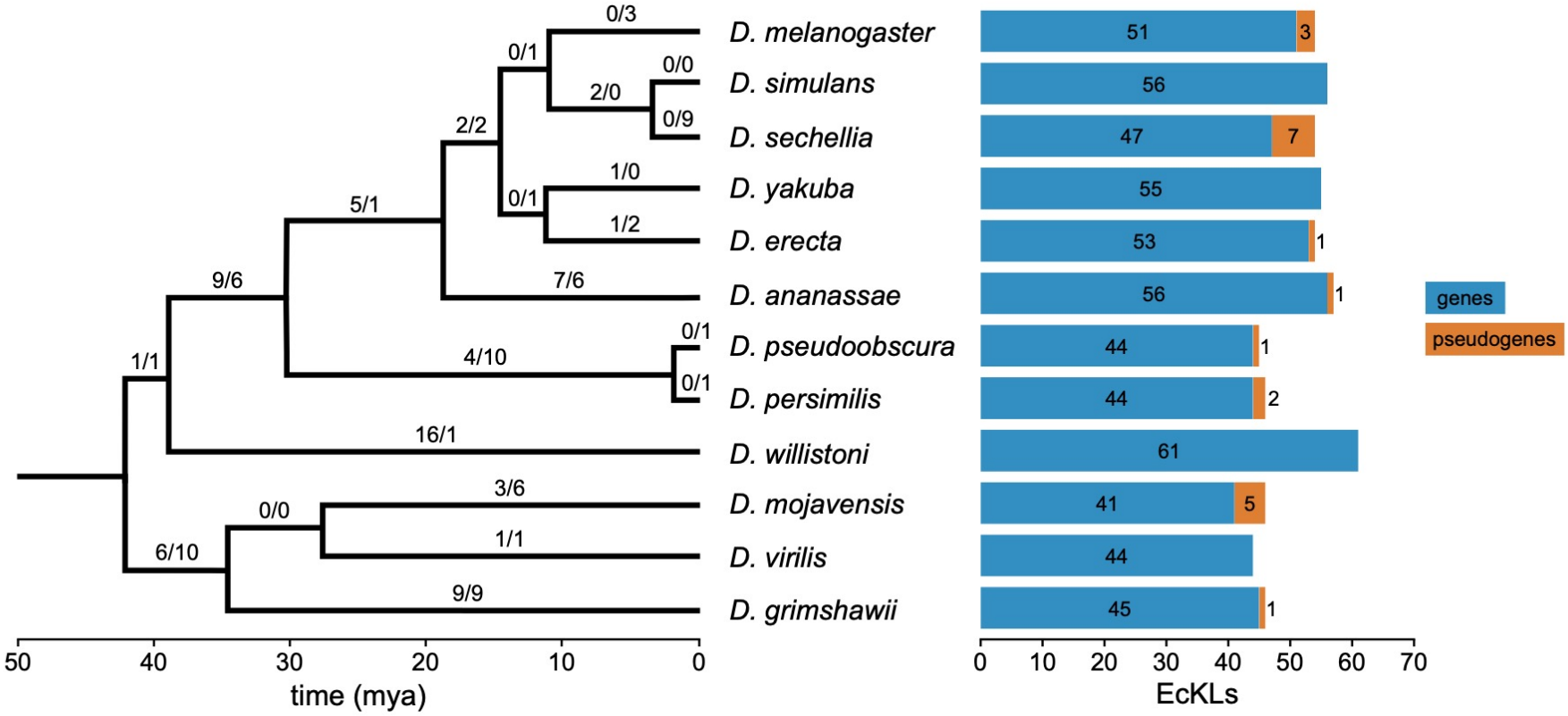


Figure 2.1. The number of EcKL genes (blue) and pseudogenes (orange) in the genomes of 12 *Drosophila* species. Numbers above the branches of the phylogenetic tree represent inferred gene gains and losses along that branch. Species tree from FlyBase (Thurmond *et al.* 2019).

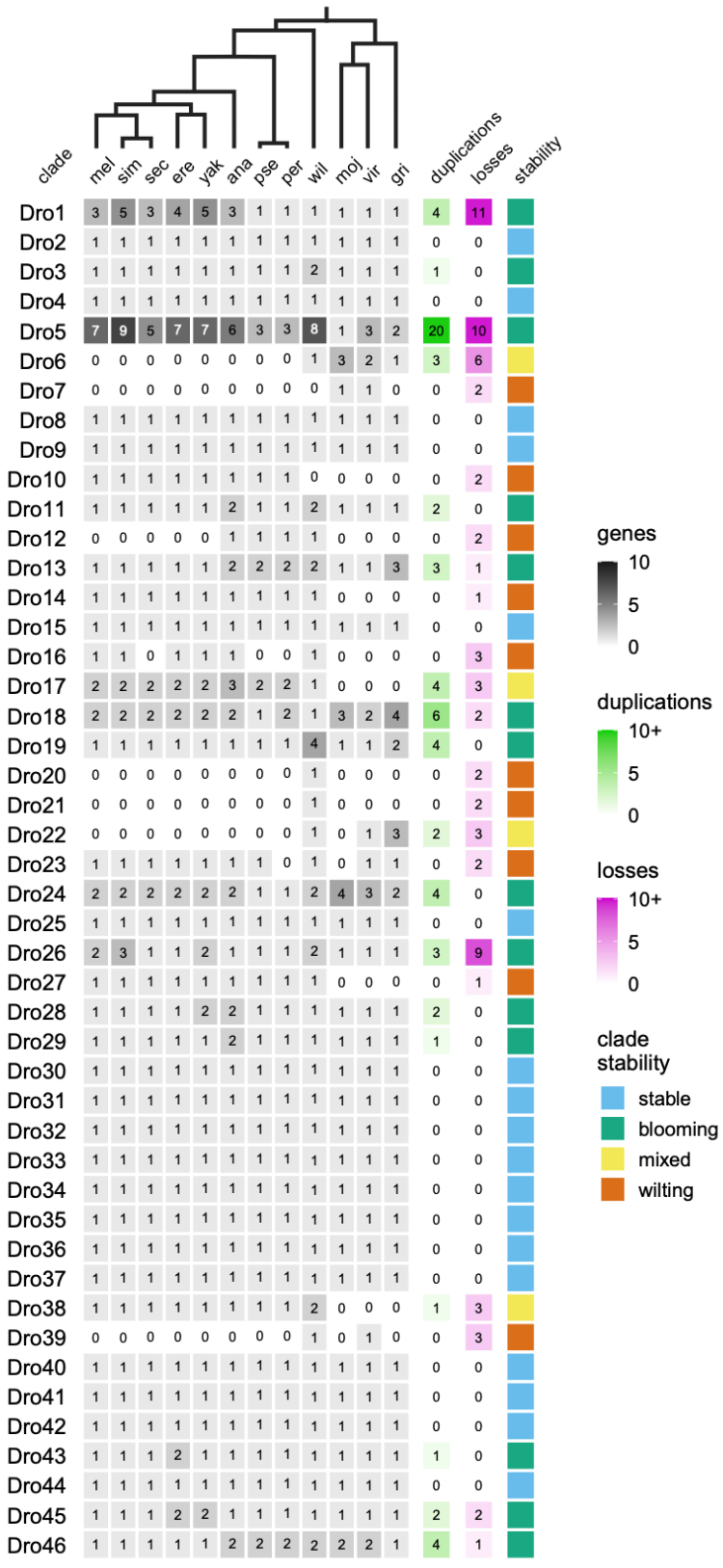


Figure 2.2. Number of EcKLs in each ancestral clade in each species of *Drosophila*, along with the inferred number of duplications and losses in—and the consequent stability of—each clade in the *Drosophila* genus.

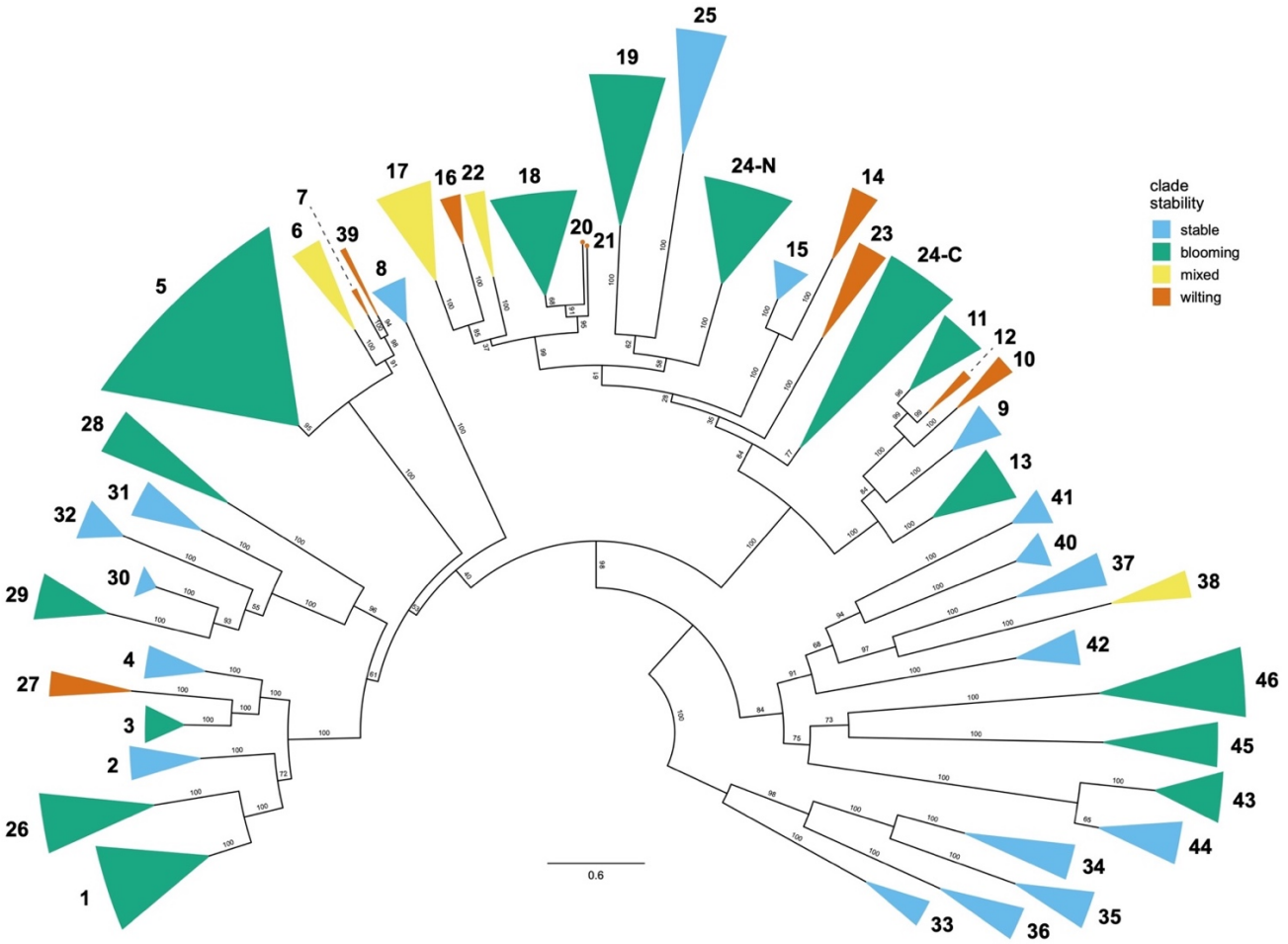


Figure 2.3. Gene tree of ancestral clades in the *Drosophila* genus, coloured by stability, and numbered by clade number; the tree was arbitrarily rooted on the branch leading to the Dro33, Dro34, Dro35 and Dro36 clades to facilitate its presentation, but the root has been omitted. Numbers on branches indicate bootstrap support values. The Dro24 clade is split into N- and C-terminal halves, 24-N and 24-C. The Dro20 and Dro21 clades comprise a single gene each, indicated by small circles.

We also found that EcKLs in *Drosophila* tend to cluster in uninterrupted tandem arrays—as is the case in other detoxification gene families (Feyereisen 2011; Friedman 2011; Robin *et al.* 1996)—the largest of which is on Muller element E (chromosome 3R in *D. melanogaster*) and contains 26 genes and two pseudogenes (encompassing all the genes in the Dro1 through Dro19 clades) in *D. melanogaster*, although this has been split into two clusters in the Drosophila subgenus, possibly by a chromosomal inversion or other rearrangement. This suggests tandem duplication is a significant cause of EcKL family expansion over time, and as many of these clusters contain genes from multiple clades, many of these duplications happened before the $MRCA_D$. However, there are also instances of genes in the same clade found on different Muller elements, suggesting other processes, such as RNA-mediated gene duplication, transposition and translocation, may also contribute to gene family dynamics in the EcKLs.

2.4.2. EcKL and P450 genes are transcriptionally enriched in detoxification tissues in *D. melanogaster*

Detailed tissue-specific gene expression data in *D. melanogaster* published in FlyAtlas 2 (Leader *et al.* 2018) allowed us to explore the patterns of gene expression in both the EcKL and P450 families. FlyAtlas 2, at the time of writing, contains data from 18 tissues (head, eye, CNS, thoracicoabdominal ganglion, crop, midgut, hindgut, Malpighian tubules, fat body, salivary gland, trachea, ovary, virgin spermatheca, mated spermatheca, testis, accessory glands, carcass and rectal pad) and whole body samples, at between one and three life stages (3rd instar larvae, adult males and adult females). Heatmaps of absolute expression (FPKM) for EcKLs and P450s can be found in Figs. S2.1 and S2.2, respectively. However, FPKM is a limited measure when comparing between genes, as some genes might require different levels of absolute expression to achieve similar functions (e.g. the enzymes they encode may have different kinetic properties, or the transcribed or translated products may have their activities post-transcriptionally or post-translationally modified). As a more useful measure of tissue-specificity, we calculated the enrichment of each gene in each tissue using whole body expression at each life stage, where a greater enrichment value indicates more specific expression in that tissue, and an enrichment value of greater than or equal to 2 (or $\log_2(\text{enrichment}) \geq 1$) indicates 'enrichment' (Figs. S2.3–2.4). As expected (Chung *et al.* 2009; Yang *et al.* 2007), many P450 genes are enriched in the midgut, Malpighian tubules and fat body at various life stages, which are widely considered detoxification

tissues (Fig. 2.4A), but more EcKLs, proportionally, are enriched in these tissues, excepting the larval midgut: 10 (20%), 30 (59%) and 29 (57%) EcKLs are enriched in the midgut, and 30 (59%), 36 (71%) and 31 (61%) EcKLs are enriched in the Malpighian tubules, in larvae, adult females and adult males, respectively (Fig. 2.4B).

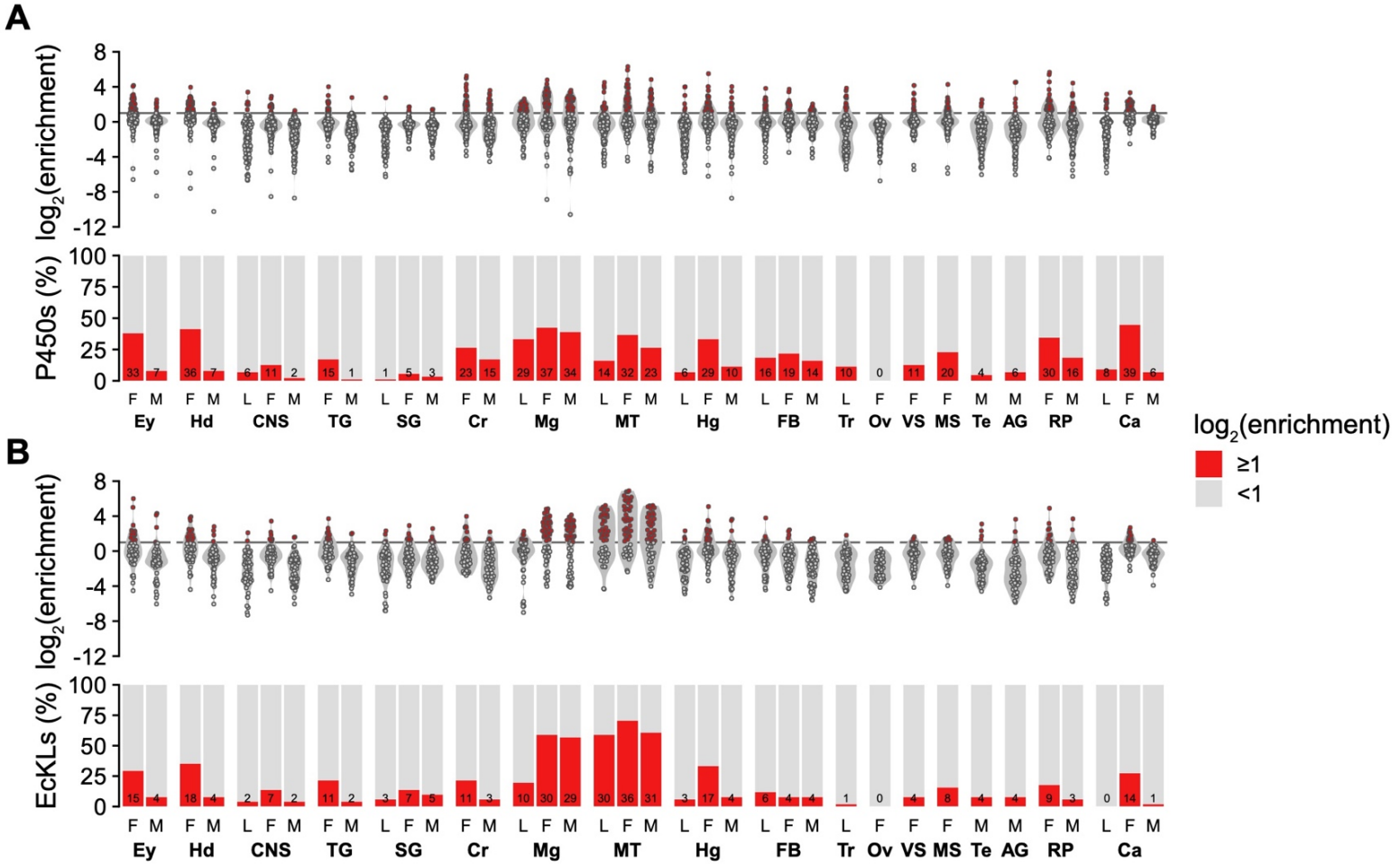


Figure 2.4. Tissue enrichment of (A) P450 and (B) EcKL genes (top) and the proportion of all genes in the family that are enriched ($\log_2$(enrichment) > 1; bottom) in *Drosophila melanogaster* across 18 tissues and three life stages. Horizontal lines on the top subfigure indicate a $\log_2$(enrichment) threshold of 1. Numbers in bars are the number of genes enriched in each tissue. L, 3rd instar larva; M, adult male; F, adult female. Ey, eye; Hd, head; CNS, central nervous system; TG, thoracicoabdominal ganglion; SG, salivary gland; Cr, crop; Mg, midgut; MT, Malpighian tubules; Hg, hindgut; FB, fat body; Tr, trachea; Ov, ovary; VS, virgin spermatheca; MS, mated spermatheca; Te, testis; AG, accessory gland; RP, rectal pad; Ca, carcass.

2.4.3. Associations between xenobiotic induction, CncC induction, evolutionary stability and detoxification tissue enrichment

We used a collection of differential gene expression datasets to explore the transcriptional response of EcKL and P450 genes in *D. melanogaster* to the ingestion of xenobiotic compounds (phenobarbital, piperonyl butoxide, methamphetamine, tunicamycin and *Aspergillus nidulans* secondary metabolites) and to the ectopic expression of the xenobiotic-response transcription factor CncC. We also used this data to test associations between the 'detoxification properties' of xenobiotic induction (defined as induction in at least one dataset), CncC induction and clade instability; we hypothesised that these two gene families will have significant groups of genes that have two or more of these properties and therefore that these properties will be associated gene family-wide. 34 (39%) P450s are induced in at least one xenobiotic dataset, compared with 23 (45%) EcKLs, while 20 (23%) P450s are induced by CncC, compared with 15 (29%) EcKLs (Fig. 2.5A,C). Family-wide, there are significant interactions between clade instability and xenobiotic induction ($p = 0.005$) and CncC induction and xenobiotic induction ($p = 9\times10^{-4}$) in the P450s (Fig. 2.5B); while in EcKLs, there is significant interaction only between clade instability and xenobiotic induction ($p = 0.005$; Fig. 2.5D).

We then integrated these data with the tissue enrichment values calculated earlier to test if genes in each family with these detoxification properties have greater median enrichment in detoxification-related tissues than genes in the same family without these properties. P450s belonging to unstable clades have greater enrichment in the adult midguts and the adult female Malpighian tubules, while EcKLs belonging to unstable clades have greater enrichment in only the adult female midgut (Fig. S2.5). P450s induced by xenobiotics have greater enrichment in adult midguts and the adult female Malpighian tubules and fat body. In contrast, EcKLs induced by xenobiotics show greater enrichment in only the adult midguts and the adult female carcass (Fig. S2.6). P450s induced by CncC have greater enrichment in only the adult female Malpighian tubules, while EcKLs induced by CncC have greater enrichment in only the adult male Malpighian tubules (Fig. S2.7). Perhaps unsurprisingly, both EcKLs and P450s with detoxification properties tend to have reduced enrichment in a number of tissues not thought to play large roles in detoxification, including the larval CNS, salivary glands, hindgut and carcass, and the adult ovary and testes. However, we also

noticed greater enrichment for EcKLs and P450s with detoxification properties in some tissues not typically associated with detoxification, such as the adult female eye, head, virgin and mated spermathecae and carcass, and the adult male carcass, particularly for the P450s (Figs. S2.5–2.7).

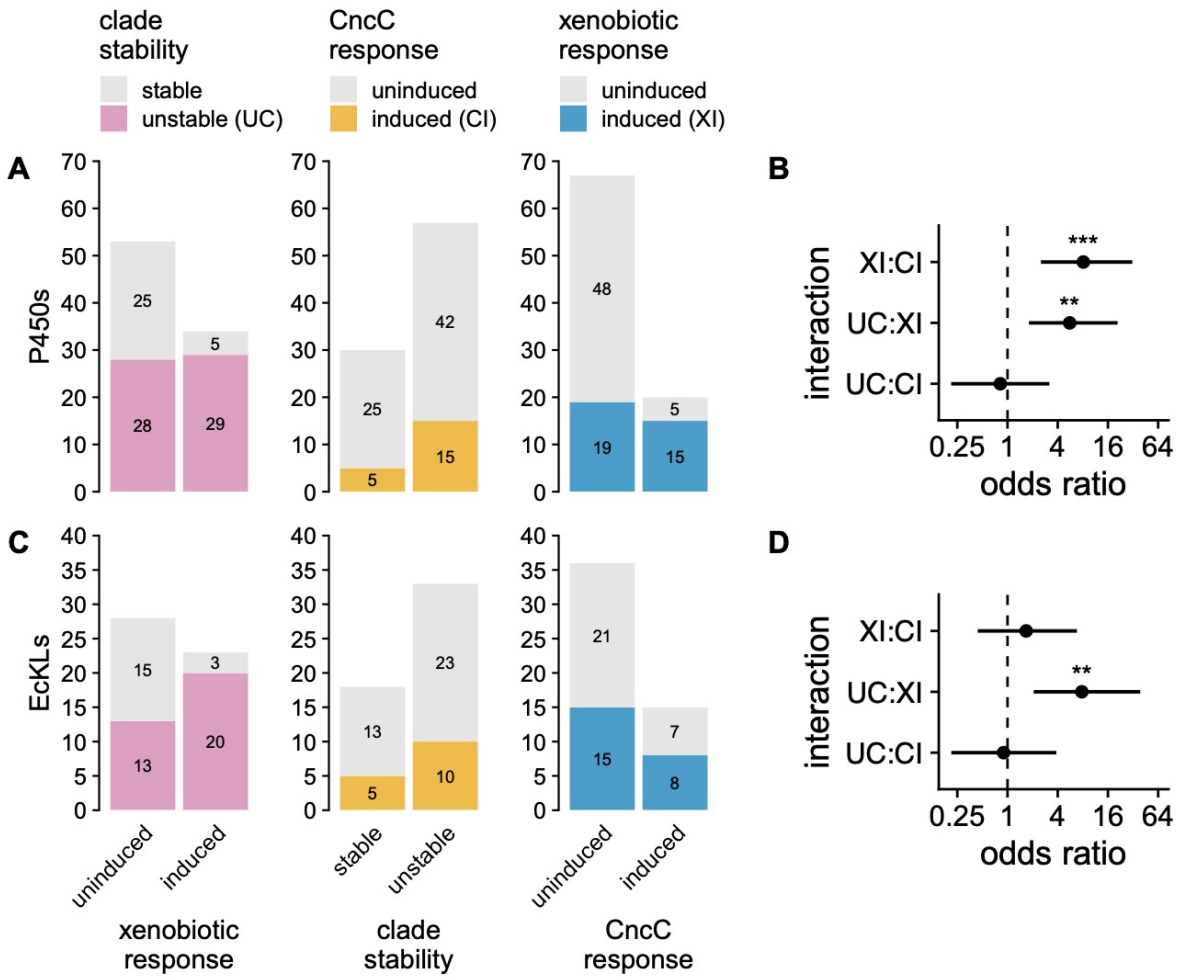


Figure 2.5. (A,C) The number of genes with sets of detoxification characteristics in *D. melanogaster*, for the (A) P450s and (C) EcKLs. (B,D) Homogenous association loglinear models were fitted for the detoxification characteristics for both the P450s and the EcKLs. Odds ratios of interactions between pairs of characteristics for the (B) P450s and (D) EcKLs are shown; significant interaction is evidence of association between characteristics within the gene family. Errors bars are 95% confidence intervals. XI, xenobiotically induced; CI, CncC induced; UC, unstable clade. p-values: * < 0.05, ** < 0.01, *** < 0.001. Genes were considered CncC induced if they were up-regulated by ectopic CncC (Misra *et al.* 2011); repression in *cncC* mutant embryos (Deng & Kerppola 2013) was not considered.

2.4.4. The detoxification score method is concordant with known functions of *D. melanogaster* P450s

As detoxification genes have generally accepted evolutionary and transcriptomic characteristics, we developed a 'detoxification score' (DS) that scores genes (0–4) based on the criteria of evolutionary instability, xenobiotic induction, CncC induction and detoxification tissue enrichment. Genes with scores of 3 or 4 were considered detoxification candidates. We tested the accuracy of the DS method by applying it to the P450 gene family in *D. melanogaster* (Fig. 2.6) and comparing the scores with published functions of 31 P450 genes (Table 2.2).

Fifteen P450s have X-class (detoxification) functions described in the literature, of which only one, *Cyp28d2* (DS = 1), had a score less than 3. *Cyp28d2* has been implicated in nicotine tolerance, along with its paralog *Cyp28d1* (DS = 3; Highfill *et al.* 2017; Marriage *et al.* 2014). The evidence for both genes' roles in nicotine detoxification comes from quantitative complementation with loss-of-function alleles, as well as ubiquitous RNAi knockdown (Highfill *et al.* 2017), but they have not been characterised further. *Cyp12a5* (DS = 3) is another interesting case, given that it has been implicated in the bioactivation—not detoxification—of the neonicotinoid insecticide nitenpyram (Harrop *et al.* 2018). However, given that it putatively metabolises a xenobiotic compound, even though that metabolism increases its toxicity, it seems likely that it detoxifies other similar compounds. *Cyp6a17* (DS = 3) has characterised roles in pyrethroid insecticide tolerance (Battlay *et al.* 2018; Duneau *et al.* 2018), but also has a connection to temperature preference behaviour (Kang *et al.* 2011). This is not the only behavioural phenotype linked to a P450: *Cyp6a20* (DS = 3) has a putatively S-class (secondary metabolic) function due to its links to male aggression (Dierick & Greenspan 2006; Robin *et al.* 2007; Wang *et al.* 2008).

Fifteen P450s have E-class (essential physiological or developmental) functions described in the literature, of which only one, *Cyp6t3* (DS = 3), had a score greater than 2. *Cyp6t3* has a proposed role in ecdysteroid biosynthesis based on RNAi evidence (Ou *et al.* 2011), but the gene has undergone a complete loss event in the *Drosophila* genus (Good *et al.* 2014), which is inconsistent with a conserved function in hormone metabolism. Nevertheless, *Cyp6t3* is not enriched in any of the three detoxification tissues and so may not be a detoxification gene. Additionally, *Cyp6g2* (DS = 1) is an interesting case, as its proposed role in juvenile hormone metabolism (Christesen *et al.*

2017) and corpora allata-specific expression (Chung *et al.* 2009) are at odds with its ability to detoxify multiple insecticides when ectopically expressed in detoxification tissues (Daborn *et al.* 2007; Denecke *et al.* 2017b).

Based on these data, the accuracy of the DS method was evaluated with a diagnostic test. The sensitivity (true positive rate) was 93.3% (95% CI: 68.1–99.83), while the specificity (true negative rate) was 87.5% (95% CI: 61.65–98.45). Considering the small sample sizes available, this suggests the DS method is reasonably accurate at predicting known detoxification functions of P450s and has the potential to identify candidate detoxification genes for follow-up analyses. Indeed, 16 P450s that had scores of 3 or 4 have not yet been characterised—we predict that between 11–16 of these genes encode detoxification enzymes (Table 2.3).

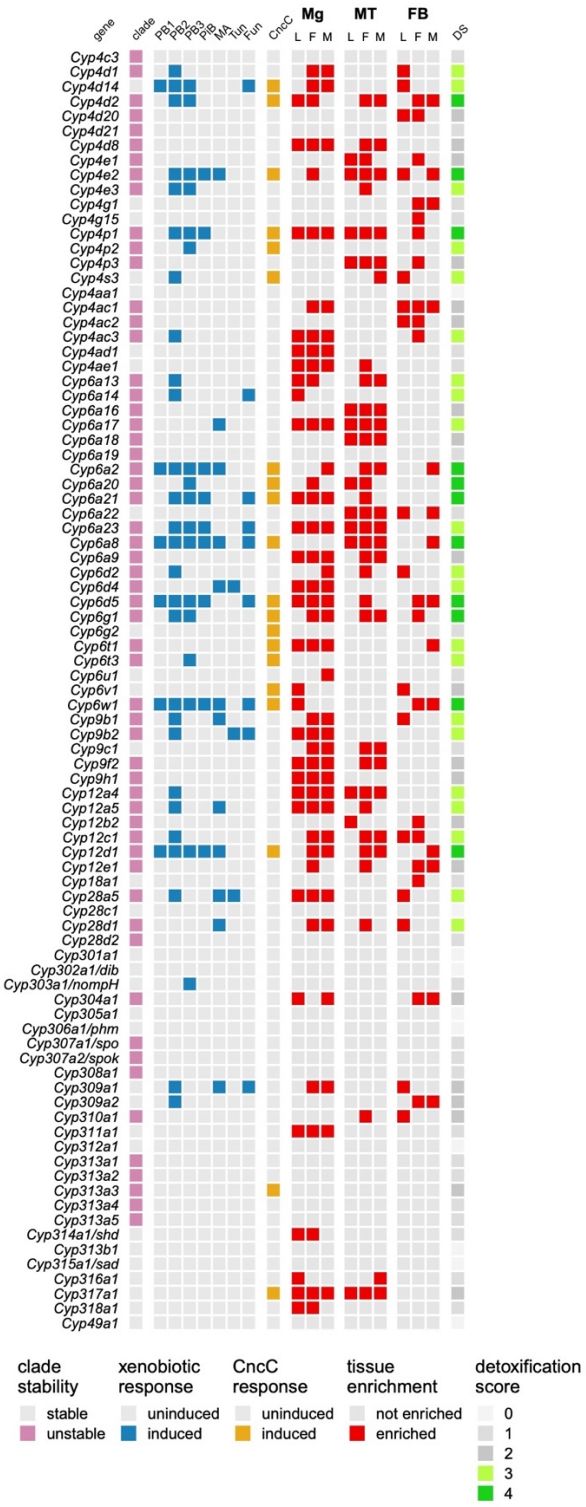


Figure 2.6. Application of the detoxification score (DS) to P450 genes in *Drosophila melanogaster*, wherein genes receive one point for meeting each detoxification criterion, indicated by non-grey coloured squares. Genes with names other than P450 nomenclature designation have both indicated, separated by a slash. PB1, phenobarbital dataset 1 (Sun *et al.* 2006); PB2, phenobarbital dataset 2 (King-Jones *et al.* 2006); PB3, phenobarbital dataset 3 (Misra *et al.* 2011); PiB, piperonyl butoxide (Willoughby *et al.* 2007); MA, methamphetamine (Sun *et al.* 2011); Tun, tunicamycin (Chow *et al.* 2013); Fun, *Aspergillus nidulans* toxins (Trienens *et al.* 2017); CncC, induction by ectopic CncC (Misra *et al.*, 2011); Mg, midgut; MT, Malpighian tubules; FB, fat body; L, 3rd instar larvae; F, adult females; M, adult males; DS, detoxification score.

Table 2.2. Comparison of detoxification score (DS) values with known P450 functions in *Drosophila melanogaster*, with genes sorted by DS.

Gene	DS	Class	Evidence[a]	Function and/or reaction	References
Cyp301a1	0	E	moderate	cuticle formation	Sztal *et al.* 2012
Cyp302a1/dib	0	E	strong	ecdysteroid 22-hydroxylase	Warren *et al.* 2002
Cyp306a1/phm	0	E	strong	ecdysteroid 25-hydroxylase	Warren *et al.* 2004; Niwa *et al.* 2004
Cyp307a1/spo	0	E	moderate	ecdysteroid biosynthesis (Black Box reaction)	Ono *et al.* 2006; Chavez *et al.* 2000; Namiki *et al.* 2005
Cyp307a2/spok	0	E	moderate	ecdysteroid biosynthesis (Black Box reaction)	Ono *et al.* 2006
Cyp315a1/sad	0	E	strong	ecdysteroid 2-hydroxylase	Warren *et al.* 2002
Cyp18a1	1	E	strong	ecdysteroid 26-hydroxylase/carboxylase	Guittard *et al.* 2011; Rewitz *et al.* 2010
Cyp303a1/nompH	1	E	moderate	function in adult eclosion and development of sensory organs	Wu *et al.* 2019; Willingham & Keil 2004
Cyp314a1/shd	1	E	strong	ecdysteroid 20-hydroxylase	Petryk *et al.* 2003
Cyp4d21	1	E	moderate	possible fatty acid gamma-hydroxylase	Fujii *et al.* 2008
Cyp4g1	1	E	strong	cuticular hydrocarbon biosynthesis	Qiu *et al.* 2012; Niwa *et al.* 2011
Cyp6g2	1	E	weak	probable juvenile hormone metabolism, but can detoxify nitenpyram, dicylanil and imidacloprid if expressed ectopically	Christesen *et al.* 2017; Daborn *et al.* 2007; Denecke *et al.* 2017
Cyp6u1	1	E	weak	ecdysteroid metabolism	Christesen *et al.* 2017
Cyp28d2	1	X	moderate	nicotine tolerance	Marriage *et al.* 2014; Highfill *et al.* 2017
Cyp310a1	2	E	weak	negative regulator of Wg signalling	Mohit *et al.* 2006
Cyp6t3	3	E	weak	ecdysteroid metabolism	Ou *et al.* 2011
Cyp6a20	3	S	moderate	aggression/pheromone sensitivity	Dierick & Greenspan 2006; Wang *et al.* 2008
Cyp12a4	3	X	moderate	lufenuron resistance	Bogwitz *et al.* 2005; Good *et al.* 2014
Cyp12a5	3	X	moderate	bioactivation/toxication of nitenpyram	Harrop *et al.* 2018
Cyp28d1	3	X	moderate	nicotine tolerance	Marriage *et al.* 2014; Highfill *et al.* 2017
Cyp4e3	3	X	weak	permethrin tolerance	Terhzaz *et al.* 2015

Gene	DS	Class	Evidence[a]	Function and/or reaction	References
Cyp6a17	3	X	moderate	permethrin and deltamethrin survivorship, temperature preference	Battlay *et al.* 2018; Duneau *et al.* 2018; Kang *et al.* 2011; Chakraborty *et al.* 2018
Cyp6a23	3	X	weak	deltamethrin tolerance	Duneau *et al.* 2018
Cyp6d2	3	X	moderate	camptothecin tolerance	Thomas *et al.* 2013
Cyp9b2	3	X	weak	boric acid tolerance	Najarro *et al.* 2017
Cyp12d1	4	X	moderate	DTT, dicyclanil, malathion, chlorantraniliprole and cyantraniliprole phenotypes, caffeine metabolism	Daborn *et al.* 2007; Najarro *et al.* 2015; Coelho *et al.* 2015; Battlay *et al.* 2018; Green *et al.* 2019
Cyp6a2	4	X	strong	aldrin, heptachlor and diazinon metabolism, bioactivation/toxication of aflatoxin B_1, DMBA and Trp-P-2	Saner *et al.* 1996; Dunkov *et al.* 1997
Cyp6a8	4	X	strong	aldrin and caffeine metabolism, but also lauric acid metabolism	Coelho *et al.* 2015; Helvig *et al.* 2004
Cyp6d5	4	X	weak	caffeine metabolism	Coelho *et al.* 2015
Cyp6g1	4	X	strong	DDT, azinphos-methyl, malathion, diazinon, nitenpyram, dicylanil and imidacloprid resistance	Daborn *et al.* 2007; Fusetto *et al.* 2017; Schmidt *et al.* 2010; Daborn *et al.* 2002; Joußen *et al.* 2008; Battlay *et al.* 2016; Battlay *et al.* 2018
Cyp6w1	4	X	moderate	DDT survivorship	Schmidt *et al.* 2017

[a] Refer to Chapter 2.3.7 for explanation of evidence levels.

Table 2.3. *Drosophila melanogaster* P450 genes with detoxification scores (DS) greater than 3 and with currently unknown functions.

DS	Genes
3	*Cyp4ac3, Cyp4d1, Cyp4d14, Cyp4p2, Cyp4s3, Cyp6a13, Cyp6a14, Cyp6d4, Cyp6t1, Cyp12c1, Cyp28a5*
4	*Cyp4d2, Cyp4e2, Cyp4p1, Cyp6a21*

2.4.5. DS method suggests many *D. melanogaster* EcKLs are candidate detoxification genes

Applying the DS method to the EcKL family in *D. melanogaster* (Fig. 2.7), 24 genes (47%) have a DS ≥ 3 (Fig. 2.8B), compared with 32 genes (36.8%) in the P450 family (Fig. 2.8A). Based on the aforementioned sensitivity of the DS method that used the P450 family as a 'truth set,' we estimate between 16–24 EcKLs are involved in detoxification. Seven EcKLs have a DS of 4: *CG31288* (Dro11-0), *CG10550* (Dro13-1), *CG10553* (Dro17-1), *CG10560* (Dro17-3), *CG10562* (Dro18-1), *CG6908* (Dro23-0) and *CG9498* (Dro29-0). Interestingly, *CG10562* is a direct paralog of *CHKov1* (Dro18-2, DS = 2), which has been implicated in resistance to sigma virus infection (Magwire *et al.* 2012; 2011). Five and two genes in the highly unstable Dro5 and Dro26 clades, respectively, were also detoxification candidates (DS = 3).

2.4.6. Genomic and transcriptomic variation in EcKL and P450 genes is associated with toxic stress phenotypes in *D. melanogaster*

A large amount of phenotypic, genomic and transcriptomic data is available for the Drosophila Genetic Reference Panel (DGRP), a collection of inbred lines of *D. melanogaster* that houses substantial naturally occurring genetic variation. Many of the phenotypes studied with the DGRP are related to toxic stress and can be used to detect candidate detoxification genes. We used a targeted phenome-wide association study (PheWAS) approach to detect associations between 146 phenotypes, including 55 toxic stress phenotypes (Table S2.2), and genomic and (adult sex-specific) transcriptomic variation in *D. melanogaster* P450 and EcKL genes, at a significance threshold of $p < 10^{-5}$ for genomic variation, and $p < 10^{-3}$ for transcriptomic variation. Summaries of PheWAS results can be found in Table S2.3A–F, while detailed outputs of the PheWAS analyses can be found in Table S2.4.

164 genomic variants in or near P450 genes were associated with 35 phenotypes, 26 of which related to toxic stress (Table S2.3A); most of these are linked together in haplotypes and have been previously reported in publications studying insecticide phenotypes (Battlay *et al.* 2018; 2016; Denecke *et al.* 2017b; Duneau *et al.* 2018; Green *et al.* 2019; Schmidt *et al.* 2017)—however, some are novel. Four SNPs in or near *Cyp4d20* (DS = 2) are associated with chlorantraniliprole survival; an intronic SNP in *Cyp4d8* (DS = 2) is associated with DDT knockdown; a haplotype containing 10 linked SNPs

in *Cyp4s3* (DS = 3), including a non-synonymous SNP (W260S), is associated with developmental caffeine survival; and a SNP producing a premature stop codon in *Cyp6a16* (DS = 2)—which is annotated as a pseudogene in the *D. melanogaster* reference genome but rarely pseudogenized in natural populations (Good *et al.* 2014)—is associated with 3 hr malathion mortality in adult females. SNPs in or near five genes—*Cyp4ac1* (DS = 2), *Cyp4ac3* (DS = 3), *Cyp4c3* (DS = 1), *Cyp4e3* (DS = 3) and *Cyp28d2* (DS = 1)—were associated with ethanol tolerance in males and/or females. P450 transcript levels in males were associated with 36 phenotypes (25 toxic stress phenotypes) for 23 genes (Table S2.3B), while P450 transcript levels in females were associated with 37 phenotypes (29 toxic stress phenotypes) for 23 genes (Table S2.3C). Like the genomic variants, most of these have previously been reported. *Cyp4d8* (DS = 2) transcript levels are associated with both boric acid and caffeine survival in adult females. *Cyp6a21* (DS = 4) transcript levels in females are associated with developmental methylmercury survival.

70 genomic variants in or near EcKL genes were associated with 10 phenotypes, eight variants of which were associated with six toxic stress phenotypes (Table S2.3D); the bulk (56) of the total associations are part of the previously noted *CHKov1*-TE (Dro18-2, DS = 2) haplotype associated with sigma virus resistance (Magwire *et al.* 2012; 2011). A total of three SNPs near two paralogous but genetically unlinked EcKLs—*CG16898* (Dro26-1, DS = 3) and *CG33301* (Dro26-2, DS = 3)—are associated with developmental methylmercury survival, while a non-synonymous SNP (T4I) in *CG33301* is also associated with ethanol tolerance in adult females, and two downstream SNPs in *CG33301* are associated with larval activity during exposure to the neonicotinoid imidacloprid. EcKL transcript levels in males were associated with 10 phenotypes (five toxic stress phenotypes) for nine genes (Table S2.3E), while transcript levels in females were associated with 10 phenotypes (seven toxic stress phenotypes) for five genes (Table S2.3F). Many of these were associations between *CG6908* (Dro23-0, DS = 4) and chlorantraniliprole and malathion survival. *CG10560* (Dro17-3, DS = 4) expression is also associated with chlorantraniliprole survival in males and females, while *CG11878* (Dro1-5, DS = 3) expression in females is associated with ethanol tolerance in males.

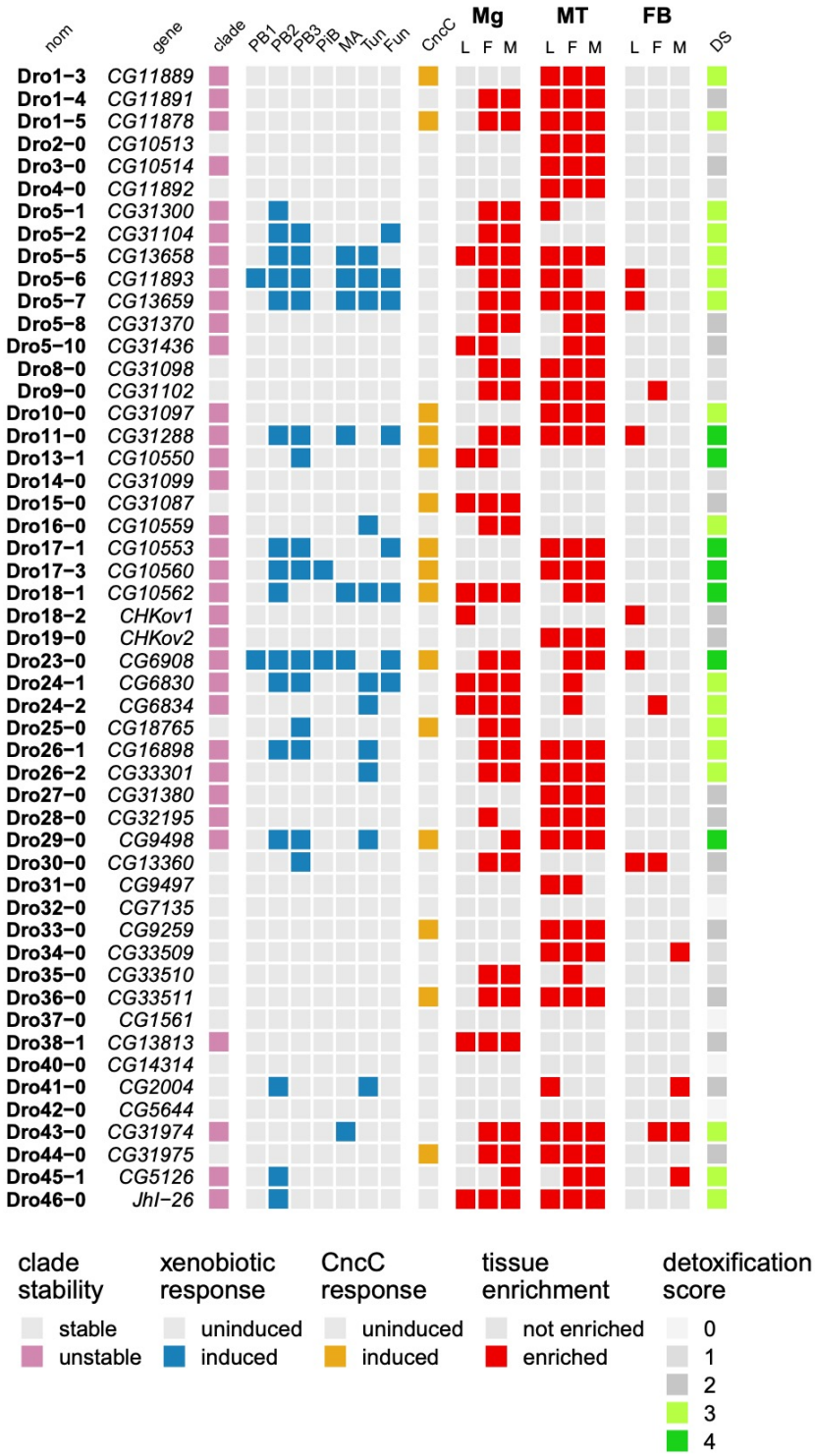


Figure 2.7. Application of the detoxification score (DS) to EcKL genes in *Drosophila melanogaster*, wherein genes receive one point for meeting each detoxification criterion, indicated by non-grey coloured squares. Genes are identified by both their *Drosophila* ancestral clade nomenclature, and their name or annotation symbol, as appropriate. PB1, phenobarbital dataset 1 (Sun *et al.* 2006); PB2, phenobarbital dataset 2 (King-Jones *et al.* 2006); PB3, phenobarbital dataset 3 (Misra *et al.* 2011); PiB, piperonyl butoxide (Willoughby *et al.* 2007); MA, methamphetamine (Sun *et al.* 2011); Tun, tunicamycin (Chow *et al.* 2013); Fun, *Aspergillus nidulans* toxins (Trienens *et al.* 2017); CncC, induction by ectopic CncC (Misra *et al.*, 2011); Mg, midgut; MT, Malpighian tubules; FB, fat body; L, 3rd instar larvae; F, adult females; M, adult males; DS, detoxification score.

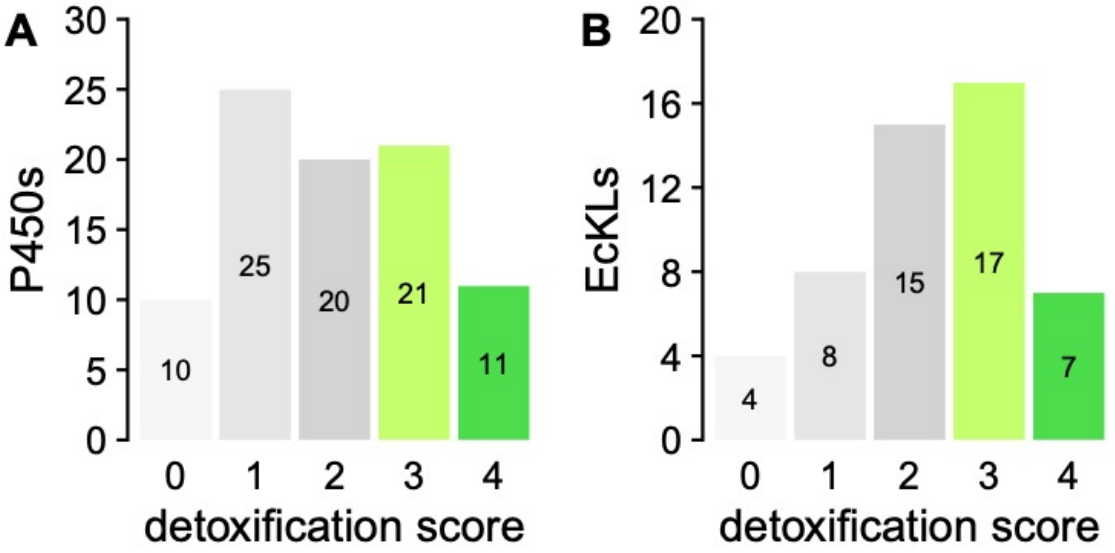


Figure 2.8. The number of genes in the (A) P450 and (B) EcKL gene families in *Drosophila melanogaster* with a given detoxification score. Scores of 3 and 4 indicate candidate detoxification genes.

2.4.7. Ubiquitous RNAi knockdown of EcKLs in *D. melanogaster*

As a preliminary attempt to find EcKLs essential for development (i.e. a subset of E-class enzymes), we crossed UAS-dsRNA responder males for 41 EcKLs (and three control lines) to females containing a *tub*-GAL4 driver over a balancer chromosome, and phenotypically scored the eclosing adult offspring (Fig. S2.8). Putative knock-down of eight EcKLs resulted in significant developmental lethality that was attributable to their reduced expression: *CG31102* (Dro9-0, DS = 1), *CG31099* (Dro14-0, DS =1), *CG10562* (Dro18-1, DS = 4), *CHKov1* (Dro18-2, DS = 2), *CG9497* (Dro31-0, DS = 1), *CG13813* (Dro38-1, DS = 2), *CG2004* (Dro41-0, DS = 2) and *CG31975* (Dro44-0, DS = 2).

2.5. Paper Discussion

2.5.1. EcKL evolution in *Drosophila*

The evolutionary pattern seen in the EcKL gene family in this study matches that seen in well-established detoxification gene families in *Drosophila*, such as the P450s, glutathione S-transferases (GSTs) and carboxylcholinesterases (CCEs): a mix of stable clades that have 1:1 orthologs in all species, and unstable clades with high rates of gene gain, gene loss or both (Good *et al.* 2014; Low *et al.* 2007; Robin *et al.* 2009; 1996). We have found that a few EcKLs belonging to stable clades in *Drosophila* are also highly conserved across insects as a whole (Chapter 3); these are good candidates for genes with essential E-class functions. It is important to note, however, that not all genes with E-class functions are necessarily stable over evolutionary time: *Cyp4g1* and *Cyp4g15* orthologs in insects encode enzymes responsible for alkane and alkene biosynthesis, yet this physiologically essential function is maintained despite seemingly random duplications and losses in many lineages (Feyereisen 2020); similarly, orthologs of the Halloween genes *spook* (*Cyp307a1*) and *spookier* (*Cyp307a2*) encode genes essential for the biosynthesis of ecdysteroids, yet also have experienced elevated rates of duplication and loss in insects and other arthropods (Perry *et al.* 2019; Rewitz *et al.* 2007; Sezutsu *et al.* 2013; Sztal *et al.* 2007).

Of interest in this study is the nine-gene difference in EcKL number between *D. simulans* and *D. sechellia*, which is due to gene losses in the latter species in only four clades: Dro5 (four losses), Dro1 (two losses), Dro26 (two losses) and Dro16 (one loss). In *D. melanogaster*, many of the genes in these clades are detoxification candidates, suggesting genes in these clades in other species may also function in detoxification. *D. sechellia* also has fewer P450 genes compared to *D. simulans*—74 versus 88 (Good *et al.* 2014)—and gene losses have also been observed in the GST and odorant receptor gene families in this species (Low *et al.* 2007; McBride *et al.* 2007). Two competing hypotheses have been proposed to explain this concerted gene loss: the first is that moving from a generalist niche, as seen with *D. simulans*, to a specialist niche feeding on morinda fruit (Dworkin & Jones 2008) has led to relaxed selection on genes involved in detoxification and olfaction, producing an elevated rate of pseudogenisation in the *D. sechellia* lineage; the second is that possible severe population reduction after the

divergence of *D. sechellia* and *D. simulans* resulted in elevated fixation of slightly deleterious null alleles (Good *et al.* 2014). If the first is true, this may be further evidence for the hypothesised detoxification functions of genes in the Dro1, Dro5, Dro16 and Dro26 clades.

2.5.2. Detoxification gene properties and tissue-specific enrichment

Detoxification genes are thought to have various properties, including evolutionary instability, transcriptional induction by toxins and toxin-response pathways, and enrichment in detoxification-related tissues (Kawashima & Satta 2014; Misra *et al.* 2011; Thomas 2007; Willoughby *et al.* 2006; Yang *et al.* 2007). If this is true, these properties should associate together in detoxification families, but to the best of our knowledge this has yet to be tested in any detail. The results in this study generally support this hypothesis, with some detoxification characteristics associating in the P450s, a classical detoxification gene family. We note that Giraudo *et al.* (2010) previously explored the overlap in P450 induction between a large number of xenobiotic datasets, finding that a small number of genes (*Cyp6a2*, *Cyp6a8*, *Cyp6d5*, *Cyp6w1* and *Cyp12d1*) are induced by virtually all xenobiotic compounds analysed—remarkably, all of these genes have DS values of 4 in this study, suggesting broad inducibility may predict the presence of other detoxification traits; it also suggests that the inducibility of these genes is primarily due to regulation by CncC, which responds to many types of chemical stressors (Wilding 2018).

The tissue enrichment data analysed here show EcKLs are strongly enriched in the Malpighian tubules, with comparatively little enrichment (proportional or otherwise) in the larval midgut (Fig. 2.4B). While both tissues are thought to be involved in detoxification, it is unclear how detoxification processes differ between the midgut and the Malpighian tubules. A reasonable assumption is that the midgut is involved in immediate 'first-pass' xenobiotic metabolism, while the Malpighian tubules are involved in reactions that promote the excretion of xenobiotic compounds, such as conjugations, and detoxify compounds that are already circulating in the hemolymph. Detoxicative kinases, while conceptually catalysing a phase II conjugation reaction, may also act on toxins in the midgut directly if they contain hydroxyl groups (such as phytoecdysteroids), or on the immediate products of P450-mediated hydroxylation. Interplay between detoxification reactions within and between tissues is a topic that

deserves greater study in insects.

It is important to note that the tissue expression data from FlyAtlas 2 (Leader *et al.* 2018) considered in this paper consist of basal levels of expression, and given that xenobiotic compounds may induce expression of detoxification genes—particularly in detoxification tissues (Willoughby *et al.* 2006)—the tissue-specific enrichment of detoxification genes during xenobiotic exposure is likely to be different to what is observed here. Comparatively low levels of basal enrichment of EcKLs in the larval midgut might be offset by extensive midgut-specific induction of these genes upon exposure to relevant toxins. In contrast, we hypothesise that the high levels of basal enrichment of EcKLs in adult midguts might be due to relatively high constitutive expression of detoxification genes in these tissues, which may be adaptive for the intermittent feeding behaviour of adults (Xu *et al.* 2008). The FlyAtlas 2 data used here are also derived from a single fly line, Canton S, which might show transcriptomic differences with other lines; however, our DS method can reliably detect detoxification genes identified in other genetic backgrounds (Table 1), suggesting the method is robust across genotypes.

A drawback of analysing tissue enrichment at a family-wide level is that it elides differences between genes—some detoxification genes have highly tissue-specific expression patterns (i.e. expressed in the midgut but not the Malpighian tubules, or vice versa; Yang *et al.* 2007). Some non-detoxification genes may also have high enrichment in these 'detoxification' tissues—this may partially explain the large confidence intervals seen for differential Malpighian tubule enrichment in the EcKLs (Figs. S2.5–2.7), as some genes with extremely high enrichment in this tissue (e.g. *CG10513*, *CG10514*, *CG11892* and *CG9259*) do not display other detoxification characteristics and consequently are poor detoxification candidates (Fig. 2.7).

2.5.3. Suitability of the DS method for identifying candidate detoxification genes in insects

In this study, we have proposed a 'detoxification score' (DS) for the identification of candidate detoxification genes from a combination of evolutionary and transcriptomic data available in *D. melanogaster*. We chose an integrative approach because each individual detoxification property (evolutionary instability, xenobiotic induction, CncC

induction or detoxification tissue enrichment) is not reliably diagnostic of detoxification function: not all detoxification genes are likely unstable over the 40–50 m.y. period analysed in this study, while essential E-class genes can sometimes be unstable (Feyereisen 2020; Sztal *et al.* 2007); not all enzymes that detoxify a compound are induced by that compound (Willoughby *et al.* 2006); CncC regulates both detoxification and developmental gene sets, including the Halloween P450s, at different life stages (Deng & Kerppola 2013; Misra *et al.* 2011); and 'detoxification tissues' have roles in non-detoxification processes, such as digestion, immunity, osmoregulation and energy metabolism (Beyenbach *et al.* 2010; Buchon *et al.* 2009; Li *et al.* 2019). S-class enzymes may also share evolutionary instability with X-class enzymes, as seen in plants and dinoflagellates, where the evolution of secondary metabolic pathways is highly dynamic (Beedessee *et al.* 2019; Kliebenstein 2008; Mukherjee *et al.* 2016). Given this, any integrative method for detecting detoxification genes will produce some level of false positives and false negatives. It is also worth noting that the DS method as described here does not weight the importance of detoxification properties, even though, for example, detoxification tissue expression could arguably be more indicative of detoxification function than regulation by CncC. A more sophisticated, future version of this method should attempt this.

The nuclear receptor DHR96 has also been linked to the transcriptional response to xenobiotic compounds in *D. melanogaster* (King-Jones *et al.* 2006; Shah *et al.* 2012). Despite this, we did not use induction by ectopically expressed DHR96 (King-Jones *et al.* 2006; Shah *et al.* 2012) as a factor in our DS method, as the role of DHR96 in detoxification appears complex. DHR96 mutant animals are less tolerant of phenobarbital and DDT (Afschar *et al.* 2016; King-Jones *et al.* 2006) but DHR96 knockdown animals are more tolerant of imidacloprid (Shah *et al.* 2012); DHR96 also positively regulates some detoxification genes but negatively regulates others (King-Jones *et al.* 2006; Shah *et al.* 2012). As such, induction by DHR96 may not be indicative of detoxification function for individual genes.

Loss of function alleles of *Cyp6a20* (DS = 3) result in aggressive behaviour in males, suggesting it may have a role in pheromone sensitivity (Dierick & Greenspan 2006; Robin *et al.* 2007; Wang *et al.* 2008), although its enzymatic substrate is not known. This gene's possible misclassification as a detoxification candidate is due to its induction by both phenobarbital and methamphetamine, and its enrichment in the adult male

midgut and the larval and adult female Malpighian tubules; it is also enriched in other tissues (Fig. S2.3), suggesting that it may have functions outside of the antennae. It is possible *Cyp6a20* encodes a bi-functional enzyme that belongs to both the S and X classes.

Indeed, the DS method might be limited more generally by the fact that some enzymes often fall into multiple functional classes simultaneously, such as the glutathione S-transferase GST16 in the cotton bollworm *Helicoverpa armigera*, which has roles in both detoxification and larval development (Shabab *et al.* 2014). There is experimental evidence that some P450 genes in *D. melanogaster* belong to multiple functional classes. *Cyp6g2* (DS = 1) may be involved in juvenile hormone metabolism (Christesen *et al.* 2017), but also has the capacity to detoxify multiple classes of insecticides if expressed ectopically (Daborn *et al.* 2007; Denecke *et al.* 2017b). *Cyp6g2* is a paralog of *Cyp6g1* (DS = 4), which encodes a detoxification enzyme with very broad substrate specificity (Battlay *et al.* 2018; 2016; Daborn *et al.* 2002; 2007; Fusetto *et al.* 2017; Joußen *et al.* 2008; Schmidt *et al.* 2010), and so *Cyp6g2* may have a pre-adapted detoxification capacity. *Cyp6a17* (DS = 3) may detoxify insecticides (Battlay *et al.* 2018; Duneau *et al.* 2018) but also regulates temperature preference (Chakraborty *et al.* 2018; Kang *et al.* 2011). *Cyp6a8* can metabolise the insecticide aldrin and possibly caffeine (Coelho *et al.* 2015; Helvig *et al.* 2004b), but it can also use the lipid lauric acid as a substrate, suggesting it may have both E- and X-class functions. From an evolutionary perspective, the transition between E-, S- and X-class functions must occur with some frequency (Sezutsu *et al.* 2013), and so it is likely that at any given time a non-zero proportion of enzymes in a gene family have multiple functions.

It has been previously noted that P450s harbouring polymorphic structural variation (SV; duplications, deletions or rearrangements) in *D. melanogaster* tend to belong to clades that are evolutionarily unstable between species (Good *et al.* 2014). We note that many high DS P450s with known detoxification functions identified in this study harbour SVs in the DGRP and the DSPR panels (Chakraborty *et al.* 2019; Good *et al.* 2014), including *Cyp6g1* (DS = 4), *Cyp12d1* (DS = 4), *Cyp28d1* (DS = 3), *Cyp6a17* (DS = 3) and *Cyp6a23* (DS = 3); some detoxification candidate EcKLs, such as *CG13659* (Dro5-7, DS = 3), *CG31097* (Dro10-0, DS = 3), *CG31288* (Dro11-0, DS = 4), *CG6834* (Dro24-2, DS = 3) and *CG16898* (Dro26-1, DS = 3), also harbour SVs in the DGRP (data not shown). These data suggest that within-species SV might be a feature of detoxification genes

more generally, complementing between-species instability.

Due to the abundance of genome- and transcriptome-wide datasets in *D. melanogaster*, there are a number of other promising gene families that could be analysed by the DS method. The GSTs, CCEs and UDP-glycosyltransferases (UGTs) are all well-known detoxification families in *Drosophila* and other insects (Ahn *et al.* 2012; Low *et al.* 2007; Oakeshott *et al.* 2005; Rane *et al.* 2019) that have yet to be systematically studied this way and likely contain many undiscovered detoxification genes. Other gene families that appear in the detoxification-related genome-wide datasets used here could also be analysed using this method to explore whether or not they have detoxification functions.

2.5.4. Integration of EcKL and P450 PheWAS associations and candidate detoxification genes

Phenome-wide association studies (PheWAS) are powerful approaches to detect previously unknown associations between genes and phenotypes (Denny *et al.* 2010), and are particularly useful to study detoxification in *D. melanogaster*, as a large number of toxic stress phenotypes have been previously determined in this species. This study has uncovered a number of phenotypic associations with P450s and EcKLs that make intriguing candidates for future study.

In the P450 family, the association between a *Cyp4s3* (DS = 3) haplotype and developmental caffeine survival deserves particular attention, especially because the most strongly associated variant in the haplotype is a non-synonymous SNP. Of note, *Cyp4s3* is not known to be induced by exposure to caffeine in adults or larvae (Coelho *et al.* 2015; Willoughby *et al.* 2006), which is perhaps why it has yet to be studied in relation to caffeine detoxification. Caffeine is thought to be detoxified by multiple P450s in *D. melanogaster*, raising the possibility that *Cyp4s3* is involved in this process (Coelho *et al.* 2015).

In the EcKL family, *CG33301* (Dro26-2, DS = 3) has unique genomic variation associated with distinct phenotypic responses to three chemically disparate toxins: ethanol (adult tolerance; associated with a T4I amino acid substitution), imidacloprid (larval activity) and methylmercury (developmental tolerance). This suggests that either

CG33301 has a very broad substrate specificity or is involved in a stress-response process unrelated to direct detoxification. Its paralog, *CG16898* (Dro26-1, DS = 3), is also associated with methylmercury survival, suggesting the Dro26 clade as a whole may have toxic stress-related functions. *CG6908* (Dro23-0, DS = 4) transcript levels in male and female flies are strongly associated with both chlorantraniliprole and malathion survival. This may be due to its strong (66.5-fold) induction by CncC, which responds to oxidative stress and regulates a core subset of genes transcriptionally associated with resistance to both insecticides (Battlay *et al.* 2018; Green *et al.* 2019). Because CncC positively regulates the transcription of hundreds of genes, and its constitutive up-regulation in a subset of DGRP lines has produced a co-transcriptional module (Green *et al.* 2019), it is possible *CG6908* is not functionally connected to chlorantraniliprole and malathion detoxification and the associations may simply be due to its co-regulation with *Cyp12d1*, which transgenic experiments suggest directly affects resistance to these insecticides (Battlay *et al.* 2018; Green *et al.* 2019). Regardless, *CG6908*'s high detoxification score suggests the gene is involved in detoxification in some capacity and requires further study. *CG10560* (Dro17-3, DS = 4) expression is also associated with chlorantraniliprole survival, but it is also strongly induced by CncC (16.2-fold), suggesting this may also be a non-causal association. *CG11878* (Dro1-5) transcript levels in female flies are associated with ethanol tolerance in male flies, which might suggest the association is spurious, but this gene is also up-regulated in response to ethanol exposure in male flies (Morozova *et al.* 2006), which tends to strengthen the link.

2.5.5. Functional characterisation of EcKLs

Prior to the current study, very few EcKLs in *D. melanogaster* have been functionally characterised. *JhI-26* (Dro46-0, DS = 3) is positively regulated by juvenile hormone (Dubrovsky *et al.* 2000) and forms a putatively causal link between both the increased juvenile hormone titre and increased expression of the male accessory gland protein CG10433 found during *Wolbachia* infection (Liu *et al.* 2014). Overexpression of *JhI-26* in the testes of male flies results in paternal-effect lethality and a reduction in the mating receptivity of females with which they have mated, suggesting it plays a role in *Wolbachia*-mediated cytoplasmic incompatibility (Liu *et al.* 2014). Curiously, *JhI-26*'s DS value determined in this study suggests it may also have a role in detoxification.

CHKov1 (Dro18-2, DS = 2) and *CHKov2* (Dro19-0, DS = 2) have also been studied

previously; a *CHKov1* TE-insertion allele has undergone a selective sweep in the last 200 years, which was putatively linked to organophosphate insecticide resistance by Aminetzach *et al.* (2005). However, this allele—and another containing duplications of *CHKov1* and *CHKov2*—also clearly confer strong resistance to the sigma virus (Magwire *et al.* 2012; 2011). While Aminetzach *et al.* (2005) provide evidence the original TE allele is associated with survivorship on the organophosphate insecticide azinphos-methyl using a single backcrossed line, no association between the *CHKov1*-TE allele and azinphos-methyl survivorship was found in a recent genome-wide association study using a large panel of inbred lines (Battlay *et al.* 2016). Our results here suggest neither *CHKov1* nor *CHKov2* is a detoxification gene, although the paralog of *CHKov1*, *CG10562* (Dro18-1), is a detoxification candidate.

Although we have presented preliminary evidence in this study for the hypothesis that the EcKLs are a detoxification family, it is likely that some genes in the family have E-class functions, such as in ecdysteroid metabolism (Sonobe & Ito 2009). Genes with low detoxification scores (0 or 1) may have important developmental or physiological functions, as seen in the P450s (Table 1), in which case up to 12 EcKLs (Fig. 2.8) could be considered candidate E-class genes. Of these, *CG31102* (Dro9-0), *CG31099* (Dro14-0) and *CG9497* (Dro31-0) show developmental lethality upon putative knockdown in this study (Fig. S2.8) and are the strongest candidates for E-class EcKLs in *D. melanogaster*. However, it is important to note that some RNAi constructs fail to knock down mRNA transcript levels sufficiently to observe a phenotype, while others can produce substantial off-target effects (Heigwer *et al.* 2018); given this, phenotypes observed with only a single RNAi construct should be treated as tentative until further characterisation is conducted, and lack of a phenotype cannot be considered evidence for a lack of developmental essentiality. With this in mind, we assert that little can be concluded from the RNAi knockdown data presented in this study without follow-up work—indeed, since these experiments were performed, we note that other potentially better RNAi libraries have become available, such as TRiP and NIG-FLY; these knockdown experiments should ideally be performed with multiple, independent RNAi constructs to validate the phenotypes seen here. Other good E-class candidate EcKLs are *CG7135* (Dro32-0), *CG1561* (Dro37-0), *CG14314* (Dro40-0) and *CG5644* (Dro42-0), which all have detoxification scores of 0 and are highly conserved in the *Drosophila* genus.

If EcKLs are responsible for detoxicative phosphorylation observed in insects, this would imply their X-class substrates may include phenols, steroids and glycosides (Table 2.1). Phytoecdysteroids and mycoecdysteroids are present in various plant and fungal species as anti-insect secondary metabolites (Dinan 2001; Kovganko 1999), and some are likely detoxified by phosphorylation (Rharrabe *et al.* 2007). Other hydroxylated toxins that could be X-class EcKL substrates include withanolides, cardenolides, cucurbitacins and flavonoids (Agrawal *et al.* 2012; Dinan *et al.* 1997; Glotter 1991; Wang *et al.* 2017). X-class EcKLs may also phosphorylate the products of phase I hydroxylation reactions or phase II glycosidation reactions, as is speculated to occur in locusts and moths (Boeckler *et al.* 2016; Olsen *et al.* 2015; 2014).

2.6. Paper Summary

In this study, we have shown that it is possible to use the abundant genomic and transcriptomic resources in the *Drosophila* genus to test functional hypotheses about a poorly understood gene family, the EcKLs. We have also highlighted that there is much more to discover about the functions of the P450 family in insects, particularly with respect to xenobiotic metabolism. We hope that this work can be used as a springboard for further characterisation of both gene families, and that it might inspire similar work in other insect taxa.

2.7. Chapter Discussion

2.7.1. PheWAS methylmercury phenotypes

In this chapter, two paralogous EcKLs—*CG16898* (Dro26-1) and *CG33301* (Dro26-2)—were found to be associated with developmental survival on methylmercury (MeHg; raw data from Montgomery *et al.* 2014). Both genes are also detoxification candidates (DS = 3) based on their tissue expression patterns, evolutionary instability and induction by phenobarbital and/or tunicamycin; neither gene is induced in the larval brain by exposure to MeHg (Mahapatra *et al.* 2010; Rand *et al.* 2012), but this is unsurprising, given these genes are basally expressed primarily in the midgut and Malpighian tubules (Figs. S2.2 & S2.4). That *CG33301* and *CG16898* are genetically unlinked (located on chromosome arms 2L and 2R, respectively) strongly suggests that the Dro26 clade in *Drosophila* may be involved with tolerance to MeHg.

Interestingly, SNPs in or near *CG33301* are also associated with adult ethanol tolerance (a T4I substitution; data from Morozova *et al.* 2015) and larval movement after sublethal exposure to the insecticide imidacloprid (data from Denecke *et al.* 2017b), raising the possibility that the role of this gene may be to ameliorate the damage caused by these toxins, rather than metabolise the toxins themselves. Indeed, developmental survival on MeHg can be improved in *D. melanogaster* by transgenic expression of either *Cyp6g1*, which is a known detoxification gene (Chung *et al.* 2007; Daborn *et al.* 2002; Fusetto *et al.* 2017; Joußen *et al.* 2008; Schmidt *et al.* 2010), and a human P450 gene with high sequence similarity, *Cyp3a4*, in the fat body or neurons (Rand *et al.* 2012). Similar effects can be seen with overexpression of various GST enzymes (Vorojeikina *et al.* 2017). Demethylation of MeHg does not seem to occur in *D. melanogaster*, at least during development (Rand *et al.* 2019), suggesting that the protective effects of P450s may not be due to MeHg detoxification; the protective effects of GSTs may be due to the formation of MeHg-glutathione conjugates, but these conjugates can form very efficiently without enzymatic action (Vorojeikina *et al.* 2017). It is likely that these P450s and GSTs protect against the neurotoxic effects of MeHg by metabolising toxic by-products of oxidative stress; given that ethanol and imidacloprid also induce oxidative stress (Balieira *et al.* 2018; Logan-Garbisch *et al.* 2015), *CG33301* and *CG16898* may have a similar function.

The molecular consequences of these SNPs in or near *CG33301* and *CG16898* are currently unclear. Five of the six SNPs are downstream of the coding sequence of their respective gene, suggesting they may modulate gene expression or post-transcriptional stability. The non-synonymous substitution in *CG33301* is at the start of the protein and converts a hydrophilic residue (T) to a hydrophobic residue (I). This could conceivably affect the catalytic function of the enzyme, but given its likely distance from the active site, it may be more likely to affect some other property of the enzyme, such as protein-protein interactions or post-translational processing. It is also possible that these SNPs are linked to uncharacterised yet causal structural variation.

I note that the transcription level of a gene from the sister Dro1 clade, *CG11878* (Dro1-5), is associated with ethanol tolerance in male flies (data from Morozova *et al.* 2015) and induced by ethanol exposure in female flies (Morozova *et al.* 2006), raising the possibility that this part of the EcKL gene family might be involved in such processes more generally. Curiously, while *CG11878* is induced by ectopic CncC, *CG33301* and *CG16898* are not (Misra *et al.* 2011), suggesting the latter two genes may not be part of the CncC-mediated oxidative stress response. However, they are both induced by tunicamycin (Chow *et al.* 2013), a nucleoside antibiotic that induces the unfolded protein response, which can in turn result in oxidative stress (Malhotra & Kaufman 2007)—this raises the possibility that these genes are part of an alternative oxidative stress response pathway.

These associations between toxic stress phenotypes and *CG33301* and *CG16898* are, unfortunately, not explored further in this thesis, but would likely be fruitful avenues for future research.

2.7.2. Detoxification candidate EcKLs

The results of this chapter have identified 24 detoxification candidate EcKLs in *Drosophila melanogaster*; these genes should ideally be studied further with the powerful genetic toolkit available in this species to determine if they indeed function in detoxification. Of particular interest in this respect are large clades in which many or most genes are detoxification candidates, as these may be promising areas to explore the recent evolution of detoxification functions through gene duplication. The Dro1 and

Dro5 clades, which have undergone extensive gene duplication and loss in the *Drosophila* genus (Chapter 2.4.1), have three and seven members in *D. melanogaster*, of which two and five are detoxification candidates. Interestingly, the Dro1 candidate genes are not induced in any of the seven xenobiotic exposure datasets analysed in this chapter (although, as previously mentioned, *CG11878* is induced by ethanol exposure; Morozova *et al.* 2006), while the Dro5 candidate genes are all induced by phenobarbital, and many by other toxins (except piperonyl butoxide). The Dro26 clade, discussed above in relation to toxic stress phenotypes, also contains good detoxification candidate genes.

Ultimately, due to the large number of detoxification candidate genes that it contains, and preliminary connections to caffeine detoxification, the Dro5 clade was chosen to be further characterised with respect to the detoxification of caffeine and other natural xenobiotic compounds in *D. melanogaster*. In Chapter 4, Dro5 EcKLs are characterised using CRISPR-Cas9 mutagenesis, UAS/GAL4 misexpression and pre-existing gene disruption lines.

2.7.3. Developmental EcKLs

While this chapter has focused on testing the hypothesis that some EcKLs encode detoxicative kinases, some data presented here (evolutionary conservation and RNAi knockdowns) might be used to generate candidates for EcKLs that are important for development (Chapter 2.5.5), with the caveat that—due to a lack of independent validation—the RNAi knockdown experiments in this chapter do not allow concrete conclusions to be drawn about the developmental functions of any particular *D. melanogaster* EcKLs.

Putative RNAi knockdown of *CG13813* (Dro38-1; named in this thesis *Wallflower*) results in a significant reduction in egg-to-adult viability (Fig. S2.8), suggesting it could be essential for development. The possible developmental functions of *CG13813*, and a related gene, *CG1561* (Dro37-0; recently named *Pinkman* by Santana *et al.* 2020), are explored further in Chapter 5, using multiple RNAi constructs, CRISPR-Cas9 mutagenesis and UAS/GAL4 misexpression.

2.8. Supplementary Materials

2.8.1. Supplementary Figures

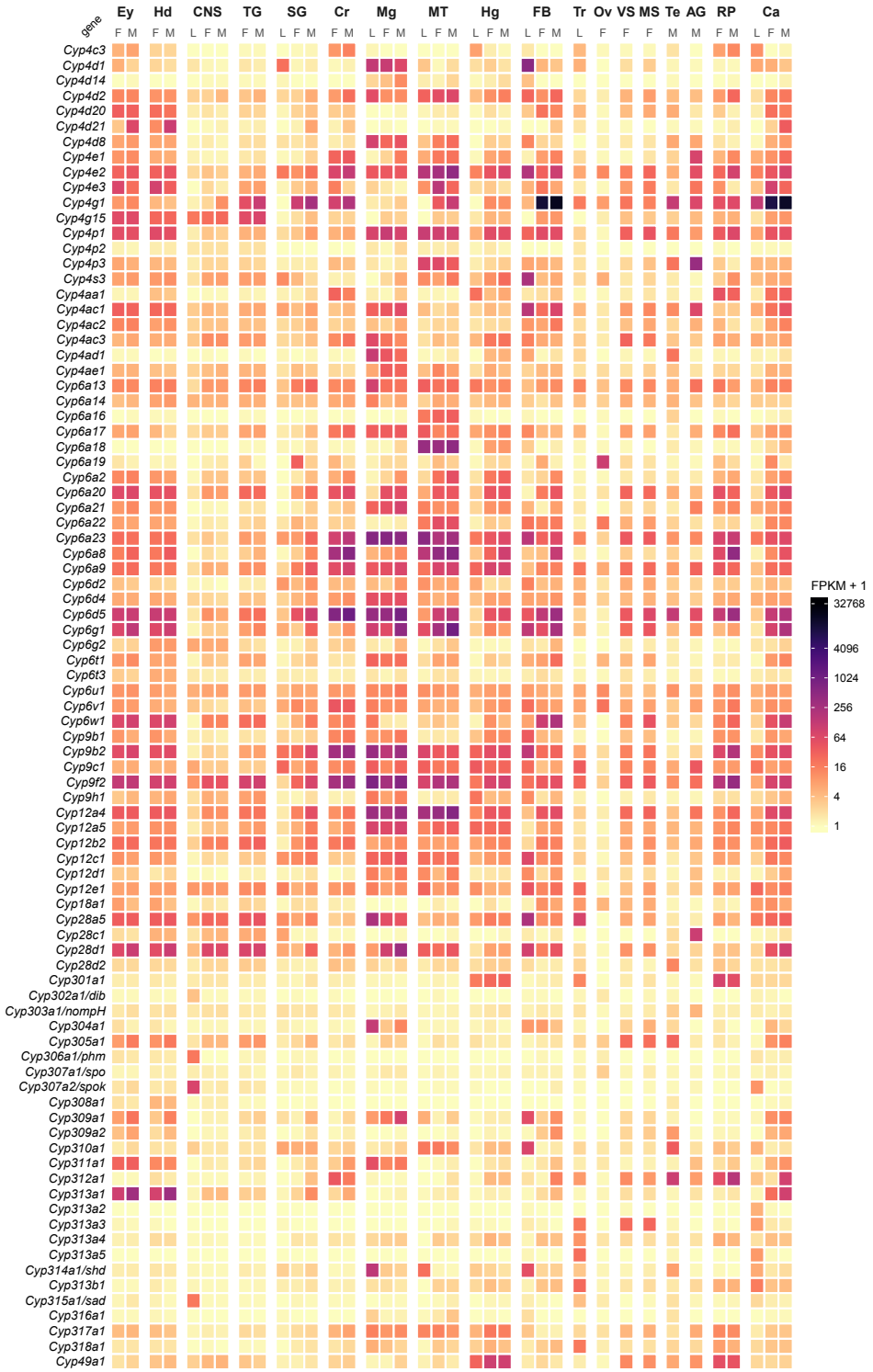


Figure S2.1. Heatmap of absolute expression (FPKM + 1) of P450 genes in *Drosophila melanogaster* across 18 tissues and three life stages from FlyAtlas 2 (Leader *et al.* 2018). Tissue and life stage abbreviations are the same as Fig. 2.4.

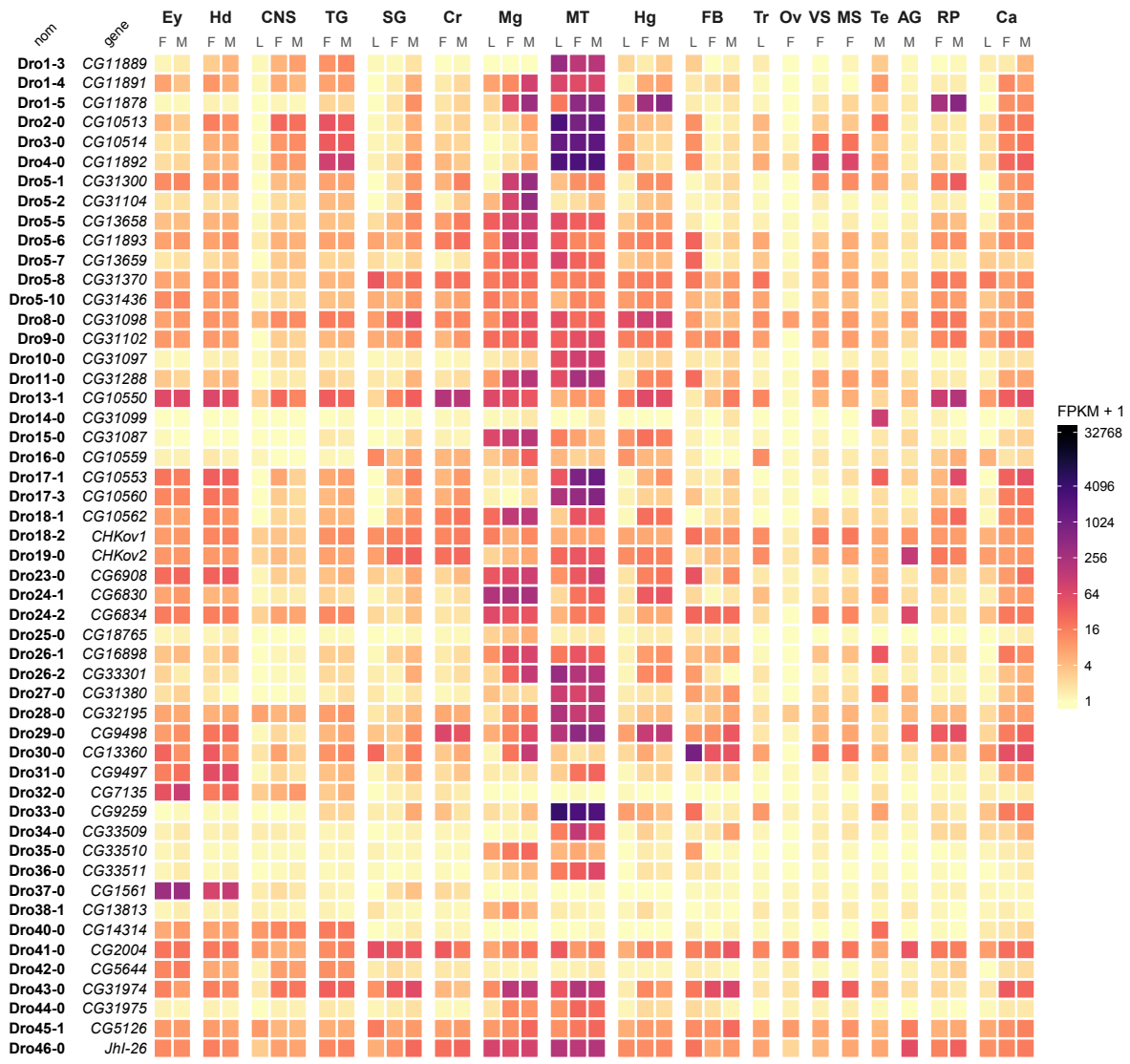
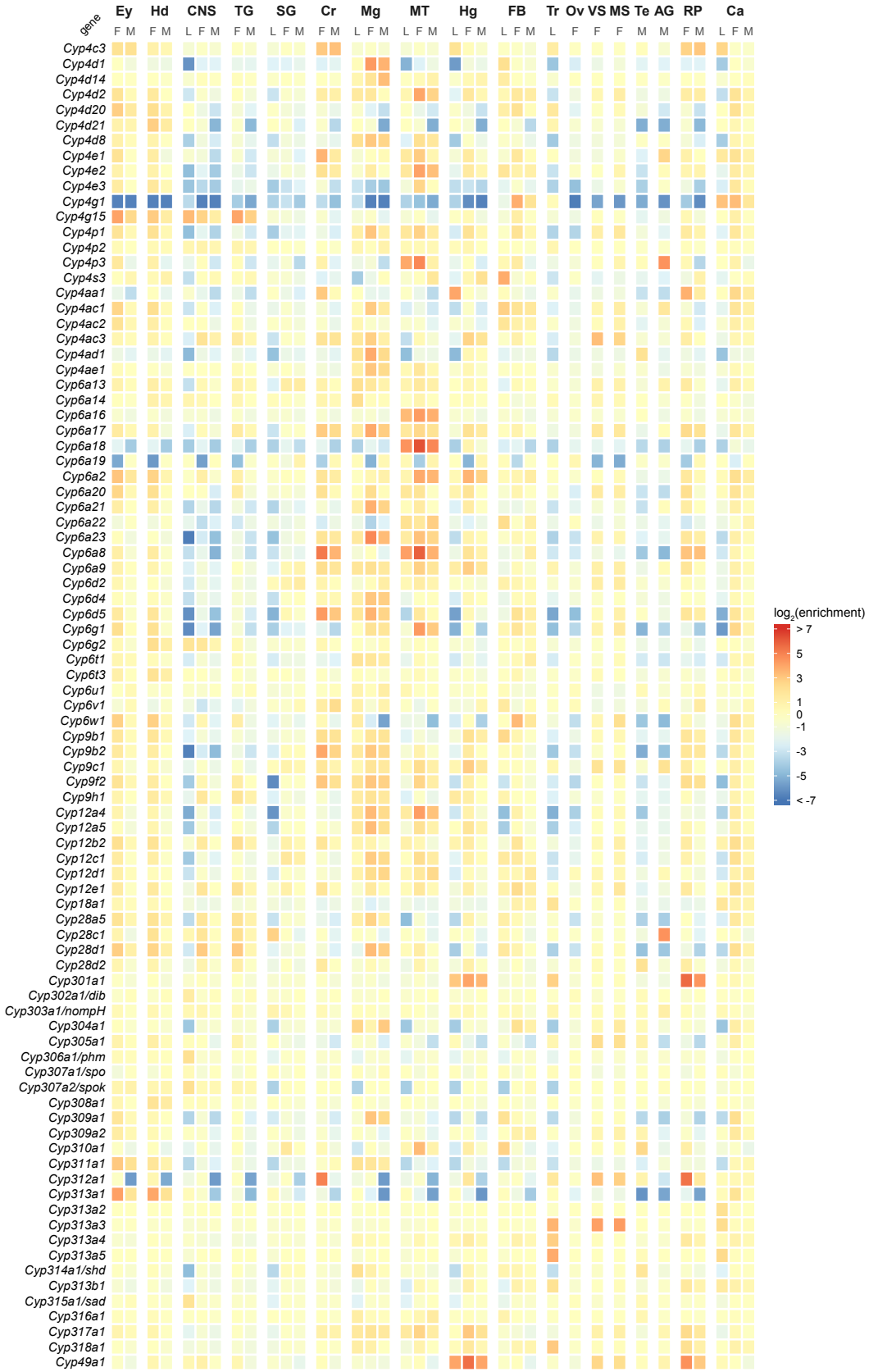


Figure S2.2. Heatmap of absolute expression (FPKM + 1) of EcKL genes in *Drosophila melanogaster* across 18 tissues and three life stages from FlyAtlas 2 (Leader *et al.* 2018). Tissue and life stage abbreviations are the same as Fig. 2.4.


gene
Ey Hd CNS TG SG Cr Mg MT Hg FB Tr Ov VS MS Te AG RP Ca
F M F M L F M F M L F M F M L F M L F M L

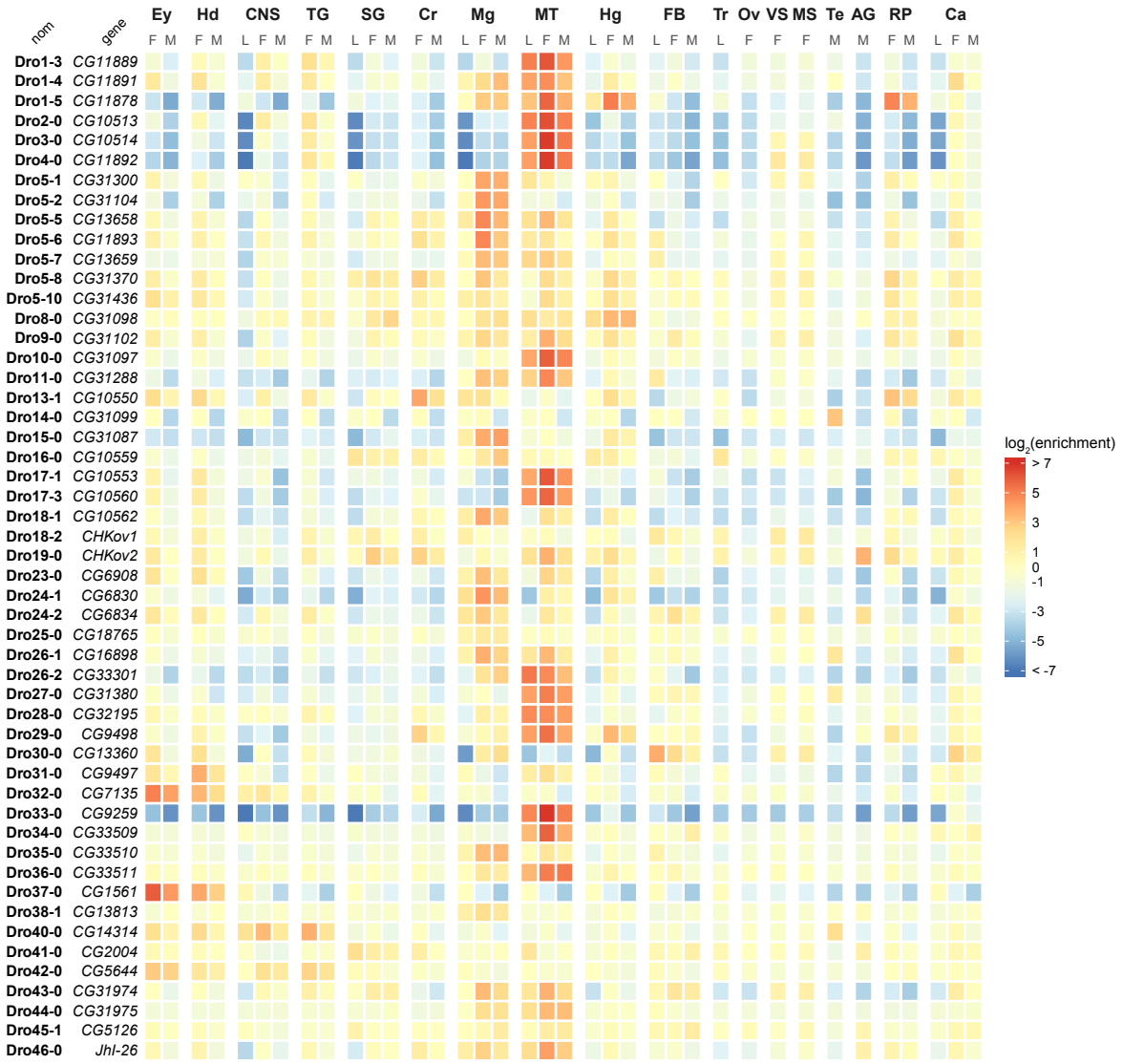


Figure S2.4. Heatmap of $\log_2$(enrichment) of EcKL genes in *Drosophila melanogaster* across 18 tissues and three life stages from FlyAtlas 2 (Leader *et al.* 2018). Tissue and life stage abbreviations are the same as Fig. 2.4.

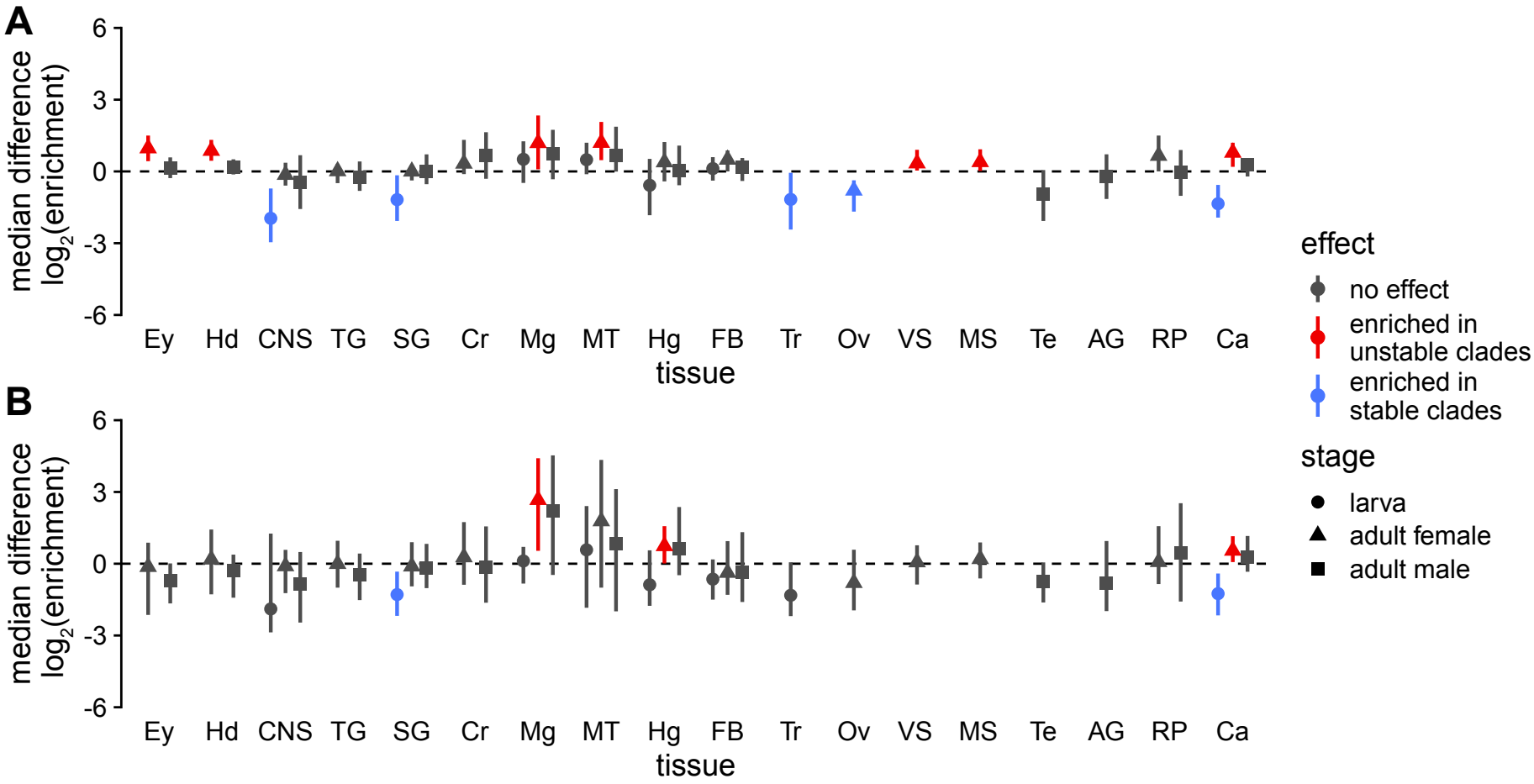


Figure S2.5. Median $\log_2$(enrichment) differences between unstable and stable genes in the (A) P450 and (B) EcKL gene families for specific tissues and life stages; data from FlyAtlas 2 (Leader *et al.* 2018). Error bars are 95% confidence intervals; effects were considered significant if the interval did not overlap with 0. Tissue and life stage abbreviations are the same as Fig. 2.4.

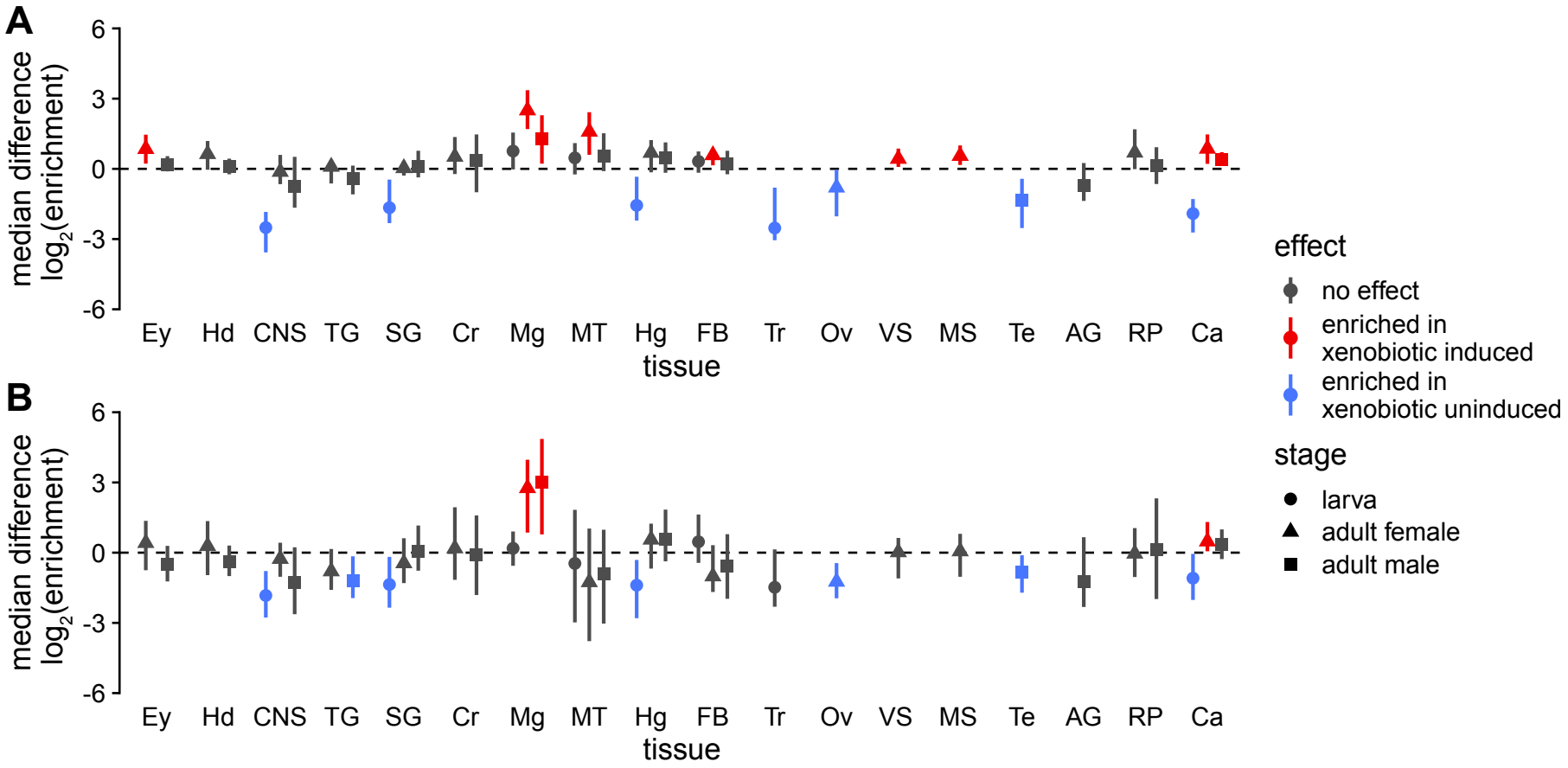


Figure S2.6. Median $\log_2$(enrichment) differences between xenobiotically induced and uninduced genes in the (A) P450 and (B) EcKL gene families for specific tissues and life stages; data from FlyAtlas 2 (Leader *et al.* 2018). Error bars are 95% confidence intervals; effects were considered significant if the interval did not overlap with 0. Tissue and life stage abbreviations are the same as Fig. 2.4.

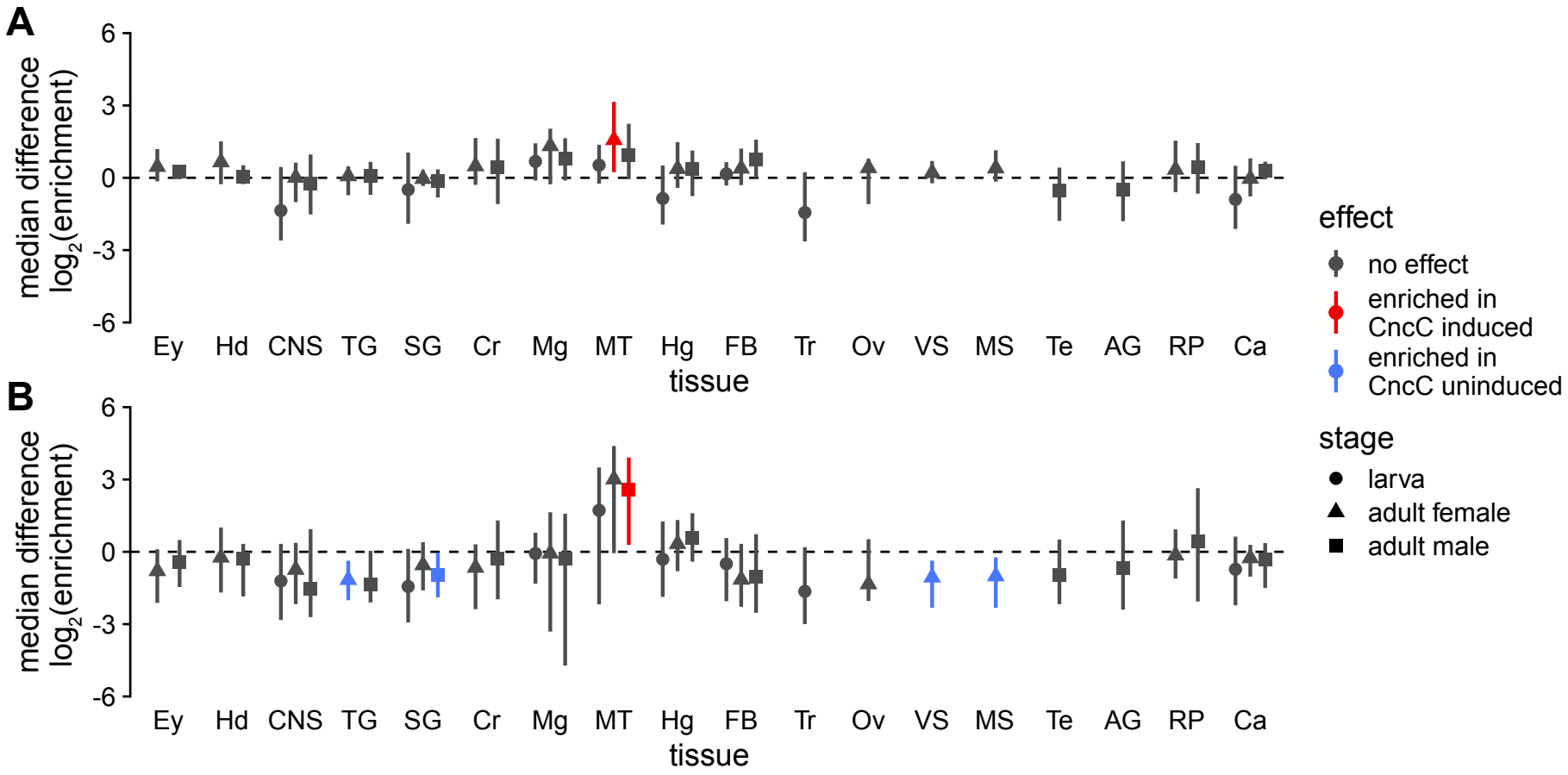


Figure S2.7. Median $\log_2$(enrichment) differences between CncC induced and uninduced genes in the (A) P450 and (B) EcKL gene families for specific tissues and life stages; data from FlyAtlas 2 (Leader *et al.* 2018). Error bars are 95% confidence intervals; effects were considered significant if the interval did not overlap with 0. Tissue and life stage abbreviations are the same as Fig. 2.4.

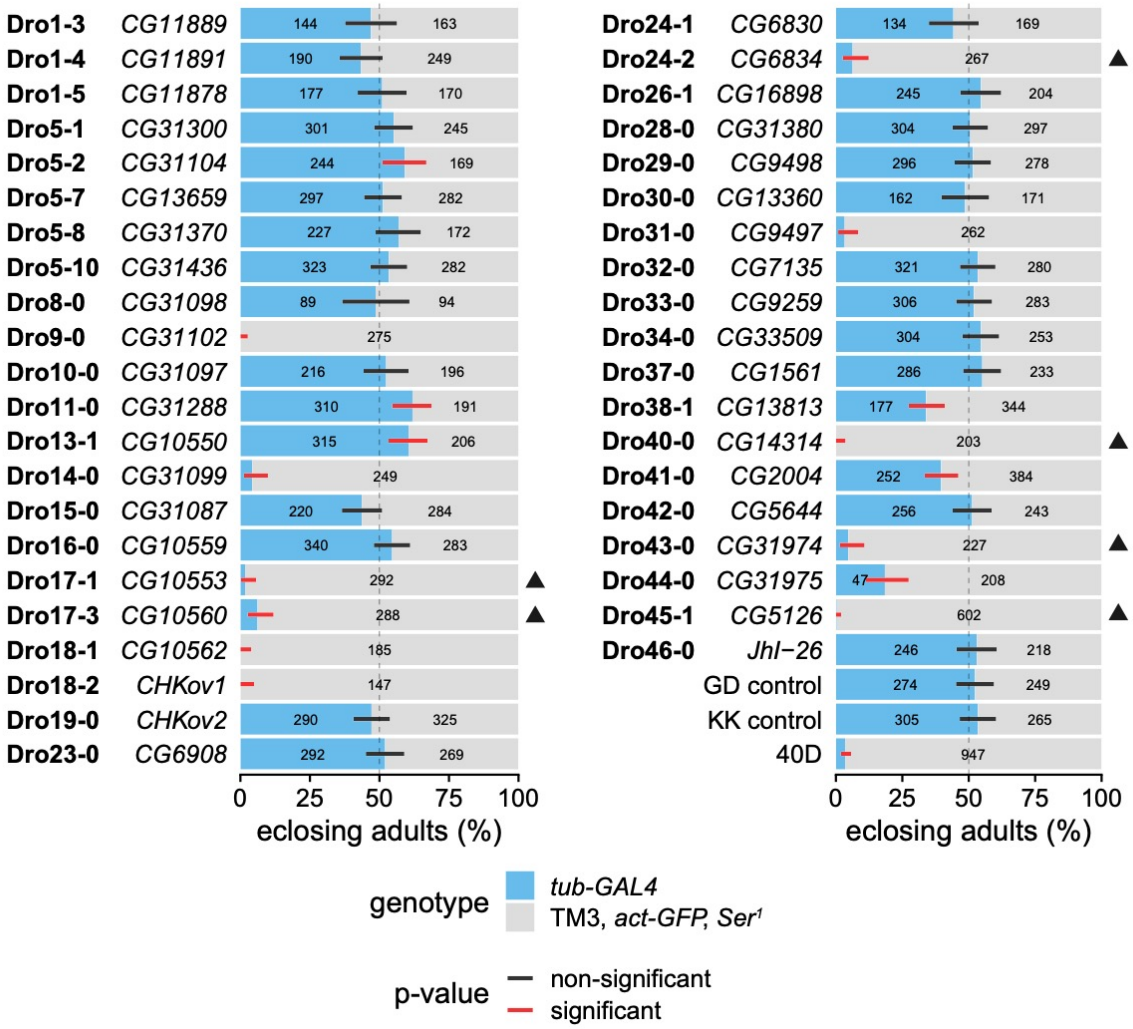


Figure S2.8. RNAi knockdown of 41 EcKL genes in *Drosophila melanogaster* and three control crosses, scoring offspring genotypes. Error bars are 99.89% confidence intervals (95% confidence interval adjusted for 44 tests) for the proportion of *tub-GAL4* individuals; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. For the *CG31098* cross, the UAS-dsRNA construct was over a CyO balancer, so only non-CyO individuals have been scored. For the *CHKov1* cross, the UAS-dsRNA construct was on the X-chromosome, so only female individuals (inheriting the driver and responder) have been scored. The number of eclosed adults of each genotype is indicated by the number within each bar (numbers less than 40 are not shown). 14 of the 41 EcKL knockdown crosses produced significantly fewer putative knockdown individuals than expected, indicating some degree of developmental lethality. However, six of these crosses involved lines from the KK library containing annotated hairpin insertions (black triangles), which can produce developmental defects due to ectopic expression of the gene *tiptop* and activation of the Hippo pathway (Green *et al.* 2014; Vissers *et al.* 2016). A control cross with the 40D responder line, which contains a UAS-only insertion at the annotated site, produced very few driver-containing offspring, strongly suggesting that the developmental lethality observed with the six annotated KK lines is due to *tiptop* misexpression and not the knockdown of the EcKL targeted by the dsRNA. dsRNA VDRC line IDs can be found in Table S2.4.

2.8.2. Supplementary Tables

The supplementary tables in this chapter are available online in the supplementary materials of Scanlan *et al.* (2020), with the following mapping (thesis to journal article):

- Table S2.1 is Supplementary Table 2
- Table S2.2 is Supplementary Table 3
- Table S2.3 is Supplementary Table 4
- Table S2.4 is Supplementary Table 5
- Table S2.5 is Supplementary Table 6

Chapter 3

Evolution and comparative phylogenomics of the EcKL gene family in insects

3.1. Introduction

In the previous chapter, the evolution of the EcKL gene family was explored across 12 species in the genus *Drosophila* (Diptera: Drosophilidae). These analyses provided some strong phylogenomic evidence supporting the hypothesis that specific EcKLs may play a role in detoxification in certain species of *Drosophila*, but to test this hypothesis more broadly, the evolutionary history of the EcKLs needs to be examined in a wider selection of insects.

3.1.1. What is known about the evolution of EcKLs across insects?

Previously published information on the evolution of the EcKL gene family is extremely limited. Only three phylogenetic trees of the EcKL family have been published so far (Fig. 3.1; Ito *et al.* 2008; Ito & Sonobe 2009; Sonobe & Ito 2009), all of which were conducted at a time when the number of annotated insect genomes was small, focus on *BmEc22K* in *Bombyx mori*, and were conducted by the same research group. A major problem with these phylogenetic explorations of the EcKL family is that they all include EcKL sequences from bacteria, teleosts and fungi, as well as DUF1679-containing sequences from nematodes—due to the very low sequence identity between these distantly related sequences, it is unlikely that the alignments generated will be accurate (Rost 1999); the limited and non-systematic selection of insect sequences also increases the chances of long-branch lengths that may attract diverged sequences (Bergsten 2005; Felsenstein 1978). In addition, the alignment program used in all cases was CLUSTAL X, which has been replaced by more accurate programs in recent years (Pais *et al.* 2014; Sievers & Higgins 2019). They also fail to report branch support values, making the interpretation of the trees difficult, and all three papers produce different relationships between sequences they all share (Fig. 3.1).

It is important to note that none of these papers attempted to group EcKL genes into orthologous clades, and so there is currently no information about how individual EcKLs can be classified or named in relation to the gene family as a whole. This also precludes any analyses about the evolutionary dynamics of the gene family or whether other insects have clear orthologs of *BmEc22K*, which remains the only EcKL to be biochemically characterised (Sonobe *et al.* 2006).

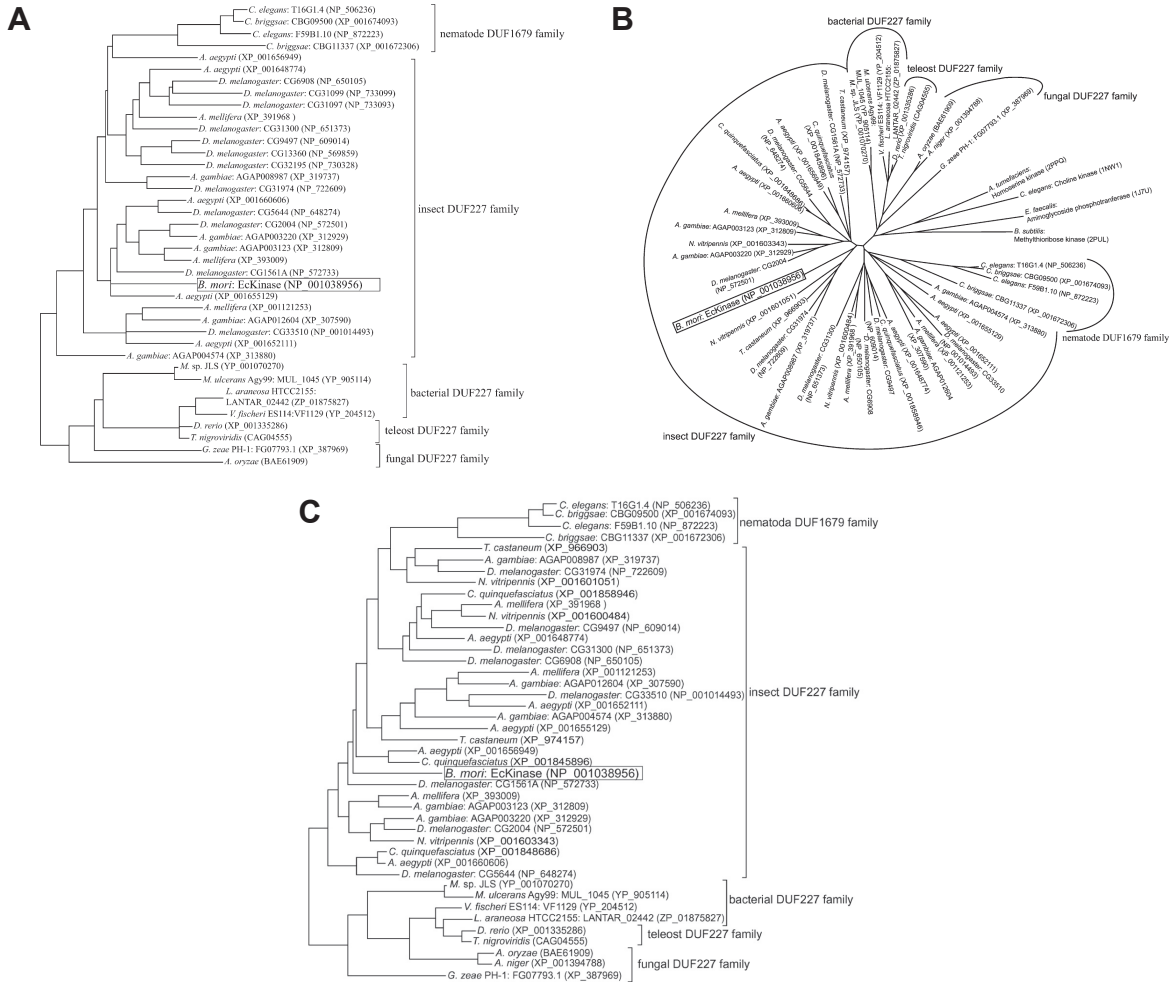


Figure 3.1. Three published phylogenetic trees of the EcKL gene family (here called DUF227): (A) a maximum likelihood tree from Ito *et al.* (2008); (B) a neighbour-joining tree from Ito & Sonobe (2009); and (C) a maximum likelihood tree from Sonobe & Ito (2009). *BmEc22K* (here called '*B. mori*: EcKinase') is indicated by a box around its tip label.

3.1.2. Why conduct comparative phylogenomic analyses of the EcKL family?

Comparative analyses of genomic data can lead to insights into the biochemistry and higher-level phenotypes of insects, such as sterol metabolism, endosymbiosis, immunity and sex determination systems (reviewed by Perry *et al.* 2019). Tied into a robust phylogenomic and comparative phylogenetic framework, which I call 'comparative phylogenomics' (CPG; Chapter 1.1.7), this approach can be extended to discover associations between multiple discrete and continuous traits, in order to generate and test hypotheses about the causal relationships between genomic and phenotypic changes (Cornwell & Nakagawa 2017). CPG could be valuable for exploring the possible functions of poorly characterised gene families, such as the EcKLs, which has only a handful of known functions (Chapter 1.4).

In Chapter 2, the hypothesis that EcKLs function in detoxification was tested through an analysis of the genes in the *Drosophila* genus, where the evolutionary and transcriptomic traits of particular EcKLs and the gene family as a whole were integrated to produce both lists of detoxification gene candidates and evidence for the detoxification hypothesis more generally; this method was also validated by the parallel analysis of the P450 gene family, which has known functions in detoxification. However, this genus-focused approach misses some of the advantages of a broader phylogenomic comparative analysis across all insects. First, the statistical power to detect genotype-phenotype associations increases as more taxa are analysed. Second, there is naturally far more phenotypic diversity across insects than there is in a single genus; all the *Drosophila* species annotated in Chapter 2 are saprophagous (Markow & O'Grady 2007), so there was no ability to robustly compare the evolution of the EcKLs between diets or detoxification breadth (DB; Chapter 1.2.1) groups. Third, large-scale phylogenetic analysis can allow for the development of a higher-level classification and nomenclature system for the EcKLs beyond what was possible with genus-level ancestral clades in Chapter 2, which can aid in predicting the functions of orthologs and paralogs in different taxa. Fourth, a larger taxonomic sampling can detect conservation over much deeper evolutionary timescales, which may be indicative of conserved functions.

Conservation is a central concept in molecular evolutionary biology, where it is used to detect the functional importance of genomic sequences at the level of gene families

down to specific nucleotides (Dolinski & Botstein 2007; Kimura 2012). However, there are multiple senses in which conservation can be meant, particularly when talking about how a gene or a group of genes might be conserved over a phylogeny:

1. sequence conservation—sequence-level identity between homologs;
2. presence conservation—the presence or absence of members of a group of genes in taxa of interest (also called gene retention; Waterhouse *et al.* 2011); or
3. copy-number conservation—the maintenance of orthologs as single-copy, rather than multi-copy (also called duplicability; Waterhouse *et al.* 2011).

These three senses of conservation can imply slightly different things about the genes to which they apply: high sequence conservation implies the specific sequence of the ancestral gene is very important for function (Cooper & Brown 2008); high presence conservation implies a gene clade's ancestral and/or derived functions are important for the biology of the taxa in which they are conserved; and high conservation of single-copy status (low duplicability) implies the gene may be highly sensitive to changes in dosage (Waterhouse *et al.* 2011) and may have retained an ancestral expression pattern (Kryuchkova-Mostacci & Robinson-Rechavi 2016). These conservation senses also tend to be related: single-copy orthologs tend to exhibit high sequence conservation, and universally retained genes tend to remain single-copy orthologs (Waterhouse *et al.* 2011).

Integrating phylogenomic associations with gene conservation data is a powerful way to explore the possible functions of members of a gene family, how important those functions might be, and how those functions may have evolved between taxa; this will be the focus of this chapter.

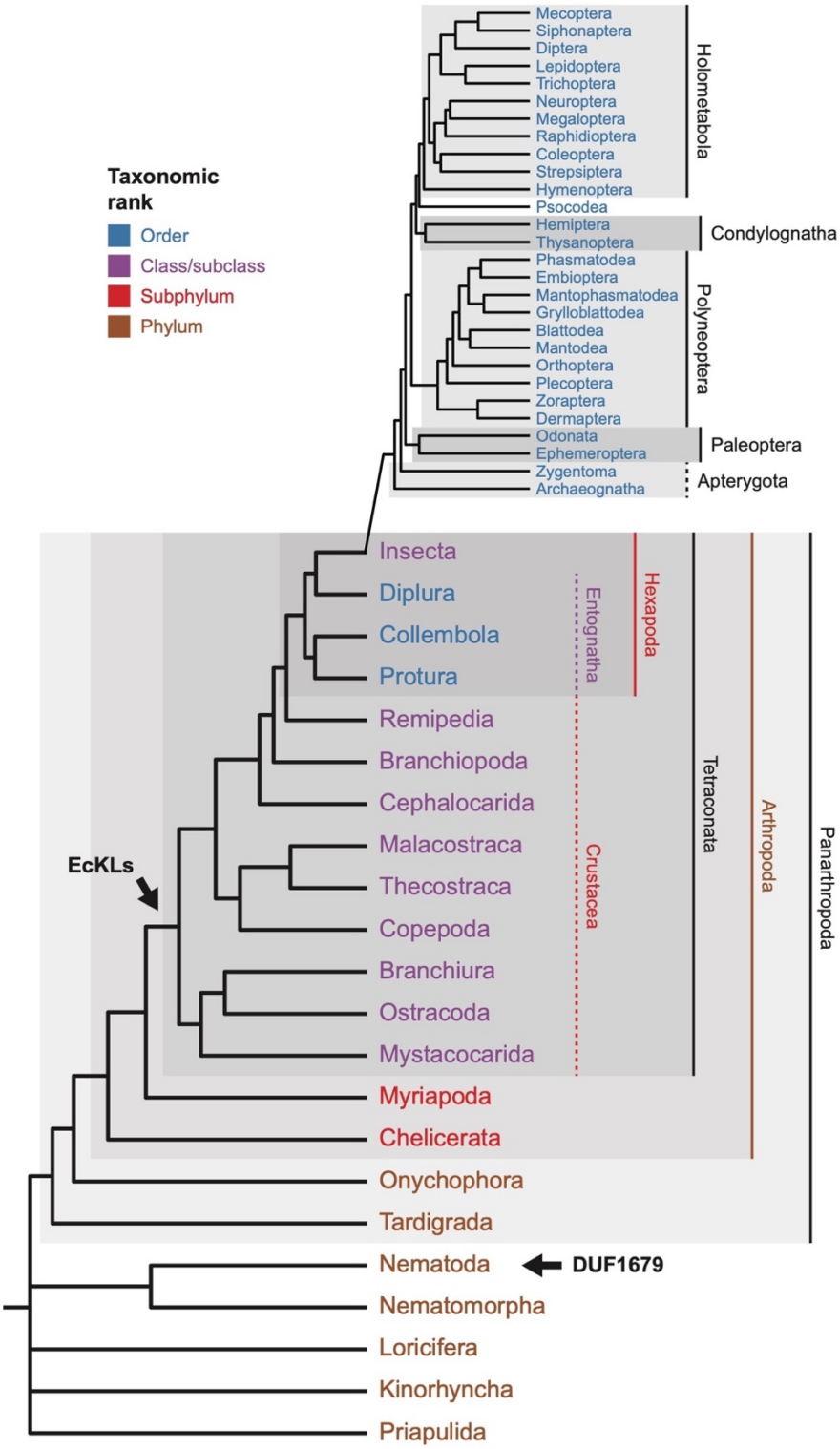


Figure 3.2. Representative phylogeny of the superphylum Ecdysozoa, with a focus on Arthropoda, Tetraconata and Insecta; insect orders are at the top. Grey boxes encompass higher taxonomic groupings centered on Insecta, with names to the right; paraphyletic groupings (Apterygpta, Entognatha and Crustacea) are noted with dashed lines. Taxonomic ranks are indicated by colours; unranked groups are black. Proposed origin of the EcKL gene family is marked, as is the presence of the related DUF1679 gene family. Phylogenetic relationships are derived from Misof *et al.* (2014), Giribet & Edgecombe (2017), Schwentner *et al.* (2017) and Schwentner *et al.* (2018); branch lengths are illustrative only and are not necessarily to scale.

3.1.3. Chapter Aim

The aim of this chapter is to follow on from the *Drosophila*-specific evolutionary analyses in Chapter 2 and explore the evolution of the EcKL gene family across insects, and to specifically test possible associations with detoxification-related traits using phylogenetic comparative methods. I manually annotated over 100 insect and arthropod genome assemblies, as well as numerous transcriptomes, for EcKL gene models and conducted detailed phylogenetic analyses to categorise individual EcKL genes into insect-wide subfamilies, as well as order-specific ancestral clades for the three best-sampled orders: Diptera, Lepidoptera and Hymenoptera. Using phylogenetic comparative methods, the size of the EcKL family in each genome, as well as the sizes of known detoxification gene families, were regressed against each other and two separate but related traits: insect diet and estimated DB. Mining a host plant database for herbivorous Lepidoptera allowed host plant diversity to be used as a quantitative proxy for DB, and its relationships to EcKL family size and lepidopteran ancestral EcKL clade sizes were explored. In addition, the stability of EcKLs in two lineages with small DB—the tsetse flies and bees—was estimated and compared with the *Drosophila* genus. Here, I integrate these data and discuss how they support the hypothesis that the EcKL gene family encodes detoxification enzymes and how they might be used to predict which EcKLs are involved in detoxification in different insect taxa.

3.2. Materials and Methods

3.2.1. Genome and transcriptome annotation

Arthropod genome assemblies uploaded to NCBI Assembly (Kitts *et al.* 2016), VectorBase (Giraldo-Calderón *et al.* 2015) or Lepbase (Challi *et al.* 2016) were queried with EcKL protein sequences using tblastn (Altschul *et al.* 1990) and matching scaffolds/contigs were downloaded and annotated manually in Artemis (Carver *et al.* 2012). RefSeq proteins and/or transcriptome shotgun assemblies (TSAs) from the closest available taxonomic group were queried with rough annotated gene model translations to iteratively inform intron-exon boundaries. Gene models were classified as 'full', 'partial' or 'pseudogenous,' with the latter defined as containing two or more inactivating mutation (frameshift, splice donor/acceptor loss, premature stop codon etc.); pseudogenous models were not rigorously annotated or collated. Gene models with only a single inactivating mutation were considered null alleles, had the mutation conservatively reverted, and otherwise treated as 'full' models. Only full or partial gene models were counted towards EcKL totals for each species. Rarely, there was evidence of alternatively spliced gene models that affected the protein sequence, in which case isoforms were counted as individual genes towards gene totals if alternative protein-coding exons accounted for more than 25% of the coding sequence. For genome assembly information, including the total number of full and partial gene models per assembly, see Table S3.1.

To obtain EcKL subfamily sequences from insect orders not represented in the genome-annotation dataset, NCBI TSAs were tblastn queried with putative EcKL subfamily sequences from closely related orders, and the putative open reading frames from selected transcripts were translated with the ExPASy Translate tool (Artimo *et al.* 2012)—up to three sequences with the highest sequence similarity per subfamily were collated per order.

Apoidea transcriptomes were annotated by using *Apis mellifera* EcKLs as tblastn queries against assembled transcriptomic contigs in the NCBI database and translating open reading frames with the ExPASy Translate tool (Artimo *et al.* 2012). Overlapping partial transcripts were manually merged into larger contigs where possible. EcKLs

from transcriptomes were assigned to ancestral hymenopteran clades by aligning with *A. mellifera*, *Nasonia vitripennis* and *Neodiprion lecontei* EcKLs with MAFFT (Katoh & Standley 2013) and constructing phylogenetic trees with IQ-TREE (Nguyen *et al.* 2015), ala. the methods described in Chapter 3.2.2.

3.2.2. Multiple sequence alignment, phylogenetic inference and ancestral clade assignment

Before alignment, sequences with multiple EcKL domains (Chapter 1.4) were split into smaller sequences each containing one EcKL domain. Sequences with N- or C-terminal regions that did not share homology with the rest of the EcKL family (such as the N-terminal disordered domain in *Drosophila melanogaster* CG1561) were removed before family-wide sequence alignments. Sequences were aligned with MAFFT 7.402, using the L-INS-i setting unless otherwise specified (recommended for sequences with a single conserved domain; Katoh & Standley 2013). MSAs were manually trimmed in AliView (Larsson 2014) to remove obviously poorly aligned columns (typically at the N- and C-terminal ends of the alignment), but over-trimming (removal of >20% of columns) was avoided, given that it often substantially reduces the accuracy of phylogenetic inference (Dessimoz & Gil 2010; Tan *et al.* 2015). Phylogenetic inference was performed on trimmed MSAs with IQ-TREE 1.6.10 (Nguyen *et al.* 2015), using the ModelFinder program to find the best model for each MSA (Kalyaanamoorthy *et al.* 2017) and UFBoot2 for bootstrapped branch support values (Hoang *et al.* 2018). IQ-TREE runs were performed 5–10 times per MSA and the tree with the highest log-likelihood was used. MAFFT and IQ-TREE runs were performed in the CIPRES Science Gateway computing environment on the XSEDE platform (Miller *et al.* 2010).

For ancestral clade phylogenetics where the total number of sequences was greater than approx. 500, all sequences in the taxon were aligned, then the MSA was pared down to a subset of representative sequences (i.e. groups of sequences with very high similarity were reduced to a single representative, and partial sequences were also removed) in order to improve phylogenetic parameter estimation (recommended in Burnham & Anderson 2004). Clade designations for excluded sequences (i.e. those not in the subset) were determined by their grouping with similar sequences in the guide tree produced by MAFFT during alignment. Where ambiguities arose, small groups of sequences from the total sequence set (e.g. all sequences thought to be in a particular

clade, plus a small number of appropriate outgroup sequences) were aligned and a tree was produced with IQ-TREE.

To determine insect EcKL subfamilies, a rough initial tree was constructed of a single sequence from each ancestral clade from Diptera, Lepidoptera and Hymenoptera, as well as representative sequences (ala. ancestral clade phylogenetics above) from all other genome-annotated insect taxa. To produce the final EcKL subfamily tree seen in Fig. 3.5, a single sequence from each putative subfamily per order (including transcriptome-only orders) was selected, except for Lepidoptera, where a single representative sequence from each of the Lep1–8/16–17 ancestral clades (subfamily A) was included in order to break up long branch lengths associated with these sequences, and for Odonata, Ephemeroptera, Zygentoma and Archaeognatha, where up to three sequences were selected from these orders per subfamily to help break up long branches. In total, 241 sequences were selected and aligned using the E-INS-i setting (for better deep node support; Wilding *et al.* 2017) in MAFFT, and then imported into IQ-TREE for analysis. To classify individual EcKLs into subfamilies, groups of sequences from each order were independently aligned to the 241 EcKL sequences described above and a tree produced with IQ-TREE.

For the arthropod EcKL tree, the selection of 241 representative insect EcKLs described above were combined with the 107 sequences from *Catajapyx aquilonaris* (Hexapoda: Diplura), *Hyalella azteca* (Malacostraca: Amphipoda) and *Oithona nanai* (Copepoda: Cyclopoida; see Fig. 3.2 for relationships to insects), aligned with the E-INS-i setting in MAFFT, then imported into IQ-TREE for analysis.

3.2.3. An estimated time tree for insects

A dated phylogenetic tree with branch lengths in units of millions of years was manually constructed in Mesquite (Maddison & Maddison 2019) for all 140 insect species annotated for EcKLs, using phylogenetic relationships and node ages published in the literature (Fig. 3.3). Where node ages conflicted between studies, the most recent study was used. Nodes with unknown age were estimated based on the known divergence times of similar taxonomic ranks in other parts of the tree (inaccuracy of these node ages is thought to have relatively minor impact on downstream comparative analyses; see Stone 2011).

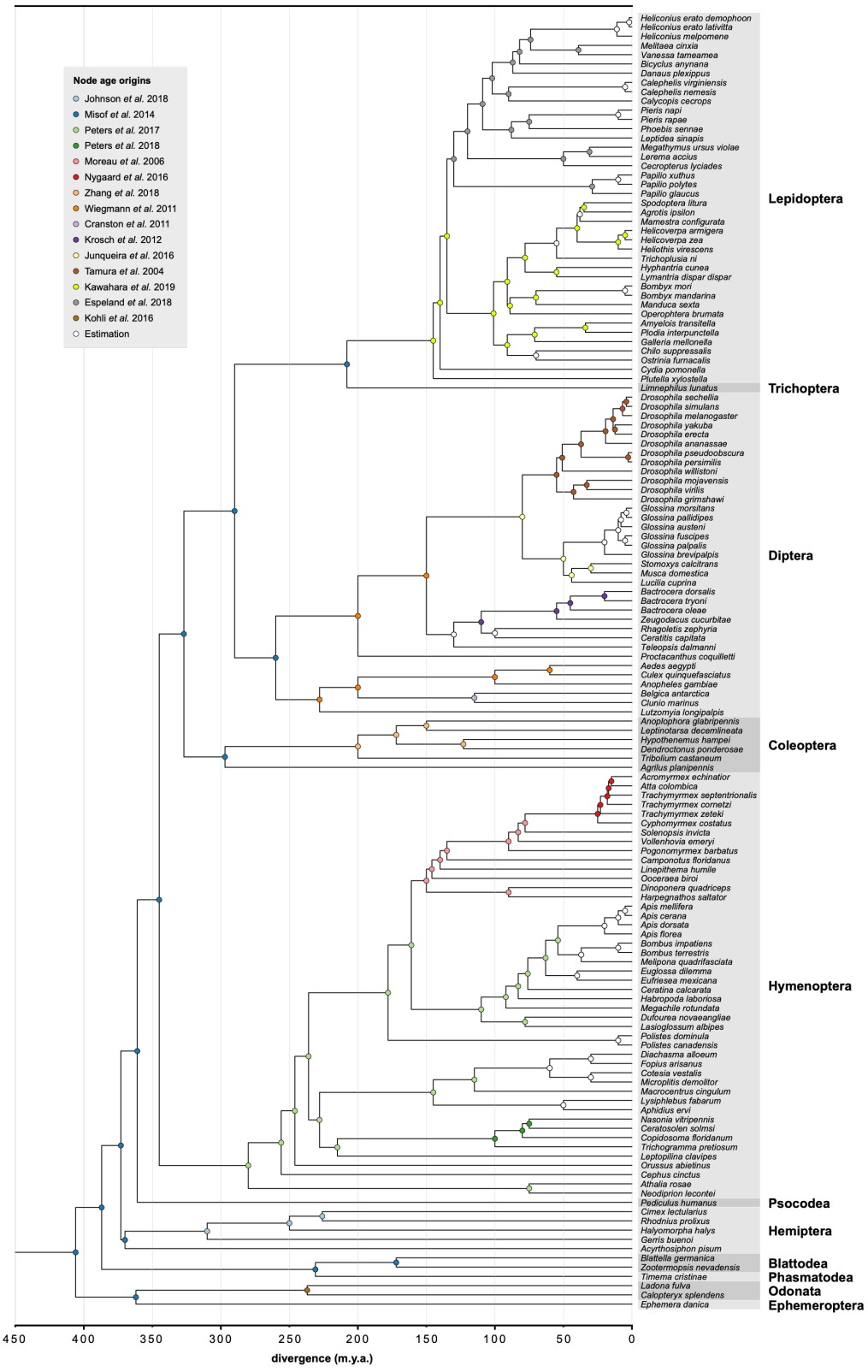


Figure 3.3. Master phylogeny of all insect species that were genomically annotated for EcKL gene models in this study, used in phylogenetic comparative methods. Coloured nodes indicate the studies from which their age was determined (Cranston *et al.* 2011; Espeland *et al.* 2018; Johnson *et al.* 2018a; Junqueira *et al.* 2016; Kawahara *et al.* 2019; Kohli *et al.* 2016; Krosch *et al.* 2012; Misof *et al.* 2014; Moreau *et al.* 2006; Nygaard *et al.* 2015; Peters *et al.* 2017; 2018; Tamura *et al.* 2004; Wiegmann *et al.* 2011; Zhang *et al.* 2018); white nodes were estimated without the use of published data.

3.2.4. Detoxification gene family sizes, dietary groups and DB groups

EcKL gene number per genome (n = 140 species) were determined earlier (Chapter 3.2.1). Cytochrome P450 (P450), glutathione S-transferase (GST) and carboxylcholinesterase (CCE) gene family sizes for species with EcKL annotations determined in this study (n = 98 species) were taken from Rane *et al.* (2019); the UDP-glycosyltransferase (UGT) and ABC transporter (ABC) family sizes were not used, as the data may be inaccurate (Rane *et al.* 2019). P450, GST and CCE data was also not used for *Gerris buenoi*, as the unrealistically low number of GST genes (six) cast doubt on the reliability of the annotation for this genome. Phylogenetic signal was determined on the insect phylogeny (Fig. 3.3) for gene family size traits with the R package *phylosignal* (Keck *et al.* 2016).

Dietary groups for each insect were defined as carnivorous (carnivorous or parasitoidal; n = 19), herbivorous (herbivorous or fungivorous; n = 54), haematophagous (n = 9), pollen-feeding (n = 15) or detritivorous (detritivorous, saprophagous or omnivorous; n = 43). If juvenile and adult diets were different, the diet with the higher chemical complexity was used, in the preference order detritivorous > herbivorous > carnivorous > pollen-feeding > haematophagous. *Calephelis nemesis*, *Calephelis virginiens* and *Calycopis cecrops* (Lepidoptera) were coded as detritivorous (see Chapter 3.2.6).

An estimation of detoxification breadth (DB; Chapter 1.2.1) was assigned for every species in the dataset, with three levels—small (DB_S), intermediate (DB_I) and large (DB_L)—based on the estimated overall chemical complexity of their diets during juvenile and adult life stages (Table 3.1). Detritivores and saprophages were coded as DB_L, except for *Drosophila mojavensis* and *Drosophila sechellia*, which are relatively specialised in their diets compared to related species (Date *et al.* 2013; Dworkin & Jones 2008) and were coded as DB_I. Pollen feeders and haematophages were consistently coded with DB_S. For lepidopteran species, DB was coded based on the host plant diversity (number of host plant families) already established: one family, DB_S; two to nine families, DB_I; ten or more families, DB_L. Non-lepidopteran herbivores were typically coded as DB_I unless their diets were known to be particularly chemically simple (e.g. the wood-eating *Zootermopsis nevadensis*; DB_S) or complex (e.g. the generalist pest *Halyomorpha halys*; DB_L). Coding DB for some ants and all parasitoid wasps was challenging and the coded values are of low confidence—most ants are omnivorous but

can have a preference for carnivory (Clay *et al.* 2017), while others are technically fungivorous but may be exposed to many different plant secondary metabolites through leaf foraging (De Fine Licht & Boomsma 2010), while the amount of exposure of parasitoids to xenobiotic compounds is poorly understood and is hard to estimate. The uncertainty of these DB assignments was explored during analyses.

Gene family sizes, juvenile and adult diets, dietary grouping and DB coding for each insect species can be found in Table 3.1.

Table 3.1. The taxonomic position, gene family sizes, diet, diet coding (Code) and estimated detoxification breadth (DB) for all 140 insects genomically annotated for the EcKL gene family in this chapter. Species without P450, GST or CCE gene family data in Rane *et al.* (2019) have the appropriate cells left blank.

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Blattella germanica	Ectobiidae	Blaberoidea	Blattodea	105	152	50	82	varied	varied	det	L
Zootermopsis nevadensis	Termopsidae	Termopsidae	Blattodea	19	78	20	40	wood	wood	her	S
Agrilus planipennis	Buprestidae	Buprestoidea	Coleoptera	24	65	25	47	wood and phloem	leaves	her	I
Anoplophora glabripennis	Cerambycidae	Chrysomeloidea	Coleoptera	59	98	36	47	wood	leaves and bark	her	I
Leptinotarsa decemlineata	Chrysomelidae	Chrysomeloidea	Coleoptera	49	90	32	102	leaves	leaves	her	I
Dendroctonus ponderosae	Curculionidae	Curculionoidea	Coleoptera	44	92	35	64	bark, phloem	bark, spores	her	I
Hypothenemus hampei	Curculionidae	Curculionoidea	Coleoptera	28				coffee berry	coffee berry	her	I
Tribolium castaneum	Tenebrionidae	Tenebrionoidea	Coleoptera	39	126	40	54	varied	varied	det	L
Proctacanthus coquillettii	Asilidae	Asiloidea	Diptera	26				unknown	insects	car	S
Belgica antarctica	Chironomidae	Chironomoidea	Diptera	40				moss, detritus	moss, detritus	det	I
Clunio marinus	Chironomidae	Chironomoidea	Diptera	34				plankton, algae	unknown	det	I
Aedes aegypti	Culicidae	Culicoidea	Diptera	50	152	40	67	organic detritus	blood, nectar	det	L
Anopheles gambiae	Culicidae	Culicoidea	Diptera	43	113	40	54	organic detritus	blood, nectar	det	L
Culex quinquefasciatus	Culicidae	Culicoidea	Diptera	60	193	46	86	organic detritus	blood, nectar	det	L
Teleopsis dalmanni	Diopsidae	Diopsoidea	Diptera	60				decaying plant matter	decaying plant matter	det	L
Drosophila ananassae	Drosophilidae	Ephydroidea	Diptera	56	95	49	40	decaying fruit	decaying fruit	det	L
Drosophila erecta	Drosophilidae	Ephydroidea	Diptera	53	100	53	45	decaying fruit	decaying fruit	det	L
Drosophila grimshawi	Drosophilidae	Ephydroidea	Diptera	45	71	37	48	decaying bark	decaying bark	det	L
Drosophila melanogaster	Drosophilidae	Ephydroidea	Diptera	51	87	52	40	decaying fruit	decaying fruit	det	L
Drosophila mojavensis	Drosophilidae	Ephydroidea	Diptera	41	75	36	45	decaying cactus	decaying cactus	det	I
Drosophila persimilis	Drosophilidae	Ephydroidea	Diptera	44	91	49	50	decaying fruit	decaying fruit	det	L
Drosophila pseudoobscura	Drosophilidae	Ephydroidea	Diptera	44	89	42	44	decaying fruit	decaying fruit	det	L

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Drosophila sechellia	Drosophilidae	Ephydroidea	Diptera	47	94	47	47	decaying *Morinda* fruit	decaying *Morinda* fruit	det	I
Drosophila simulans	Drosophilidae	Ephydroidea	Diptera	56	85	44	41	decaying fruit	decaying fruit	det	L
Drosophila virilis	Drosophilidae	Ephydroidea	Diptera	44	79	38	42	decaying bark	decaying bark	det	L
Drosophila willistoni	Drosophilidae	Ephydroidea	Diptera	61	97	46	57	decaying fruit	decaying fruit	det	L
Drosophila yakuba	Drosophilidae	Ephydroidea	Diptera	55	85	42	56	decaying fruit	decaying fruit	det	L
Glossina austeni	Glossinidae	Hippoboscoidea	Diptera	26	68	24	43	parental milk	blood	hae	S
Glossina brevipalpis	Glossinidae	Hippoboscoidea	Diptera	27	65	20	36	parental milk	blood	hae	S
Glossina fuscipes	Glossinidae	Hippoboscoidea	Diptera	26	82	23	39	parental milk	blood	hae	S
Glossina morsitans	Glossinidae	Hippoboscoidea	Diptera	27	62	21	38	parental milk	blood	hae	S
Glossina pallidipes	Glossinidae	Hippoboscoidea	Diptera	28	66	20	41	parental milk	blood	hae	S
Glossina palpalis	Glossinidae	Hippoboscoidea	Diptera	26	69	22	40	parental milk	blood	hae	S
Musca domestica	Muscidae	Muscoidea	Diptera	63	150	39	47	varied	varied	det	L
Stomoxys calcitrans	Muscidae	Muscoidea	Diptera	54	202	46	54	decaying matter	blood	det	L
Lucilia cuprina	Calliphoridae	Oestroidea	Diptera	56	93	37	43	carrion, decaying matter	decaying matter	det	L
Lutzomyia longipalpis	Psychodidae	Psychodoidea	Diptera	31	100	37	43	decaying matter	blood, nectar	det	L
Bactrocera dorsalis	Tephritidae	Tephritoidea	Diptera	47	96	41	43	fruit, yeast	fruit, yeast	det	L
Bactrocera oleae	Tephritidae	Tephritoidea	Diptera	39	103	45	49	olives, yeast	olives, yeast	det	I
Bactrocera tryoni	Tephritidae	Tephritoidea	Diptera	45				fruit, yeast	fruit, yeast	det	L
Ceratitis capitata	Tephritidae	Tephritoidea	Diptera	46	105	38	46	fruit, yeast	fruit, yeast	det	L
Rhagoletis zephyria	Tephritidae	Tephritoidea	Diptera	46	199	62	89	fruit, yeast	fruit, yeast	det	L
Zeugodacus cucurbitae	Tephritidae	Tephritoidea	Diptera	53	99	41	53	fruit, yeast	fruit, yeast	det	L
Ephemera danica	Ephemeridae	Ephemeroidea	Ephemeroptera	35				organic detritus	does not feed	det	L
Acyrthosiphon pisum	Aphididae	Aphidoidea	Hemiptera	16	77	31	47	phloem	phloem	her	I
Cimex lectularius	Cimicidae	Cimicoidea	Hemiptera	19	56	20	50	blood	blood	hae	S
Gerris buenoi	Gerridae	Gerroidea	Hemiptera	36				insects	insects	car	S
Halyomorpha halys	Pentatomidae	Pentatomoidea	Hemiptera	36	130	34	82	various plants	various plants	her	L

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Rhodnius prolixus	Reduviidae	Reduvioidea	Hemiptera	33	117	23	69	blood	blood	hae	S
Apis cerana	Apidae	Apoidea	Hymenoptera	12	41	18	40	pollen, honey	pollen, nectar	pol	S
Apis dorsata	Apidae	Apoidea	Hymenoptera	12	41	17	23	pollen, honey	pollen, nectar	pol	S
Apis florea	Apidae	Apoidea	Hymenoptera	12	42	20	32	pollen, honey	pollen, nectar	pol	S
Apis mellifera	Apidae	Apoidea	Hymenoptera	12	48	19	29	pollen, honey	pollen, nectar	pol	S
Bombus impatiens	Apidae	Apoidea	Hymenoptera	12	48	19	25	pollen, honey	pollen, nectar	pol	S
Bombus terrestris	Apidae	Apoidea	Hymenoptera	12	52	23	25	pollen, honey	pollen, nectar	pol	S
Ceratina calcarata	Apidae	Apoidea	Hymenoptera	12	81	23	47	pollen, honey	pollen, nectar	pol	S
Eufriesea mexicana	Apidae	Apoidea	Hymenoptera	12	50	18	31	pollen, honey	pollen, nectar	pol	S
Euglossa dilemma	Apidae	Apoidea	Hymenoptera	12				pollen, honey	pollen, nectar	pol	S
Habropoda laboriosa	Apidae	Apoidea	Hymenoptera	12	41	17	33	pollen, honey	pollen, nectar	pol	S
Melipona quadrifasciata	Apidae	Apoidea	Hymenoptera	15	58	17	28	pollen, honey	pollen, nectar	pol	S
Dufourea novaeangliae	Halictidae	Apoidea	Hymenoptera	12	46	23	32	pollen, honey	pollen, nectar	pol	S
Lasioglossum albipes	Halictidae	Apoidea	Hymenoptera	16				pollen, honey	pollen, nectar	pol	S
Megachile rotundata	Megachilidae	Apoidea	Hymenoptera	12	53	20	24	pollen, honey	pollen, nectar	pol	S
Cephus cinctus	Cephidae	Cephoidea	Hymenoptera	18	88	16	30	grass stems	does not feed	her	I
Ceratosolen solmsi	Agaonidae	Chalcidoidea	Hymenoptera	13	38	19	22	figs	figs	her	S
Copidosoma floridanum	Encyrtidae	Chalcidoidea	Hymenoptera	28	82	21	50	moth larvae	unknown, possibly pollen or nectar	car	I
Nasonia vitripennis	Pteromalidae	Chalcidoidea	Hymenoptera	47	95	28	49	flesh fly pupae	host juices, nectar	car	I
Trichogramma pretiosum	Trichogrammatidae	Chalcidoidea	Hymenoptera	27	68	25	49	moth eggs	nectar, pollen	car	S
Leptopilina clavipes	Figitidae	Cynipoidea	Hymenoptera	24				Drosophila larvae	unknown	car	I
Acromyrmex echinatior	Formicidae	Formicoidea	Hymenoptera	27	71	24	28	farmed fungi (indirect herbivory)	farmed fungi (indirect herbivory)	her	L
Atta colombica	Formicidae	Formicoidea	Hymenoptera	24	59	22	28	farmed fungi (indirect herbivory)	farmed fungi (indirect herbivory)	her	L
Camponotus floridanus	Formicidae	Formicoidea	Hymenoptera	20	173	20	34	various	various	det	I

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Cyphomyrmex costatus	Formicidae	Formicoidea	Hymenoptera	24	67	20	28	farmed fungi (indirect omnivory)	farmed fungi (indirect omnivory)	det	L
Dinoponera quadriceps	Formicidae	Formicoidea	Hymenoptera	13	126	19	23	various	various	det	I
Harpegnathos saltator	Formicidae	Formicoidea	Hymenoptera	12	96	18	26	various	various	det	I
Linepithema humile	Formicidae	Formicoidea	Hymenoptera	14	110	14	30	various	various	det	I
Ooceraea biroi	Formicidae	Formicoidea	Hymenoptera	22				other ants	other ants	car	S
Pogonomyrmex barbatus	Formicidae	Formicoidea	Hymenoptera	17	80	21	35	varied	varied	det	I
Solenopsis invicta	Formicidae	Formicoidea	Hymenoptera	24	197	22	32	varied	varied	det	I
Trachymyrmex cornetzi	Formicidae	Formicoidea	Hymenoptera	34	77	21	28	farmed fungi (indirect herbivory)	farmed fungi (indirect herbivory)	her	L
Trachymyrmex septentrionalis	Formicidae	Formicoidea	Hymenoptera	25	62	20	29	farmed fungi (indirect herbivory)	farmed fungi (indirect herbivory)	her	L
Trachymyrmex zeteki	Formicidae	Formicoidea	Hymenoptera	26	58	18	28	farmed fungi (indirect herbivory)	farmed fungi (indirect herbivory)	her	L
Vollenhovia emeryi	Formicidae	Formicoidea	Hymenoptera	18	129	19	27	varied	varied	det	I
Aphidius ervi	Braconidae	Ichneumonoidea	Hymenoptera	14				aphids (endoparasitoid)	nectar, honeydew	car	I
Cotesia vestalis	Braconidae	Ichneumonoidea	Hymenoptera	19				*Plutella xylostella* larvae (endoparasitoid)	nectar, honeydew	car	I
Diachasma alloeum	Braconidae	Ichneumonoidea	Hymenoptera	24	71	20	40	*Rhagoletis pomonella* larvae (endoparasitoid)	nectar, honeydew	car	I
Fopius arisanus	Braconidae	Ichneumonoidea	Hymenoptera	19	56	19	33	Tephritid eggs and larvae (endoparasitoid)	nectar, honeydew	car	S
Lysiphlebus fabarum	Braconidae	Ichneumonoidea	Hymenoptera	16				aphids (endoparasitoid)	nectar, honeydew	car	S
Macrocentrus cingulum	Braconidae	Ichneumonoidea	Hymenoptera	13				*Ostrinia nubilalis* larvae (endoparasitoid)	nectar, honeydew	car	I

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Microplitis demolitor	Braconidae	Ichneumonoidea	Hymenoptera	21	50	20	44	Noctuidae larvae (endoparasitoid)	nectar, honeydew	car	I
Orussus abietinus	Orussidae	Orussoidea	Hymenoptera	14	38	20	27	Beetle larvae (ectoparasitoid)	nectar, honeydew	car	S
Neodiprion lecontei	Diprionidae	Tenthredinoidea	Hymenoptera	33				pine trees	nectar	her	I
Athalia rosae	Tenthredinidae	Tenthredinoidea	Hymenoptera	29	63	25	46	Brassica	nectar	her	I
Polistes canadensis	Vespidae	Vespoidea	Hymenoptera	16	64	19	26	insects	insects	car	S
Polistes dominula	Vespidae	Vespoidea	Hymenoptera	16	89	19	25	insects	insects	car	S
Bombyx mandarina	Bombycidae	Bombycoidea	Lepidoptera	15				mulberry	does not feed	her	S
Bombyx mori	Bombycidae	Bombycoidea	Lepidoptera	15	90	30	96	mulberry	does not feed	her	S
Manduca sexta	Sphingidae	Bombycoidea	Lepidoptera	38	93	24	93	MPF[a]	nectar	her	L
Operophtera brumata	Geometridae	Geometroidea	Lepidoptera	29	129	30	92	MPF	does not feed	her	L
Hyphantria cunea	Erebidae	Noctuoidea	Lepidoptera	23				MPF	does not feed	her	L
Lymantria dispar dispar	Erebidae	Noctuoidea	Lepidoptera	36				MPF	does not feed	her	L
Agrotis ipsilon	Noctuidae	Noctuoidea	Lepidoptera	36				MPF	nectar	her	L
Helicoverpa armigera	Noctuidae	Noctuoidea	Lepidoptera	38	114	49	103	MPF	nectar	her	L
Helicoverpa zea	Noctuidae	Noctuoidea	Lepidoptera	36	103	47	95	MPF	nectar	her	L
Heliothis virescens	Noctuidae	Noctuoidea	Lepidoptera	42	101	35	89	MPF	nectar	her	L
Mamestra configurata	Noctuidae	Noctuoidea	Lepidoptera	28				MPF	nectar	her	L
Spodoptera litura	Noctuidae	Noctuoidea	Lepidoptera	46	122	46	115	MPF	nectar	her	L
Trichoplusia ni	Noctuidae	Noctuoidea	Lepidoptera	24				MPF	nectar	her	L
Cecropterus lyciades	Hesperiidae	Papilionoidea	Lepidoptera	28				Fabaceae	nectar	her	S
Lerema accius	Hesperiidae	Papilionoidea	Lepidoptera	23				Gramineae	nectar	her	S
Megathymus ursus violae	Hesperiidae	Papilionoidea	Lepidoptera	16				Agavaceae	does not feed	her	S
Calycopis cecrops	Lycaenidae	Papilionoidea	Lepidoptera	54				MPF, dead leaves	nectar	det	L
Bicyclus anynana	Nymphalidae	Papilionoidea	Lepidoptera	20	115	30	91	Gramineae	fruit	her	S
Danaus plexippus	Nymphalidae	Papilionoidea	Lepidoptera	25	81	30	55	MPF	nectar	her	I

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Heliconius erato demophoon	Nymphalidae	Papilionoidea	Lepidoptera	20				Passifloraceae	pollen	her	S
Heliconius erato lativitta	Nymphalidae	Papilionoidea	Lepidoptera	17				Passifloraceae	pollen	her	S
Heliconius melpomene	Nymphalidae	Papilionoidea	Lepidoptera	19	122	41	77	Passifloraceae	pollen	her	S
Melitaea cinxia	Nymphalidae	Papilionoidea	Lepidoptera	23	83	24	65	MPF	nectar	her	I
Vanessa tameamea	Nymphalidae	Papilionoidea	Lepidoptera	13				Urticaceae	sap	her	S
Papilio glaucus	Papilionidae	Papilionoidea	Lepidoptera	21				MPF	nectar	her	L
Papilio polytes	Papilionidae	Papilionoidea	Lepidoptera	27	115	38	89	Rutaceae	nectar	her	S
Papilio xuthus	Papilionidae	Papilionoidea	Lepidoptera	26	90	31	78	Rutaceae	nectar	her	S
Leptidea sinapis	Pieridae	Papilionoidea	Lepidoptera	16				Fabaceae	nectar	her	S
Phoebis sennae	Pieridae	Papilionoidea	Lepidoptera	14				MPF	nectar	her	I
Pieris napi	Pieridae	Papilionoidea	Lepidoptera	18				MPF	nectar	her	I
Pieris rapae	Pieridae	Papilionoidea	Lepidoptera	17	85	33	77	MPF	nectar	her	I
Calephelis nemesis	Riodinidae	Papilionoidea	Lepidoptera	44				MPF, maybe rotten leaves	nectar	det	L
Calephelis virginiensis	Riodinidae	Papilionoidea	Lepidoptera	39				Compositae, maybe rotten leaves	nectar	det	L
Chilo suppressalis	Crambidae	Pyraloidea	Lepidoptera	21	76	18	57	MPF	nectar	her	I
Ostrinia furnacalis	Crambidae	Pyraloidea	Lepidoptera	36				MPF	nectar	her	L
Amyelois transitella	Pyralidae	Pyraloidea	Lepidoptera	31	99	44	75	MPF	nectar	her	L
Galleria mellonella	Pyralidae	Pyraloidea	Lepidoptera	26				wax, pollen, honey	does not feed	pol	S
Plodia interpunctella	Pyralidae	Pyraloidea	Lepidoptera	37	60	21	60	MPF	does not feed	her	L
Cydia pomonella	Tortricidae	Tortricoidea	Lepidoptera	28				MPF	nectar, possibly fruit	her	I
Plutella xylostella	Plutellidae	Yponomeutoidea	Lepidoptera	17	98	44	71	MPF	nectar	her	L
Calopteryx splendens	Calopterygidae	Calopterygidae	Odonata	30	199	49	144	insects	insects	car	S
Libellula fulva	Libellulidae	Libellulidae	Odonata	29				insects	insects	car	S
Timema cristinae	Timematidae	Timematodea	Phasmatodea	53				MPF	MPF	her	L

Species	Family	Superfamily	Order	EcKL	P450	GST	CCE	Juvenile Diet	Adult Diet	Code	DB
Pediculus humanus	Pediculidae	Anoplura	Psocoptera	12				blood	blood	hae	S
Limnephilus lunatus	Limnephilidae	Limnephiloidea	Trichoptera	47	122	35	68	fresh/decaying leaves, detritus	does not feed	det	L

[a] multiple plant families

3.2.5. PGLS regression of gene family sizes in insects

Phylogenetic generalised least squares (PGLS) regression between gene family sizes was performed with the 'pgls' function in the package *caper* (version 1.0.1) in R, with the λ parameter estimated with maximum likelihood. The *ape* package (version 5.0; Paradis & Schliep 2018) was used to manipulate phylogenetic trees in R.

3.2.6. PGLS regression of EcKL family size and host plant diversity in Lepidoptera

Caterpillar/host plant relationships were extracted from the HOSTS Database (Robinson *et al.* 2010) in early 2019, and the number of unique host plant families, genera and species, as well as polyphagy designation, were compiled for each annotated Lepidopteran species (Fig. 3.4A). Those without an entry in the HOSTS Database (*Papilio xuthus*) or no associated plant taxa (*Galleria mellonella*) were not included in further analyses. Detritivory at adult or larval stages may be a confounding variable in this analysis, as rotting food substrates are a likely source of bacterial and fungal toxins, which may influence detoxification gene family evolution. To account for this, we conducted subsequent analyses twice: once on the full dataset, and once on a subset that contained only non-detritivorous species. As related members of the Riodinini tribe are detritivorous as larvae and adults (Hall & Willmott 2000) and species of in the genus *Calephelis* can feed on partially rotten leaves (Kendall 1959), the species *Calephelis nemesis* and *Calephelis virginiens* were marked as detritophagous, as was *Calycopis cecrops*, which feeds on fallen leaves and detritus as larvae (references in Cong *et al.* 2016).

Direct polyphagy designation in the HOSTS Database was not adequately consistent with host plant diversity measures based on the number of associated host plant families, genera or species (Fig. 3.4B), so the latter were used as proxies for dietary toxin breadth. These measures were highly correlated in the full dataset (Table 3.3), so family number was reasoned the best measure of dietary toxin breadth and chosen for clade-specific analyses.

Phylogenetic relationships between lepidopteran species were extracted from the master phylogeny previously estimated for all insects (Fig. 3.3). Host plant families, genera and species counts ('host counts') were $\log_2$-transformed (as their distributions

were heavily right-skewed), except where noted. PGLS regression of EcKL number to host counts was performed with the 'pgls' function in the package *caper*, with the λ parameter estimated with maximum likelihood. The *ape* package (version 5.0; Paradis & Schliep 2018) was used to manipulate phylogenetic trees.

3.2.7. PGLS regression of gene family sizes and diet/detoxification breadth (DB)

Phylogenetic ANOVAs for gene family size against diet or DB were performed with the 'gls' function in the *nlme* package (version 3.1-141) in R, with the master phylogeny (pruned to overlap species for P450s, GSTs and CCEs) used as a correlation structure with 'corPagel' from the *ape* package (version 5.0; Paradis & Schliep 2018) and a λ value estimated by maximum likelihood. The *ape* package was also used to manipulate phylogenetic trees. The 'glht' function from the *multcomp* package (version 1.4-10; Hothorn *et al.* 2008) in R was used to perform two-sided multiple comparison of means post-hoc tests between each pair of diets or DB levels, with p-values adjusted for multiple comparisons with the 'single-step' argument. For DB analyses, two datasets were analysed: the full dataset with all species (140 or 98 species, for EcKLs and P450s/GSTs/CCEs, respectively), and a restricted dataset without omnivorous/fungivorous ants or parasitoid wasps.

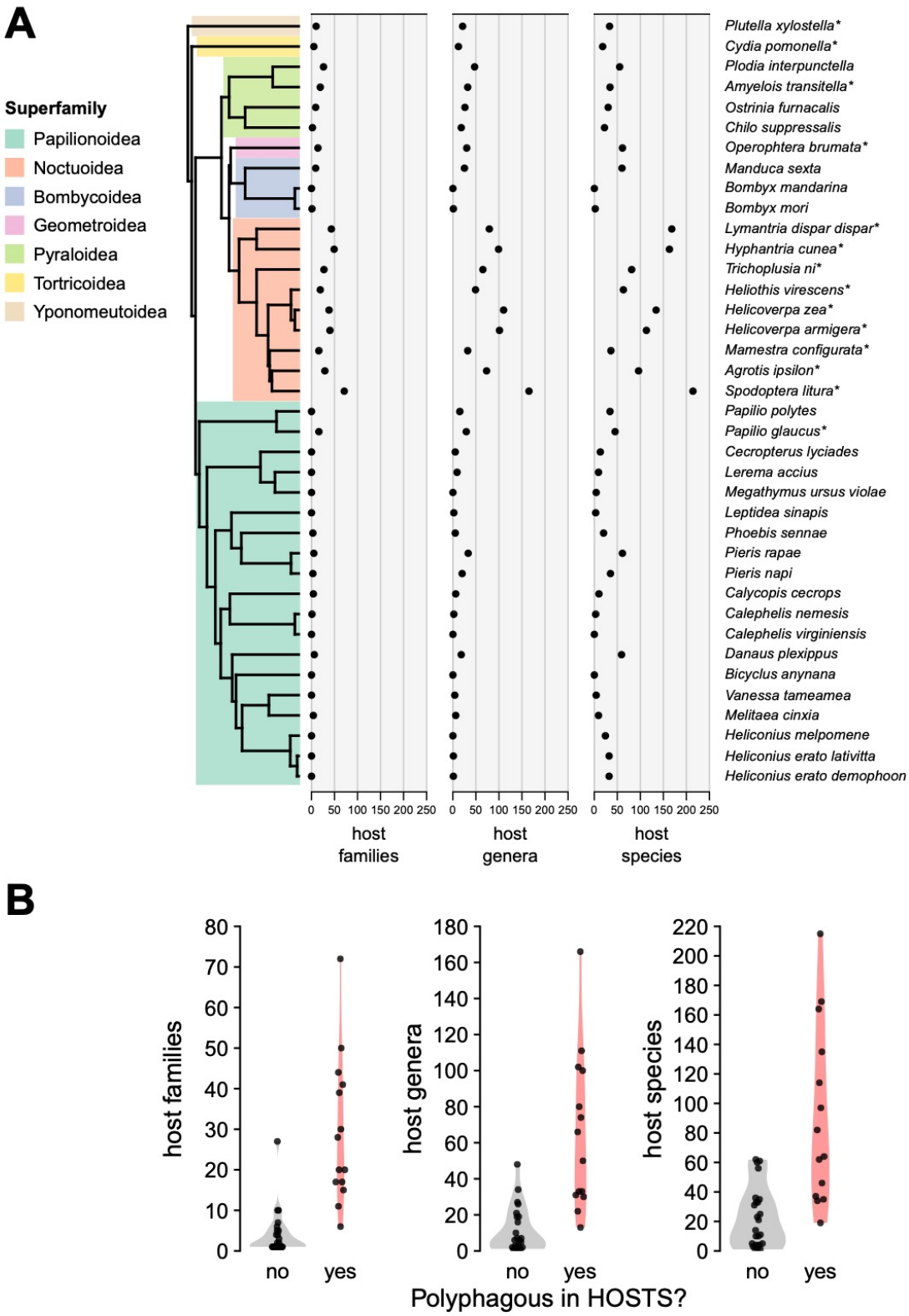


Figure 3.4. Lepidopteran host plant data from the HOSTS Database (Robinson *et al.* 2010). (A) Unique host plant families, genera and species for 38 lepidopteran species, taken from the HOSTS Database. Insect superfamilies are coloured on the phylogenetic tree to the left, derived from Fig. 3.1. Species designated 'polyphagous' in the HOSTS Database are denoted by asterisks. (B) The number of unique host plant families (left), genera (middle) and species (right) of lepidopteran species/subspecies designated 'polyphagous' (red; n = 14) or 'non-polyphagous' (grey; n = 24) in the HOSTS Database.

3.3. Results

3.3.1. Annotation of the EcKL gene family in insect and arthropod genomes

The EcKL gene family was annotated in 140 insect genome assemblies, representing 139 species (including two subspecies of *Heliconius erato*) from 58 families, 43 superfamilies and 11 orders, including the 12 *Drosophila* genomes from Chapter 2. The number of putatively functional EcKL genes per genome ranged substantially, from 12 in the louse *Pediculus humanus*, the ant *Harpegnathos saltator* and 12 bee species, to 105 in the cockroach *Blattella germanica*. Three non-insect genomes were also annotated: the hexapod *Catajapyx aquilonaris* (Diplura: Japygidae; 71 EcKLs) and two crustaceans, *Hyalella azteca* (Amphipoda: Hyalellidae; 17 EcKLs) and *Oithona nana* (Cyclopoida: Oithonidae; 19 EcKLs). BLAST searches against TSAs from all major groups of crustaceans indicated the presence of EcKL domain-containing transcripts, although these were not analysed further. In total, 4,319 EcKL gene models were annotated, comprising 3,999 (92.6%) full gene models and 320 (7.4%) partial gene models (Table S3.1). The proportion of total gene models that were partial gene models from any particular assembly was significantly predicted by measures of assembly contiguity (N50 and L50; Fig. S3.1), suggesting more fragmented genome assemblies produce less accurate total EcKL gene family sizes (Denton *et al.* 2014).

The annotation of non-Tetraconata genomes was not seriously attempted, given that the only EcKL sequences found in the TSAs of arthropods outside the Tetraconata (n = 16 TSAs for Myriapoda, n = 144 TSAs for Chelicerata) were in mites in the genus *Varroa* (Chelicerata: Varroidae), which are obligate bee ectoparasites and feed on bee fat body tissue (Ramsey *et al.* 2019). These *Varroa* sequences have near-identity to EcKLs from species in the bee genus *Apis* and are likely derived from bee RNA contamination present in the guts of *Varroa* samples and not transcribed from the mite's genome. In addition, no EcKL protein sequences were found in the NCBI 'nr' database when restricted to Chelicerata and Myriapoda when using a wide variety of EcKL sequences as queries. Overall, these data strongly suggest the EcKL gene family is absent from the genomes of myriapods and chelicerates, and is restricted to the Tetraconata in arthropods. EcKLs were also not found in the protein or TSA databases of Onychophora (velvet worms) or Tardigrada (tardigrades), phyla that form the

Panarthropoda with arthropods, suggesting the EcKL gene family arose in the Tetraconata (Fig. 3.2). In addition, the related DUF1679 gene family was not found to be present in panarthropod protein and TSA databases; if EcKL and DUF1679 genes arose from an ancestral gene family in the ancestor of Nematoda and Panarthropoda, at least four loss events must have occurred in the Ecdysozoa (Fig. 3.2).

3.3.2. EcKL subfamilies in insects

Phylogenetic analysis of EcKLs from all 28 insect orders (Misof *et al.* 2014) strongly suggests that insect EcKLs can be classified into 13 ancestral subfamilies, labelled A through M (Fig. 3.5). The sizes of these subfamilies in individual genomes varies widely, from zero in many cases to 76 in subfamily H in *Blattella germanica* (Blattodea); there are also clear patterns of some subfamilies being preferentially conserved as single-copy orthologs in most taxa (subfamilies C, D, G and I), while others have expanded in numerous orders (subfamilies A, B, E, F, H, J, K, L; Figs. 3.6–7). In annotated genomes, where false negative rates of gene presence are thought to be low (although it is likely fragmented assemblies are missing some genes; see Fig. S3.1), subfamilies had highly variable conservation, with some subfamilies (I and J) present—if not fully retained—in all orders and others present in one or two orders (K, L and M; Fig. 3.8). Across genomes and queried TSAs, there was additional evidence for the widespread presence of subfamilies I and J across insects, while subfamily M—previously eight orphaned genomic sequences in *B. germanica* (Blattodea)—was present in the transcriptomes of other polyneopteran orders (Fig. 3.9). Overall, no EcKL subfamily was fully retained across all insects.

Branch support for subfamilies B, C, E, F, I, J, K and M were very high (Fig. 3.5). While basal branch support for subfamily D was low (57), it was much higher (95) for an internal branch that excluded apterygotan (Zygentoma and Archaeognatha) sequences; subfamilies G and H also had relatively low basal branch support values due to the inclusion of apterygotan sequences. Phylogenetic signal from these basal insect lineages may be poor, and so inclusion of apterygotan EcKLs in the gene subfamilies as currently defined should be regarded as tentative until better sampling can be achieved. Subfamilies A and L also had low basal branch support—in the case of L, this is likely due to poor sampling from Odonata and Ephemeroptera, and in the case of A, it may be because many sequences in this subfamily are rapidly evolving,

particularly in the Holometabola. While many deep nodes in the subfamily tree were poorly resolved, there is evidence for some higher groupings of subfamilies: A, B, C, L and M form a strongly supported monophyletic clade, as do J and K. There is also weaker support for a grouping of E, G and, to a lesser extent, F (Fig. 3.5).

During tblastn searches for EcKL subfamily members in TSAs of all insect orders, EcKLs from subfamilies B, D and J were found in the transcriptomes of termites (Blattodea, infraorder Isoptera), even though they are missing from the genome of *Zootermopsis nevadensis* (Blattodea: Archotermopsidae), suggesting the genome assembly (scaffold N50 = 751,105; scaffold L50 = 194) of this species is poorly assembled.

During the course of phylogenetic analyses of all EcKLs, the subfamily membership of previously characterised EcKLs was determined. The juvenile hormone-inducible EcKL *Ipi10G08* (Genbank accession: AY875646.1; Bearfield *et al.* 2008) from *Ips pini* (Coleoptera: Scolytidae) was identified as a member of subfamily E; the ecdysteroid 22-kinase *BmEc22K* (Sonobe *et al.* 2006) from *Bombyx mori* (Lepidoptera: Bombycidae) is a member of subfamily A (and see Chapter 3.3.5); the viral resistance-associated genes *CHKov1* and *CHKov2* (Magwire *et al.* 2011) in *Drosophila melanogaster* (Diptera: Drosophilidae) are members of subfamily H; and the juvenile hormone-inducible and *Wolbachia*-associated gene *JhI-26* (Dubrovsky *et al.* 2000; Liu *et al.* 2014) in *D. melanogaster* is a member of subfamily F (Fig. 3.5). The relationships between EcKL subfamilies and the ancestral EcKL clades of *Drosophila* can be found in Table 3.2.

Significantly, even though ecdysteroid 2- and 22-kinases have been biochemically identified in *Schistocerca gregaria* (Caelifera: Acrididae; Kabbouh & Rees 1993; 1991) and are likely EcKLs (Chapter 1.3.7), subfamily A (of which *BmEc22K* is a member) was not found in 30 TSAs of Orthoptera (Fig. 3.9), including one from *S. gregaria* itself (accession PRJNA524786). To further confirm this absence, four Orthoptera genomes were partially annotated to search for subfamily A orthologs: *Laupala kohalensis* (Ensifera: Gryllidae; NCBI assembly ASM231320v1), *Teleogryllus occipitalis* (Ensifera: Gryllidae; NCBI assembly Tocci_1.0), *Xenocatantops brachycerus* (Caelifera: Acrididae; NCBI assembly ASM90024965v1) and *Locusta migratoria* (Caelifera: Acrididae; NCBI assembly LocustGenomeV1). All EcKL genes annotated in these genomes with the highest sequence similarity to subfamily A were in subfamily B, strongly suggesting the subfamily A clade may indeed be lost from Orthoptera and that *S. gregaria*

ecdysteroid 2- and 22-kinases may not be direct orthologs of the ecdysteroid 22-kinase BmEc22K.

3.3.3. Evolutionary relationships between EcKLs in insects and other arthropods

An unrooted phylogeny of EcKLs from *Catajapyx aquilonaris* (Hexapoda: Diplura), *Hyalella azteca* (Malacostraca: Amphipoda) and *Oithona nanai* (Copepoda: Cyclopoida), along with representatives of each insect EcKL subfamily, was produced (Fig. 3.10); rooting the tree is impossible without knowing which sequences are true outgroups. Despite low branch support values for deep internal nodes of the tree, a few conclusions can be drawn: the only well-supported clade containing insect and non-insect EcKLs is *H.a.* 2 (which contains a single sequence) with subfamily I, suggesting *H. azteca* may have a direct ortholog of this insect subfamily; insect subfamilies A, B, C, D, E, F, G, H, L and M form a clade separate from J–K and I, strongly suggesting insect EcKLs are not monophyletic; *C. aquilonaris*, *O. nana* and *H. azteca* sequences form a number of intra-taxa clades, suggesting their EcKLs are also not monophyletic; subfamilies J and K may group with sequences from all three non-insect arthropods (*C.a.* 3, *O.n.* 2 and *H.a.* 1) but support values are low; and insect, *C. aquilonaris*, *O. nana* and *H. azteca* sequences do not tend to form well-supported inter-taxa clades on the tree, suggesting that EcKL lineages are generally not well conserved between Copepoda, Malacostraca and Hexapoda.

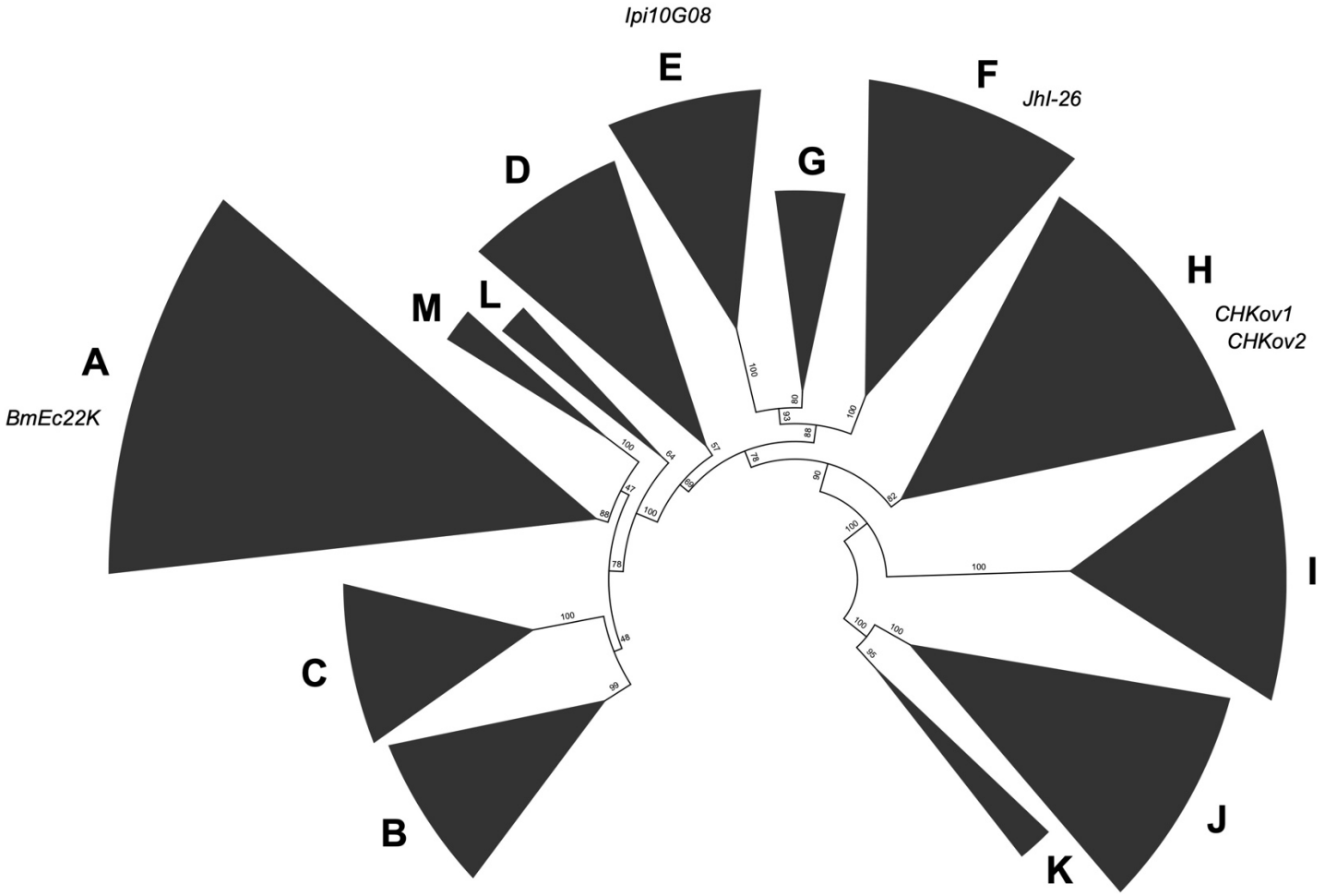


Figure 3.5. Maximum likelihood phylogenetic tree of the 13 inferred EcKL subfamilies in insects, using sequences from all 28 insect orders (see Fig. 3.9). Clades are collapsed for clarity. Branch numbers are ultrafast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018). Tree is arbitrarily rooted, with the root branch removed.

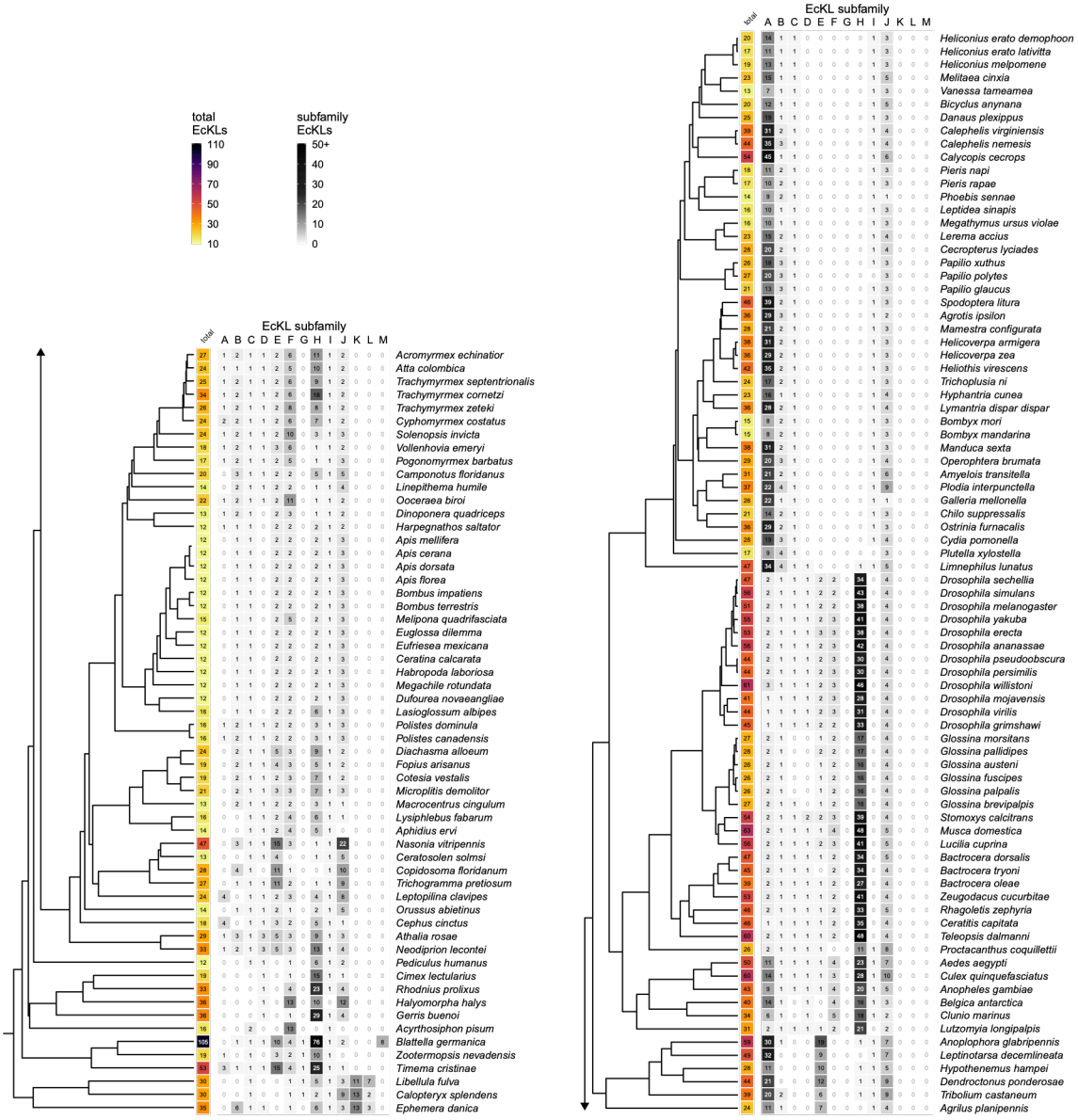


Figure 3.6. The number of EcKL genes per insect species/subspecies with annotated genomes (140 assemblies from 11 orders; coloured), as well as the number of EcKL genes in each of the 13 inferred insect EcKL subfamilies (shades of grey; see Fig. 3.5).

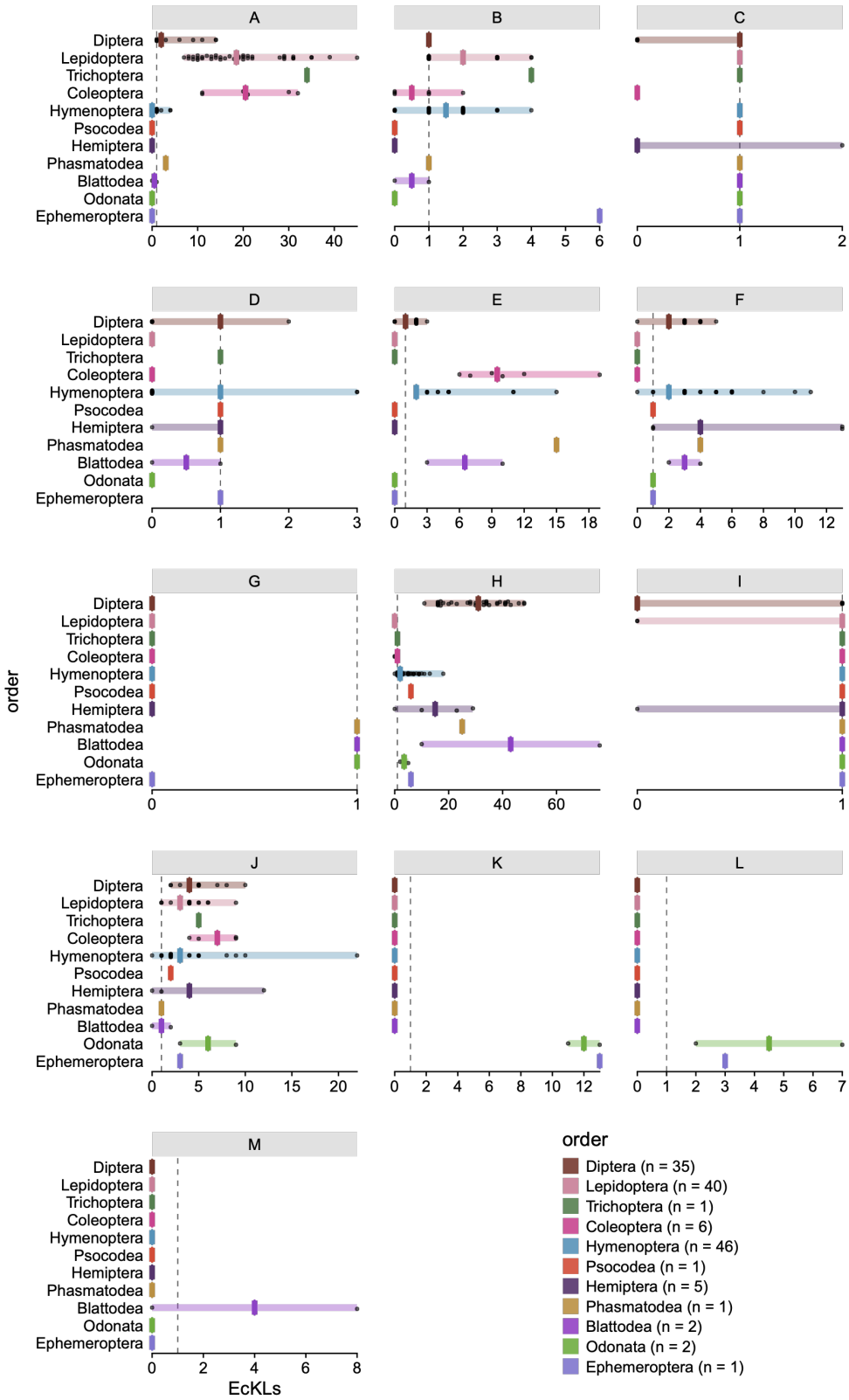


Figure 3.7. The number of EcKL genes per subfamily per insect order, using only genomic annotations (140 species from 11 orders). Individual species or subspecies are represented as dots; coloured vertical bars are order medians, while coloured horizontal bars are order ranges. The vertical dashed line is the position of 1 EcKL on the x-axis, to help highlight where subfamilies are absent from individual assemblies or orders.

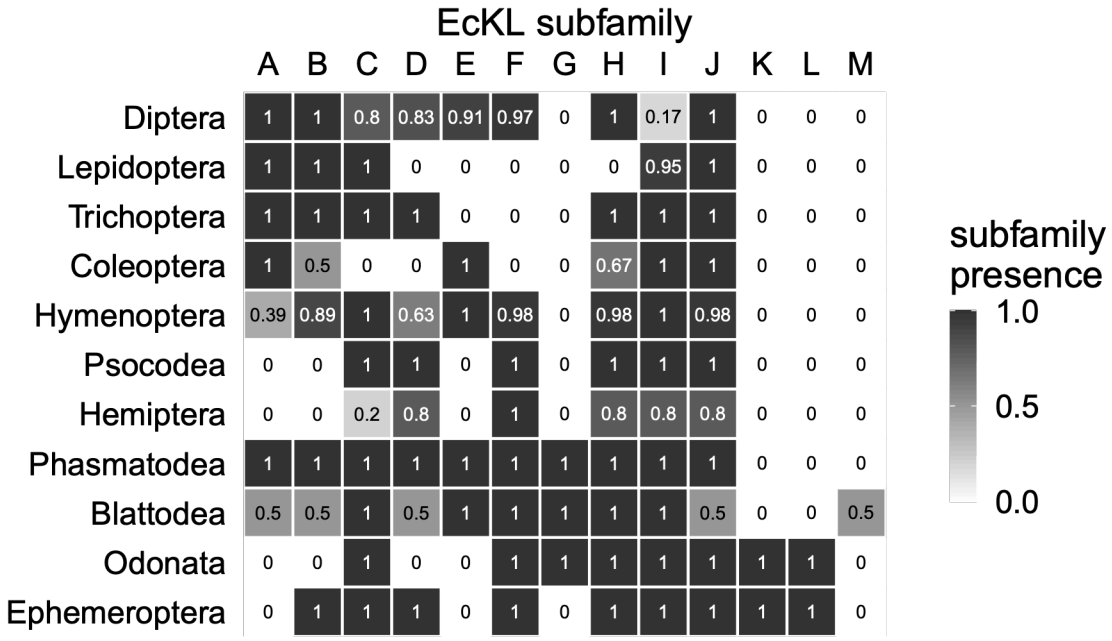


Figure 3.8. The proportion of species (from 0 to 1) in each insect order that possess at least one gene from each EcKL subfamily, using only genomic annotations (140 assemblies from 11 orders).

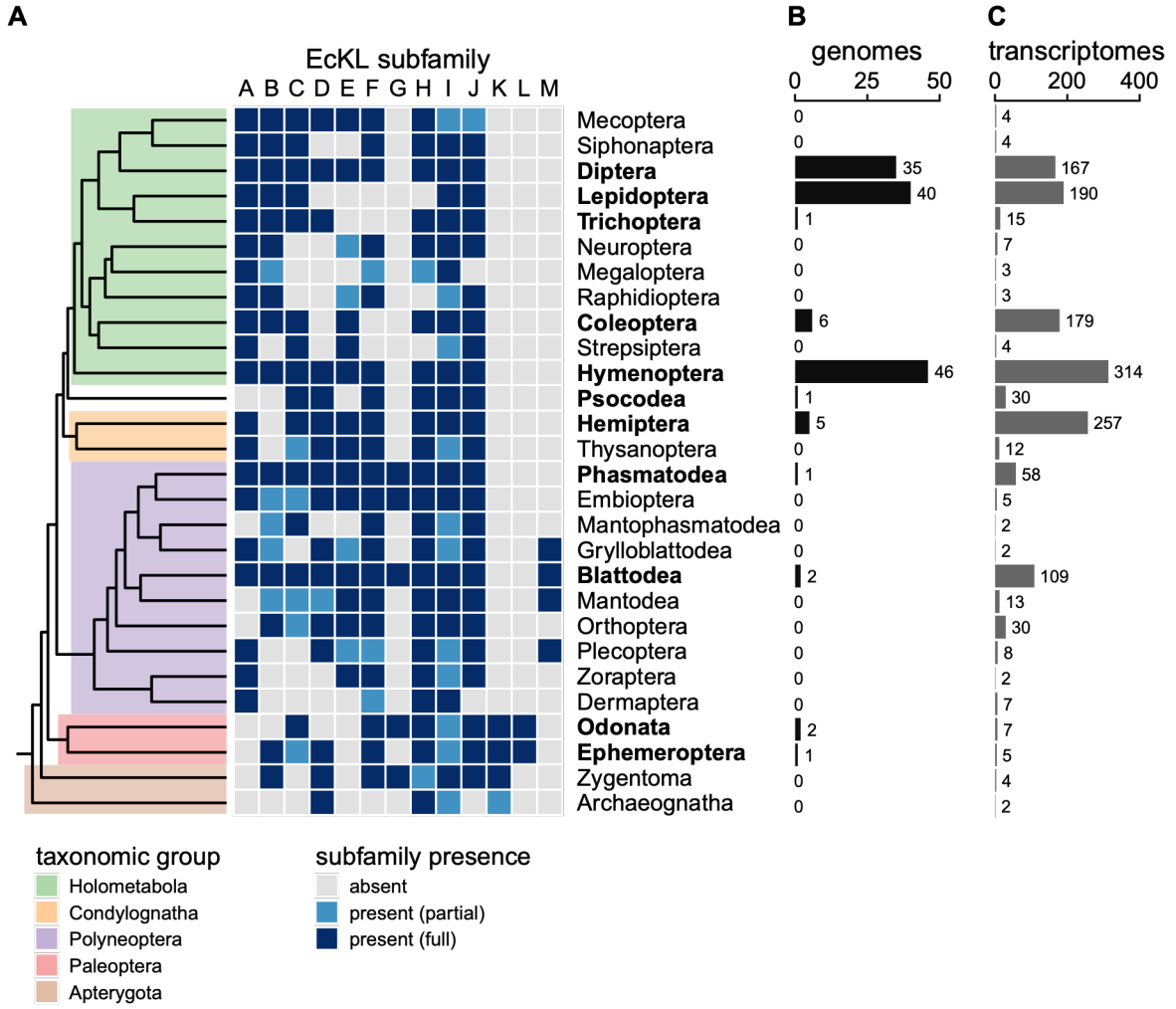


Figure 3.9. Presence and absence of EcKL subfamilies across the genomes and transcriptomes of 28 insect orders. (A) Subfamilies were considered 'absent' (no sequences in either genomes or transcriptomes), fully present (at least one full-length sequence in at least one genome or transcriptome) or partially present (at least one identifiable sequence in at least one genome or transcriptome, but no full-length sequences). (B) Number of annotated genomes per order. (C) Number of NCBI transcriptome assemblies queried per order. Insect order phylogeny from Misof *et al.* (2014). Orders in bold have at least one annotated genome.

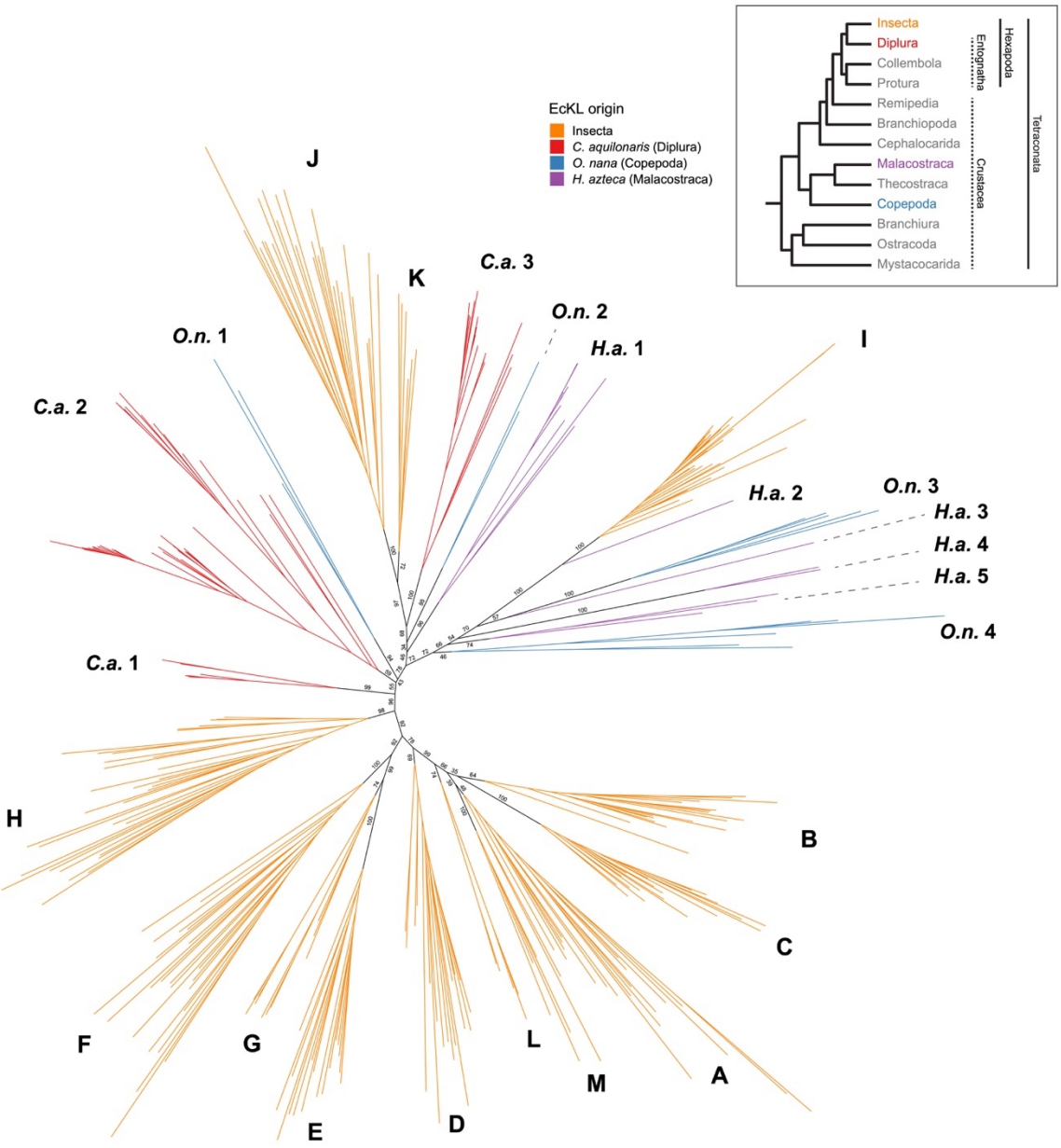


Figure 3.10. Unrooted maximum likelihood phylogenetic tree of the 13 insect EcKL subfamilies (A–M; orange) and EcKLs from three non-insect arthropods within Tetraconata: *Catajapyx aquilonaris* (*C.a.*; red), *Hyalella azteca* (*H.a.*; purple) and *Oithona nana* (*O.n.*; blue). Non-insect EcKLs are grouped into clades by species and numbered and coloured. Branch numbers are ultrafast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018); support values for branches within coloured clades are omitted for clarity. Inset, top right: relationships between taxa within Tetraconata, from Fig. 3.2. Paraphyletic groupings (Entognatha and Crustacea) are denoted with dashed lines.

I will now consider the evolution of EcKLs in three holometabolan orders—Diptera (Chapter 3.3.4), Lepidoptera (Chapter 3.3.5) and Hymenoptera (Chapter 3.3.6)—for which taxon sampling was sufficiently large to allow the development of order-specific EcKL nomenclatures (based on 'order-ancestral' clades). These clade designations are useful for identifying groups of genes that may be associated with phenotypic diversity within each order, but are ultimately mapped to deeper insect subfamilies and are not intended to be official gene names (Box 3.1).

3.3.4. Ancestral EcKL clades in Diptera

There are 20 inferred ancestral clades of EcKLs in Diptera (Fig. 3.11). Only the Dip3, Dip4, Dip5, Dip12 and Dip19 clades are conserved in every species examined, with only Dip12 considered as single-copy orthologs—although Dip5 is conserved as single-copy orthologs in every dipteran species except *Culex quinquefasciatus* (Fig. 3.12). Dip1, Dip7, Dip9, Dip15, Dip16, Dip17 and Dip20 are nearly completely conserved across all species, raising the possibility that their absences are due to incomplete genome assemblies. Much of the variation in EcKL family size is due to only a handful of clades: Dip1, Dip3, Dip4 and Dip7. Genes in the Dip1–6 clades are typically found together in large, multi-clade clusters of EcKLs in the genomes of most species—for example, one such cluster in *D. melanogaster* contains 26 genes, over half of the species' total. That these clusters are so large and yet variable in size between even closely related species suggests that their constituent EcKLs were formed by sustained, localised tandem gene duplication that has persisted for many millions of years, even before the ancestor of all Diptera. Not all genes present in these clusters show highly variable patterns of evolution: genes in the Dip5 and Dip6 clades appear more resistant to copy number variation than other genes in these clusters, suggesting that the fixation of duplicates of these genes is rarely favoured by positive selection and they may have evolved important functions during a period of rapid gene birth, or perhaps retained an important ancestral function as their paralogs have diverged in function.

In addition, the 46 ancestral EcKL clades identified in the *Drosophila* genus (superfamily Ephydroidea) in Chapter 2 were assigned to their respective ancestral dipteran clades (Table 3.2).

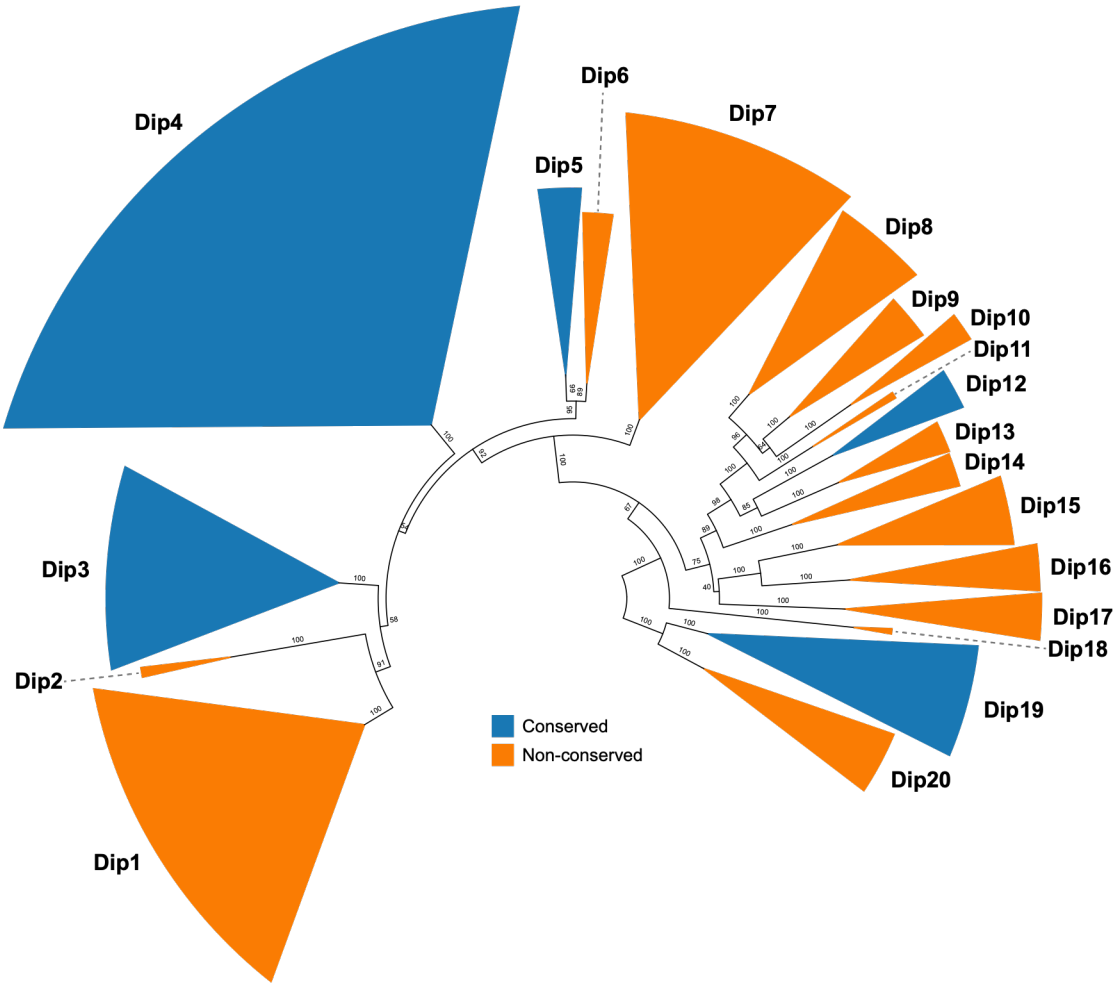


Figure 3.11. Maximum likelihood phylogenetic tree of the 20 inferred ancestral EcKL clades in 35 species of Diptera. Clades are collapsed for clarity and coloured by conservation: a clade is 'conserved' (blue) if every dipteran species possesses at least one member of the clade, otherwise it is 'non-conserved' (orange). Branch numbers are ultrafast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018). Tree is arbitrarily rooted, with the root branch removed.

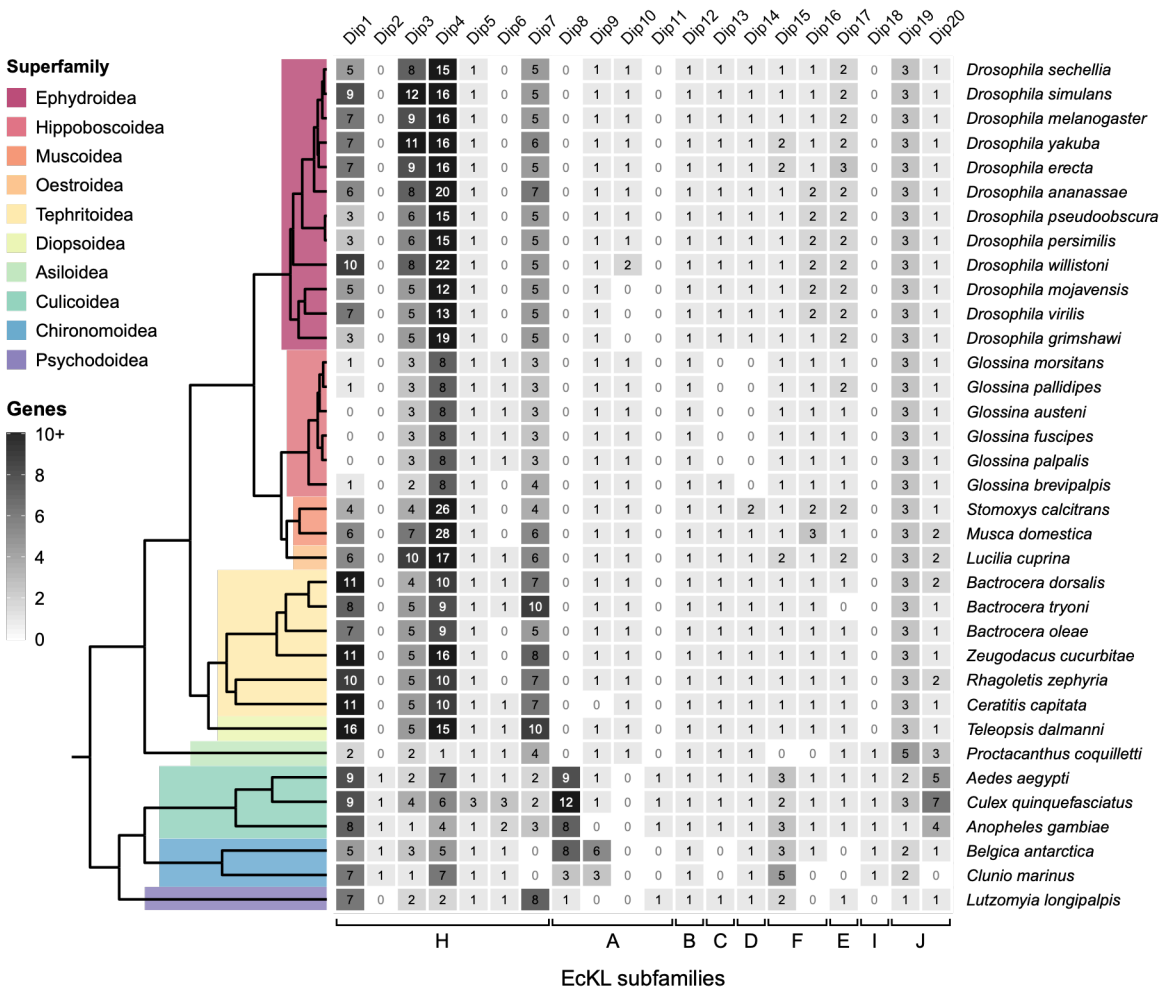


Figure 3.12. Number of genes per ancestral EcKL clade in 35 species of Diptera. Dipteran superfamilies are coloured on the phylogenetic tree to the left. EcKL subfamilies are indicated above each dipteran ancestral clade.

Table 3.2. Ancestral clade mapping between *Drosophila* genus ancestral clades (from Chapter 2), ancestral clades in Diptera and insect EcKL subfamilies.

Subfamily	Dipteran clade	*Drosophila* clade
A	Dip8	-
	Dip9	Dro37
	Dip10	Dro38
	Dip11	-
B	Dip12	Dro41
C	Dip13	Dro40
D	Dip14	Dro42
E	Dip17	Dro43, Dro44
F	Dip15	Dro45
	Dip16	Dro46
G	-	-
H	Dip1	Dro5, Dro6, Dro7, Dro39
	Dip2	-
	Dip3	Dro1, Dro2, Dro3, Dro4, Dro26, Dro27
	Dip4	Dro9, Dro10, Dro11, Dro12, Dro13, Dro14, Dro15, Dro16, Dro17, Dro18, Dro19, Dro20, Dro21, Dro22, Dro23, Dro24, Dro25
	Dip5	Dro8
	Dip6	-
	Dip7	Dro28, Dro29, Dro30, Dro31, Dro32
I	Dip18	-
J	Dip19	Dro34, Dro35, Dro36
	Dip20	Dro33

3.3.5. Ancestral EcKL clades in Lepidoptera

There are 17 inferred ancestral clades of EcKLs in Lepidoptera (Fig. 3.13). Only the Lep2, Lep9 and Lep10 clades are conserved in every species examined, with Lep9 conserved as single-copy orthologs (Fig. 3.14). The Lep1, Lep3, Lep4 and Lep6 clades are nearly completely conserved—only missing in a single species each—which might be explained by genome assembly errors. Some clades, such as Lep1, Lep3 and Lep8, have extreme changes in size between taxa. Two clades, Lep16 and Lep17, are only present in a single basal or near-basal species—*Plutella xylostella* and *Cydia pomenella*, respectively—suggesting these clades were lost relatively early in the evolution of some lepidopteran lineages.

All ancestral clades had high bootstrap support, except for Lep1 and Lep2, which are sister clades and are typically located next to each other on genomic scaffolds. A revised phylogeny with improved taxon sampling may more appropriately resolve the relationship between these two clades, if indeed they are distinct.

BmEc22K, the only EcKL with a known biochemical function (ecdysteroid 22-phosphorylation), is the sole member of the Lep1 clade in *Bombyx mori*, suggesting Lep1—and related Lep2, Lep3 and Lep16—genes in other species may also encode ecdysteroid kinases.

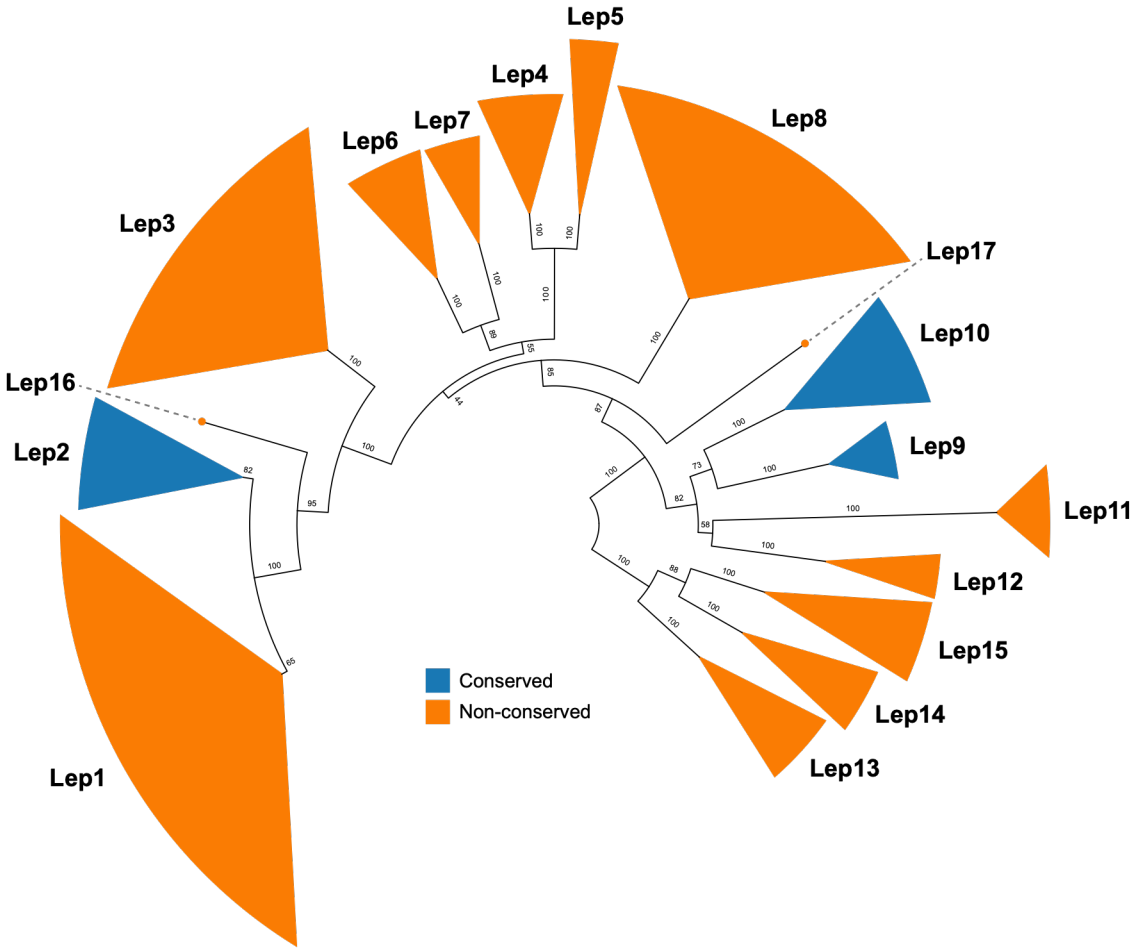


Figure 3.13. Maximum likelihood phylogenetic tree of the 17 inferred ancestral EcKL clades in 40 species of Lepidoptera. Clades are collapsed for clarity and coloured by conservation: a clade is 'conserved' (blue) if every lepidopteran species possesses at least one member of the clade, otherwise it is 'non-conserved' (orange). Clades containing a single sequence are represented by small circles. Branch numbers are ultrafast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018). Tree is arbitrarily rooted, with the root branch removed.

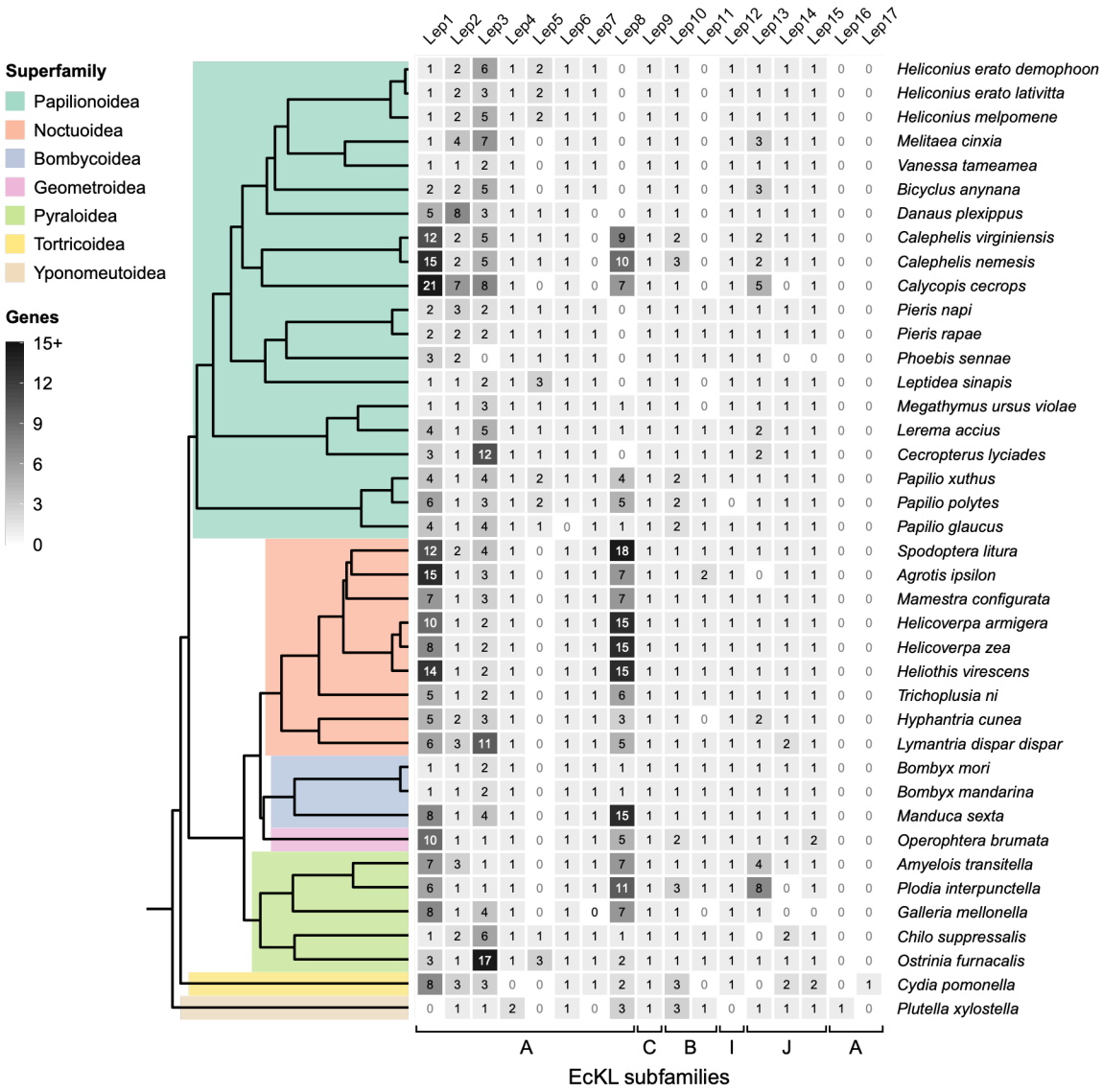


Figure 3.14. Number of genes per ancestral EcKL clade in 40 species of Lepidoptera. Lepidopteran superfamilies are coloured on the phylogenetic tree to the left. EcKL subfamilies are indicated above each lepidopteran ancestral clade.

3.3.6. Ancestral EcKL clades in Hymenoptera

There are 22 inferred ancestral clades of EcKLs in Hymenoptera (Fig. 3.15). Only the Hym17, Hym18, Hym19 and Hym21 clades are conserved in every species examined, with Hym19 and Hym21 conserved as single-copy orthologs (Fig. 3.16). The Hym1 clade is nearly completely conserved—only missing in *Copidosoma floridanum*—which raises the possibility that its absence in *C. floridanum* is due to an incomplete genome assembly. Some clades, such as Hym1, Hym14 and Hym17, have extreme changes in size between taxa. Hym7, Hym8, Hym9 and Hym10 are curiously only present in wasps in the superfamily Chalcidoidea. Bees only possess 12 ancestral clades, which are stably conserved as single-copy orthologs, except for Hym1 in *Lasioglossum albipes* (5 orthologs) and Hym14 in *Melipona quadrifasciata* (4 orthologs). *Ceratosolen solmsi*, a fig wasp, has very few EcKLs compared to parasitoid chalcid wasps—matching a general pattern of gene loss in this species (Xiao *et al.* 2013)—suggesting that the clades lost in this species compared to its closest relative, *Nasonia vitripennis* (Hym7, Hym10, Hym11, Hym14, Hym16 and Hym20), are not essential for development in chalcid wasps; it also suggests that some genes in the large Hym17 clade, as well as the taxonomically restricted Hym7, Hym10 and Hym11 clades, may play roles in parasitoid biology.

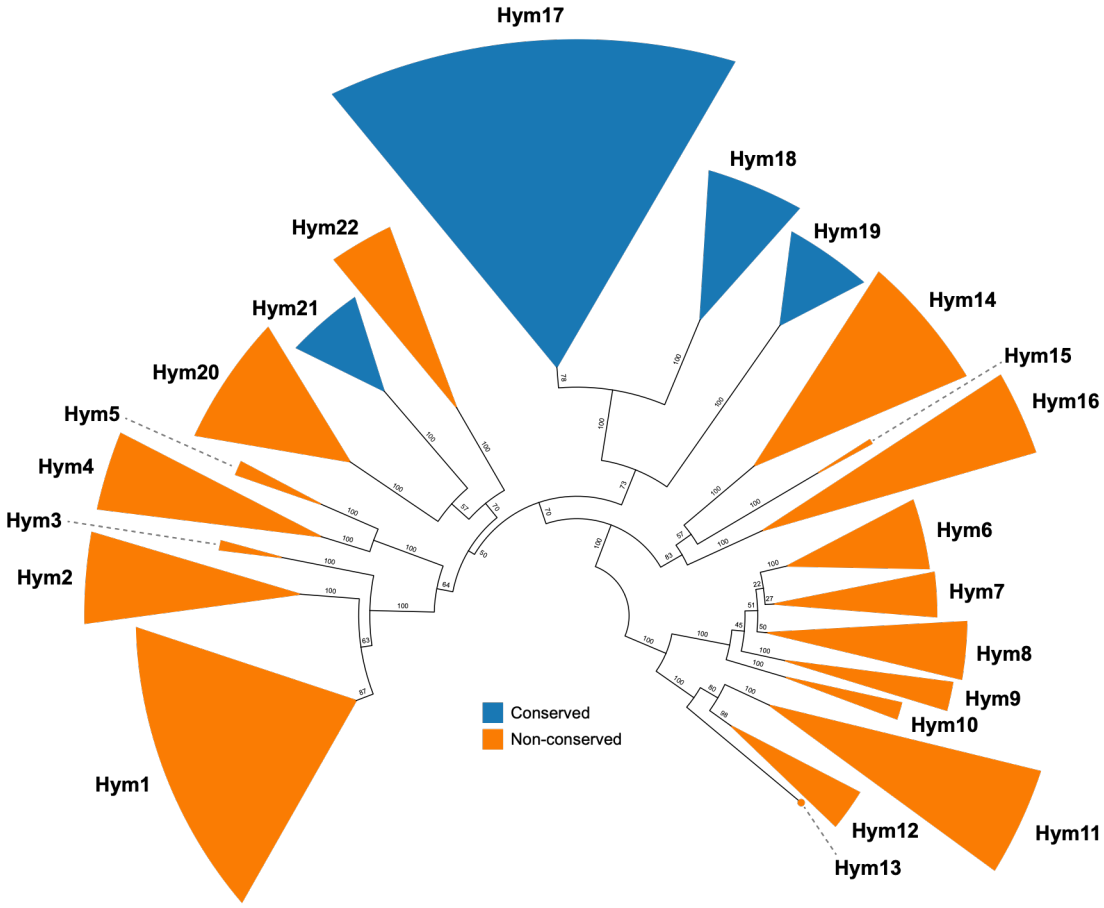


Figure 3.15. Maximum likelihood phylogenetic tree of the 22 inferred ancestral EcKL clades in 46 species of Hymenoptera. Clades are collapsed for clarity and coloured by conservation: a clade is 'conserved' (blue) if every hymenopteran species possesses at least one member of the clade, otherwise it is 'non-conserved' (orange). Clades containing a single sequence are represented by small circles. Branch numbers are ultrafast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018). Tree is arbitrarily rooted, with the root branch removed.

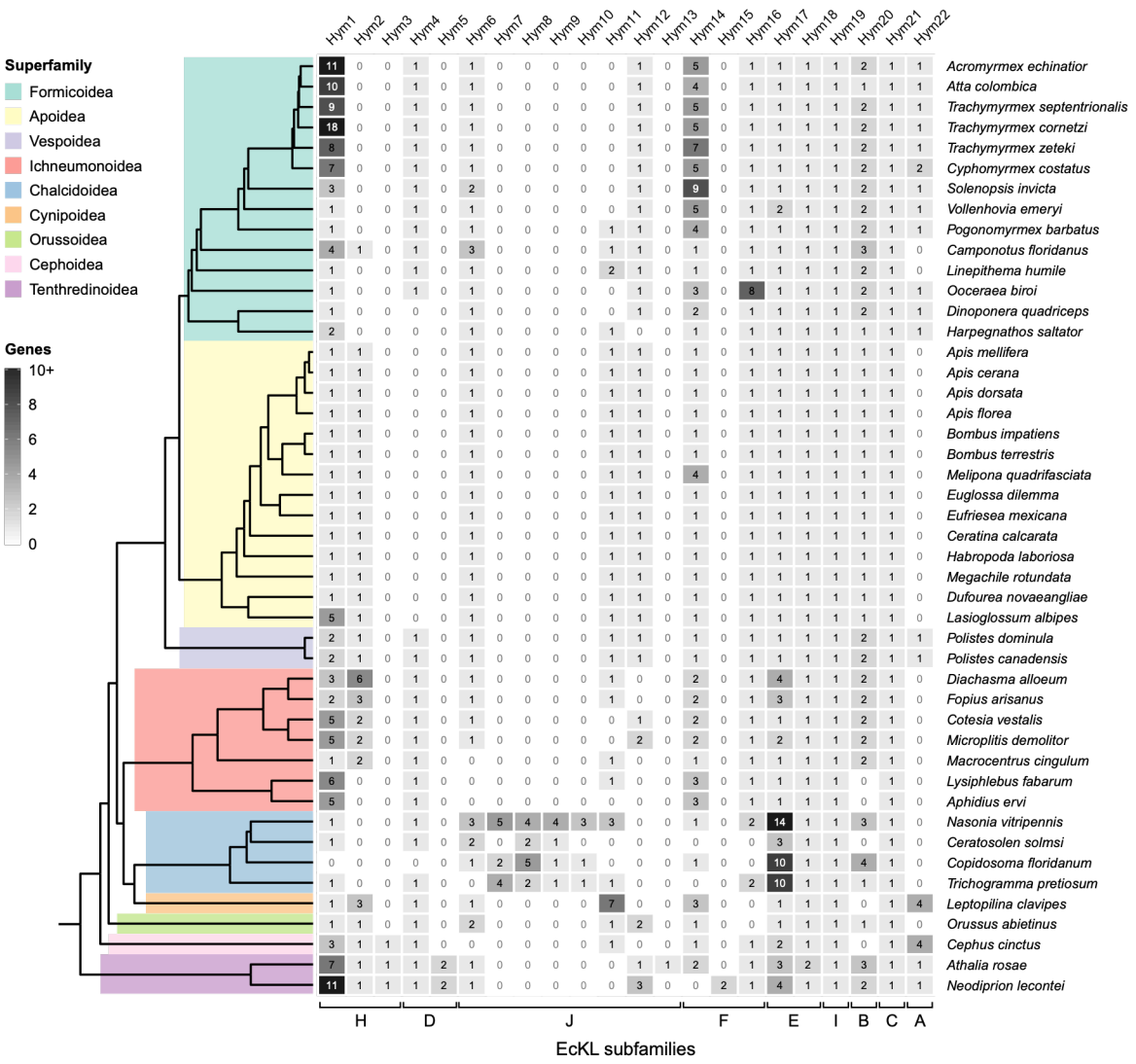


Figure 3.16. Number of genes per ancestral EcKL clade in 46 species of Hymenoptera. Hymenopteran superfamilies are coloured on the phylogenetic tree to the left. EcKL subfamilies are indicated above each hymenopteran ancestral clade.

3.3.7. Non-standard sequence features of insect EcKL proteins

The vast majority (98.6%) of EcKL gene models annotated encoded predicted protein sequences that comprise a single EcKL domain flanked by N- and C-terminal regions with no unique sequence features. However, a small number (1.4%) of insect EcKLs deviate from this general structure and possess multiple EcKL domains or additional non-EcKL regions; others possess novel transcriptional characteristics.

D. melanogaster CG1561 (Dro37-0/Dip9/subfamily A) encodes a protein with an intrinsically disordered N-terminus that may be involved in protein-protein interactions, as predicted by IUPred2A (Mészáros *et al.* 2018); this feature is shared with eight direct orthologs in the *Drosophila* subgenus Sophophora, but is not conserved in other dipteran species. Curiously, seven Dip10 (subfamily A) proteins have a C-terminal region that is also intrinsically disordered, although it does not have predicted protein-protein interaction properties, and this feature is not seen in *Drosophila*, *Glossina* or *Proctacanthus* orthologs. In addition, the *Musca domestica* Dip10 protein has an intrinsically disordered N-terminus similar to *Drosophila* Dip9 proteins that may also participate in protein-protein interactions. The functional significance of these disordered regions will be discussed in Chapter 5.

24 EcKLs from Diptera (Dip4/subfamily H) and nine EcKLs from Coleoptera (subfamily J) possess two full EcKL domains within the same polypeptide. As these genes are from separate subfamilies, it is likely the dual-domain structure arose independently in both lineages. The functional significance of this dual-domain structure will be discussed in Chapter 6.

Two *Drosophila melanogaster* EcKLs—*CG11889* (Dro1-3/Dip3/subfamily H) and *CG11891* (Dro1-4/Dip3/subfamily H)—are known to be transcribed as a dicistronic transcript (Thurmond *et al.* 2019); in addition, their upstream pseudogenous paralog *CR13656* (Dro1-3/Dip3/subfamily H) is contained in a cDNA sequence that spans all three genes (Genbank accession BT150065), suggesting they may be part of a functional tricistronic unit in other *Drosophila* species. Another *D. melanogaster* EcKL, *CG31975* (Dro44-0/Dip17/subfamily E), is transcribed as a discistronic transcript with *oven mitt* (*ovm*), a downstream DUF1091-encoding gene identified and named in an RNAi screen exploring genes required for normal thermal nociception (Honjo *et al.*

2016). The significance of polycistronic transcription in these cases is unclear, although it may allow for coordinated expression of EcKL paralogs or unrelated genes that have cooperative function (Blumenthal 1998; Karginov *et al.* 2017). Given this, it is possible *CG31975* may play a role in thermal nociception, although it is primarily expressed in the midgut and Malpighian tubules and not the nervous system (Leader *et al.* 2018).

Three Lep1 / Lep2 EcKL loci in *Agrotis ipsilon*, *Spodoptera litura* and *Trichoplusia ni* (Lepidoptera: Noctuoidea) have evidence of alternative splicing, where the first of three exons (encompassing ~50% of the total coding sequencing) appear to have been duplicated multiple times (four Lep1 exons in *A. ipsilon*, seven Lep1 and two Lep2 exons in *S. litura*, and five Lep1 exons in *T. ni*). These duplicated exons have diverged amino acid and nucleotide sequences, but do not contain any nonsense mutations, suggesting they are functional, and transcripts from these three species or very closely related species in the NCBI TSA database support multiple transcripts mapping to the same loci, suggesting these genes are truly alternatively spliced. Due to the divergence of these exons, it is likely that the encoded protein isoforms have also diverged in function. That most of these genes are in the Lep1 clade, which may be involved in detoxification (see Chapters 3.3.9 & 3.4.9), raises the possibility that each of these isoforms have differences in substrate specificity.

3.3.8. EcKL family size is positively associated with the size of known detoxification gene families across insects

As the first comparative phylogenomic test of the hypothesis that the EcKL gene family is involved in detoxification in insects, EcKL family size was regressed against the sizes of three known detoxification gene families: the P450s, the GSTs and the CCEs (Figs. 3.17–18). There is substantial phylogenetic signal in EcKL, P450, GST and CCE family sizes across the insect phylogeny (Fig. S3.3), necessitating the use of phylogenetic generalised least squares (PGLS) regression due to the non-independence of each data point (Cornwell & Nakagawa 2017). PGLS regression found that the sizes of all four gene families were significantly associated with each other (Fig. 3.19; Table 3.2. for model details). The best-associated gene families were the GSTs and the P450s (adjusted R^2 between 0.283 and 0.306) and the poorest associated gene families were the EcKLs and the CCEs (adjusted R^2 between 0.053 and 0.074).

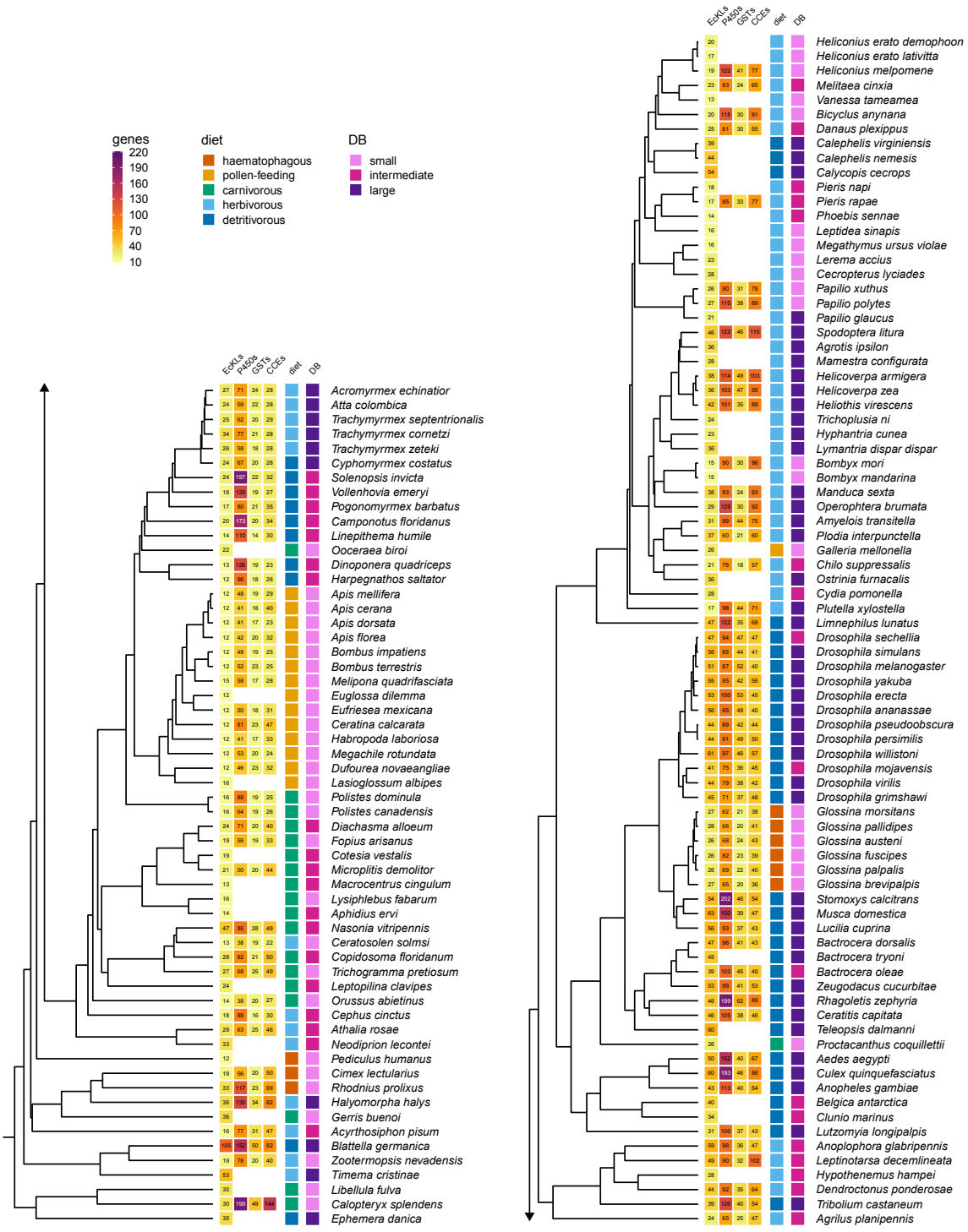


Figure 3.17. Number of EcKL, P450, GST and CCE genes per species, along with diet and DB designations. For higher-level taxonomic ranks, see Figs. 3.3, 3.12 (for Diptera), 3.14 (for Lepidoptera) and 3.16 (for Hymenoptera).

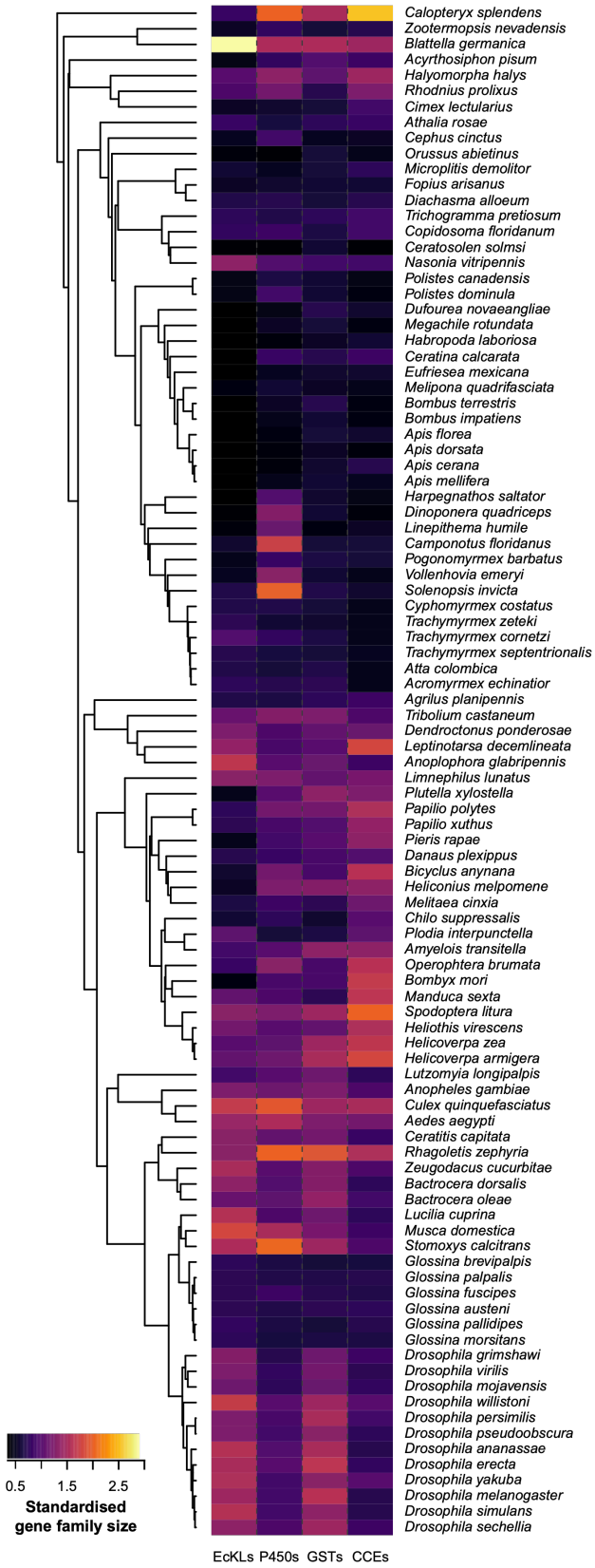


Figure 3.18. Heatmap comparison of the size of the EcKL, P450, GST and CCE clades across in the insect phylogeny (n = 98 species). Plot generated with the *phylosignal* package (version 1.3; Keck *et al.* 2016) in R; family sizes are normalised (scale = T, center = F) for ease of comparison and encoded by colour on the same scale.

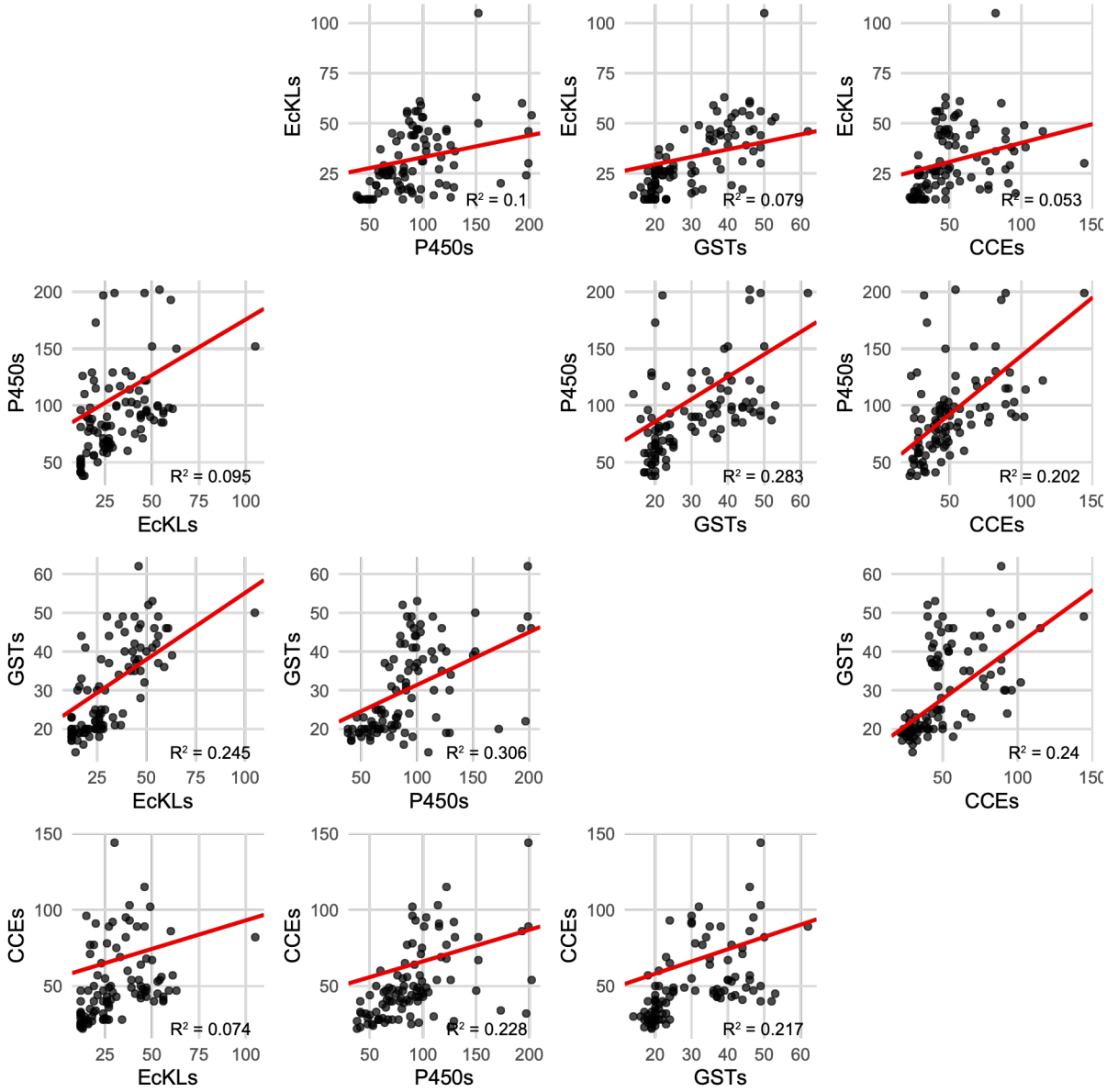


Figure 3.19. Scatterplots of PGLS regressions of EcKL, P450, GST and CCE gene family sizes (n = 98 species). Red trendlines are defined by slope and intercept from PGLS models. R^2 values are adjusted R^2. Full model outputs can be found in Table 3.2.

Table 3.3. PGLS outputs for models with regarding the co-evolution of EcKL, P450, GST and CCE gene families (see Fig. 3.19).

Response	Predictor	λ	Estimation	SE	t-value	p-value	df	Adjusted R^2
EcKLs	P450s	0.990	0.109	0.032	3.426	9.02×10^{-4}	96	0.100
EcKLs	GSTs	0.983	0.375	0.123	3.055	2.91×10^{-3}	96	0.079
EcKLs	CCEs	0.991	0.189	0.074	2.539	1.27×10^{-2}	96	0.053
P450s	EcKLs	0.993	0.979	0.293	3.339	1.20×10^{-3}	96	0.095
P450s	GSTs	0.993	1.975	0.315	6.272	1.02×10^{-8}	96	0.283
P450s	CCEs	0.996	1.031	0.204	5.057	2.04×10^{-6}	96	0.202
GSTs	EcKLs	0.762	0.344	0.060	5.700	1.32×10^{-7}	96	0.245
GSTs	P450s	0.923	0.135	0.020	6.609	2.17×10^{-9}	96	0.306
GSTs	CCEs	0.895	0.282	0.050	5.631	1.78×10^{-7}	96	0.241
CCEs	EcKLs	0.969	0.374	0.126	2.965	3.82×10^{-3}	96	0.074
CCEs	P450s	0.969	0.207	0.038	5.452	3.86×10^{-7}	96	0.229
CCEs	GSTs	0.963	0.803	0.152	5.281	8.01×10^{-7}	96	0.217

3.3.9. EcKL family size is positively associated with host plant diversity in Lepidoptera

The number of host plants a herbivorous insect can consume is a reasonable proxy for detoxification breadth (DB), given that a large phylogenetic diversity of host plants means an insect is likely tolerant of a wide variety of plant secondary metabolites. As the number of host plant species that caterpillars can consume is of great interest to lepidopterists and agricultural entomologists, a large volume of data exists on this topic, collated in the HOSTS Database (Robinson *et al.* 2010). As a test of the relationship between EcKLs and DB, I explored, using PGLS regression and a lepidopteran subset of the phylogeny in Fig. 3.3 (n = 38 species), whether the number of genes in the EcKL family a species possesses is positively associated with the number of unique host plant taxa they consume. P450, GST and CCE family sizes were not used in a similar analysis due to the small number of lepidopteran species (n = 16) with data available (Rane *et al.* 2019).

In the full dataset (n = 38 species), EcKL family size is significantly, albeit weakly, positively associated with the $\log_2$-transformed number of host plant genera ($p = 0.031$), but not families ($p = 0.075$) or species ($p = 0.052$), consumed by caterpillars (Fig. 3.20A–C). However, *Calephelis nemesis*, *Calephelis virginiens* and *Calycopis cecrops*—which are all likely detritivorous species—have some of the highest EcKL family sizes in Lepidoptera, yet feed on comparatively few host plant taxa; these outliers may be confounding possible stronger associations in the data. Indeed, when these three detritivorous species were removed from the dataset, EcKL gene family size in Lepidoptera is significantly positively associated with the number of host plant families ($p = 2.1\times10^{-6}$), genera ($p = 5.3\times10^{-4}$) and species ($p = 0.0025$) consumed, with substantially increased model fits (Fig. 3.20D–F; Table 3.4). This subset (n = 35 species) was used for all subsequent analyses.

Given that it is likely that not all EcKLs are involved in detoxification, I asked if the sizes of any specific ancestral clades of lepidopteran EcKLs are associated with host plant diversity. The sizes of the Lep1 ($p = 0.0058$) and Lep12 ($p = 0.018$) clades are positively associated with transformed host family counts, although the Lep12 clade association is likely spurious, given only two species have zero genes in the clade and every other species has one. However, the size of the Lep8 clade is, curiously, strongly

associated ($p = 2.7\times10^{-8}$ and 3.1×10^{-5}) with the size of the Lep1 clade (Fig. 3.21A–B), despite not being significantly associated with transformed host family count ($p = 0.19$), and Lep8 clade size is also weakly associated with both transformed host genera ($p = 0.043$) and species ($p = 0.035$) counts, as well as more strongly with untransformed counts (families, $p = 0.0062$; genera, $p = 0.0032$; species, $p = 0.0047$; Table 3.4). These data suggest that Lep8 size may be truly associated with host plant diversity in the Lepidoptera along with Lep1, but this may be obscured depending on the analysis. To further test that Lep1 and Lep8 may be associated with host plant diversity, I asked whether total EcKL size excluding Lep1 and Lep8 counts would still be associated with host family counts. Subtracting both Lep1 and Lep8 clade sizes together—but not individually—from total EcKL counts removes the association with host family count, further suggesting these two clades are both responsible for the EcKL family-wide association with host plant diversity in the Lepidoptera (Fig. 3.21C–E, Table 3.4). Genes in the Lep1 and Lep8 clades are therefore strong candidates for detoxification EcKLs in Lepidoptera.

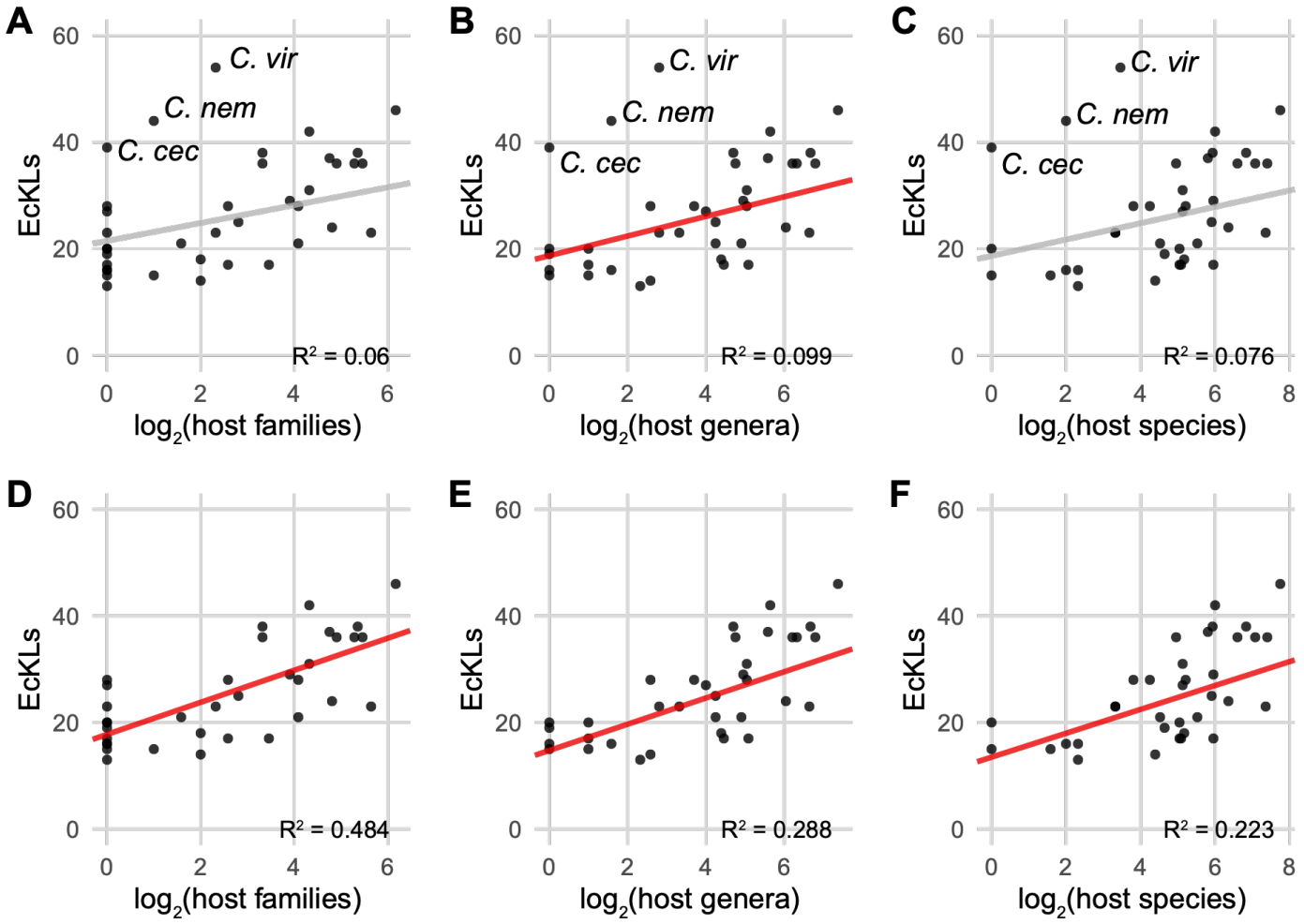


Figure 3.20. Scatterplots of PGLS regressions of EcKL gene family size against host plant diversity in the Lepidoptera for the full dataset (A–C, n = 38 species) and a subset of non-detritivorous species (D–F, n = 35 species). Data points of detritivorous species are marked: *C. nem*, *Calephelis nemesis*; *C. vir*, *Calephelis virginiensis*; *C. cec*, *Calycopis cecrops*. Red trendlines indicate significant association ($p < 0.05$), grey trendlines indicate non-significant association ($p > 0.05$). R^2 values are adjusted R^2. Full model outputs can be found in Table 3.4.

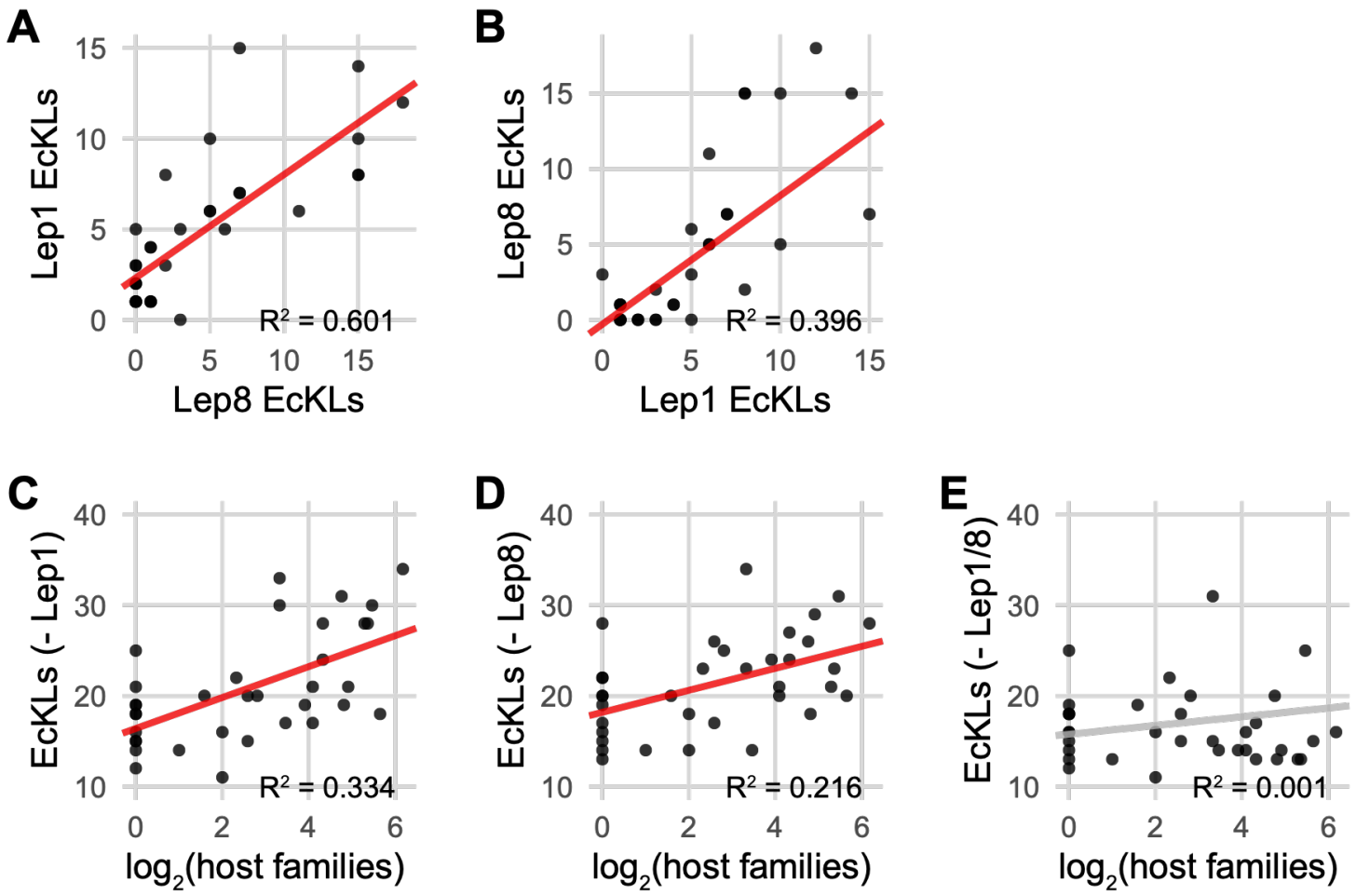


Figure 3.21. Lep1 and Lep8 clade sizes are strongly correlated, and removing the sizes of both clades from total EcKL family size removes the latter's association with host plant diversity. (A) Lep1 clade size regressed against Lep8 clade size. (B) Lep8 clade size regressed against Lep1 clade size. (C) Total EcKL family size minus Lep1 clade size regressed against $\log_2$(host families). (D) Total EcKL family size minus Lep8 clade size regressed against $\log_2$(host families). (E) Total EcKL family size minus Lep1 and Lep8 clade sizes regressed against $\log_2$(host families). Red trendlines indicate significant association ($p < 0.05$), grey trendlines indicate non-significant association ($p > 0.05$). R^2 values are adjusted R^2. Full model outputs can be found in Table 3.4.

Table 3.4. PGLS outputs for models with Lepidoptera EcKLs and host plant diversity counts (Figs. 3.20–21).

Dataset	Response	Predictor	λ	Estimation	SE	t-value	p-value	df	Adjusted R^2
full	log_2(families)	log_2(genera)	0.286	0.784	0.078	10.048	5.46×10^{-12}	36	0.73
full	log_2(families)	log_2(species)	1	0.620	0.093	6.695	8.27×10^{-8}	36	0.542
full	log_2(genera)	log_2(species)	1	0.803	0.073	11.05	4.04×10^{-13}	36	0.766
full	total EcKLs	log_2(families)	0.99	1.677	0.914	1.835	7.48×10^{-2}	36	0.060
full	total EcKLs	log_2(genera)	0.992	1.842	0.820	2.247	3.09×10^{-2}	36	0.099
full	total EcKLs	log_2(species)	0.993	1.535	0.764	2.009	5.21×10^{-2}	36	0.052
subset	total EcKLs	log_2(families)	0	3.011	0.525	5.736	2.10×10^{-6}	33	0.484
subset	total EcKLs	log_2(genera)	0.581	2.460	0.641	3.841	5.27×10^{-4}	33	0.288
subset	total EcKLs	log_2(species)	0.814	2.243	0.684	3.280	2.45×10^{-3}	33	0.223
subset	Lep1 EcKLs	log_2(families)	0.468	0.914	0.310	2.950	5.80×10^{-3}	33	0.185
subset	Lep2 EcKLs	log_2(families)	1	0.222	0.136	1.632	1.12×10^{-1}	33	0.047
subset	Lep3 EcKLs	log_2(families)	0.947	0.178	0.397	0.449	6.57×10^{-1}	33	-0.024
subset	Lep4 EcKLs	log_2(families)	1	0.006	0.020	0.274	7.86×10^{-1}	33	-0.028
subset	Lep5 EcKLs	log_2(families)	1	-0.142	0.070	-2.018	5.18×10^{-2}	33	0.083
subset	Lep6 EcKLs	log_2(families)	0.46	-0.031	0.018	-1.761	8.75×10^{-2}	33	0.058
subset	Lep7 EcKLs	log_2(families)	1	-0.023	0.020	-1.140	2.62×10^{-1}	33	0.009
subset	Lep8 EcKLs	log_2(families)	1	0.606	0.419	1.445	1.58×10^{-1}	33	0.031
subset	Lep8 EcKLs	log_2(genera)	1	0.828	0.393	2.105	4.30×10^{-2}	33	0.092
subset	Lep8 EcKLs	log_2(species)	1	0.828	0.376	2.206	3.45×10^{-2}	33	0.102
subset	Lep8 EcKLs	families	1	0.115	0.039	2.927	6.16×10^{-3}	33	0.182
subset	Lep8 EcKLs	genera	1	0.052	0.016	3.186	3.15×10^{-3}	33	0.212
subset	Lep8 EcKLs	species	1	0.037	0.012	3.034	4.67×10^{-3}	33	0.195
subset	Lep9 EcKLs	log_2(families)	0	0.001	0.003	0.567	5.75×10^{-1}	33	-0.021
subset	Lep10 EcKLs	log_2(families)	1	0.050	0.057	0.882	3.84×10^{-1}	33	-0.007
subset	Lep11 EcKLs	log_2(families)	1	0.016	0.047	0.330	7.43×10^{-1}	33	-0.027
subset	Lep12 EcKLs	log_2(families)	1	0.061	0.024	2.498	1.77×10^{-2}	33	0.134
subset	Lep13 EcKLs	log_2(families)	1	0.158	0.138	1.139	2.63×10^{-1}	33	0.009
subset	Lep14 EcKLs	log_2(families)	1	-0.024	0.041	-0.577	5.68×10^{-1}	33	-0.020
subset	Lep15 EcKLs	log_2(families)	1	0.003	0.028	0.092	9.28×10^{-1}	33	-0.030
subset	Lep16 EcKLs	log_2(families)	1	0.007	0.013	0.543	5.91×10^{-1}	33	-0.021
subset	Lep17 EcKLs	log_2(families)	1	0.002	0.013	0.117	9.08×10^{-1}	33	-0.030
subset	Lep1 EcKLs	Lep8 EcKLs	0	0.571	0.079	7.229	2.74×10^{-8}	33	0.601
subset	Lep8 EcKLs	Lep1 EcKLs	0.336	0.853	0.177	4.828	3.06×10^{-5}	33	0.396
subset	total EcKLs - Lep1	log_2(families)	0	1.710	0.403	4.245	1.67×10^{-4}	33	0.334
subset	total EcKLs - Lep8	log_2(families)	0	1.215	0.378	3.218	2.90×10^{-3}	33	0.216
subset	total EcKLs - Lep1 - Lep8	log_2(families)	0.956	0.485	0.478	1.013	3.18×10^{-1}	33	0.001

3.3.10. EcKL family size varies with diet and estimated detoxification breadth (DB) in insects

If the EcKL gene family is involved in detoxification processes, I hypothesised that it should vary with diet and estimated DB among insects generally; and the same should occur with the P450s, GSTs and CCEs, known detoxification families. Associations between the sizes of gene families and insect diet (classified into five groups) and an estimation of DB (classified into three groups) were tested by PGLS regression using the phylogeny in Fig. 3.3.

There were significant differences in EcKL ($F_{4,135} = 7.56$, $p < 1.0\times10^{-4}$) , P450 ($F_{4,93} = 4.63$, $p = 0.0019$) and GST ($F_{4,93} = 6.56$, $p = 1.0\times10^{-4}$) family sizes—but not CCE family size ($F_{4,93} = 1.28$, $p = 0.28$)—between insect diets after correcting for phylogeny (Fig. 3.22A–D, Table 3.5). Pairwise comparisons between diet groups revealed the largest and most consistent mean differences were between haematophagous insects and detritivorous insects; there were no significant differences between pollen-feeding insects and other groups for any gene family, despite some large effect sizes in the EcKLs and P450s (Fig. 3.22E).

There were also significant differences in EcKL ($F_{2,137} = 29.4$, $p < 1.0\times10^{-4}$), P450 ($F_{2,95} = 3.51$, $p = 0.034$) and GST ($F_{2,95} = 9.99$, $p = 1.0\times10^{-4}$) family sizes—but not CCE family sizes ($F_{2,95} = 0.161$, $p = 0.85$)—between DB groups for the full dataset (Fig. 3.23A–D, Table 3.5), but the P450 association became non-significant ($F_{2,75} = 1.44$, $p = 0.24$) when omnivorous/fungivorous ants and parasitoid wasps were excluded from the analysis (Fig. 3.23F, Table 3.5). Increasing DB was consistently associated with increased EcKL family size regardless of dataset, while for the GSTs, only DB_L insects had larger family size than DB_I and DB_S insects regardless of dataset; the only significant pairwise difference for the P450s was DB_I vs DB_S for the subsetted data (Fig. 3.23I).

Overall, these data are strongly consistent with a role for EcKL genes in detoxification processes across all insects.

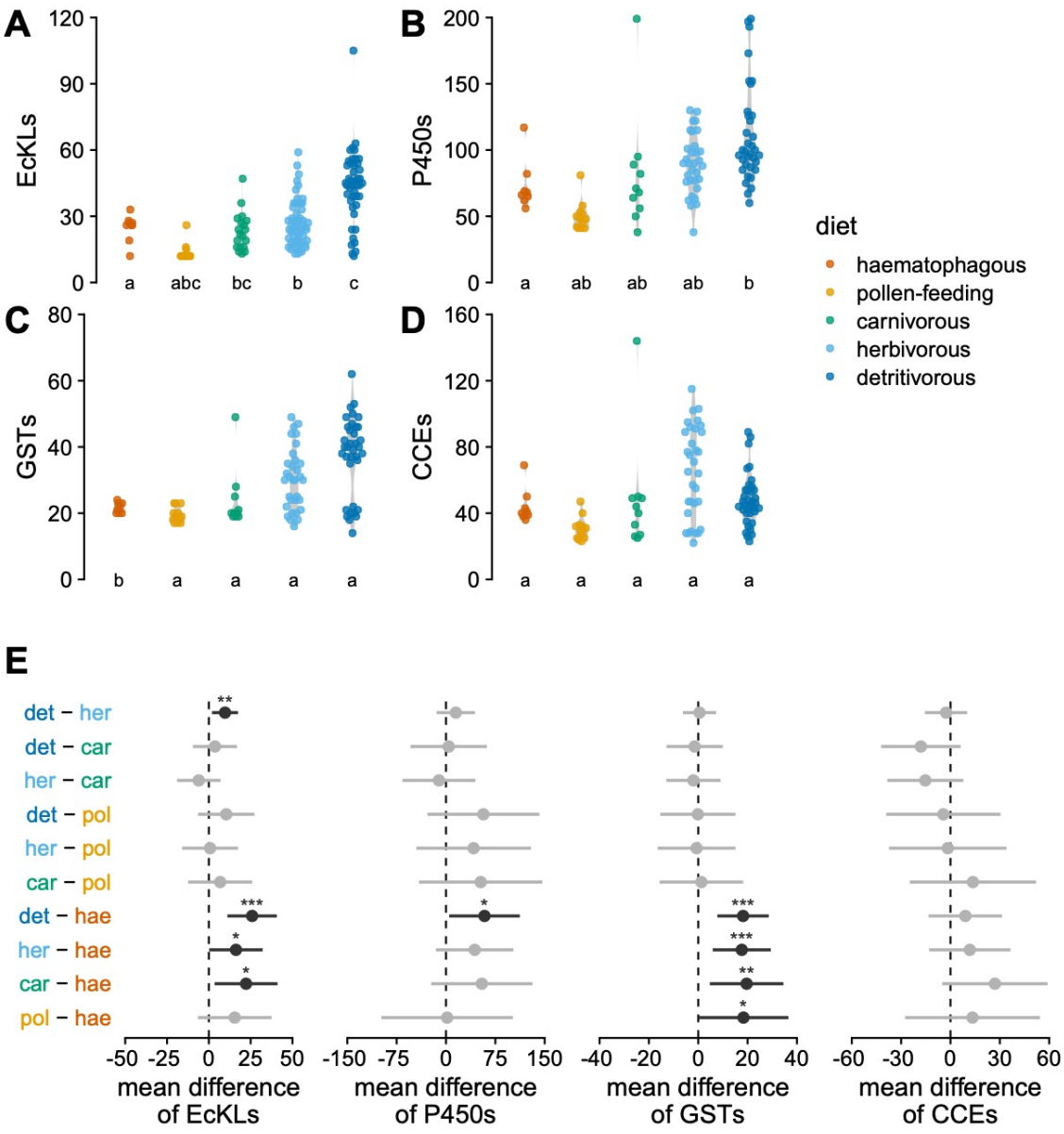


Figure 3.22. Dot (coloured) and violin (grey) plots of number of genes per genome against insect diet, for the (A) EcKL, (B) P450, (C) GST and (D) CCE gene families (n = 140 species for EcKLs, n = 98 species for P450s, GSTs and CCEs). Letters underneath each diet group encode pairwise comparisons from a phylogenetically corrected GLS model; diet groups that do not share a letter are significantly different with an adjusted p-value < 0.05. (E) Mean differences (and 95% CIs) in gene number between diet groups for the EcKL, P450, GST and CCE gene families. Effect sizes (and 95% CI bars) are coloured by significance (black if adjusted $p < 0.05$, grey if $p > 0.05$). det, detritivorous; her, herbivorous; car, carnivorous; pol, pollen-feeding; hae, haematophagous. p-value coding: * < 0.05, ** < 0.01, *** < 0.001.

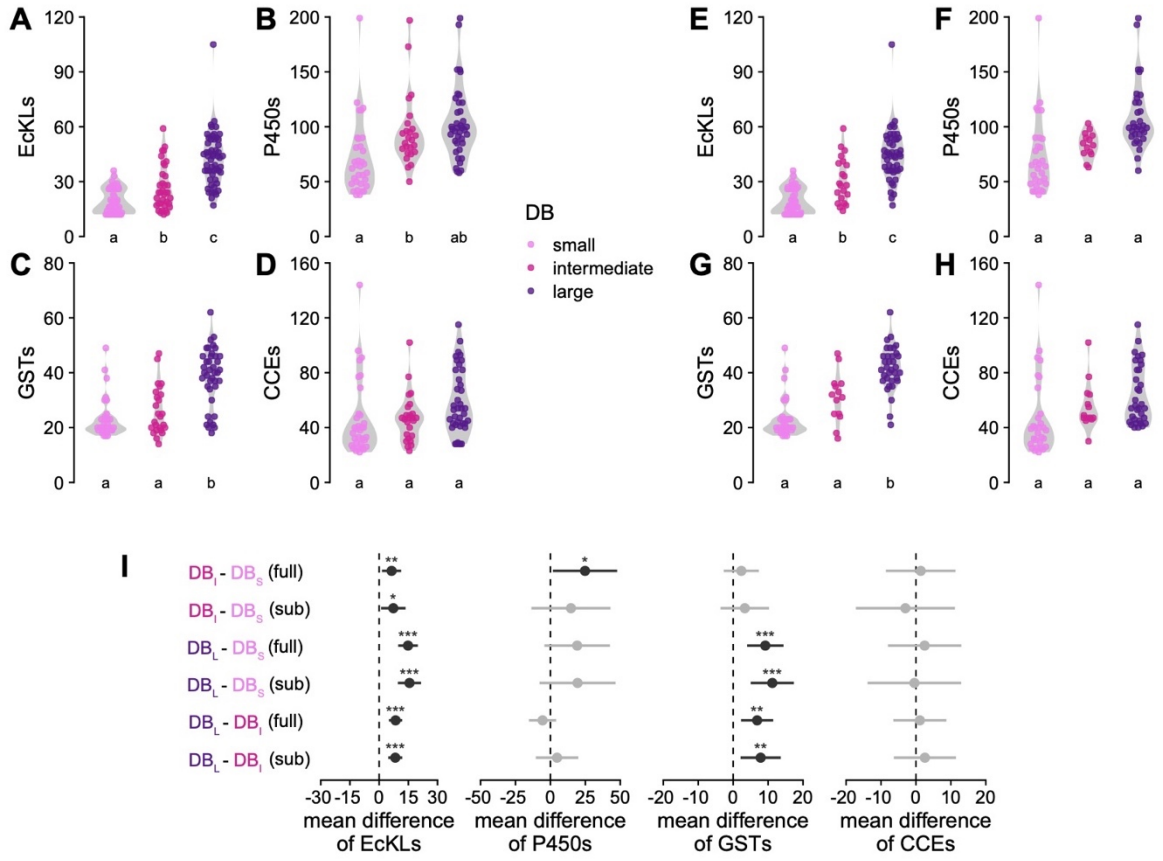


Figure 3.23. Dot (coloured) and violin (grey) plots of number of genes per genome against estimated detoxification breadth (DB) for the (A,E) EcKL, (B,F) P450, (C,G) GST and (D,H) CCE gene families. Letters underneath each diet group encode pairwise comparisons from a phylogenetically corrected GLS model; DB groups that do not share a letter are significantly different with an adjusted p-value < 0.05. (A–D) Full dataset (n = 140 for EcKLs, n = 98 for P450s, GSTs and CCEs). (E–H) Subset dataset (no omnivorous/fungivorous ants or parasitoid wasps; n = 115 for EcKLs, n = 79 for P450s, GSTs and CCEs). (I) Mean differences (and 95% CIs) in gene number between detoxification breadth groups for the EcKLs, P450s, GSTs and CCEs gene families. Effect sizes (and 95% CI bars) are coloured by significance (black if $p < 0.05$, grey if $p > 0.05$). DB_L, large DB; DB_I, intermediate DB; DB_S, small DB. full, full dataset; sub, subset dataset. p-value coding: * < 0.05, ** < 0.01, *** < 0.001.

Table 3.5. PGLS outputs for models of diet and estimated DB against EcKL, P450, GST and CCE gene family sizes (Figs. 3.22–23).

Dataset	Response	Predictor	λ	df_n	df_d	F-value	p-value
full	EcKLs	DB	0.999	2	137	29.388	$< 1.0\times10^{-4}$
full	P450s	DB	1.004	2	95	3.509	3.4×10^{-2}
full	GSTs	DB	0.834	2	95	9.989	1.0×10^{-4}
full	CCEs	DB	0.978	2	95	0.161	8.5×10^{-1}
subset	EcKLs	DB	1.000	2	112	27.111	$< 1.0\times10^{-4}$
subset	P450s	DB	0.993	2	75	1.442	2.4×10^{-1}
subset	GSTs	DB	0.783	2	75	10.809	1.0×10^{-4}
subset	CCEs	DB	0.978	2	75	0.257	7.7×10^{-1}
full	EcKLs	diet	0.991	4	135	7.566	$< 1.0\times10^{-4}$
full	P450s	diet	1.000	4	93	4.630	1.9×10^{-3}
full	GSTs	diet	0.865	4	93	6.556	1.0×10^{-4}
full	CCEs	diet	0.980	4	93	1.281	2.8×10^{-1}

3.3.11. EcKLs have low rates of duplication in tsetse flies (Diptera: Glossinidae)

Tsetse flies—species in the genus *Glossina* (Diptera: Glossinidae)—are obligate haematophagous parasites of vertebrates and have an unusual life cycle in which larvae develop—until pupariation—inside their mother, which themselves only feed on blood (Tobe 1978). Because of this, they are a rare example of a taxon where all nutrition is derived from a single chemically stable food source, which should relax selection on DB breadth. The six annotated *Glossina* genomes analysed above have some of the smallest EcKL gene family sizes (26–28 genes) of all dipterans, along with the carnivorous *Proctacanthus coquillettii* (26 genes; Fig. 3.6), consistent with the small expected DB of both taxa. To explore whether the small DB of tsetse flies has also resulted in stable EcKL genes across the six annotated species in the *Glossina* genus, I generated a phylogenetic tree of all *Glossina* EcKLs and inferred the number of duplication and loss events by parsimony.

Glossina contains an inferred 30 ancestral EcKL clades, of which eight have experienced a duplication or loss event on the phylogeny (Fig. 3.24). Overall, only one duplication and eight losses were mapped to the *Glossina* phylogeny; this is substantially less than the 67 duplications and 71 losses inferred on the phylogeny of 12 species in *Drosophila* genus (Fig. 2.1). Tentatively normalised to the number of ancestral clades (30 for *Glossina*, 46 for *Drosophila*) and the total branch lengths of the phylogeny of each genus (67 m.y. for *Glossina*, 363 m.y. for *Drosophila*; from Fig. 3.3), *Glossina* has experienced 5.0×10^{-4} duplications/clade/m.y. and 4.0×10^{-3} losses/clade/m.y., while *Drosophila* has experienced 4.0×10^{-3} duplications/clade/m.y. and 4.3×10^{-3} losses/clade/m.y. Note that this normalisation of duplications and losses is contingent on the branch lengths of the *Glossina* phylogeny, which were estimated due to a lack of data on this genus in the literature. If this normalisation is accurate, the EcKL gene family in *Glossina* is duplicating at a rate 8-fold less (12.5%) than that in *Drosophila*, while its gene loss rate is roughly the same (93%).

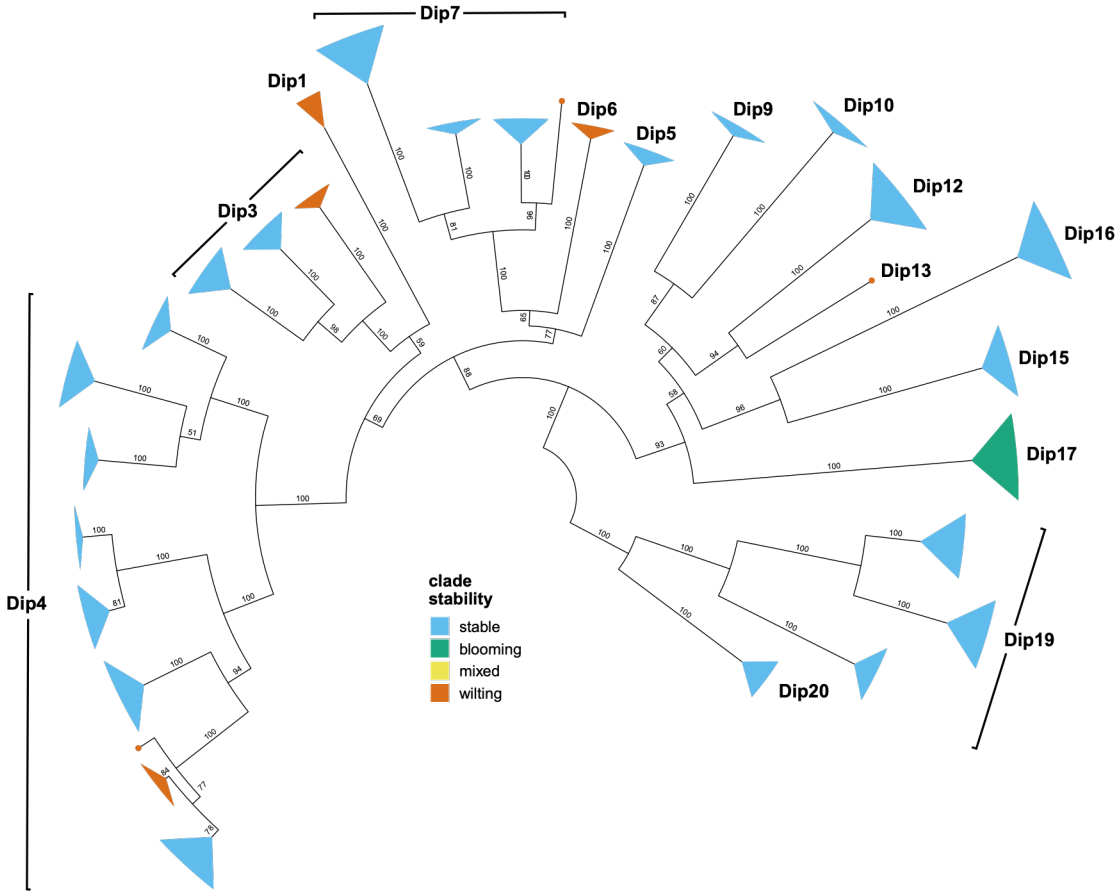


Figure 3.24. Maximum likelihood phylogenetic tree of 30 inferred ancestral EcKL clades in the genus *Glossina* (Diptera), labelled by the dipteran ancestral clades they belong to (see Fig. 3.10). Clades are collapsed for clarity and coloured by stability: 'stable' clades have 1:1 orthologs in all species; 'blooming' clades have at least one gene duplication with no gene losses; 'wilting' clades have at least one gene loss with no gene duplications; and 'mixed' clades have at least one gene duplication and no gene losses. Clades containing a single sequence are represented by small circles. Branch numbers are ultrafast bootstrap support values from UFBoot2 (Hoang *et al.* 2018). Tree is arbitrarily rooted, with the root branch removed.

3.3.12. EcKLs have low rates of duplication and loss in bees and sphecoid wasps (Hymenoptera: Apoidea)

Apoidea is a lineage within Hymenoptera that contains the bees (Anthophila) and the paraphyletic sphecoid wasps (Crabronidae, Sphecidae and others). The overwhelming majority of bees feed on pollen and nectar, a largely chemically stable diet that is lower in defensive secondary metabolites compared to plant tissues such as leaves (Rivest & Forrest 2020). Bees have very few EcKLs (12–16) compared to most other hymenopterans (Chapter 3.3.6); however, it is unclear if the observed stability of bee EcKLs extends to the close relatives of bees, which are carnivorous (predatory or parasitoids) and therefore should also have diets relatively low in toxins. The large number of Apoidea TSAs in the NCBI database—largely deposited by Peters *et al.* (2017) and San *et al.* (2018)—offered an opportunity to test this hypothesis.

I annotated the EcKLs from 75 Apoidea TSAs—39 from six bee families and 36 from eight wasp families—in addition to the 14 bee genomes already annotated (Fig. 3.25). Assuming that clade absences in bee transcriptomes (of the 12 clades present in bee genomes) were due to transcript absence and not true gene loss, the estimated false negative rate in bees was only 3.6% (17/468). Across all bees—including species with genomes—1.1% (7/636) of ancestral hymenopteran EcKL clades contained greater than one gene, while in wasps, 2.6% (11/432) of EcKL clades had greater than one gene; this difference between groups is not significant ($p = 0.092$, Fisher's exact test). Focusing just on the 14 annotated bee genomes (total branch length = 788 m.y.; Fig. 3.3) to minimise false negatives, the EcKL family in this taxon has experienced seven duplications and zero losses. Using the same normalisation method as for *Drosophila* and *Glossina* (Chapter 3.3.11), bee EcKLs have experienced an estimated 7.4×10^{-4} duplications/clade/m.y. (18% the rate of *Drosophila*) and 0 losses/clade/m.y. (0% the rate of *Drosophila*).

Overall, these data suggest that the transition from carnivory to pollen-feeding in bees did not substantially change the evolutionary stability of the EcKL family, and that the EcKL family in Apoidea is very stable relative to the *Drosophila* genus. In addition, the presence of EcKLs in the Hym22 clade (subfamily A) in the families Crabronidae and Sphecidae, but not Psenidae, Pemphredonidae or Philanthidae, is consistent with the Hym22 clade having been lost before the most recent common ancestor of bees.

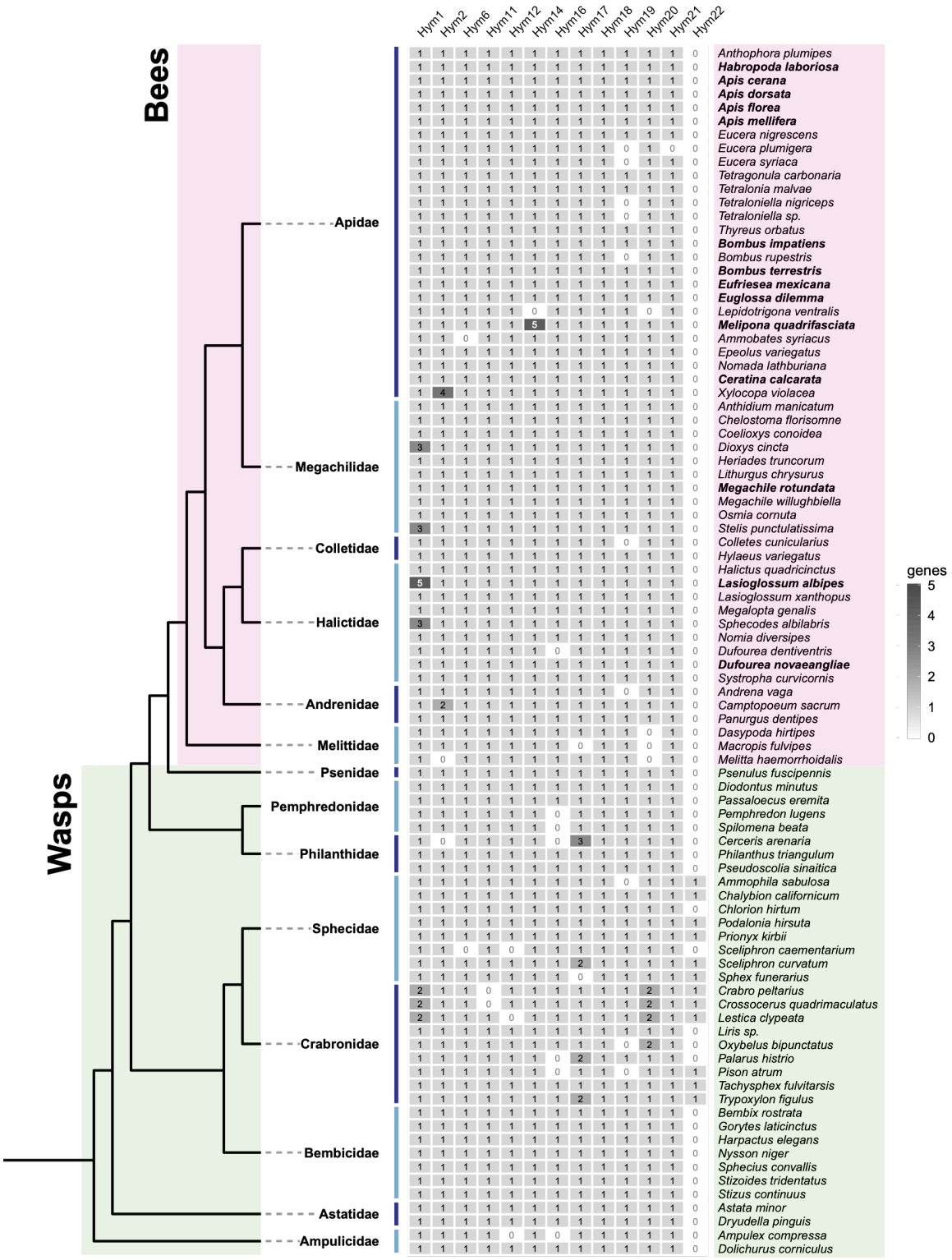


Figure 3.25. The number of EcKLs in the 13 relevant ancestral hymenopteran clades for bees (Anthophila; pink) and bee relatives (sphecoid wasps; green). The cladogram of phylogenetic relationships between Apoidea families, as well as species designations into each family, is derived from Sann *et al.* (2018). Bolded species are genomically annotated, all others are from transcriptomes. Within families, species are ordered by subfamily and then alphabetically.

3.4. Discussion

The analyses in this chapter have cast some light on the broader evolution of the EcKL gene family in insects that build on the results of Chapter 2. While much is discussed in this section, given the scope of these analyses, there are many details that cannot be realistically discussed in the thesis format. In the future, it might be useful to have a centralised electronic database set up for the purposes of integrating information about the evolution and functions of the EcKL gene family across insects. However, such a thing has not been attempted for the purposes of this thesis.

In addition, the results in this chapter inform on broader issues relating to the evolution of gene families that are partially outside the scope of this chapter's discussion. As such, these have been placed in Box 3.1 (EcKL nomenclature) and Box 3.2 (phylogenetic and annotation limitations).

3.4.1. Insect-wide EcKL conservation

Of the 13 EcKL subfamilies present in insects, I and J—and to a lesser extent, C and H—are retained in the vast majority of orders with genomic annotations (Fig. 3.8). Of these, subfamily C and I genes tend to be maintained as single-copy orthologs; subfamilies D and G, when present, are similar (Fig. 3.7). This makes the C, H, I and J subfamilies likely to contain EcKLs with widely important functions across insects, perhaps in development or physiology—interestingly, this implies that subfamily H, which has expanded substantially in some orders (Diptera, Hymenoptera, Hemiptera, Phasmatodea and Blattodea), may have had an important ancestral function that could be retained by one or more genes within the subfamily, even as other members functionally diverged to play roles in detoxification (see Chapter 3.4.6).

Subfamily I has the highest level of sequence identity among all its members: a subfamily I amino acid sequence from the apterygotan order Zygentoma has >60% identity with subfamily I sequences from holometabolan insects (a divergence of ~425 my; Misof *et al.* 2014); for comparison, subfamily J sequences between the same two taxa share typically <30% identity. This, combined with the other two conservation senses, suggest that subfamily I EcKLs may have a conserved and important function that

relates to a constant substrate.

It is important to note that even the best-conserved EcKLs appear relatively poorly conserved compared to some members of other large gene families in insects, such as the P450s. The Halloween P450s, which encode enzymes responsible for ecdysteroid biosynthesis and are essential for insect development, are—for the most part (Sztal *et al.* 2007)—retained as single-copy orthologs in insects (Rewitz *et al.* 2007) and even in most cases in crustaceans and other arthropods (Schumann *et al.* 2018). By comparison, subfamily I EcKLs are missing from 32 insect genomes, including 28 genomes from the dipteran section Schizophora (Fig. 3.6), demonstrating their function cannot be universally required in all insects in the same way as ecdysteroid biosynthesis.

3.4.2. Intra-order EcKL conservation

Despite a lack of broad conservation, lineage-specific genes can still be essential for processes such as development and reproduction (Bedell *et al.* 2012; Isoe *et al.* 2019; Katsube *et al.* 2009; Smith *et al.* 2018). As such, focusing on the conservation of EcKLs across insects as a whole may obscure the high conservation of EcKL clades within specific insect taxa, which may function in biological processes or traits that are specific to these taxa. Of course, the taxonomic group in question could be from superorder clades all the way down to genera or even species complexes—I am using the orders Diptera, Lepidoptera and Hymenoptera here because the species sampling in this study is only of a substantial size for these taxa; EcKL conservation across 12 species in the genus *Drosophila* has been previously explored in Chapter 2.

Diptera has five completely retained ancestral clades: Dip3, Dip4 and Dip5 (subfamily H), Dip12 (subfamily B) and Dip19 (subfamily J). Of these, only Dip12 is conserved as a single-copy ortholog in all species, while Dip5 is similar, with multiple copies only in the mosquito *Culex quinquefasciatus*. In *Drosophila melanogaster*, the Dip12 ortholog *CG2004* (Dro41-0) has a very broad expression pattern across tissues and life stages, offering few hints as to its higher-level functions; the *D. melanogaster* Dip5 ortholog *CG31098* (Dro8-0) is transcribed in a similar pattern (Graveley *et al.* 2011; Leader *et al.* 2018). However, *CG31098* may be involved in oocyte maturation and eggshell formation (Appendix 2) raising the possibility that it may play a similar role in other dipteran species. *CG9259* (Dro33-0) is the sole Dip19 ortholog in *D. melanogaster* and

is extremely highly and specifically expressed in the Malpighian tubules (Leader *et al.* 2018), raising the possibility that Dip19 genes could play essential roles in the physiology of this insect organ. It is interesting to note that Dip1 is conserved in every dipteran taxon except three species of *Glossina*—as will be discussed and explored later (Chapters 3.4.6 & 4), this EcKL clade is strongly linked to detoxification, and its loss in some *Glossina* species is consistent with a very small DB in these haematophagous insects. Also of note is the loss of Dip13 (subfamily C) in all *Glossina* species except the basal *G. brevipalpis*, as well as two chironomid species—Dip13 is otherwise well conserved and its absence in these species (if not due to genome assembly incompleteness) may reflect divergences in their biology with other dipterans related to the function of this EcKL; the same could be hypothesised about the restricted number of losses of the Dip7 (subfamily H), Dip14 (subfamily D), Dip15 and Dip16 (subfamily F), Dip17 (subfamily E) and Dip20 (subfamily J) clades.

Lepidoptera has three completely retained ancestral clades: Lep2 (subfamily A), Lep9 (subfamily C) and Lep10 (subfamily B). Of these, only Lep9 is conserved as a single-copy ortholog in all species, with Lep2 and Lep10 experiencing multiple duplications across the phylogeny. Lep2 is a sister clade to Lep1, which contains *BmEc22K*, the only biochemically identified ecdysteroid kinase (Sonobe *et al.* 2006); it is possible that the ancestral Lep2 gene encoded an ecdysteroid kinase, but it has also expanded to up to eight genes in some species, which tend to be DB_I or DB_L taxa (Fig. 3.17), suggesting it could be involved in detoxification in some taxa. If this is true, the retention of Lep2 may be due to a developmentally or reproductively important ecdysteroid kinase function. The closely related Lep1, Lep3, Lep4 and Lep6 clades are all nearly retained in all Lepidoptera—if their absences are due to genome incompleteness, they may also have conserved ecdysteroid kinase functions.

Hymenoptera has four completely retained ancestral clades: Hym17 and Hym18 (subfamily E), Hym19 (subfamily I) and Hym21 (subfamily C). Of these, Hym19 and Hym21 are conserved as single-copy orthologs in all species. Hym17 has greatly expanded in the chalcid wasps (Chalcidoidea; up to 14 genes in *Nasonia vitripennis*) but many of these genes have not been retained in *Ceratosolen solmsi*, a fig wasp in the family Agaonidae. Fig wasps are fig mutualists and are not parasitoidal (Xiao *et al.* 2013)—it is possible the large expansion of Hym17 genes in Chalcidoidea is related to parasitoid biology and these genes are largely under relaxed selection in *C. solmsi*, but

given the clade's complete retention in Hymenoptera, it likely has an important ancestral function that has been retained in one of *C. solmsi*'s remaining three Hym17 genes. Curiously, while bees have completely retained all 12 clades present in their common ancestor, overwhelmingly as single-copy orthologs, these clades aren't always well retained outside of bees—this suggests these clades may have become essential in the bee lineage.

3.4.3. Co-evolution of detoxification gene family sizes

In this chapter, positive relationships were found between the size of the EcKL gene family and the sizes of the P450, GST and CCE families in insects, all of which have known roles in the detoxification of natural and/or synthetic toxins (Li *et al.* 2007); the sizes of these three 'classical' detoxification families were also positively associated with each other, which recapitulates the findings of Rane *et al.* (2019)—albeit in a robust phylogenetic comparative framework with a smaller sample size (98 species vs. 160 species)—with R^2 values of between 0.21 and 0.31 (Fig. 3.19, Table 3.3). These data are consistent with the expectation that a population's toxicological environment exerts a selective pressure (or relaxes selection) to increase or decrease the sizes of detoxification gene families in concert with the population's required DB (Calla *et al.* 2017; Rane *et al.* 2019; 2016), and therefore also support the hypothesis that the EcKLs are generally involved in detoxification across insects.

Curiously, the sizes of the three classical detoxification families were, in general, more strongly associated with each other than they are with the EcKLs. This may be due to systematic errors (e.g. inflated gene size in poorly assembled genomes due to gene model fragmentation; Denton *et al.* 2014) introduced by the automated method (Rane *et al.* 2017) used to generate the P450, GST and CCE data, which produced unreliable UGT and ABC family size data in the same study (Rane *et al.* 2019). Ideally, future studies of associations between gene family sizes and other traits should use a variety of gene family annotation methods to explore the effect of different estimates of family size on the associations detected. Another possible explanation for the weaker EcKL associations is that the EcKL family has a higher proportion of S-class genes to X-class genes than the other three families, reducing the amount of variation in the family that is due to detoxification response, but this is inconsistent with the associations observed between EcKL family size and diet/DB, which were stronger than those of the

other three gene families (Chapter 3.3.10).

While positive associations between the sizes of detoxification gene families may be explained by detoxification-dependent co-evolution, it is possible that the sizes of the EcKL, P450, GST and CCE gene families correlate due to detoxification-independent genome-scale evolutionary forces, such as some lineages having higher overall rates of gene duplication and fixation than others. This could be tested by comparing the sizes of these four families with large gene families not thought to have roles in detoxification, but this was not done here due to a lack of accessible data on the scale needed for analysis.

3.4.4. Associations between dietary phenotypes and detoxification gene family sizes

If the sizes of their detoxification gene families partially determine how many toxin species a particular insect can detoxify, it would follow that the sizes of these families should vary between insect taxa with differing diets, with complex, unpredictable or otherwise toxin-rich diets promoting the expansion of detoxification gene families, and simple, stable or otherwise toxin-poor diets promoting the contraction of detoxification gene families. This hypothesis has been proposed numerous times (Calla *et al.* 2017; Rane *et al.* 2019; 2016) but has never been robustly tested with phylogenetic comparative methods, which are necessary to correct for phylogenetic relationships when associating traits among collections of species. In this chapter, my own annotations of the EcKL gene family, along with gene family size data determined by Rane *et al.* (2019), were used to test whether the sizes of known detoxification gene families (the P450s, GSTs and CCEs) or a hypothesised detoxification gene family (the EcKLs) varied between insect diets or an estimated level of detoxification breadth (DB).

Unexpectedly, the most consistent signals of association between gene family size and either diet or DB were for the EcKLs, with similar signals seen for the GSTs and only weakly for the P450s (Figs. 3.22–23). This is good evidence that the functions of a large subset of the EcKL family across insects are related to changes in diet and/or exposure to toxins, and that there are few other variable functions of the gene family that could confound this signal. These data are also consistent with the known roles of GSTs in detoxification in insects and other taxa (Shabab *et al.* 2014; Tan & Low 2018), and also

suggest that many insect GST genes may function in detoxification, with the known examples of GST involvement in development and physiology (Chanut-Delalande *et al.* 2014; Enya *et al.* 2014) in a relative minority.

P450 size was only significantly different between detritivores and haematophages (Fig. 3.22) and between DB_I and DB_S species (Fig. 3.23); however, the association between P450 family size and estimated DB failed to be significant when parasitoid wasps and omnivorous/fungivorous ants were removed from the data, suggesting this association is not robust to taxon sampling. The P450s are one of the gene families best associated with detoxification across eukaryotes (Kawashima & Satta 2014; Thomas 2007), let alone insects (Daborn *et al.* 2002; 2007; Denecke *et al.* 2017b; Schuler 2010), so it is surprising to see a general lack of association between the family's size and diet or DB. It is possible some outlier taxa—such as *Calopteryx splendens* (Odonata; 199 P450s), *Solenopsis invicta* (Hymenoptera; 197 P450s) and *Camponotus floridanus* (Hymenoptera; 173 P450s)—are confounding a true association between P450 size and diet/DB. But it is also possible that the many non-detoxification functions of P450s, which include secondary metabolism and pheromone degradation (Beran *et al.* 2019; Wojtasek & Leal 1999), produce enough additional variability in the size of the family that signals of detoxification become confounded. It is also possible that subsetting the data into specific P450s clades might produce better resolution, as some are thought to be differentially expanded in insects with larger DB (Calla *et al.* 2017; Feyereisen 2011; Li *et al.* 2007).

A lack of association between CCE family size and insect diet or DB suggests that this family may contribute little to the metabolism of natural toxins—indeed, CCEs are best known for the detoxification of synthetic insecticides, such as pyrethroids and organophosphates (see references in Oakeshott *et al.* 2005), which cannot have contributed to the evolution of insect detoxification enzymes before their initial uses. Pyrethroids are synthetic derivatives of pyrethrins, which are natural toxins synthesised by species of the plant genus *Chrysanthemum* (Staudinger & Ruzicka 1924; Ujihara 2019) but are not likely to be encountered by most insect taxa in nature. Some natural organophosphate toxins can be synthesised by bacteria and algae (Fiore *et al.* 2020; Neumann & Peter 1987) but it is unclear how widespread they are in insect habitats. As such, it remains to be seen if any insect CCEs have roles detoxifying natural toxins in the same manner as P450s and GSTs; however, CCEs are clearly involved in the

degradation of volatile compounds in odorant-sensing organs, which may be a form of localised detoxification (Durand *et al.* 2011; Leal 2013).

A significant contrast was only found between pollen-feeding insects and another diet (haematophagy) for a single gene family (Fig. 3.22); this is unsurprising given the poor phylogenetic sampling of pollen-feeding insects—all pollen-feeding insects in this dataset were bees, with the exception of the lepidopteran *Galleria mellonella* (Fig. 3.17)—which made the statistical power for comparisons with other groups extremely low after phylogenetic correction. Haematophagy was the diet group with the best evidence of reduced detoxification gene size compared to other diet groups, with significantly fewer EcKLs, P450s and GSTs compared to detritivores, and fewer EcKLs and GSTs compared to herbivores and carnivores (Fig. 3.22). This is consistent with vertebrate blood—the food source of all haematophagous insects in this study—lacking the capability of conducting a chemical 'arms-race' with the insect, in strong contrast with a herbivorous diet (de Castro *et al.* 2017; Edger *et al.* 2015; Ibanez *et al.* 2012), and lacking the toxin complexity present in detritus and rotting substrates (Nielsen *et al.* 2013; Trienens *et al.* 2017; 2010).

The association between EcKL family size and DB was further supported in Lepidoptera, where the taxonomic breadth of the host plants consumed by moth and butterfly larvae was used as a measure of the degree of polyphagy and hence a quantitative proxy for DB, with the expectation that the larger the variety of plants consumed, the larger the variety of plant secondary metabolites a species is exposed to and must detoxify. Total EcKL family size was significantly associated with host plant diversity (Fig. 3.20), as was the size of two ancestral clades, Lep1 and Lep8 (both subfamily A; Fig. 3.21), suggesting that genes in these clades may be involved in the detoxification of plant secondary metabolites. However, these results are contingent on the quality of the host plant data analysed, which came from the HOSTS Database (Robinson *et al.* 2010), a convenient resource that has been used to study dietary breadth in the past (Hardy *et al.* 2018; Lancaster 2020; Singer *et al.* 2014) but nevertheless is likely incomplete and may contain human-derived errors. In addition, records in HOSTS are either 'natural records' (larvae observed feeding on a plant in nature) or 'captive records' (larvae feed on a plant in the laboratory/captivity and development is successfully completed). This means that natural records in this database may not truly reflect the ability of a species to complete development on a given plant (a measure of toxicity),

as lepidopteran larvae can engage in 'diet-mixing' by switching between favourable and unfavourable host plants throughout development, which can improve growth, survival and fecundity even if a particular host plant is of low nutritional quality or contains unpalatable toxins (Barbosa *et al.* 1986; Mody *et al.* 2007); the observation of a species on a given plant in nature is therefore compatible with the plant containing toxins the larvae cannot easily tolerate. Given this, follow-up studies on the relationship between EcKL gene family size and host plant diversity in Lepidoptera should include information on mortality, growth rate (which may be more sensitive to dietary toxicity than mortality; Cornell & Hawkins 2003) and adult fecundity (Awmack & Leather 2002), to test if the associations seen in these analyses are still observed.

3.4.5. Is detoxification breadth (DB) associated with EcKL stability as well as size?

A greater number of neofunctionalised detoxification enzymes is thought to be linked to larger DB (Calla *et al.* 2017; Rane *et al.* 2019; 2016), but the evolutionary stability of a detoxification gene family may also vary with DB. The logic is as follows: if a detoxification family has contracted to a small size in a DB_S lineage, it likely still contains E- and S-class genes that are essential for development, reproduction and viability outside of a detoxification context, as well as a relatively small number of X-class genes that are important for general or niche-specific detoxification purposes; all of these genes should be under purifying selection and therefore conserved as single-copy orthologs in all closely related species. In a DB_L lineage, the complexity of the dietary exposure to toxins may be variable and the higher rate of gene duplication, fixation and pseudogenisation may lead to rapid changes in copy number even in closely related species. As such, DB_S and DB_L lineages may differ in the average stability of the members of their detoxification gene families, not just the overall size of the families. Another way to think about this is that detoxification genes are already thought to be more evolutionarily unstable than housekeeping genes in the same family, as seen in the P450s (Kawashima & Satta 2014; Thomas 2007), and relaxed selection on detoxification gene number due to relaxed selection on DB in 'low toxin' lineages will tend to reduce the number of detoxification genes through pseudogenisation, increasing the average stability of the detoxification gene families once these genes have been removed.

In this chapter, estimates of EcKL duplication and loss in *Drosophila* (Diptera: Ephydroidea), tsetse flies (Diptera: Hippoboscoidea) and bees (Hymenoptera: Anthophila), normalised by ancestral family size and time, are consistent with the expectation that lineages with large DB (*Drosophila*) have high rates of duplication, while lineages with small DB (tsetse flies and bees) have lower rates of duplication (Chapters 3.3.11–12). Curiously, while EcKL loss was not detected in bees, tsetse flies have a similar rate of EcKL loss to *Drosophila*. This could be explained by most or all the 12 ancestral clades in bees having important E-class functions, with unneeded X-class genes already lost in the carnivorous apoid wasp lineage leading to bees, while some of the 30 ancestral clades in tsetse flies are X-class genes retained from a possibly saprophagous ancestor (Wiegmann *et al.* 2011), now under relaxed selection due to obligate haematophagy.

While not quantified due to much smaller taxon sampling, DB_S lineages such as *Polistes* (paper wasps; carnivores) and *Bombyx* (silkworm moths; specialist herbivores), and DB_I and DB_L lineages such as *Calephelis* (metalmark butterflies; herbivores/detritivores), *Bactrocera* (fruit flies; frugivores/saprophages) and Heliothinae (moths; generalist herbivores), also follow the pattern of more stable EcKLs in smaller DB lineages and less stable EcKLs in larger DB lineages (Figs. 3.12, 3.14 & 3.16). More focused analyses with greater taxon sampling would ideally be conducted to explore this association further.

3.4.6. Candidate detoxification EcKLs in Diptera

The detoxification score method developed and applied in Chapter 2 suggests that up to 24 EcKLs in *Drosophila melanogaster* are good detoxification candidates—the evolutionary analysis of the EcKL gene family in the current chapter allows for the identification of orthologs of these *D. melanogaster* detoxification candidates in other dipteran species. 11 *D. melanogaster* detoxification candidate EcKLs belong to the Dip4 clade, five belong to the Dip1 clade, four belong to the Dip3 clade and one each belong to the Dip7, Dip15, Dip16 and Dip17 clades (Table 3.5). Given that the Dip1, Dip3 and Dip4 have such divergent sizes between dipteran taxa (Fig. 3.12), it is tempting to speculate that these genes have undergone gene blooms in different lineages in response to changing detoxification requirements. In addition, the fact that species with very small expected DB, such as *Glossina* species (obligate haematophages) and *Proctacanthus coquilletti* (an insect carnivore), have very few Dip1, Dip3 and Dip4 genes (Fig. 3.12) is

consistent with a general role for these clades in detoxification processes. The Dip1 and Dip4 clades in *Glossina* have also experienced multiple gene loss events (Fig. 3.24), suggesting they are in the process of contracting in this taxon.

It is important not to overgeneralise and claim that every member of these clades is involved in detoxification. Indeed, for example, five Dip4 genes in *D. melanogaster* have no or very few detoxification characteristics, three of which show developmental lethality phenotypes upon putative RNAi knockdown (Chapter 2); it is likely that some members of the Dip1, Dip3 and Dip4 clades have evolved developmental or physiological functions in certain lineages and more work is required to test the detoxification functions of any particular gene of interest.

It is also possible that clades not present in *D. melanogaster*, which therefore could not be analysed with the detoxification score method, may also have members involved in detoxification. The Dip8 clade has expanded in the Culicomorpha (mosquitos and midges), which are expected to have relatively large DB due to detritivorous life stages (Fig. 3.17), raising the possibility that these genes fulfil some of the roles played by Dip3 and Dip4 genes in other dipterans, especially since the latter clades are relatively small in the Culicomorpha compared to other species (Fig. 3.12).

3.4.7. Candidate detoxification EcKLs in Lepidoptera

In this chapter, a positive relationship was found between the size of the EcKL gene family and the diversity of the host plants consumed within the order Lepidoptera (Fig. 3.20). Furthermore, this association was linked to the sizes of only two ancestral clades of EcKLs—Lep1 and Lep8 (subfamily A)—which appear to have undergone multiple independent expansions through gene duplication within different moth and butterfly superfamilies (Figs. 3.21 & S3.4). Based on these data, it is reasonable to hypothesise that some Lep1 and Lep8 genes are involved in the detoxification of plant secondary metabolites in Lepidoptera and have been repeatedly neofunctionalised throughout the order in response to changes in host plant range. Lep1 and Lep8 genes may also be involved in the detoxification of microbial toxins too, given their substantial expansions in the detritivorous species *Calephelis virginiensis*, *Calephelis nemesis* and *Calycopis cecrops* (Fig. 3.14).

The possible association of Lep1 genes with detoxification is surprising given the known function of the sole Lep1 member in *Bombyx mori*—*BmEc22K*—which encodes an ecdysteroid 22-kinase involved in development and reproduction (Ito & Sonobe 2009; Sonobe *et al.* 2006; Sonobe & Ito 2009). However, a plausible substrate transition from ecdysteroid kinases to xenobiotic kinases could have been through phytoecdysteroids, which are produced by many plant taxa as insecticidal defensive compounds and are in some cases chemically identical to endogenous ecdysteroids in insects (Dinan 2001). There are other steroidal secondary metabolites produced by plants as well that could have been within the substrate specificity of an ancestral Lep1 ecdysteroid kinase, including withanolides, cucurbitacins and cardiac glycosides (Dinan *et al.* 1997; Dobler *et al.* 2011; Glotter 1991).

A genomic locus that contains four Lep1 EcKLs in the Australian *Helicoverpa armigera* subspecies *H. armigera conferta* appears to have recently undergone a selective sweep (Anderson *et al.* 2018). Given the evidence presented in this chapter that Lep1 genes may be involved in adaptation to a broad range of host plants, and given *H. armigera*'s broad host plant range and large EcKL family size, an attractive hypothesis is that least one of these Lep1 EcKLs has been selected in response to a novel detoxification challenge and is in the process of neofunctionalization. Alternatively, given the role of *BmEc22K* in reproduction and development within the ovarian/embryonic ecdysteroid-phosphate (OEEP) system in *B. mori*, it is also possible that selection on these Lep1 genes could be involved in developmental adaptations to more specialised environmental conditions, given *H. armigera conferta*'s more limited geographic range (Anderson *et al.* 2018).

It is possible that EcKLs in other lepidopteran ancestral clades are also involved in detoxification, even if they were not detected as associated with host plant diversity in this study, particularly those closely related to Lep1 and Lep8 in subfamily A. The Lep3 clade has undergone large expansions in some taxa (Fig. 3.14) and could conceivably be involved in detoxification in taxa where the Lep1 and Lep8 clades have not sufficiently neofunctionalised. The Lep2 clade may also be involved in detoxification, as previously mentioned (Chapter 3.4.2)—the largest number of Lep2 genes are in *Danaus plexippus*, a milkweed specialist, raising the possibility that some of them might be involved in cardiac glycoside metabolism or storage (Mebs *et al.* 2005).

Boeckler *et al.* (2016) found that salicinoid phosphorylation after feeding on poplar leaves was only detected in two species, *Lymantria dispar* (subspecies unspecified) and *Orgyia antiqua* (Noctuoidea: Erebidae), and not in species from the Noctuoidea families Notodontidae and Noctuidae, nor from the families Lasiocampidae (superfamily Lasiocampoidea), Geometridae (superfamily Geometrioidea) or Sphingidae (superfamily Bombycoidea). They were also absent from the beetle *Chrysomela populi* (Coleoptera: Chrysomelidae). If EcKLs are responsible for salicinoid phosphorylation, this suggests that they may have arisen in the Erebidae. Two Erebidae species were annotated in this study: *Lymantria dispar dispar*, the European gypsy moth, and *Hyphantria cunea*, the fall webworm. *L. dispar dispar* and *O. antiqua* belong to the Erebidae subfamily Lymantriinae, while *H. cunea* belongs to the subfamily Arctiinae (Zahiri *et al.* 2011); while it is unknown if *H. cunea* can phosphorylate salicinoids, poplar trees are among its known host plants (Robinson *et al.* 2010). If *H. cunea* cannot phosphorylate salicinoids, the EcKLs responsible in *L. dispar* might be in the Lep3 clade, which has substantially expanded in this species; otherwise, neofunctionalised Lep1 or Lep8 EcKLs may be involved.

3.4.8. Candidate detoxification EcKLs in Hymenoptera

Unlike in Diptera and Lepidoptera, there are no additional data analysed in this study that allow for associations between specific EcKL clades in Hymenoptera and detoxification properties to be inferred—but this does not mean hypotheses cannot be generated about which ancestral clades in Hymenoptera might be involved in detoxification. The most dynamic EcKL clade in Hymenoptera is Hym1 (subfamily H), which has expanded substantially in the fungus-cultivating ants (Attini), parasitoid wasps (Ichneumonoidea) and some basal sawflies (Tenthredinoidea and Cephoidea); it is also one of the only unstable clades in bees, with five genes in *Lasioglossum albipes*. This makes Hym1 a possible candidate for a detoxification clade in Hymenoptera.

Attine ants consume a single food source—fungal symbionts they 'farm' in their nests in 'gardens'—which would superficially appear to make their DB quite small. However, these fungal gardens are supplied with a number of food substrates foraged by their ant hosts, which can vary depending on the ant taxon; higher attine genera (*Atta*, *Acromyrmex* and *Trachymyrmex*)—some of which are called 'leaf-cutter ants'—predominantly or exclusively forage fresh plant material (De Fine Licht & Boomsma

2010), effectively making them indirect herbivores. While fungal symbionts may detoxify plant-derived phenolic compounds (De Fine Licht *et al.* 2013), it is likely that not all plant secondary metabolites are effectively detoxified this way, given that attine ants selectively forage plant material based on secondary chemistry (Howard 1988; 1987; Hubbell *et al.* 1984) and can be deterred by the presence of terpenes, phenolics and tannins (Howard 1990; Howard *et al.* 1989). As such, there are likely selective pressures on attine ants to be able to tolerate a wide variety of plant secondary metabolites, as these appear to limit their foraging range. As attine ants possess 7–18 EcKLs in the Hym1 clade, this leads to an attractive hypothesis that the expansion of Hym1 EcKLs in this taxon may reflect a role in detoxifying plant secondary metabolites still present in their fungal gardens during consumption.

Sawflies are basal Hymenoptera that are predominantly herbivorous, in contrast to most of the rest of the order. Hym1 genes have expanded in *Athalia rosae* and *Neodiprion lecontei* (Tenthredinoidea), both of which consume leaves, and to a lesser extent *Cephus cinctus* (Cephoidea), a grass stem feeder. This is consistent with a role for basal Hym1 genes in plant secondary metabolite detoxification, although more data is needed from sawflies to see if this trend is sustained.

The Hym1 clade has also expanded in parasitoid wasps in the superfamily Ichneumonoidea. As previously mentioned (Chapter 3.2.4), the expected DB of parasitoid wasps as a group is not clear and likely depends on the life stage parasitised and its exposure to toxins—for example, an egg parasitoid is likely less exposed to toxins than a larval parasitoid with a generalist herbivore host. There is some evidence that parasitoid wasps with herbivorous hosts are exposed to, and negatively affected by, plant toxins (reviewed in Ode 2006), suggesting some parasitoids may be under selective pressure to develop adequate detoxification processes. Unfortunately, the small sample size of species in this taxon (n = 7) means that relationships between parasitoid host DB and EcKL clade size are unable to be determined—however, this would make for a worthwhile hypothesis to test in future work.

The Hym1 clade is also one of the only clades to be repeatedly (albeit relatively mildly) expanded in bees, despite the EcKL gene family's high overall stability in this taxon (Fig. 3.25). While the diet of bees—pollen and nectar—is generally very low in plant secondary metabolites compared to that of herbivores that feed on leaves and stems,

all bees are naturally exposed to compounds such as flavonoids and alkaloids that they must detoxify (Berenbaum & Johnson 2015; Gegear *et al.* 2007; Jacobsen & Raguso 2018; Mao *et al.* 2017). In addition, some pollen is chemically defended and preferentially consumed by specialist (mono- or oligolectic) bees over generalist (polylectic) ones (Francis *et al.* 2019; Rivest & Forrest 2020; Vanderplanck *et al.* 2016). Bees are hypothesised to have 'social detoxification' systems, whereby pollen and nectar toxins are reduced by microbial fermentation, as well as pollen mixing, which dilutes toxins present in chemically defended pollen to palatable levels (Berenbaum & Johnson 2015; Eckhardt *et al.* 2014). However, there has likely been selective pressures on certain bee lineages to evolve slightly larger DB, which may explain the small expansions of Hym1 EcKLs observed here, similar to the small expansions found in some P450 clades in bees (Johnson *et al.* 2018b).

Overall, while not statistically robust, these data may suggest a possible association between Hym1 genes and detoxification processes in Hymenoptera, and this deserves to be explored in further detail in future work.

3.4.9. Candidate detoxification EcKLs in other insect orders

Due to poor sampling in other insect orders, it is hard to rigorously associate specific EcKLs with detoxification in these taxa. However, the associations previously discussed for Diptera, Lepidoptera and Hymenoptera make it possible to speculate about similar associations in other orders. Subfamily A, which may contain detoxification genes in Lepidoptera, has expanded in the coleopteran species annotated in this study, all of which are herbivores or detritivores, as well as in *Limnephilus lunatus*, a detritivore (Trichoptera). Likewise, subfamily H—which contains the bulk of detoxification candidate genes and clades in Diptera (and could also be involved in detoxification in Hymenoptera)—is very large in the stick insect *Timema cristinae* (Phasmatodea) and the cockroach *Blattella germanica* (Blattodea), species with very large expected DB. Conversely, it is interesting to note that subfamily H is very small or absent in Coleoptera, Lepidoptera and *L. lunatus*, while subfamily A is absent or relatively small in Diptera, Hymenoptera, *T. cristinae* and *B. germanica* (Fig. 3.7)—this suggests that different taxa may have expanded different subfamilies of EcKLs to increase their DB.

Box 3.1. EcKL gene family nomenclature

Appropriate gene naming is a long-standing problem in genetics and bioinformatics, with implications for biomedicine and other applied fields (Giguère 2015; Hall & Schwarz 2016; Schwarz *et al.* 2018; Shows & McAlpine 1978). To prevent confusion, individual gene names should ideally be unique and refer only to a particular gene and its 1:1 orthologs in other species (Seringhaus *et al.* 2008). Gene family nomenclatures that can be used to derive unique gene names are desirable because they allow names to be purely derived from a gene's evolutionary classification within a family, without resorting to knowing their specific biochemical or higher-level functions, experimental evidence for which is typically lagging far behind the sequencing of an ever greater number of genomes (Baric *et al.* 2016; El-Gebali *et al.* 2018). Another desirable property of gene family nomenclatures—previously discussed in Chapter 1.1.5—is the cladistic ('ortholog-based') grouping of genes within the family, which allows for the accurate and intelligible representation of ortholog and paralog relationships. Existing nomenclatures for large, complex gene families are typically very successful for the former property and variably successful for the latter property—for example, the largely sequence identity-based P450 and UGT nomenclatures (Mackenzie *et al.* 2005; Nelson 2006) successfully generate unique gene names (e.g. *Cyp6g1* and *Ugt36D1*), but can fail to capture higher-level cladistic relationships when large numbers of paralogs with divergent sequences are present, and can even fail to group 1:1 orthologs if sequence divergence is high enough. In addition, there are many thousands of P450 families reserved in various taxa (Nelson 2009), but it is likely that many of these families could be collapsed together if cladistic groups were used.

Ideally, the phylogenomic analyses in this chapter would allow for the development of an orthology-based nomenclature for the EcKL gene family that allows for unique gene naming, but in practice, this appears to be quite challenging. Various cladistic groups have been defined in this chapter, as well as Chapter 2: subfamilies (A–M), order-specific ancestral clades (e.g. Dip1) and genus-specific ancestral clades and subclades (e.g. Dro1-1). In the P450 system, unique gene names are derived by using a sequentially larger set of family numbers (e.g. *Cyp1a1*, *Cyp2a1*, *Cyp3a1*), subfamily letters (e.g. *Cyp1a1*, *Cyp1b1*, *Cyp1c1*) and terminal (gene-specific) numbers (e.g. *Cyp1a1*, *Cyp1a2*, *Cyp1a3*). Could the aforementioned EcKL groupings be used this way? Subfamily information seems reasonable but insufficient in the case of large subfamilies

(e.g. *EcKL-H* would not be a unique identifier in Diptera or Blattodea), adding order-specific clades is useful but also insufficient (e.g. there could be 16 genes named *EcKL-H-Dip1* in *Teleopsis dalmanni*), which suggests genus-level clades may be required (e.g. *EcKL-H-Dip1-Dro5-1* for *CG31300* in *Drosophila melanogaster*). By this point, the unique names produced have become overly long, clunky and hard to parse; it also requires that detailed order- and genus-level phylogenetic analyses have been performed before genes can be named—how might the genes in *Timema cristinae* be named without any other Phasmatodea or *Timema* species annotated? The root issue appears to be that each additional symbol in a unique gene name necessarily asserts a grouping—ideally a cladistic one—but evolutionary relationships are best described by graphs (Huson & Bryant 2005) rather than strings of symbols. As such, no elegant orthology-based gene-naming system can be perfectly suited to its task if sufficient complexity exists in the relationships between genes.

Ultimately, the task of family-based gene naming may be quixotic—in 2009, there were 11,512 P450s officially named (Nelson 2009), in 2013 that number had nearly doubled (https://drnelson.uthsc.edu/CytochromeP450.html), and many thousands of P450s likely remain officially unnamed given the recent explosion of sequenced and annotated genomes. It seems unlikely that genes in large, complex gene families like the EcKLs will ever be individually named in a satisfactory way through the rigid application of a defined nomenclature.

So, what is the alternative? The official gene sets of annotated genomes assign every gene a symbol (e.g. *CG13813* in *D. melanogaster*, *GMOY009792* in *Glossina mortisans*). For some, these symbols are unsatisfactory names for genes, given their typically long strings of unmemorable digits, and some of these symbols are eventually supplanted by 'true' names based on experimentally determined functions or clear 1:1 orthology with named genes in other taxa; in addition, gene symbols contain no information about the family to which the genes belong. But is this a problem? Whenever a gene in *D. melanogaster* or another species is named by its phenotype over its gene family name (e.g. *phantom* instead of *Cyp306a1*), the gene family is superficially obscured. In the age of widely accessible online bioinformatic databases (Coordinators 2016; Harris *et al.* 2020; Karimi *et al.* 2017; Thurmond *et al.* 2019), information about specific genes is easy to collect, parse and analyse, including gene family nomenclature and orthology data. Perhaps gene family nomenclature, including all relevant cladistic groups,

could be collated in these databases alongside experimentally determined functions, transcriptomic and proteomic expression data and citing literature.

Box 3.2. Limitations of annotation and phylogenetic methodologies

This study used manual annotation to curate a large set of EcKL gene models from 140 insect genome assemblies, as well as three non-insect arthropod assemblies. Manual curation was used to limit the gene model errors that can arise from automated gene model annotation methods, such as missing, fragmented or concatenated genes (Denton *et al.* 2014; Zhang *et al.* 2012). While for some metrics, such as genome-wide gene structural properties, automated annotated gene models may be similar to manually curated models (Wilbrandt *et al.* 2019), careful manual curation can undoubtedly minimise errors that may affect downstream phylogenetic analyses of protein sequences (McDonnell *et al.* 2018; Pfeiffer & Oesterhelt 2015; Stroehlein *et al.* 2018). However, a major limitation of manual annotation is low throughput. At the time of writing, there are 977 Tetraconata genome assemblies publicly available in NCBI (906 from Insecta, 20 from other lineages of Hexapoda and 51 from Crustacea) but only 143 assemblies were annotated in this study over a 4+ year period—while some genome assemblies were rejected for annotation due to poor quality (low scaffold N50, high scaffold L50), most were passed over due to time constraints. As automated annotation methods continue to improve (Keilwagen *et al.* 2018; König *et al.* 2016) along with the quality of genome assemblies through long-read assembly methods (Dudchenko *et al.* 2017; Richards & Murali 2015; Saha 2018), future analyses of the EcKLs and other gene families could adopt a faster hybrid annotation approach, where automated annotations are manually polished after a series of quality checks involving sequence alignments, cross-species comparisons and transcriptome sequencing evidence.

Regardless of the annotation method used, it is likely some small proportion of EcKL gene models remain missing from some of the annotated genomes due to assembly incompleteness—this is particularly likely for older genome assemblies that have a greater degree of fragmentation. Such false negatives may explain the small number of absent genes in otherwise well-conserved subfamilies or order ancestral clades, such as Dip15 and Dip20 in Diptera (Fig. 3.12). Ideally, these species should be

reannotated once higher-quality assemblies are available.

In this study, transcriptome shotgun assemblies (TSAs) across all 28 insect orders were also queried and mined to explore the retention of EcKL subfamilies across all insects, as well as improve sequence sampling for subfamily phylogenetic analyses. Only partial sequences of some subfamilies could be found for some orders (Fig. 3.9) and orders with subfamilies with only partial sequences have very low numbers of annotated genomes and queried TSAs (Fig. S3.2). This suggests that sufficient taxon sampling is required before one can be confident that subfamily absences are not false negatives. As such, for orders with poor genome and transcriptome sampling (essentially all orders except Diptera, Lepidoptera, Coleoptera, Hymenoptera, Hemiptera, Phasmatodea and Blattodea), the absence of any particular subfamily should be treated as tentative.

The construction of phylogenetic trees of amino acid sequences requires the production of multiple sequence alignments (MSAs), which assert the homology of residues aligned to the same column; these columns are then used (in likelihood-based methods) to find a well-fitting tree of the relationships between the sequences, given a specific model of evolution (Yang 1996). This makes phylogenetic inference sensitive to the MSA produced: if aligned residues are not truly homologous, the best tree found will not be accurate (Wong *et al.* 2008). This is particularly a problem for groups of highly divergent sequences, where low sequence identity makes the possibility of misalignment high (Rost 1999). One approach to improve the phylogenetic analyses of such diverged protein sequences is to use 3D structural information to guide alignments, which has been successfully applied to a number of gene/protein families and groups of families (Breitling *et al.* 2001; Bujnicki 2000; Cvicek *et al.* 2016; Garau *et al.* 2005; Lundin *et al.* 2012). Unfortunately, the EcKL gene family in insects has a very low (16–25%) overall sequence identity between its most divergent members, and structural information is not available, making the deep evolutionary relationships between divergent EcKLs difficult to determine, as evidenced by the low support values for most deep nodes in Fig. 3.5. Despite this, most EcKL sequences can be confidently grouped into clades that have been labelled subfamilies. Exceptions are some sequences from basal insect taxa (Odonata, Ephemeroptera, Zygentoma and Archaeognatha), whose assignment to some subfamilies were difficult due to poor branch support—this may have been due to long branch lengths, which can cause the

erroneous grouping of sequences through long-branch attraction (Bergsten 2005; Felsenstein 1978). This makes their assignment to subfamilies such as D, G, H and L only tentative—future phylogenetic analyses of the family should use a greater number of sequences from these taxa (if available), which may break up long branches and increase phylogenetic accuracy.

3.5. Conclusion

This chapter has provided the first robust and broad phylogenetic analysis of the EcKLs across insects, which will allow categorisation of members of the family in new genomes and leads to numerous hypotheses about the functions of highly conserved clades across insects and in specific taxa. This chapter has also provided multiple lines of phylogenomic evidence that the EcKL gene family is involved in detoxification processes, including associations with chemically complex diets, estimated DB, the sizes of known detoxification gene families and host plant diversity. Integration of this chapter's data with those from the *Drosophila* genus in Chapter 2 indicate specific clades of EcKLs in other Diptera may be involved in detoxification and may guide future work on detoxification in Diptera.

The next two chapters of this thesis deal with experimentally testing the functions of specific genes in *Drosophila melanogaster* using its powerful molecular genetic toolkit—Chapter 4 will explore the hypothesised detoxification functions of multiple Dro5 (Dip1/subfamily H) EcKLs, while Chapter 5 will explore the hypothesised ecdysteroid kinase functions of *CG13813* (Dro38-1/Dip10, named *Wallflower*) and *Pinkman/pkm* (Dro37-0/Dip9), both of which belong to the same subfamily (A) as the ecdysteroid 22-kinase gene *BmEc22K* in *Bombyx mori*.

3.6. Supplementary Materials

3.6.1. Supplementary Figures

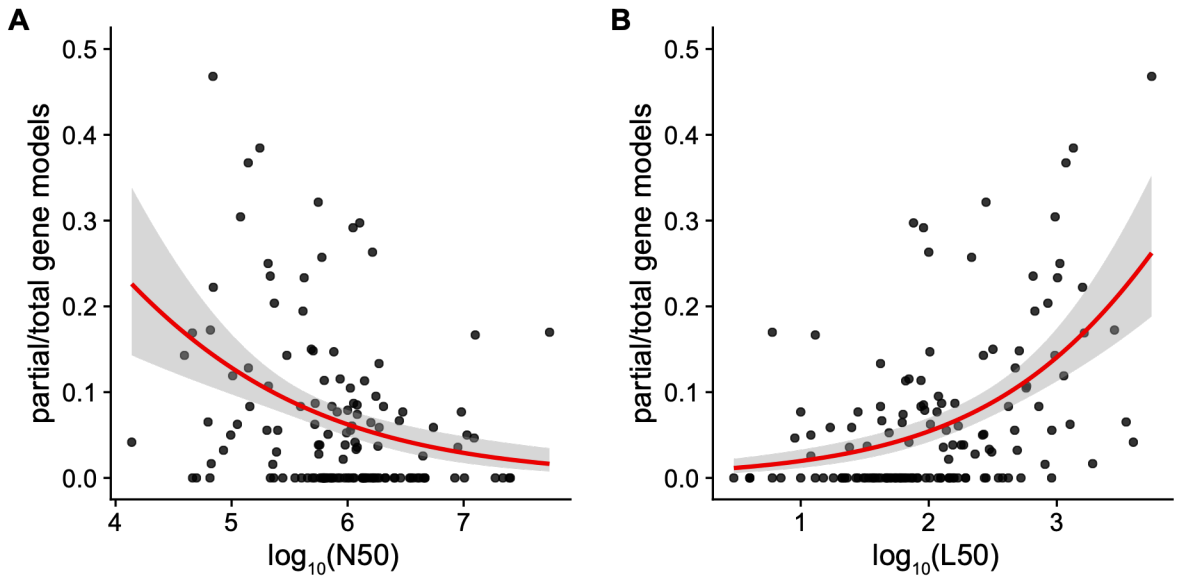


Figure S3.1. The proportion of partial to total EcKL gene models per annotated genome is (A) negatively associated with assembly scaffold N50 (length such that scaffolds of this length or longer include half the bases of the assembly; larger is better), and (B) positively associated with assembly scaffold L50 (number of scaffolds that are longer than, or equal to, the N50 length and therefore include half the bases of the assembly; smaller is better). Generalised linear models fit in R with the 'glm' function (family, quasibinomial; link function, logit; weight, total gene models): partial proportion ~ $\log_{10}$(N50), p = 1.0×10^{-5}; partial proportion ~ $\log_{10}$(L50), p = 7.0×10^{-10}. Shaded area is the standard error of the curve.

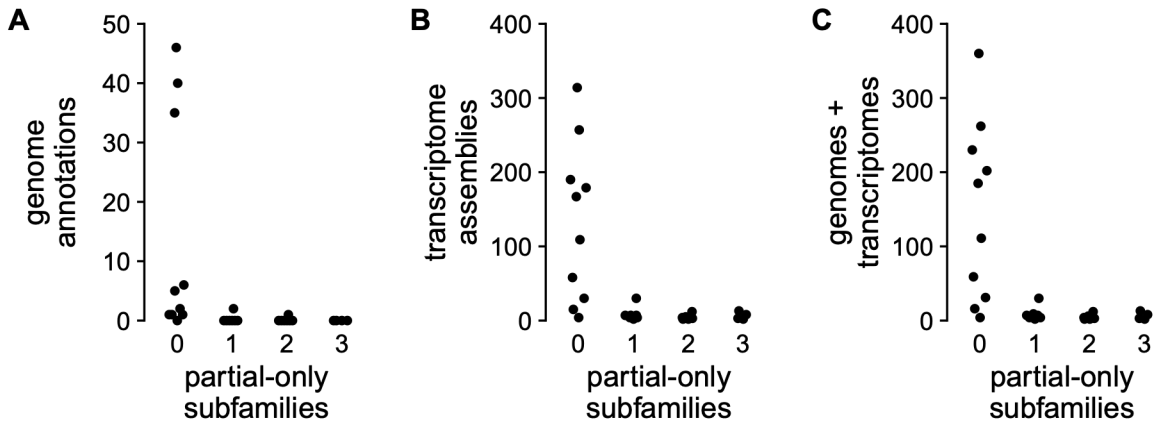


Figure S3.2. The number of (A) annotated genomes, (B) queried transcriptome shotgun assemblies (TSAs) and (C) combined annotated genomes and TSAs for each insect order per the number of insect EcKL subfamilies that contain only partial sequences. Data derived from Fig. 3.9.

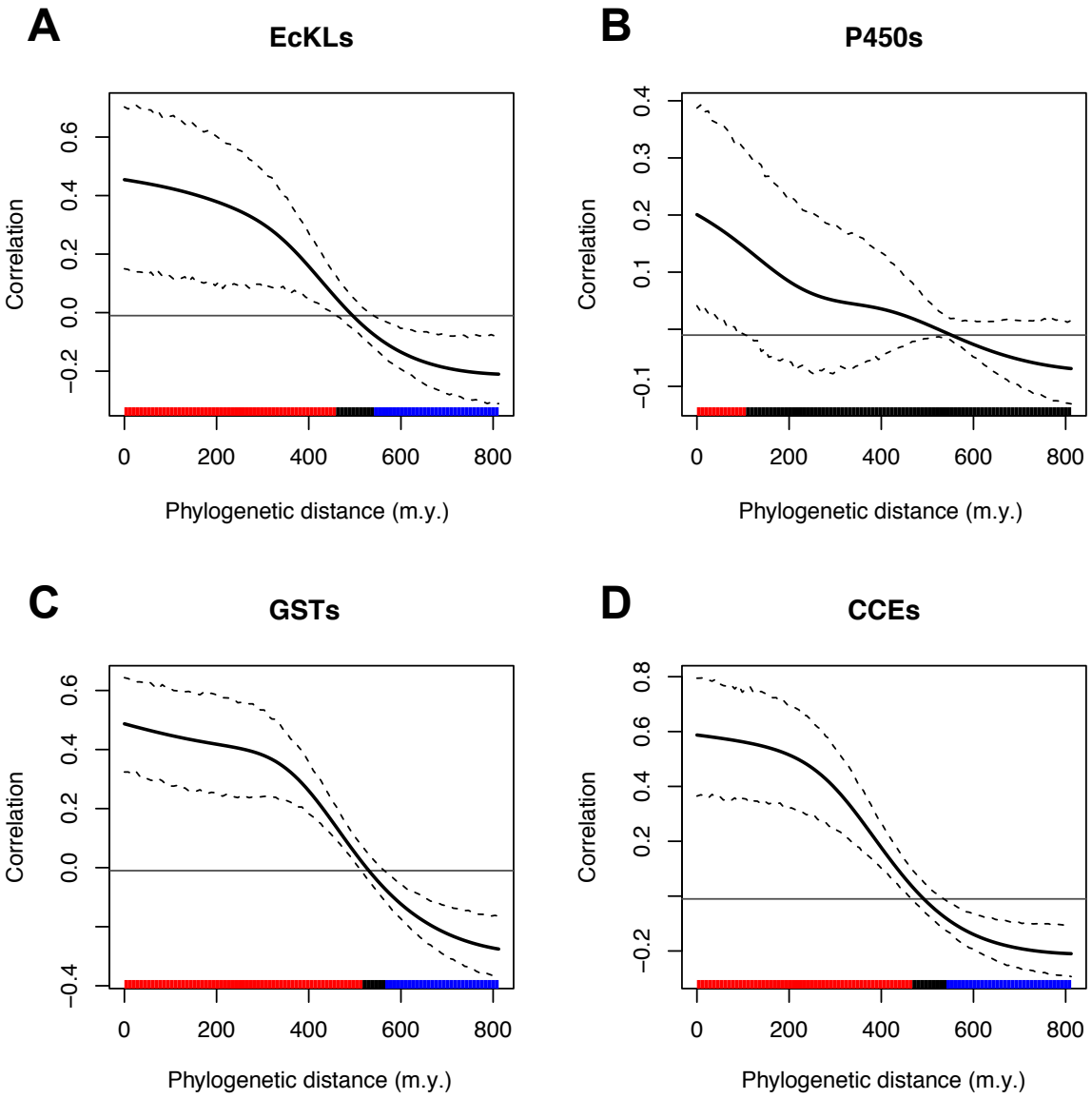


Figure S3.3. Phylogenetic correlograms for gene family sizes on the insect phylogeny, for the (A) EcKLs, (B) P450s, (C) GSTs and (D) CCEs (n = 140 species for EcKLs, n = 98 species for P450s, GSTs and CCEs). The solid black curves represent Moran's I index of autocorrelation at different tree depths, and the dashed black lines represent the 95% confidence envelop. The horizontal black line indicates the expected value of Moran's I under the null hypothesis of no phylogenetic autocorrelation. The coloured bars show whether the autocorrelation is significant: red for significant positive autocorrelation, black for nonsignificant autocorrelation, and blue for significant negative autocorrelation. Graphs generated with the R package *phylosignal* (Keck *et al.* 2016).

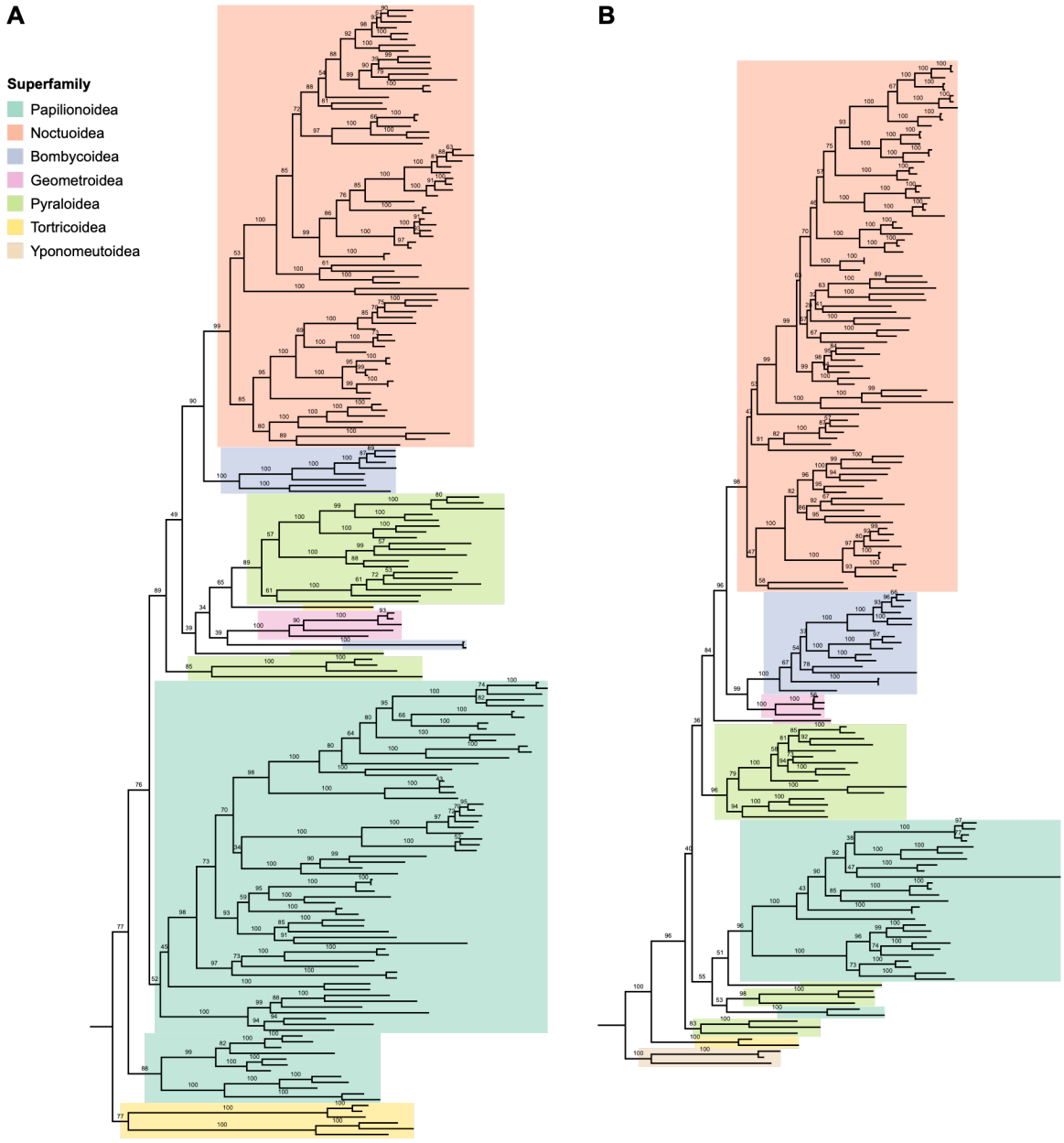


Figure S3.4. Maximum likelihood phylogenetic trees of the (A) Lep1 and (B) Lep8 ancestral clades in Lepidoptera. Clades are coloured by taxonomic superfamily, where all sequences (tips) within a coloured clade comes from a species in that superfamily. Partial sequences were excluded from the alignment and the tree. Branch numbers are ultrafast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018). These trees are consistent with multiple independent expansions of Lep1 and Lep8 genes within different lepidopteran superfamilies, due to low branch support (<95) where sequences from the same superfamily are not monophyletic.

3.6.2. Supplementary Tables

Table S3.1. All arthropod genome assemblies completely annotated for the EcKL gene family.

Species	Order	Database	Assembly Name	Date[a]	N50[b]	L50[c]	FGM[d]	PGM[e]	TGM[f]
Blattella germanica	Blattodea	NCBI	Bger_1.1	26/03/18	1,056,071	576	94	11	105
Zootermopsis nevadensis	Blattodea	NCBI	ZooNev1.0	22/07/14	751,105	194	19	0	19
Agrilus planipennis	Coleoptera	NCBI	Apla_2.0	3/01/18	1,113,421	91	17	7	24
Anoplophora glabripennis	Coleoptera	NCBI	Agla_2.0	12/12/17	678,234	269	56	3	59
Leptinotarsa decemlineata	Coleoptera	NCBI	Ldec_2.0	31/10/17	139,046	1,179	31	18	49
Dendroctonus ponderosae	Coleoptera	NCBI	DendPond_male_1.0	10/04/13	628,732	87	39	5	44
Hypothenemus hampei	Coleoptera	NCBI	ASM101285v1	22/05/15	39,211	966	24	4	28
Tribolium castaneum	Coleoptera	NCBI	Tcas5.2	10/03/16	4,456,720	12	38	1	39
Proctacanthus coquilletti	Diptera	NCBI	200kmer_750.trimmed	4/01/17	862,345	70	23	3	26
Belgica antarctica	Diptera	NCBI	ASM77530v1	3/09/14	98,263	263	38	2	40
Clunio marinus	Diptera	NCBI	CLUMA_1.0	28/11/16	1,871,155	17	32	2	34
Aedes aegypti	Diptera	NCBI	AaegL3	30/04/14	1,547,048	272	50	0	50
Anopheles gambiae	Diptera	NCBI	AgamP3	16/10/06	12,309,988	9	41	2	43
Culex quinquefasciatus	Diptera	NCBI	CulPip1.0	23/04/07	486,756	317	51	9	60
Teleopsis dalmanni	Diptera	NCBI	Tel_dalmanni_2A_v1.0	28/07/17	66,701	1,905	59	1	60
Drosophila ananassae	Diptera	NCBI	dana_caf1	30/06/06	4,599,533	10	56	0	56
Drosophila erecta	Diptera	NCBI	dere_caf1	12/07/06	18,748,788	4	53	0	53
Drosophila grimshawi	Diptera	NCBI	dgri_caf1	30/06/06	8,399,593	7	45	0	45
Drosophila melanogaster	Diptera	NCBI	Release 6 plus ISO1 MT	1/08/14	25,286,936	3	51	0	51
Drosophila mojavensis	Diptera	NCBI	dmoj_caf1	30/06/06	24,764,193	4	41	0	41
Drosophila persimilis	Diptera	NCBI	dper_caf1	18/04/06	1,869,541	21	44	0	44
Drosophila pseudoobscura	Diptera	NCBI	ASM14949v1	6/12/04	1,459,550	23	44	0	44

Species	Order	Database	Assembly Name	Date[a]	N50[b]	L50[c]	FGM[d]	PGM[e]	TGM[f]
Drosophila sechellia	Diptera	NCBI	dsec_caf1	18/04/06	2,123,299	13	47	0	47
Drosophila simulans	Diptera	NCBI	dsim_caf1	2/11/06	857,818	36	56	0	56
Drosophila virilis	Diptera	NCBI	dvir_caf1	12/07/06	10,161,210	6	44	0	44
Drosophila willistoni	Diptera	NCBI	dwil_caf1	3/08/06	4,511,350	15	61	0	61
Drosophila yakuba	Diptera	NCBI	dyak_caf1	27/06/06	21,770,863	4	55	0	55
Glossina austeni	Diptera	NCBI	Glossina_austeni-1.0.3	12/05/14	812,585	116	24	2	26
Glossina brevipalpis	Diptera	NCBI	Glossina_brevipalpis_1.0.3	8/05/14	1,209,507	63	25	2	27
Glossina fuscipes	Diptera	NCBI	Glossina_fuscipes-3.0.2	8/05/14	561,190	179	25	1	26
Glossina morsitans	Diptera	VectorBase	GmorY1	1/12/10	49,769		27	0	27
Glossina pallidipes	Diptera	NCBI	Glossina_pallidipes-1.0.3	12/05/14	1,038,751	95	28	0	28
Glossina palpalis	Diptera	NCBI	Glossina_palpalis_gambiensis-2.0.1	15/01/15	575,037	187	25	1	26
Musca domestica	Diptera	NCBI	Musca_domestica-2.0.2	22/04/13	226,573	809	62	1	63
Stomoxys calcitrans	Diptera	NCBI	Stomoxys_calcitrans-1.0.1	31/05/15	504,651	509	46	8	54
Lucilia cuprina	Diptera	NCBI	Lcup_2.0	15/12/17	275,862	350	56	0	56
Lutzomyia longipalpis	Diptera	NCBI	Llon_1.0	11/06/12	85,093	491	30	1	31
Bactrocera dorsalis	Diptera	NCBI	ASM78921v2	3/12/14	1,206,000	91	43	4	47
Bactrocera oleae	Diptera	NCBI	gapfilled_joined_lt9474.gt500.covgt10	29/07/15	139,566	474	34	5	39
Bactrocera tryoni	Diptera	NCBI	Assembly 2.2 of Bactrocera tryoni genome	21/05/14	69,551	1,589	35	10	45
Ceratitis capitata	Diptera	NCBI	Ccap_2.1	6/11/17	1,665,634	75	46	0	46
Rhagoletis zephyria	Diptera	NCBI	Rhagoletis_zephyria_1.0	19/07/16	62,643	3,479	43	3	46
Zeugodacus cucurbitae	Diptera	NCBI	ASM80634v1	23/12/14	1,399,015	66	47	6	53
Ephemera danica	Ephemeroptera	NCBI	Edan_2.0	12/12/17	599,527	216	26	9	35
Acyrthosiphon pisum	Hemiptera	NCBI	Acyr_2.0	24/06/10	518,546	280	16	0	16
Cimex lectularius	Hemiptera	NCBI	Clec_2.1	12/12/17	1,637,644	100	14	5	19

Species	Order	Database	Assembly Name	Date[a]	N50[b]	L50[c]	FGM[d]	PGM[e]	TGM[f]
Gerris buenoi	Hemiptera	NCBI	Gbue_2.0	12/02/18	412,278	674	29	7	36
Halyomorpha halys	Hemiptera	NCBI	Hhal_2.0	15/12/17	393,089	724	33	3	36
Rhodnius prolixus	Hemiptera	NCBI	Rhodnius_prolixus-3.0.3	8/04/15	1,088,772	170	31	2	33
Apis cerana	Hymenoptera	NCBI	ACSNU-2.0	12/11/15	1,421,626	42	12	0	12
Apis dorsata	Hymenoptera	NCBI	Apis dorsata 1.3	24/09/13	732,052	87	11	1	12
Apis florea	Hymenoptera	NCBI	Aflo_1.1	7/12/12	2,863,240	22	12	0	12
Apis mellifera	Hymenoptera	NCBI	Amel_4.5	14/01/11	997,192	65	12	0	12
Bombus impatiens	Hymenoptera	NCBI	BIMP_2.2	22/09/11	1,399,493	54	12	0	12
Bombus terrestris	Hymenoptera	NCBI	Bter_1.0	3/06/11	3,506,793	18	12	0	12
Ceratina calcarata	Hymenoptera	NCBI	ASM165200v1	25/05/16	632,424	82	12	0	12
Eufriesea mexicana	Hymenoptera	NCBI	ASM148370v1	9/05/16	351,926	377	12	0	12
Euglossa dilemma	Hymenoptera	NCBI	Edil_v1.0	20/06/17	143,590	419	11	1	12
Habropoda laboriosa	Hymenoptera	NCBI	ASM126327v1	13/08/15	1,784,116	47	12	0	12
Melipona quadrifasciata	Hymenoptera	NCBI	ASM127656v1	4/09/15	1,864,352	42	13	2	15
Dufourea novaeangliae	Hymenoptera	NCBI	ASM127255v1	21/08/15	2,549,405	32	12	0	12
Lasioglossum albipes	Hymenoptera	NCBI	ASM34657v1	8/10/13	628,061	148	16	0	16
Megachile rotundata	Hymenoptera	NCBI	MROT_1.0	19/07/11	1,699,680	47	12	0	12
Cephus cinctus	Hymenoptera	NCBI	Ccin1	9/07/14	622,163	56	18	0	18
Ceratosolen solmsi	Hymenoptera	NCBI	CerSol_1.0	4/12/13	9,558,897	10	12	1	13
Copidosoma floridanum	Hymenoptera	NCBI	Cflo_2.0	12/12/17	1,210,516	130	27	1	28
Nasonia vitripennis	Hymenoptera	NCBI	Nvit_2.1	28/11/12	708,988	71	47	0	47
Trichogramma pretiosum	Hymenoptera	NCBI	Tpre_2.0	15/12/17	1,825,723	33	26	1	27
Leptopilina clavipes	Hymenoptera	NCBI	ASM185565v1	28/10/16	13,761	3,959	23	1	24
Acromyrmex echinatior	Hymenoptera	NCBI	Aech_3.9	3/05/11	1,110,580	74	27	0	27
Atta colombica	Hymenoptera	NCBI	Acol1.0	25/03/16	2,037,154	42	22	2	24

Species	Order	Database	Assembly Name	Date[a]	N50[b]	L50[c]	FGM[d]	PGM[e]	TGM[f]
Camponotus floridanus	Hymenoptera	NCBI	CamFlo_1.0	17/09/10	451,320	144	20	0	20
Cyphomyrmex costatus	Hymenoptera	NCBI	Ccosl1.0	25/03/16	1,159,032	70	23	1	24
Dinoponera quadriceps	Hymenoptera	NCBI	ASM131382v1	13/10/15	1,361,239	59	13	0	13
Harpegnathos saltator	Hymenoptera	NCBI	HarSal_1.0	17/09/10	601,965	144	12	0	12
Linepithema humile	Hymenoptera	NCBI	Lhum_UMD_V04	10/06/11	1,402,257	40	14	0	14
Ooceraea biroi	Hymenoptera	NCBI	CerBir1.0	8/04/14	1,350,650	48	22	0	22
Pogonomyrmex barbatus	Hymenoptera	NCBI	Pbar_UMD_V03	4/02/11	819,605	81	17	0	17
Solenopsis invicta	Hymenoptera	NCBI	Si_gnH	1/08/18	621,039	124	24	0	24
Trachymyrmex cornetzi	Hymenoptera	NCBI	Tcor1.0	25/03/16	760,749	102	29	5	34
Trachymyrmex septentrionalis	Hymenoptera	NCBI	Tsep1.0	25/03/16	2,520,094	31	25	0	25
Trachymyrmex zeteki	Hymenoptera	NCBI	Tzet1.0	31/03/16	1,333,945	51	26	0	26
Vollenhovia emeryi	Hymenoptera	NCBI	V.emery_V1.0	6/03/15	1,346,088	61	18	0	18
Aphidius ervi	Hymenoptera	NCBI	ASM1142645v1	18/03/20	581,355	55	14	0	14
Cotesia vestalis	Hymenoptera	NCBI	ASM95615v1	18/03/15	46,055	911	19	0	19
Diachasma alloeum	Hymenoptera	NCBI	Dall2.0	6/03/19	657,001	124	24	0	24
Fopius arisanus	Hymenoptera	NCBI	ASM80636v1	23/12/14	978,588	49	18	1	19
Lysiphlebus fabarum	Hymenoptera	NCBI	ASM1142643v1	18/03/20	216,143	158	16	0	16
Macrocentrus cingulum	Hymenoptera	NCBI	MCINOGS1.0	22/05/17	65,089	525	13	0	13
Microplitis demolitor	Hymenoptera	NCBI	Mdem2	7/10/15	1,139,389	50	21	0	21
Orussus abietinus	Hymenoptera	NCBI	Oabi_2.0	15/12/17	612,083	63	14	0	14
Neodiprion lecontei	Hymenoptera	NCBI	Nlec1.1	21/06/18	243,810	310	32	1	33
Athalia rosae	Hymenoptera	NCBI	Aros_2.0	3/01/18	943,070	46	29	0	29
Polistes canadensis	Hymenoptera	NCBI	ASM131383v1	13/10/15	521,566	103	15	1	16
Polistes dominula	Hymenoptera	NCBI	Pdom r1.2	14/12/15	1,625,592	37	16	0	16
Bombyx mandarina	Lepidoptera	NCBI	ASM398793v1	31/12/18	2,789,315	43	14	1	15

Species	Order	Database	Assembly Name	Date[a]	N50[b]	L50[c]	FGM[d]	PGM[e]	TGM[f]
Bombyx mori	Lepidoptera	NCBI	ASM15162v1	28/04/08	4,008,358	38	15	0	15
Manduca sexta	Lepidoptera	NCBI	Msex_1.0	24/05/12	664,006	169	38	0	38
Operophtera brumata	Lepidoptera	NCBI	ASM126657v1	11/08/15	65,630	2,821	24	5	29
Hyphantria cunea	Lepidoptera	NCBI	ASM370950v1	1/11/18	1,126,406	126	21	2	23
Lymantria dispar dispar	Lepidoptera	NCBI	ASM411510v1	28/01/19	249,594	914	34	2	36
Agrotis ipsilon	Lepidoptera	NCBI	ASM419385v1	13/02/19	1,057,295	137	34	2	36
Helicoverpa armigera	Lepidoptera	NCBI	Harm_1.0	22/05/17	1,000,414	93	35	3	38
Helicoverpa zea	Lepidoptera	NCBI	Hzea_1.0	19/05/17	201,477	469	34	2	36
Heliothis virescens	Lepidoptera	NCBI	K63_refined_pacbio	27/09/17	102,124	1,134	37	5	42
Mamestra configurata	Lepidoptera	NCBI	ASM219265v2	2/08/19	207,752	582	25	3	28
Spodoptera litura	Lepidoptera	NCBI	ASM270686v2	25/10/17	915,465	143	45	1	46
Trichoplusia ni	Lepidoptera	NCBI	ASM360422v1	1/10/18	4,648,132	27	24	0	24
Cecropterus lyciades	Lepidoptera	NCBI	3311_assembly_v1_with_mito	20/02/18	558,064	280	19	9	28
Lerema accius	Lepidoptera	NCBI	lac_assembly_V1	2/09/15	525,349	160	21	2	23
Megathymus ursus violae	Lepidoptera	NCBI	mvi_v1	22/10/18	4,153,133	28	16	0	16
Calycopis cecrops	Lepidoptera	NCBI	3306_assembly_v2	21/04/16	233,537	852	43	11	54
Bicyclus anynana	Lepidoptera	NCBI	Bicyclus_anynana_v1.2	2/01/18	638,282	194	20	0	20
Danaus plexippus	Lepidoptera	NCBI	Dpv3	23/06/17	715,714	102	25	0	25
Heliconius erato demophoon	Lepidoptera	Lepbase	Heliconius erato demophoon v1	11/03/16	10,688,973	12	19	1	20
Heliconius erato lativitta	Lepidoptera	Lepbase	Heliconius erato lativitta v1	12/09/16	5,483,780	25	16	1	17
Heliconius melpomene	Lepidoptera	Lepbase	Heliconius melpomene melpomene Hmel2	17/07/15	2,102,720	34	19	0	19
Melitaea cinxia	Lepidoptera	NCBI	MelCinx1.0	9/07/14	119,328	970	16	7	23
Vanessa tameamea	Lepidoptera	NCBI	ASM293899v1	23/02/18	2,988,984	28	12	1	13
Papilio glaucus	Lepidoptera	NCBI	pgl_assembly_v1	23/02/15	230,841	418	21	0	21

Species	Order	Database	Assembly Name	Date[a]	N50[b]	L50[c]	FGM[d]	PGM[e]	TGM[f]
Papilio polytes	Lepidoptera	NCBI	Ppol_1.0	2/02/15	3,672,263	21	27	0	27
Papilio xuthus	Lepidoptera	NCBI	Pap_xu_1.0	28/09/15	3,432,602	22	26	0	26
Leptidea sinapis	Lepidoptera	NCBI	ASM90019944v1	11/10/17	112,092	1,270	15	1	16
Phoebis sennae	Lepidoptera	NCBI	3314_assembly_v1	10/03/16	299,140	267	12	2	14
Pieris napi	Lepidoptera	Lepbase	Pieris napi v1.1	13/02/17	12,597,868	13	15	3	18
Pieris rapae	Lepidoptera	NCBI	P_rapae_3842_assembly_v2	16/10/16	617,301	118	17	0	17
Calephelis nemesis	Lepidoptera	NCBI	Cne_v1	10/08/17	206,312	1,057	33	11	44
Calephelis virginiensis	Lepidoptera	NCBI	Cvi_v1	10/08/17	175,106	1,347	24	15	39
Chilo suppressalis	Lepidoptera	NCBI	ASM400044v1	8/01/19	1,753,049	119	19	2	21
Ostrinia furnacalis	Lepidoptera	NCBI	ASM419383v1	13/02/19	564,364	230	35	1	36
Amyelois transitella	Lepidoptera	NCBI	ASM118610v1	22/07/15	1,586,980	62	29	2	31
Galleria mellonella	Lepidoptera	NCBI	ASM258982v1	18/10/17	952,598	154	25	1	26
Plodia interpunctella	Lepidoptera	NCBI	plodia_v1	23/02/15	1,270,674	76	26	11	37
Cydia pomonella	Lepidoptera	NCBI	Cpom.V2	29/03/19	8,915,549	24	27	1	28
Plutella xylostella	Lepidoptera	NCBI	DBM_FJ_V1.1	2/10/14	737,182	155	17	0	17
Calopteryx splendens	Odonata	NCBI	Calsple1.0	14/04/17	422,252	1,013	23	7	30
Libellula fulva	Odonata	NCBI	Lful_2.0	31/10/17	1,183,267	297	29	1	30
Timema cristinae	Phasmatodea	NCBI	tcristinae_1.3c2	15/02/18	55,026,842	6	44	9	53
Pediculus humanus	Psocodea	NCBI	JCVI_LOUSE_1.0	23/04/07	497,057	67	12	0	12
Limnephilus lunatus	Trichoptera	NCBI	Llun_2.0	31/10/17	69,049	5,506	25	22	47
Catajapyx aquilonaris	Diplura	NCBI	Caqu_2.0	15/12/17	45,890	1,638	59	12	71
Oithona nana	Cyclopoida	NCBI	O_NANA_1	17/02/17	400,614	60	19	0	19
Hyalella azteca	Amphipoda	NCBI	Hazt_2.0.1	9/09/19	215,427	653	13	4	17

[a] Date submitted to database; [b] Scaffold N50 (or contig N50); [c] scaffold L50 (or contig L50); [d] full EcKL gene models; [e] partial EcKL gene models; [f] total EcKL gene models.

Chapter 4

Functional genetic analyses of detoxification candidate EcKLs in *Drosophila melanogaster*

4.1. Introduction

As seen in Chapters 2 and 3 of this thesis, multiple approaches can lead to the generation of detoxification gene candidates, even through methods that do not directly test the function of these genes. Ultimately, however, these candidates need to be validated by functional experiments to robustly demonstrate their involvement in detoxification processes.

Functional detoxification experiments can be broadly classified into two types: biochemical and genetic. Biochemical experiments of *in vitro* enzymatic activity are powerful because they directly test whether the protein product of a gene can catalyse a known or toxicologically plausible detoxification reaction. However, there are drawbacks to this approach: some proteins cannot be easily heterologously expressed for *in vitro* assays, and the demonstration of a particular reaction *in vitro* does not guarantee it has biological relevance *in vivo*. Functional genetic experiments nicely complement biochemical experiments by dealing with both of these drawbacks: in principle, genetic manipulation is achievable for any gene in a genome, particularly in model organisms; and genetic experiments can directly link a gene to higher-level phenotypes, typically physiological effects of a toxin such as mortality.

Genetic manipulation can be done in multiple ways. Overexpression (or misexpression) of a gene can lead to increased detoxification of a particular toxin and therefore increase the dose required to produce a specific effect (an 'effective dose'); likewise, gene disruption (knockdown, knockout or residue-specific substitutions) can lead to decreased detoxification and a consequent reduction in effective dose. The observation of both effects is strong evidence for the involvement of a candidate gene in the detoxification of the toxin used.

Ideally, combinations of genetic experiments are needed to fully understand how a particular gene works in detoxification processes, and the data generated must be interpreted carefully. For example, Cyp6g2 can detoxify many toxins when expressed ectopically (Daborn *et al.* 2007; Denecke *et al.* 2017b), but where it is expressed (the corpora allata, the site of juvenile hormone biosynthesis) likely precludes it from acting on xenobiotic substrates, and RNAi knockdown of the gene suggests it is involved

in juvenile hormone metabolism (Christesen *et al.* 2017; Chung *et al.* 2009), meaning it is unlikely to be a detoxification gene in the typical sense of the term.

In vitro biochemical assays of EcKL enzyme activity, at the time of writing, have been frustrated by problems with heterologous expression, leaving genetic experiments as the only way to test the functions of detoxification candidate EcKLs. Due to the identification of 24 such candidate genes in *Drosophila melanogaster* (Diptera: Drosophilidae) in Chapter 2, I decided to undertake the functional characterisation of EcKLs in this species.

4.1.1. Detoxification in *Drosophila*

From a practical perspective, *D. melanogaster* is well suited to use as a model to test the detoxification functions of candidate genes (reviewed by Scott & Buchon 2019). It has a high-quality, well-annotated reference genome assembly (Adams *et al.* 2000; Thurmond *et al.* 2019), and access to advanced gene disruption techniques, with publicly available RNAi, CRISPR-Cas9, TE-insertion and chromosomal deficiency libraries (Bellen *et al.* 2011; Dietzl *et al.* 2007; Ni *et al.* 2011; Port *et al.* 2020; Ryder *et al.* 2007); gene overexpression is very powerful in this species as well, with a large and ever-growing library of UAS and GAL4 lines that can target expression of native or transgenic ORFs in a temporally and spatially specific manner (Bischof *et al.* 2013; Manseau *et al.* 1997; Venken & Bellen 2014). Using these tools, multiple genes from 'classical' detoxification gene families have been linked to the detoxification of specific xenobiotic compounds in *D. melanogaster*, including P450s (Battlay *et al.* 2018; 2016; Daborn *et al.* 2007; Green *et al.* 2019; Najarro *et al.* 2015; Schmidt *et al.* 2017), GSTs (Gonzalez *et al.* 2018; Low *et al.* 2010) and UGTs (Highfill *et al.* 2017; Marriage *et al.* 2014). Candidate detoxification genes have been identified and validated from genome-wide association studies (GWAS) for naturally occurring compounds, such as boric acid, caffeine and methylmercury (Montgomery *et al.* 2014; Najarro *et al.* 2017; 2015), and synthetic insecticides, such as DDT, malathion and chlorantraniliprole (Battlay *et al.* 2018; Daborn *et al.* 2002; Green *et al.* 2019; Schmidt *et al.* 2017).

Detoxification in *D. melanogaster* is mainly thought to occur in the midgut, Malpighian tubules and the fat body (Yang *et al.* 2007; Chapter 1.2.5); it is also thought that the cells of the blood-brain barrier (glia surrounding the CNS) express detoxification

genes that protect the CNS from neurotoxic compounds present in the haemolymph (DeSalvo *et al.* 2014; Mayer *et al.* 2009), but xenobiotic metabolism in this tissue has only started to be studied biochemically, such as in the orthopteran *Schistocerca gregaria* (Hellman *et al.* 2016). As shown in Chapter 2 with the P450s, the tissue-specific expression patterns of genes in large enzyme families can be used to accurately predict which members may be involved in detoxification (Scanlan *et al.* 2020). *D. melanogaster* also has a stable gut-associated microbiome (Pais *et al.* 2018), although it is currently unclear if it contributes to the detoxification of dietary xenobiotic compounds.

The transcriptional regulation of detoxification genes has received some attention in *D. melanogaster*, with CncC- and DHR96-mediated pathways linked to this process (King-Jones *et al.* 2006; Misra *et al.* 2011). However, both pathways appear very general, with many hundreds of genes regulated by either pathway, and the CncC pathway likely responds more to elevated levels of reactive oxygen species (ROS) than unique toxic compounds (Wilding 2018). Pathways linked more specifically to the chemical ecology of *D. melanogaster* have yet to be identified.

Unfortunately, detoxicative phosphorylation is poorly studied in *D. melanogaster*, with only one xenobiotic compound—the hydroxylated β-carboline harmol—known to be phosphorylated. The *D. melanogaster* harmol phosphotransferase enzyme has only been characterised in a cursory manner, with little information known other than its cytosolic location (Baars *et al.* 1980), which is consistent with its hypothesised identity as an EcKL enzyme.

4.1.2. The chemical ecology of *Drosophila*

While detoxification is often an important process in the evolution of resistance to synthetic insecticides (Li *et al.* 2007), the natural environment has clearly played an overwhelmingly dominant role in the evolution of detoxification processes in all insects. As such, detoxification in *D. melanogaster* cannot be easily separated from its chemical ecology; the latter has shaped the detoxification breadth (DB; Chapter 1.2.1) of this species and likely defines the specific functions of many detoxification genes in the *D. melanogaster* genome.

Species in the genus *Drosophila* are largely saprophagous as both larvae and adult flies,

feeding on plant substrates in various states of decomposition (Markow & O'Grady 2007)—some, such as *D. suzukii* and *D. sechellia*, feed on fresh or recently fallen fruits, but require the presence of yeasts to complete development (Hamby & Becher 2016; Yassin *et al.* 2016), while others, such as *D. simulans* and *D. melanogaster*, can develop on highly decomposed fruit substrates that may contain high levels of microbial secondary metabolites and fermentation products (Kim *et al.* 2018b). This tolerance of a wide variety of food substrates with unpredictable chemical compositions makes *D. melanogaster* a dietary 'generalist', although there are some naturally occurring plant compounds to which it has a low tolerance, such as sugar alcohols (O'Donnell *et al.* 2016) and octanoic acid present in the *Morinda* (noni) fruit (Legal *et al.* 1992).

While *D. melanogaster* is likely able to detoxify many secondary metabolites present in fruits, a key source of dietary toxins for this species are microbes, especially filamentous fungi, which must compete with insect larvae for access to fruit substrates, as well as defend their hyphal networks against larval grazing (Rohlfs *et al.* 2005). These fungi, such as *Aspergillus* spp., synthesise and secrete a variety of insecticidal secondary metabolites to limit the foraging of *D. melanogaster* (Kempken & Rohlfs 2010; Nielsen *et al.* 2013; Rohlfs 2013; Trienens *et al.* 2010), and feeding on fungal-colonised substrates induces detoxification gene expression in larvae (Trienens *et al.* 2017). *D. melanogaster* larvae vary strongly in tolerance to the presence of different species of fungal competitors (Rohlfs *et al.* 2005; Trienens *et al.* 2010), suggesting their DB may encompass some but not all fungal toxins and that fungal competition may be exerting a selective pressure on DB in natural populations; this is further supported by experiments that show *D. melanogaster* can respond to selection for development on substrates colonised by *Aspergillus nidulans* (Wölfle *et al.* 2009). It is therefore likely that at least some detoxification enzymes encoded in the *D. melanogaster* genome are adapted to use fungal toxins as substrates. Bacterial secondary metabolites have likely also contributed to the evolution of *D. melanogaster*'s DB, although toxicological interactions between larvae and non-entomopathogenic bacteria are currently poorly understood.

4.1.3. Dro5 EcKLs in *Drosophila melanogaster*

In Chapter 2, 24 EcKL genes in *D. melanogaster* were identified as candidate detoxification genes using a 'detoxification score' method that integrated evolutionary stability, tissue-specific expression, xenobiotic induction and CncC regulation data (Scanlan

et al. 2020). Five of these 24 EcKLs belong to the Dro5 clade, part of the hyper-diverse Dip1 clade in Diptera and the H subfamily in insects (Chapter 3). As previously discussed (Chapter 2.7.2), these genes are attractive candidates to focus on with functional experiments, and the rapid expansion of this clade in the *Drosophila* genus suggests it may be adaptively linked to the detoxification of ecologically important xenobiotic compounds, such as fungal or plant toxins.

The seven Dro5 genes in *D. melanogaster* are located within a large, contiguous cluster of 26 subfamily H EcKLs on chromosome arm 3R—which contains 14 detoxification candidate genes (Fig. 4.1)—themselves grouped into two Dro5-only loci, 'Dro5A' and 'Dro5B'; the former contains *CG31300* (Dro5-1), *CG31104* (Dro5-2), *CG13658* (Dro5-5) and *CG11893* (Dro5-6) and the latter contains *CG13659* (Dro5-7), *CG31370* (Dro5-8) and *CG31436* (Dro5-10), with four other EcKLs in-between (*CG31098*, *CG31102*, *CG31097* and *CG31288*). While all seven *D. melanogaster* Dro5 genes are transcriptionally enriched in detoxification tissues (typically adult midguts and Malpighian tubules; Fig. S2.4), only five—*CG31300*, *CG31104*, *CG13658*, *CG11893* and *CG13659*—are detoxification candidates, as *CG31370* and *CG31436* are not induced in any xenobiotic datasets collated in Chapter 2. Notably, no Dro5 genes are positively regulated by CncC (Fig. 2.7), and only *CG13659* is appreciably expressed in the blood-brain barrier (enriched 18.8-fold in surface glia compared to the whole brain; DeSalvo *et al.* 2014), another site of detoxification.

Unfortunately, there are limited existing gene disruption and overexpression resources for the Dro5 clade. TE-insertion lines are available for only five genes (*CG31300*, *CG13658*, *CG11893*, *CG31370* and *CG31436*; Bellen *et al.* 2011), but a chromosomal deficiency—*Df(3R)BSC852* (Ryder *et al.* 2007)—deletes the gene body of six of the genes, while also deleting the promoter and transcription start site of *CG31436*, likely acting as a loss-of-function allele for all seven genes; a UAS-ORF line is only available for *CG31300* (Bischof *et al.* 2013). In order to investigate the functions of these genes in detoxification, additional transgenic constructs and transformed fly lines needed to be developed.

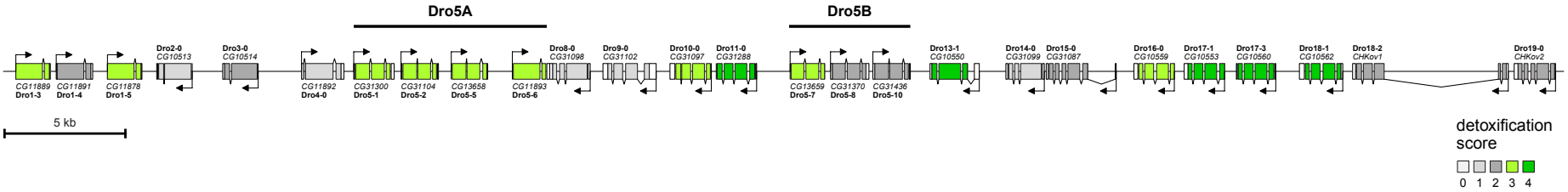


Figure 4.1. The chromosome 3R cluster of 26 subfamily H EcKLs in *Drosophila melanogaster*, with the Dro5A and Dro5B loci indicated with black lines. Exons of genes are coloured by that gene's 'detoxification score' (Chapter 2), whereby genes with scores of 3 and 4 are considered detoxification candidate genes; UTRs are white. *CG11889* (Dro1-3) and *CG11891* (Dro1-4) are part of a dicistronic transcription unit, which has not been shown for simplicity; likewise, only one transcript from each gene has been included to simplify the figure. *CHKov1* (Dro18-2) is depicted as in the *D. melanogaster* R6 reference genome, with the Doc1420 insertion producing the large intron (Aminetzach *et al.* 2005).

4.1.4. Plant and fungal toxins selected for experiments

In nature, *D. melanogaster* likely consumes many compounds of plant, fungal and bacterial origin that it has evolved to detoxify. However, in general there is little information about which specific toxins these are, and so a relatively broad selection of seven compounds were selected for single-toxin experiments in this chapter: kojic acid, salicin, esculin, escin, quercetin, curcumin and caffeine (Fig. 4.2). All toxins chosen were relatively inexpensive: some plausible candidate substrates for EcKLs in *D. melanogaster*, such as harmol and tunicamycin (Baars *et al.* 1980; Chow *et al.* 2013; see Chapter 2), were too expensive for the likely doses needed in experiments.

Six of these toxins were partially chosen because they contain one or more hydroxyl moieties, which are likely required to be an EcKL substrate. Kojic acid is synthesised and secreted by various species of filamentous fungi (El-Kady *et al.* 2014), and it has demonstrated toxicity against *D. melanogaster* and other insects (Beard & Walton 1969; Dobias *et al.* 1977). Salicin is an alcoholic glucoside synthesised by plants in the genus *Populus* and one of the few compounds known to be detoxified by phosphorylation in insects, which occurs in the gypsy moth *Lymantria dispar* (Boeckler *et al.* 2016). Quercetin is a flavonol that is found in a wide range of fruits and vegetables and is toxic to *D. melanogaster* in combination with other phytochemicals (Júnior *et al.* 2016). Escin is a mixture of saponins found in horse chestnut, *Aesculus hippocastanum*, and is cytotoxic to *D. melanogaster* cells (De Geyter *et al.* 2012), esculin is a coumarin glucoside also found in *A. hippocastanum*, and curcumin is a curcuminoid found in turmeric, *Curcuma longa*.

The last toxin, caffeine, is a methylxanthine alkaloid found in plant species from the genera *Coffea*, *Camellia*, *Theobroma*, *Paullinia*, *Cola*, *Ilex* and *Citrus* (Anaya *et al.* 2006) and is developmentally toxic to *D. melanogaster* (Li *et al.* 2013; Nigsch *et al.* 1977). Despite its lack of hydroxyl moieties, caffeine was selected for single-toxin experiments because two unpublished microarray datasets implicate Dro5 genes in the transcriptional response to caffeine consumption in *D. melanogaster*: Ran Zhuo, a PhD student at the University of Alberta, found five Dro5 genes induced by 8 mM (1,553 μg/mL) caffeine in 3rd-instar larvae after 4 hours (Zhuo 2014), while Jin Kee, a former MSc student in the Robin Lab, found three Dro5 genes induced by 1,500 μg/mL caffeine in

3rd-instar larvae after 4 hours (Kee & Robin, unpublished). These two datasets both found *CG31104* (Dro5-2), *CG13659* (Dro5-7), and *CG31436* (Dro5-10) induced by caffeine (albeit to different extents) and *CG31300* (Dro5-1) and *CG31370* (Dro5-8) not induced, while only Zhuo (2014) found *CG13658* and *CG11893* induced (Table 4.1). In addition, during the course of conducting the PheWAS analyses in Chapter 2, it was discovered that an allele of *CG31370* may be associated with a caffeine-related phenotype in the DGRP (see Chapter 4.3.3 & Table 4.1), giving further justification for including caffeine in the toxicological experiments in this chapter.

4.1.5. Chapter Aim

The aim of this chapter is to test the hypothesis that Dro5 EcKL genes in *D. melanogaster* function in the detoxification of xenobiotic compounds. To do this, a combination of gene disruption (CRISPR-Cas9 mutagenesis and complementation with existing gene knockout alleles) and gene overexpression (using an established GAL4 driver that targets detoxification tissues) experiments were performed with a selection of naturally occurring xenobiotic compounds and semi-natural citrus fruit substrates, using newly generated transgenic constructs and transformed fly lines.

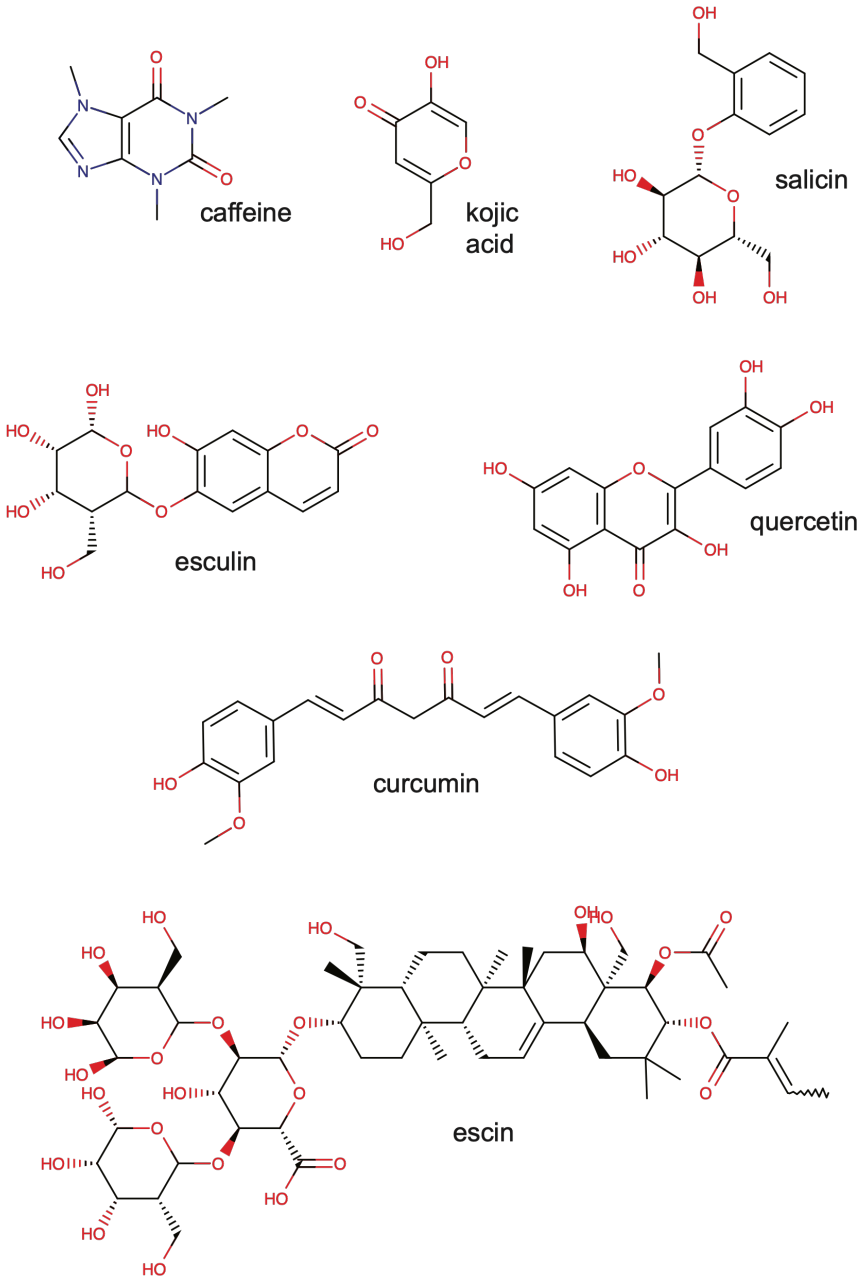


Figure 4.2. Chemical structures of the natural toxins used in toxicological experiments in this chapter. Curcumin is shown in its keto form.

Table 4.1. Transcriptional induction of Dro5 genes in 3rd-instar larvae of *Drosophila melanogaster* 4 hours after ingestion of caffeine, and PheWAS associations with caffeine phenotypes.

Nomenclature	Gene	Induction (Z)[a]	Induction (K)[b]	PheWAS[c]
Dro5-1	*CG31300*	-	-	-
Dro5-2	*CG31104*	36.7	2.9	-
Dro5-5	*CG13658*	2.7	-	-
Dro5-6	*CG11893*	2.3	-	-
Dro5-7	*CG13659*	16.6	3.0	-
Dro5-8	*CG31370*	-	-	+
Dro5-10	*CG31436*	2.6	1.7	-

[a] Positive fold-change in Zhuo (2014) caffeine dataset (8 mM/1.553 mg/mL caffeine); -, no significant fold-change

[b] Positive fold-change in Kee and Robin (unpublished) caffeine dataset (1.5 mg/mL caffeine); -, no significant fold-change

[c] +, variation in gene associated with adult survival on caffeine from Najarro *et al.* (2015); -, no association.

4.2. Materials and Methods

4.2.1. Fly lines and husbandry

The following fly lines were obtained from the Bloomington Drosophila Stock Center (BDSC): *CG31300*^{MB00063} (BL22688), *CG13658*^{MI03110} (BL37335), *CG11893*^{MB00360} (BL22775), *CG31370*^{MI07438} (BL44188), *CG31436*^{MI01111} (BL33107), *w*^{1118}; *Df(3R)BSC852*/TM6C, *Sb*^{1}, *cu*^{1} (BL27923), *w**;; *Sb*^{1}/TM3, *actGFP*, *Ser*^{1} (BL4534), *hsFLP*, *y*^{1}, *w*^{1118};; *nos-GAL4*, UAS-*Cas9* (BL54593), and *y*^{1}, *v*^{1}, *P{y*$^{+t7.7}$*=nos-phiC31\int.NLS}X*; *P{y*$^{+t7.7}$*=CaryP}attP40*; (BL25709). DGRP lines were also obtained from the BDSC. The *w*^{1118}; *Kr*$^{IF-1}$/CyO; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1} double-balancer line (also known as *w*^{1118}-DB), *6g1HR-6c-GAL4* (also known as HR-GAL4; Chung *et al.* 2007) and *tub-GAL4*/TM3, *actGFP*, *Ser*^{1} (likely derived from crosses between BL4534 and BL5138) were a kind gift of Philip Batterham and Trent Perry (The University of Melbourne).

Df(3R)BSC852/TM3, *actGFP*, *Ser*^{1} was made by crossing BL7923 to BL4534 and selecting the appropriate genotype. *w*^{1118}; *Kr*$^{If-1}$/CyO; *nos-GAL4*, UAS-*Cas9* was made by routine crosses, starting with BL54593 males and *w*^{1118}-DB females, until the desired genotype was achieved. *w*^{1118}; 25709; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1} (chr2 isogenic to BL25709) was made by routine crosses, starting with BL25709 males and *w*^{1118}-DB females, until the desired genotype was achieved. *w*^{1118}; 25709; *Sb*^{1}/TM3, *actGFP*, *Ser*^{1} (chr2 isogenic to BL25709) was made by routine crosses, starting with BL25709 males and *w*^{1118}; *Kr*$^{IF-1}$/CyO; *Sb*^{1}/TM3, *actGFP*, *Ser*^{1} females (which themselves were made by routine crosses beginning with BL4534 males and *w*^{1118}-DB females), until the desired genotype was achieved.

For routine stock maintenance, flies were kept on yeast-cornmeal-molasses media ('lab media'; http://bdsc.indiana.edu/information/recipes/molassesfood.html) at 18 ºC, 21 ºC or 25 ºC in plastic vials sealed with cotton stoppers. Bioassays that were analysed together (each represented by a different figure or sub-figure in the results) were carried out as a group on the same batch of media at the same time to minimise intra-experiment batch effects. All experiments were carried out at 25 ºC.

4.2.2. Generation of UAS-ORF lines

cDNA clones for *CG31300* (FI01822), *CG31104* (IP12282), *CG13658* (FI12013), *CG11893* (IP11926), *CG13659* (IP11858), *CG31370* (IP10876) and *CG31436* (IP12392) were obtained from the Drosophila Genomics Resource Center (DGRC) on Whatman filter paper and processed according to DGRC instructions to produce pure plasmids. Recombinant pUASTattB plasmids (Bischof *et al.* 2007) containing EcKL ORFs under the control of a UAS promoter were designed *in silico* using Benchling (http://benchling.com).

The pUASTattB vector was digested with EagI-HF (NEB) and KpnI-HF (NEB) and the 8.5 kb backbone gel-purified (28704, Qiagen). ORFs were amplified with PCR—10 sec initial denaturation (98 ºC), then 5 sec denaturation (98 ºC), 5 sec annealing (55 ºC) and 15 sec extension (72 ºC) for 32 cycles, then a final 1 min extension (72 ºC)—using Phusion Flash polymerase (NEB) from DGRC cDNA clones using primers containing an EagI restriction site (forward primers) or a KpnI restriction site (reverse primers), as well as an additional 5' sequence (5'-TAAGCA-3') to aid digestion (Table 4.2). Amplicons were column-purified (FAPCK 001, Favorgen), double-digested with EagI-HF and KpnI-HF for 8 hr, then gel-purified. EagI/KpnI-digested ORFs were ligated into the EagI/KpnI-digested pUASTattB vector backbone using a 6:1 insert:vector molar ratio and T4 DNA ligase (M0202S, NEB) in a thermocycler overnight (~16 hr), alternating between 10 ºC for 30 sec and 30 ºC for 30 sec (Lund *et al.* 1996). 5 μL of each 20 μL ligation reaction was used to transform DH5-alpha *E. coli* (C2987H, NEB), which were plated on LB+amp (100 μg/mL ampicillin) agar plates and grown overnight at 37 ºC. Using a sterile micropipette tip, each transformant colony was transferred to an LB+amp agar master plate and then directly screened with colony PCR—2 min initial denaturation (95 ºC), then 2 min denaturation (95 ºC), 45 sec annealing (58 ºC) and 1.5 min extension (72 ºC) for 32 cycles, then a 5 min final extension (72 ºC)—for successful integration using GoTaq Green Master Mix (M712, Promega) and the pUASTattB_3F/5R primers (Table 4.2), with expected amplicon sizes of 1,498 bp (*CG31300*), 1,497 bp (*CG31104*), 1,497 bp (*CG13658*), 1,480 bp (*CG11893*), 1,482 bp (*CG13659*), 1,419 bp (*CG31370*) and 1,495 bp (*CG31436*). Positive colonies were selected from the master plate and cultured overnight in LB+amp; plasmids were isolated with the FavorPrep Plasmid DNA Extraction Mini Kit (FAPDE 100, Favorgen) and Sanger sequenced using the pUASTattB_3F/5R primers at the Australian Genome Research

Facility (AGRF).

Correctly assembled plasmids were sent to TheBestGene Inc. (US) for microinjection and incorporation into the *D. melanogaster* genome at the attP40 site on chromosome 2 (BL25709). Transformed lines were received as a mixture of white-eyed (zero copies of pUASTattB), orange-eyed (one copy of pUASTattB) and red-eyed (two copies of pUASTattB) flies—virgin white-eyed flies from all seven lines were pooled and retained as a common genetic background line ('*yw*'), while the pUASTattB-transformed lines were individually kept as red-eyed homozygous stocks.

4.2.3. Generation of CRISPR-Cas9 mutant lines

gRNAs were designed with the CRISPR Optimal Target Finder tool (Gratz *et al.* 2014) to have a guide length of 20 nt, with the stringency set to 'maximum' to minimise off-target mutagenesis elsewhere in the genome. The pCFD6 vector was a gift from Simon Bullock (Addgene plasmid #73915; http://n2t.net/addgene:73915; RRID:Addgene_73915). The recombinant pCFD6 plasmids 'pCFD6-Dro5A' and 'pCFD6-Dro5B', each of which express—under the control of a UAS promoter—four gRNAs that target either the Dro5A or Dro5B locus (Fig. 4.4), were designed *in silico* using Benchling (http://benchling.com), and cloned following the protocol laid out in Port & Bullock (2016), with minor modifications below.

pCFD6 was digested with BbsI-HF (NEB) and the 9.4 kb backbone gel-purified. The intact pCFD6 vector was used as a PCR template for the production of the three overlapping gRNA-containing inserts (insert 1, 233 bp; insert 2, 204 bp; insert 3, 234 bp) per recombinant plasmid, using pairs of primers (pCFD6-Dro5A: pCFD6_D5ΔA_1F/R, pCFD6_D5ΔA_2F/R and pCFD6_D5ΔA_3F/R; pCFD6-Dro5B: pCFD6_D5ΔB_1F/R, pCFD6_D5ΔB_2F/R and pCFD6_D5ΔB_3F/R). Inserts were amplified with Phusion Flash polymerase—a ramped PCR cycle, starting with 20 sec initial denaturation (98 °C), then 1 sec denaturation (98 °C), 5 sec annealing (61 °C, increasing 0.5 °C each cycle) and 15 sec extension (72 °C) for 22 cycles, then 18 cycles without an annealing step, and a final 1 min extension (72 °C)—and gel-purified. Purified inserts were cloned into the digested pCFD6 backbone using Gibson assembly (E5520S, NEB) according to the manufacturer's instructions, except with a 3:1 molar ratio of each insert to vector and 0.3 pmol of total DNA per reaction, with a 4 hr

incubation time. 2 μL of each 20 μL assembly reaction was used to transform DH5-alpha *E. coli*, which were plated on LB+amp agar plates and grown overnight at 37 ºC. Using a sterile micropipette tip, each transformant colony was transferred to an LB+amp agar master plate and then directly screened with colony PCR—2 min initial denaturation (95 ºC), then 2 min denaturation (95 ºC), 45 sec annealing (58 ºC) and 1 min extension (72 ºC) for 32 cycles, then a 5 min final extension (72 ºC)—for successful assembly using GoTaq Green Master Mix and the pCFD6_seqfwd and pCFD6_seqrev primers (Table 4.2), with an expected amplicon size of 890 bp for both plasmids. Positive colonies were selected from the master plate and cultured overnight in LB+amp; plasmids were isolated with the FavorPrep Plasmid DNA Extraction Mini Kit and Sanger sequenced using the pCFD6_seqfwd and pCFD6_seqrev primers at AGRF.

Correctly assembled plasmids were sent to TheBestGene Inc. for microinjection and incorporation into the *D. melanogaster* genome at the attP40 site on chromosome 2 (BL25709), to produce the homozygous fly lines 'pCFD6Dro5A' and pCFD6Dro5B' (of genotype w^-, 25709; *pCFD6*; 25709).

Transgenic CRISPR-Cas9 mutagenesis of wild-type chromosomes was performed with the crossing scheme in Figure 4.3A, which can be adapted for any type of transgenic mutagenesis on chromosome 3. Single founder male flies—which were heterozygous for possibly mutagenised loci on chromosome 3—were allowed to mate with w^{1118}; 25709; Sb^1/TM3, *actGFP*, Ser^1 virgin females, and when larvae were observed in the food media, the DNA from each founder male was extracted using the squish prep protocol (Appendix 3.1) and PCR-genotyped as below.

Mutagenesis of already-mutagenised chromosomes was performed with the crossing scheme in Figure 4.3B, using homozygous mutant lines generated previously. Single founder male flies—which were heterozygous for possibly (singly- or doubly-) mutagenised loci on chromosome 3—were allowed to mate with w^{1118}; 25709; Sb^1/TM3, *actGFP*, Ser^1 virgin females, and when larvae were observed in the food media, the DNA from each founder male was extracted using the squish prep protocol (Appendix 3.1), then genotyped using primers for both the Dro5A and Dro5B loci.

Deletion loci from mutant flies (Fig. 4.3, blue boxes) were amplified using the genotyping primers above with four GoTaq Green PCR reactions per line, which were

combined before gel-purification to allow for the detection of early-cycle polymerase-derived errors by close inspection of the sequencing chromatogram output. Dro5A PCR genotyping used the primer pairs D5ΔA_1F/1R and D5ΔA_2F/2R, which produce amplicons of size 910 bp and 963 bp, respectively, from wild-type chromosomes, and D5ΔA_1F/2R, which produce (non-amplifying) 8,025 bp amplicons from wild-type chromosomes and amplicons between ~300–1,400 bp in the case of deletions between Dro5A gRNA pairs 1–3, 1–4, 2–3 or 2–4. Dro5B PCR genotyping used the D5ΔB_1F/1R and D5ΔB_2F/2R primer pairs, which produce amplicons of size 859 bp and 792 bp, respectively from wild-type chromosomes, and D5ΔB_1F/2R, which produce (non-amplifying) 4,654 bp amplicons from wild-type chromosomes and amplicons between ~600–1,200 bp in the case of deletions between Dro5B gRNA pairs 1–3, 1–4, 2–3 or 2–4. PCR—2 min initial denaturation (95 ºC), then 2 min denaturation (95 ºC), 45 sec annealing (55 ºC) and 1.5 min extension (72 ºC) for 32 cycles, then a 5 min final extension (72 ºC)—was carried out with GoTaq Green Master Mix. Gel-purified amplicons were sequenced using the appropriate genotyping primers at AGRF.

Genetic background flies (*w*^{1118}; 25709; 25709—otherwise known as '+') were generated by following the crossing scheme in Figure 4.3A, but using BL25709 as the maternal genotype in C1 instead of pCFD6Dro5A or pCFD6Dro5B.

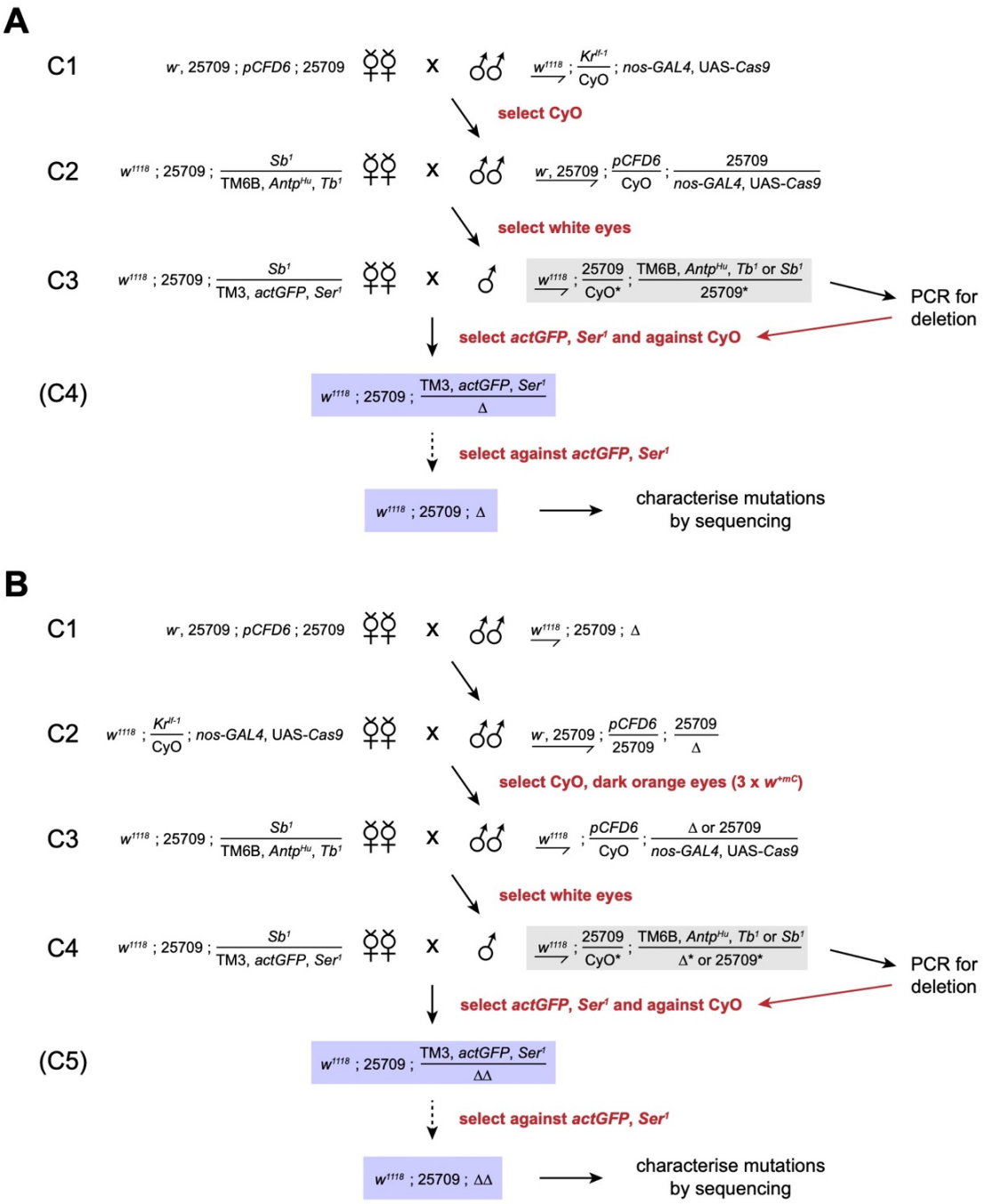


Figure 4.3. Crossing schemes for CRISPR-Cas9 mutagenesis on chromosome 3 (chr3) using *pCFD6*-transformed flies. (A) Initial mutagenesis of a wild-type chr3 locus in the BL25709 genetic background to produce a deletion allele (Δ). The single males used in C3 (grey box) are 'founder males' of each potential mutant line. (B) Mutagenesis of an already-mutagenised chr3 (Δ) to produce a double mutant line (ΔΔ). In this scheme, founder males (grey box) are the single males used in C4. Males used in C3 are selected by the colour of their eyes—as the *pCFD6*, *nos-GAL4* and UAS-*Cas9* constructs all contain a mini-white gene (w^{+mC}) that produces orange eyes in a *w-* background, individuals that inherit all three transgenic constructs (i.e. mutagenic males) can be distinguished from those that only inherit only two (*nos-GAL4* and UAS-*Cas9*). The PCR step after C4 (scheme B) needs to check for the presence of both the initial mutation as well as any new mutations produced. Dashed arrows indicate a possible homozygosing step (if the alleles generated are homozygous-viable). C3 in (A) and C4 in (B) can use either TM3, *actGFP*, Ser^1 or Sb^1 males as founders, in order to double the number of potential mutant lines that can be generated from the cross.

4.2.4. DGRP genotyping and statistical analyses

The PCR primers CG31370del_1F and CG31370del_1R (Table 4.2) were designed to flank the *CG31370*^{del} region and produce a 576 bp amplicon from *CG31370*^{wt} and a 392 bp amplicon from *CG31370*^{del}. DNA was extracted from single flies (Appendix 3.1), with three independent extractions per DGRP line. PCR—2 min initial denaturation (95 ºC), then 30 sec denaturation (95 ºC), 30 sec annealing (53 ºC) and 40 sec extension (72 ºC) for 30 cycles, then a 5 min final extension (72 ºC)—was carried out with GoTaq Green Master Mix, using 0.4 μL of DNA extract as a template per 10 μL reaction.

The preliminary PheWAS analysis was conducted as described in Chapter 2.3.3. Adult caffeine survival data was obtained from Najarro *et al.* (2015). Developmental caffeine (388 μg/mL) survival data from Montgomery *et al.* (2014) was averaged across the three replicates, and then corrected for (similarly averaged) control (0 μg/mL caffeine) survival using Abbott's formula (Abbott 1925), with corrected survival values greater than 1 (indicating greater survival than control) adjusted to 1. DGRP basal gene expression in adult female and adult male flies was obtained from Everett *et al.* (2020). Mean differences in phenotypes and $\log_2$(FPKM) between *CG31370* genotypes were determined with the *dabestr* package (Ho *et al.* 2019) in R; effect sizes with 95% confidence intervals that did not include zero were considered significant.

4.2.5. Lab media developmental viability assays

Egg-to-adult viability was estimated from the adult genotypic ratios of successfully eclosing offspring produced from crosses between a homozygous parental genotype and a heterozygous parental genotype, the latter of which had at least one phenotypic marker that revealed the genotype of the offspring. Males and females of the relevant genotypes were allowed to mate and lay eggs on vials of lab media, with at least five vials per cross, and the number of adults of each genotype were scored after development at 25 ºC for 14 days. If the adult genotypic ratio was significantly different from the Mendelian expectation (1:1), as determined by the 'binom.test' function in R, this was considered evidence that one genotype was less viable than the other.

Larval-to-adult viability was estimated by transferring particular quantities of 1st-instar larvae of known genotypes (either as the offspring of a cross between homozygous parents, or offspring sorted phenotypically by the presence or absence of a

dominant marker such as GFP expression) to vials of lab media, letting them develop at 25 ºC for 14 days, and scoring the number of individuals that reached pupariation, pupation, pharate adult development and eclosion. Fisher's exact test was used to determine if there were significant differences between genotypes, using the 'fisher.test' function in R.

4.2.6. Single-dose developmental toxicology assays

Quercetin, escin, esculin and curcumin were purchased from Sigma-Aldrich. Toxin-containing food media was prepared by adding 100 μL of toxin stock solution—quercetin, escin, esculin or curcumin dissolved in 100% EtOH—to 5mL of molten yeast-sucrose media (Appendix 3.3) in each vial and mixing with a clean plastic rod. Control food was made the same way, but with the addition of 100 μL of EtOH instead of toxin stock solution. Food was stored at 4 ºC for a maximum of three days before use to minimise toxin degradation.

Dro5$^{A3\text{-}B7}$ females were mated to *Dro5*$^{A3\text{-}B7}$ or wild-type (+; the genetic background of the CRISPR-Cas9 mutagenesis lines) males and were allowed to lay on juice plates (Appendix 3.2) topped with yeast paste. After hatching, 20 1st-instar larvae were transferred to each toxin food vial using a fine paintbrush that was washed between each transfer, and left to develop at 25 ºC for 14 days. Vials were scored for the number of individuals that had pupated (formation of the puparium) and that had successfully eclosed (complete vacation of the puparium). Mortality counts were determined as 'larval' (# of larvae – # of pupae) or 'pre-adult' (# of larvae – # of adults eclosed), and proportional mortality was calculated by dividing mortality counts by the number of larvae added to each vial.

Data manipulation and statistical analyses were performed in R (R Core Team 2019). Mean differences in proportional mortality between the two genotypes on each type of media was analysed with Welch's t-test ('t.test' function in R).

4.2.7. Dose-response developmental toxicology assays

Kojic acid, salicin and caffeine were purchased from Sigma-Aldrich. Toxin-containing food media was prepared by adding toxin stock solution—kojic acid, salicin or caffeine dissolved in dH_2O—to molten 1.25x yeast-sucrose media (made with 80% of the

volume of dH_2O, but the same quantity of solutes; Appendix 3.3): X μL of 40–50 mg/mL toxin stock solution, 1000–X μL of dH_2O (where X varied according to the final concentration of toxin) and 4 mL of 1.25x food media were added to each vial, for a total food media volume of 5 mL, then mixed with a clean plastic rod. Control food was made by mixing 4 mL of molten 1.25x media and 1 mL of dH_2O. Food was stored at 4 ºC for a maximum of three days before use to minimise toxin degradation.

Males and females of the relevant genotypes were crossed, and females were allowed to lay on juice plates (Appendix 3.2) topped with yeast paste. 20–30 1st-instar larvae were transferred to each toxin-containing food vial using a fine paintbrush that was washed between each transfer, and left to develop at 25 ºC for 14 days. Crosses involving GFP-marked balancer chromosomes had larvae sorted against GFP under a bright-field fluorescent microscope before transferal. Vials were scored for the number of individuals that had pupated (formation of the puparium) and that had successfully eclosed (complete vacation of the puparium). Three types of mortality counts were calculated: 'larval' (# of larvae – # of pupae), 'pupal' (# of pupae – # of adults eclosed) or 'total' (# of larvae – # of adults eclosed). Mortality counts were converted to proportional mortality by dividing by the number of larvae per vial, except in the case of pupal mortality counts, which were converted by dividing by the number of pupae; if the number of pupae was zero for a dose, the proportional pupal mortality count was set to 1 to avoid missing data (as doses high enough to prevent successful pupation are by definition enough to prevent successful eclosion).

Data manipulation and statistical analyses were performed in R. Proportional mortality values were corrected for control mortality using Schneider-Orelli's formula (Puntener 1981), with negative values after correction adjusted to 0. Calculation of LC_{50} (median lethal concentration) values and their 95% confidence intervals was performed with the 'LC_probit' function in the *ecotox* package (v1.4.2), with model weights as number of larvae added to the relevant vial. Statistical comparison of LC_{50} values between genotypes was performed with the 'comped' function in the *drc* package (v3.0-1; Ritz *et al.* 2015), which implements the ratio test devised by Wheeler *et al.* (2006); if the ratio of two LC_{50}s had a 95% confidence interval (CI) that included 1, the two LC_{50}s were not considered significantly different.

4.2.8. Citrus-based media developmental assays

'Delite' mandarin oranges (*Citrus reticulata*), 'Ruby Blush' grapefruits (*C.* × *paradisi*) and navel sweet oranges (*C.* × *sinensis*) were juiced with a hand juicer, juice was strained to remove large pulp particles, and yeast, agar and dH_2O were added and heated in a microwave. After cooling to 60 ºC, 10% Tegosept in EtOH was added, and 5 mL of media was aliquoted into each vial. Final concentrations of yeast and agar were 5% and 1% w/v, respectively, and 0.174% and 1.65% v/v for Tegosept and EtOH, respectively (5 g yeast, 1 g agar, 20 mL dH_2O and 80 mL juice for 100 mL of media). Propionic acid and orthophosphoric acid were not added due to the already low pH of the juices. 30 1st-instar larvae were transferred to each fruit media vial using a fine paintbrush and left to develop at 25 ºC for 14 days. Vials were each scored for the number of larvae that had pupated (formation of the puparium) and that had successfully eclosed (complete vacation of the puparium), and proportional survival was calculated by dividing by the number of larvae added to each vial. Mean differences in proportional mortality between the two genotypes on each type of media was analysed with Welch's t-test ('t.test' function in R).

Table 4.2. Primer sequences used in Chapter 4. All oligonucleotides were synthesised by Integrated DNA Technologies (IDT).

Primer ID	Sequence (5' to 3')
pUASTattB_3F	CGCAGCTGAACAAGCTAAAC
pUASTattB_5R	TGTCACACCACAGAAGTAAGG
CG31300_EagIF	TAAGCACGGCCGATGACTGACAAGTTAGATGC
CG31300_KpnIR	TAAGCAGGTACCGCTATAGACATTTAAAGTAGCC
CG31104_EagIF	TAAGCACGGCCGGAAAATGGAAGGCAAAAATATTG
CG31104_KpnIR	TAAGCAGGTACCCATTATAGATCCTTAAAGTATCC
CG13658_EagIF	TAAGCACGGCCGATGGCGGAAAACGTAGATTC
CG13658_KpnIR	TAAGCAGGTACCTTAAAGATCTTTAAAATATCCCAG
CG11893_EagIF	TAAGCACGGCCGATGCCAGAAAACGCAGATAC
CG11893_KpnIR	TAAGCAGGTACCATCAAAGATCGTTAAAGTATCCC
CG13659_EagIF	TAAGCACGGCCGATGGCCGAGGAAAGTTTC
CG13659_KpnIR	TAAGCAGGTACCTTAAAAGTCGTCAAAATATCCCG
CG31370_EagIF	TAAGCACGGCCGATGGCTGAAGATAGCTTAGC
CG31370_KpnIR	TAAGCAGGTACCTTATAAGTTCTCAAAATATCCAG
CG31436_EagIF	TAAGCACGGCCGATGTCCGGGAACCCCCAAAAC
CG31436_KpnIR	TAAGCAGGTACCATGTTTAGGCATGGAGTAATCCC
pCFD6_D5ΔA_1F	CGGCCCGGGTTCGATTCCCGGCCGATGCATCAGTTGTAACCTCTAAGGTG TTTCAGAGCTATGCTGGAAAC
pCFD6_D5ΔA_1R	TCGCGGTGGTACACTCTGCATGCACCAGCCGGGAATCGAACC
pCFD6_D5ΔA_2F	TGCAGAGTGTACCACCGCGAGTTTCAGAGCTATGCTGGAAAC
pCFD6_D5ΔA_2R	ACATAATAGAAGGCATTTCCTGCACCAGCCGGGAATCGAACC
pCFD6_D5ΔA_3F	GGAAATGCCTTCTATTATGTGTTTCAGAGCTATGCTGGAAAC
pCFD6_D5ΔA_3R	ATTTTAACTTGCTATTTCTAGCTCTAAAACCTATGACCCTTATGTTCAAGTGC ACCAGCCGGGAATCGAACC
pCFD6_D5ΔB_1F	CGGCCCGGGTTCGATTCCCGGCCGATGCAAATCGGTTGAACACGTATATG TTTCAGAGCTATGCTGGAAAC
pCFD6_D5ΔB_1R	CCCTAGCGCGAAACATAATGTGCACCAGCCGGGAATCGAACC
pCFD6_D5ΔB_2F	CATTATGTTTCGCGCTAGGGGTTTCAGAGCTATGCTGGAAAC
pCFD6_D5ΔB_2R	TGTCATCGCCGACCTGTCGATGCACCAGCCGGGAATCGAACC
pCFD6_D5ΔB_3F	TCGACAGGTCGGCGATGACAGTTTCAGAGCTATGCTGGAAAC
pCFD6_D5ΔB_3R	ATTTTAACTTGCTATTTCTAGCTCTAAAACGTCCATGGGTGTACGACTCTTG CACCAGCCGGGAATCGAACC
pCFD6_seqfwd	GTAGACATCAAGCATCGGTGG
pCFD6_seqrev	TTAGAGCTTTAAATCTCTGTAGGTAG
D5ΔA_1F	GATGGGTCATTCTGACACCGA
D5ΔA_1R	TTCTTCCTGAGCAACCGGAC
D5ΔA_2F	GAGCCTCGGCAGGTGTTAAT
D5ΔA_2R	TGCGATCAATTAGCCATGCAA
D5ΔB_1F	CTGATCCGTTTGCAGACACT
D5ΔB_1R	CTTGGAGTAGGCACTGCTGAT
D5ΔB_2F	ACCAACCGAAAAGGCGAGTT
D5ΔB_2R	TCCGGCTCCAAAAGCATGTAA
CG31370del_1F	GCTGAATGTCCCAGAATGGT
CG31370del_1R	TCCTTAACGAATTCTGGTCGCT

4.3. Results

4.3.1. CRISPR-Cas9 mutagenesis of the Dro5 EcKL clade

As the seven Dro5 EcKLs in *D. melanogaster* were all derived from a single ancestral gene within the last 40–50 million years (Chapter 2), it is possible that there may be some functional redundancy between some or all of the genes in this clade. Because of this, a CRISPR-Cas9 mutagenesis strategy was designed that aimed to generate full seven-gene deletion alleles of both Dro5A and Dro5B, which could then be placed in *trans* with other mutant alleles to explore the functions of single or multiple genes in the Dro5 clade. Dro5A and Dro5B are separated by four distantly related EcKLs (Fig. 4.4), preventing the single-step generation of 'Dro5-null' alleles—as such, I aimed to generate them by sequential mutagenesis: creating Dro5A or Dro5B deletion alleles (Fig. 4.3A) and then passing them through a second round of mutagenesis to disrupt the other region (Fig. 4.3B). Some alleles generated through this method (either single or double mutagenesis) appeared unable to be homozygosed, suggesting their chromosomes may have harboured off-target, lethal mutations; these alleles were not used in functional experiments.

Dro5A deletion was highly successful, with eight 'large' deletion alleles (putatively generated by cuts between Ag1/2 and Ag3/4; Fig. 4.4) detected through PCR from screening 24 founder males, including the $Dro5^{A3}$ allele, which contained a 7,690 bp deletion between Ag1 and Ag4 (Fig. 4.4). 40 founder males were screened for mutations at the Dro5B locus using PCR, but no large deletions encompassing the entire Dro5B locus were detected. Unfortunately, it appears that the 3rd and 4th gRNAs (Bg3 and Bg4; Fig. 4.4) from the pCFD6-Dro5B construct failed to cut, given the lack of heteroduplex bands generated after PCR with the D5ΔB_2F/D5ΔB_2R primer pair (Fig. S4.1). Mutagenesis with the 1st and 2nd gRNAs (Bg1 and Bg2; Fig. 4.4) was more successful; three alleles of *CG13659* with lesions around the 1st and 2nd gRNA target sites were detected by PCR, sequenced and named $CG13659^{19}$ (a 63 bp deletion 107 bp upstream of the transcription start site), $CG13659^{36}$ (a 38 bp deletion 16 bp upstream of the transcription start site) and $CG13659^{38}$ (a composite deletion allele consisting of a 2 bp deletion 43 bp upstream of the transcription start site and a 241 bp deletion in the first exon that deleted 81 aa and causes a frameshift in the CDS; Fig. 4.4A). Only

CG13659^{38}—likely a strong loss-of-function allele that was homozygous-viable—was used in functional experiments.

The Dro5A allele *Dro5*^{A3} was selected for a second round of mutagenesis to produce additional deletions at the Dro5B locus. 16 founder males were screened, with two putative deletions detected at the *CG13659* locus, but none across the Dro5B locus as a whole. The main allele phenotypically characterised was *Dro5*$^{A3\text{-}B7}$, which is a homozygous-viable composite allele consisting of two deletions, one 7,690 bp long in Dro5A and one 153 bp long in Dro5B, completely deleting *CG31300*, *CG31104* and *CG13658*, and deleting 1,245 bp (415 aa) of the CDS of *CG11893* and the first 85 bp (28 aa) of *CG13659*, including the transcription and translation start sites of both genes (Fig. 4.4).

4.3.2. Developmental viability of Dro5 knockout and misexpression

Loss-of-function alleles, either pre-existing (TE-insertion) or generated in this study (Chapter 4.3.2), were placed in *trans* with the homozygous-lethal chromosomal deficiency *Df(3R)BSC852*, which deletes or otherwise disrupts all seven Dro5 genes in *D. melanogaster*. Loss of individual Dro5 genes, or five genes in the case of the *Dro5*$^{A3\text{-}B7}$ allele, did not significantly affect adult genotypic ratios, suggesting Dro5 genes are generally not required for development (Fig. 4.5A). A larval-to-adult viability experiment involving just the *Dro5*$^{A3\text{-}B7}$ allele further supported this conclusion, with the vast majority of *Dro5*$^{A3\text{-}B7}$/*Df(3R)BSC852* individuals successfully completing development, and no significant difference between the developmental outcomes of *Dro5*$^{A3\text{-}B7}$/*Df(3R)BSC852* and *Dro5*$^{A3\text{-}B7}$/TM3, *actGFP*, *Ser*^{1} animals ($p = 0.56$, Fisher's exact test; Fig. 4.5B). However, it is possible that all seven genes share a redundant developmental function and that disruption of all seven Dro5 genes would result in developmental arrest, but this was not able to be tested due to the inability to generate a complete Dro5 null allele using CRISPR-Cas9 mutagenesis (Chapter 4.3.2).

All seven Dro5 genes were individually misexpressed with the strong, ubiquitous GAL4 driver *tub-GAL4*—misexpression of *CG31104* and *CG13658* resulted in no or very few successfully eclosing adults (Fig. 4.6), suggesting ectopic or excessive expression of either gene arrests development. Examination of *tub>CG31104* and *tub>CG13658* animals, using brightfield fluorescence microscopy to select against GFP-positive individuals, revealed that these genotypes are arrested during

metamorphosis—pharate adults have completely undifferentiated abdomens, which lack bristles and genitalia. Misexpression of the other five Dro5 genes did not significantly change adult genotypic ratios (Fig. 4.6) and therefore does not appear to grossly affect developmental progression.

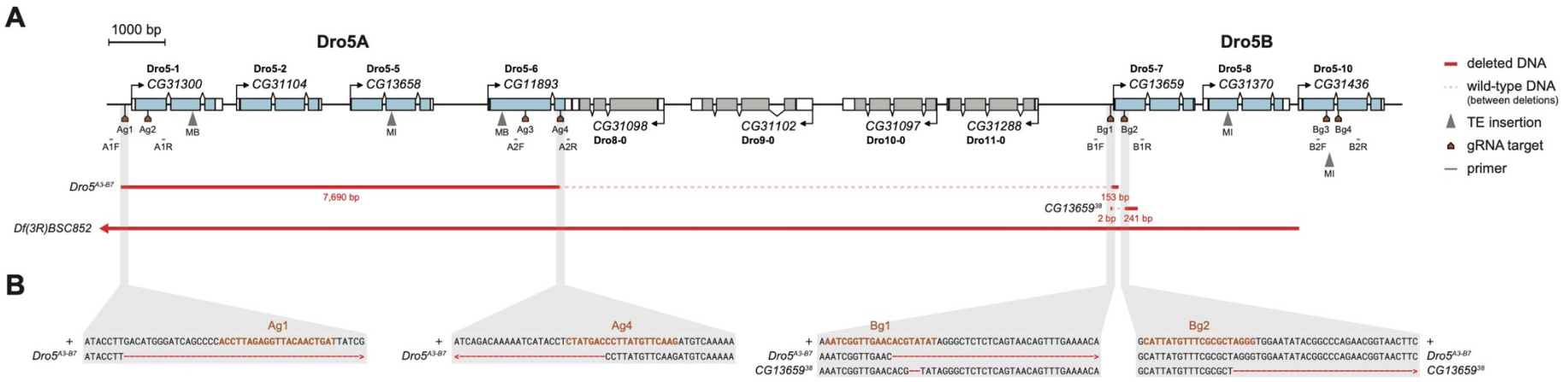


Figure 4.4. (A) The Dro5 locus on chromosome 3R of *Drosophila melanogaster*, including the seven Dro5 genes (genes with blue exons) in two clusters (Dro5A and Dro5B) and four other EcKLs (genes with grey exons; Dro8-0, Dro9-0, Dro10-0 and Dro11-0). Mapped onto the locus are the locations of gRNA target sites (Ag1–4 and Bg1–4; brown), genotyping primer-binding sites (purple), TE insertion sites (grey triangles), and two CRISPR-Cas9-induced composite deletion alleles (*Dro5*$^{A3\text{-}B7}$ and *CG13659*^{38}; deletion sizes in red text) and the chromosomal deficiency *Df(3R)BSC852* (red lines). Wild-type DNA present in-between deletions is indicated with pale dashed lines. MB, Mi{ET1} element; MI, Mi{MIC} element; A1F, D5ΔA_1F; A1R, D5ΔA_1R; A2F, D5ΔA_2F; A2R, D5ΔA_2R; B1F, D5ΔB_1F; B1R, D5ΔB_1R; B2F, D5ΔB_2F; B2R, D5ΔB_2R. (B) Sequence-level detail (grey boxes) of the deletions comprising the CRISPR alleles *Dro5*$^{A3\text{-}B7}$ and *CG13659*^{38}, with respect to their wild-type genetic background (+). gRNA target sites are highlighted in brown, and deleted bases are red dashes; > and < symbols indicate that the deletion continues out of the frame of the highlighted sequence.

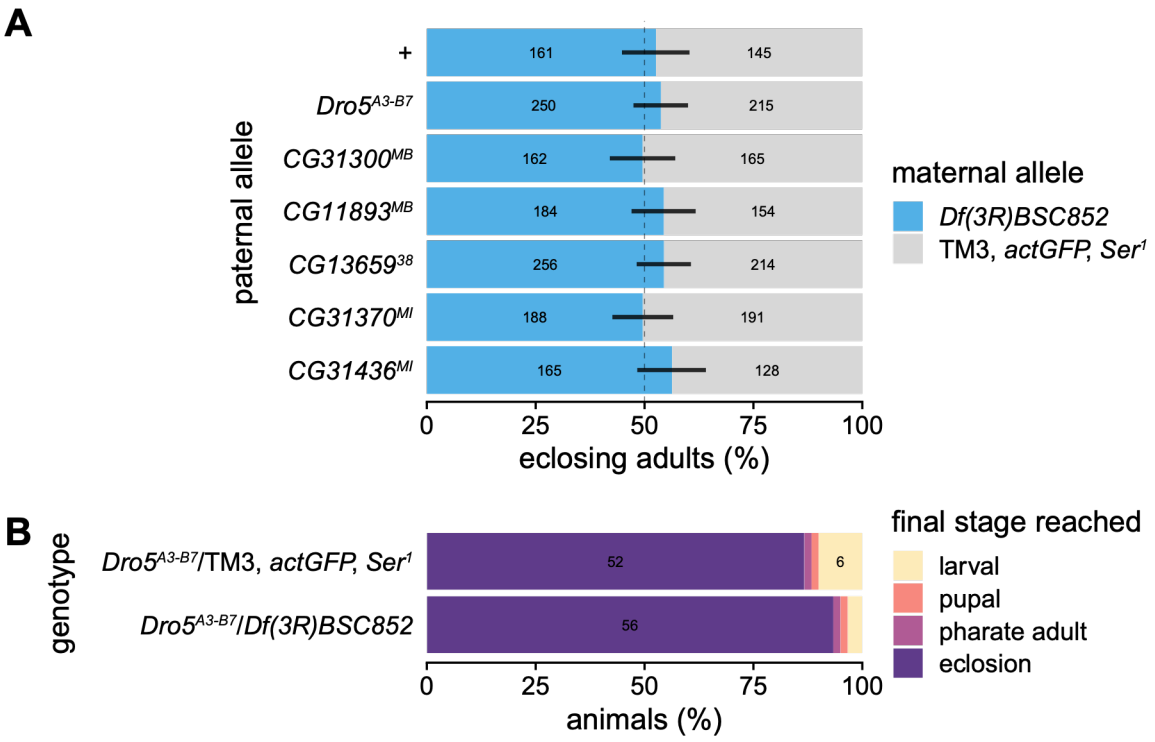


Figure 4.5. Developmental viability of Dro5 loss-of-function alleles (or the wild-type allele +) over the deficiency *Df(3R)BSC852* on lab media. (A) Egg-to-adult viability, estimated from the adult genotypic ratios of offspring from crosses between *Df(3R)BSC852*/TM3, *actGFP*, *Ser*^{1} females and males of one of seven homozygous genotypes. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; error bars are 99.29% confidence intervals (95% confidence interval adjusted for seven tests) for the proportion of *Df(3R)BSC852* heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of individuals of that genotype scored. (B) Larvae-to-adult viability of offspring from the cross between *Df(3R)BSC852*/TM3, *actGFP*, *Ser*^{1} females and homozygous *Dro5*$^{A3\text{-}B7}$ males, sorted at the 1st-instar larval stage by GFP fluorescence (n = 60 larvae—three vials of 20—per genotype). Numbers on the bars are the number of individuals in each lethal phase category (for numbers greater than five).

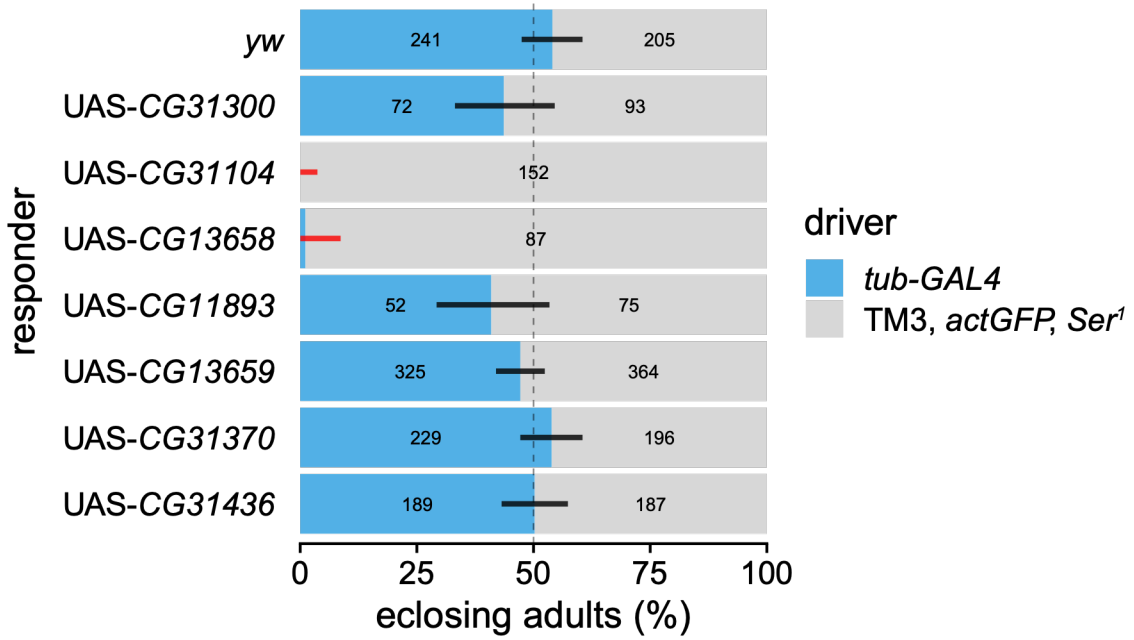


Figure 4.6. Egg-to-adult viability of the misexpression of Dro5 ORFs (or non-misexpression from the genetic background *yw*) using the strong, ubiquitous GAL4 driver *tub-GAL4*, estimated from the adult genotypic ratios of offspring from crosses between *tub-GAL4*/TM3, *actGFP*, Ser^{1} females and males of one of eight homozygous responder genotypes. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; error bars are 99.38% CIs (95% CI adjusted for eight tests) for the proportion of *tub-GAL4* heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of individuals of that genotype scored.

4.3.3. A *CG31370* loss-of-function allele is associated with adult caffeine susceptibility in the DGRP

Analysis of structural variation in the DGRP (Mackay *et al.* 2012) called *in silico* from paired-end mapped reads identified a *CG31370* haplotype not present in the official DGRP variant annotation data (http://dgrp2.gnets.ncsu.edu/data.html): two deletions 5 bp apart, of size 183 bp (3R:25,302,734..25,302,916) and 1 bp (3R:25,302,921), in the first exon of *CG31370*, the latter of which induces a frameshift in the CDS. Given that the two deletions were always present together and clearly derived (based on comparisons with *CG31370* orthologs in other *Drosophila* genomes), they were considered a single composite allele designated $CG31370^{del}$ (Fig. 4.7A), while the allele with the intact gene was designated $CG31370^{wt}$. 152 DGRP lines were successfully *in silico* genotyped at *CG31370*: 141 lines were homozygous for $CG31370^{wt}$ and 11 lines were homozygous for $CG31370^{del}$, while 39 lines were unable to be called due to uninformative read mapping depth or a lack of available mapped-read data.

A preliminary PheWAS analysis of the $CG31370^{del}$ polymorphism with a collection of DGRP phenotypes (Table S2.2) suggested it is associated with adult survival on caffeine ($p = 2.5\times10^{-3}$), although this did not meet a genome-wide significance threshold of $p < 1.0\times10^{-5}$ (see Chapter 2.3.3). The source of the phenotype data is Najarro *et al.* (2015), who phenotyped 165 lines from the DGRP for caffeine tolerance—mean lifespan of adult female flies feeding on food media containing 1% (10 mg/mL) caffeine—but found no significant genome-wide associations with any SNPs. To increase the sample size for this association test in the DGRP, PCR genotyping was performed on 46 available DGRP lines, nine of which had confident *in silico* calls, with the remaining 37 lines uncalled. PCR confirmed the genotypes of the called lines and revealed an additional 29 lines homozygous for $CG31370^{wt}$ and eight lines homozygous for $CG31370^{del}$. Of the 164 total lines with confident *CG31370* genotypes, 123 lines had been phenotyped by Najarro *et al.* (2015). The mean difference in the survival time on 10 mg/mL caffeine food between phenotyped homozygous $CG31370^{wt}$ and $CG31370^{del}$ lines was -20.2 hours (95.0% CI: -30.3, -10.4; Fig. 4.7B). In adult females, $CG31370^{del}$ homozygotes have reduced basal expression of *CG31370* compared to $CG31370^{wt}$ homozygotes (-1.16 $\log_2$(FPKM), 95% CI: -1.53, -0.681; Fig. 4.7D), and the same is true in adult males (-1.05 $\log_2$(FPKM), 95% CI: -1.4, -0.585; Fig. 4.7E)—this is consistent with nonsense-mediated decay of $CG31370^{del}$ mRNA *in vivo* and suggests this allele indeed

does induce a frameshift in the *CG31370* CDS.

I also explored whether the $CG31370^{del}$ allele is negatively associated with developmental caffeine survival (successful eclosion on media supplemented with 388 μg/mL caffeine; Montgomery *et al.* 2014) in the DGRP. 173 DGRP lines were phenotyped by Montgomery *et al.* (2014), of which 141 had known *CG31370* genotypes (125 homozygous $CG31370^{wt}$, 16 homozygous $CG31370^{del}$). The mean difference in corrected proportional survival on 388 μg/mL caffeine between phenotyped homozygous $CG31370^{wt}$ and $CG31370^{del}$ lines was -0.0561 (95% CI: -0.161, 0.0313; Fig. 4.7C), suggesting there is unlikely to be a real difference in survival between genotypes at this dose; removing the 37 lines for which survival on caffeine was greater than on the control (which might suggest a confounding hormetic effect; Calabrese & Mattson 2017) did not substantially affect the mean difference between genotypes (-0.054; 95% CI: -0.17, 0.0463). Due to the limited response to caffeine from most DGRP lines, it is likely that the dose used by Montgomery *et al.* (2014)—which was originally intended to be subtoxic and is 28.8-fold lower than that used by Najarro *et al.* (2015)—was insufficiently high to discriminate between the $CG31370^{wt}$ and $CG31370^{del}$ genotypes, if they do indeed vary in their tolerance to caffeine.

Taken together with the known induction of other Dro5 genes by caffeine (Table 4.1), these data led to a hypothesis that *CG31370* and/or other Dro5 genes may play a role in the detoxification of caffeine in *D. melanogaster*.

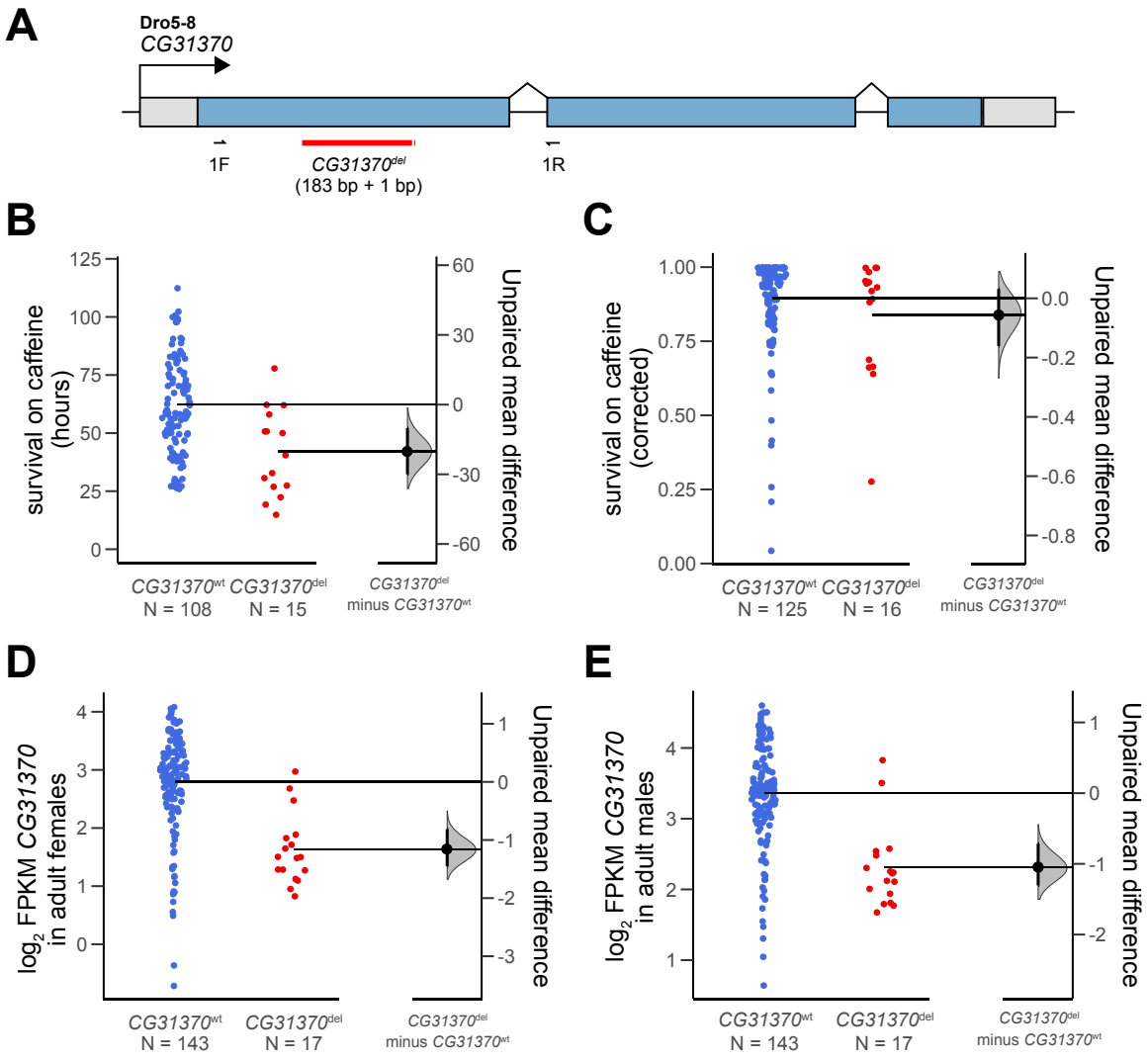


Figure 4.7. The *CG31370*^{del} composite deletion allele in the DGRP. (A) Location of the *CG31370*^{del} deletions within the first exon of *CG31370* (Dro5-8), along with the locations of the CG31370del_1F (1F) and CG31370del_1R (1R) genotyping primers. Coding sequence is blue, UTRs are grey. (B–E) Estimation plots (Ho *et al.* 2019) of mean phenotypic and gene expression differences between homozygous *CG31370*^{wt} (blue) and homozygous *CG31370*^{del} (red) DGRP lines. The right-hand axis shows the mean difference (effect size) between groups, with the 95% CI (black line) and distribution of bootstrapped means (grey curve). Effect sizes with CIs that do not include zero are considered significant. (B) Survival, in hours, of adult female flies on 10 mg/mL caffeine food, as phenotyped by Najarro *et al.* (2015). (C) Developmental survival (Abbott-corrected proportional successful eclosion) on 388 μg/mL caffeine food, as phenotyped by Montgomery *et al.* (2014). (D–E) Expression ($\log_2$(FPKM)) of *CG31370* in adult (D) females and (E) males in the DGRP, as reported by Everett *et al.* (2020).

4.3.4. Dro5 loss-of-function mutants are developmentally susceptible to caffeine

Developmental dose-response assays on caffeine-containing media were used to test the hypothesis that Dro5 genes are involved in caffeine detoxification; doses used were a minimum of 25 μg/mL and a maximum of 1,500 μg/mL, compared to 10,000 μg/mL for the adult assays of Najarro *et al.* (2015) and 388 μg/mL for the developmental assays of Montgomery *et al.* (2014). A comparison between $Dro5^{A3\text{-}B7}$ and wild-type homozygotes found that the caffeine LC_{50} (median lethal concentration) of mutant animals was significantly lower than that wild-type animals (Fig. 4.8), with mutant:wild-type LC_{50} ratios of 0.66 (95% CI: 0.59, 0.74), 0.20 (95% CI: 0.18, 0.23) and 0.26 (95% CI: 0.23, 0.3), for larval, pupal and total mortality, respectively. A notable, qualitative effect of caffeine exposure on $Dro5^{A3\text{-}B7}$ mutant animals was high pharate adult lethality—where adults attempted to eclose but remained trapped in the puparium before dying—even at relatively low doses (200–300 μg/mL). Defects in earlier stages of metamorphosis were generally not seen until doses were much higher, although this effect was not quantified, as all types of mortality after pupation were scored alike.

An initial experiment utilising an independently generated caffeine stock solution and the $+/Dro5^{A3\text{-}B7}$ heterozygote genotype also showed significant differences in pupal and total LC_{50}s between $Dro5^{A3\text{-}B7}$ homozygotes and control genotypes with at least one wild-type allele, although larval LC_{50}s could not be compared for $Dro5^{A3\text{-}B7}$ homozygotes (Fig. S4.2). $Dro5^{A3\text{-}B7}$ homozygotes had relative pupal and total LC_{50}s of 0.2 (95% CI: 0.15, 0.27) and 0.29 (95% CI: 0.21, 0.39), respectively, compared to +/+ homozygotes, and relative pupal and total LC_{50}s of 0.29 (95% CI: 0.23, 0.38) and 0.29 (95% CI: 0.23, 0.37), respectively, compared to $+/Dro5^{A3\text{-}B7}$ heterozygotes. There were no significant differences in larval and total LC_{50}s between +/+ homozygotes and $+/Dro5^{A3\text{-}B7}$ heterozygotes, consistent with the expectation that the caffeine susceptibility phenotype is recessive; the pupal LC_{50} ratio was significant—1.46 (95% CI: 1.03, 2.06)—although barely (with a wide CI). This may have been due to inappropriate dose spacing, which was concentrated towards the lower end of the dose range.

Complementation experiments with the *Df(3R)BSC852* chromosomal deficiency, which encompasses—and therefore should fail to complement—the two deleted $Dro5^{A3\text{-}B7}$ loci, failed to allow a statistical comparison between *Df(3R)BSC852*/+ and $Df(3R)BSC852/Dro5^{A3\text{-}B7}$ genotypes due to an insufficiently high dose range (Fig. 4.9).

Df(3R)BSC852/+ animals did not respond to caffeine above control levels, although *Df(3R)BSC852*/*Dro5*$^{A3\text{-}B7}$ animals showed very high pupal and total mortalities (and low larval mortality) on the highest dose (1,000 μg/mL), consistent with *Df(3R)BSC852* failing to complement *Dro5*$^{A3\text{-}B7}$, as *Dro5*$^{A3\text{-}B7}$ homozygotes have much lower pupal and total LC_{50}s than larval LC_{50} (Fig. 4.8).

I also sought to determine if loss-of-function alleles for individual Dro5 genes disrupted in the *Dro5*$^{A3\text{-}B7}$ allele (*CG31300*, *CG31104*, *CG13658*, *CG11893* and *CG13659*) fail to complement *Dro5*$^{A3\text{-}B7}$, which would indicate that the loss of that particular gene may be responsible for the decreased caffeine LC_{50}s of homozygous *Dro5*$^{A3\text{-}B7}$ animals. Loss-of-function alleles were only available for four of the five genes: TE-insertion alleles for *CG31300* (*CG31300*^{MB}), *CG13658* (*CG13658*^{MI}) and *CG11893* (*CG11893*^{MB}), and a CRISPR-Cas9-induced composite deletion allele in *CG13659* generated in this study (*CG13659*^{38}). There was curiously high larval and total mortality in some 0 μg/mL caffeine vials, meaning doses had to be corrected based on the 25 μg/mL dose, which showed lower mortality, and the dose range used was unfortunately largely too low to allow for statistical comparisons between most pairs of genotypes (Fig. 4.10). Despite this, significant differences in +:*Dro5*$^{A3\text{-}B7}$ pupal and total LC_{50} ratios could be detected for *CG11893*^{MB} in *trans*—pupal LC_{50} ratio of 0.34 (95% CI: 0.1, 0.63) and total LC_{50} ratio of 0.60 (95% CI: 0.4, 0.91)—and *CG13659*^{38} in *trans*—pupal LC_{50} ratio of 0.14 (95% CI: 0.02, 0.94) and total LC_{50} ratio of 0.28 (95% CI: 0.12, 0.62). These estimates have relatively large CIs and therefore have low precision, but suggest that both of these genes may be responsible for the increased caffeine susceptibility of *Dro5*$^{A3\text{-}B7}$ homozygotes. In addition, the high pupal and total mortalities of *Dro5*$^{A3\text{-}B7}$/*CG31300*^{MB} and *Dro5*$^{A3\text{-}B7}$/*CG13658*^{MI} genotypes compared to their control genotypes are consistent with these two genes also contributing to the caffeine susceptibility phenotype.

The *CG31370*^{MI} allele, which consists of the insertion of the Mi{MIC} element *MI07438* into the first exon of *CG31370*, is almost certainly a strong loss-of-function allele. Comparison between *CG31370*^{MI} homozygotes and +/*CG31370*^{MI} heterozygotes suggested that homozygotes have lower larval—ratio of 0.46 (95% CI: 0.31, 0.7)—and total—0.48 (95% CI: 0.35, 0.65)—LC_{50}s (pupal LC_{50}s could not be compared; Fig. 4.11). However, the wild-type allele used in this experiment (the genetic background of the CRISPR-Cas9-derived alleles) is not isogenic to the *CG31370*^{MI} line (Venken *et al.* 2011); as such, these results may be confounded by differences in genetic background.

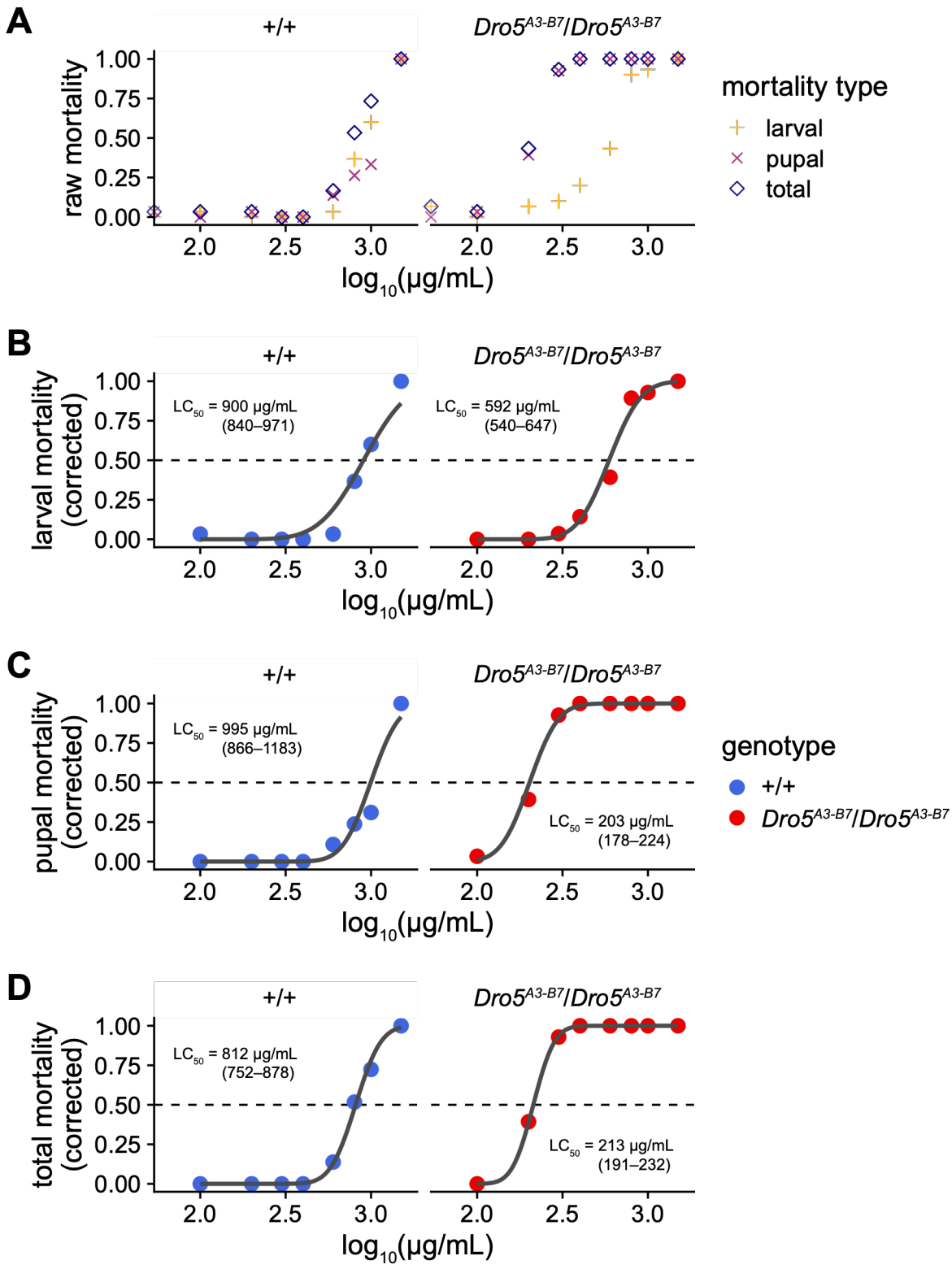


Figure 4.8. Developmental mortality of +/+ homozygotes and *Dro5*$^{A3\text{-}B7}$/*Dro5*$^{A3\text{-}B7}$ homozygotes on media containing 100–1,500 µg/mL caffeine. (A) Raw proportional mortality, with control mortality (0 µg/mL caffeine) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; larval LC_{50} calculation for +/+ animals did not use the 100 µg/mL dose to prevent undefined values.

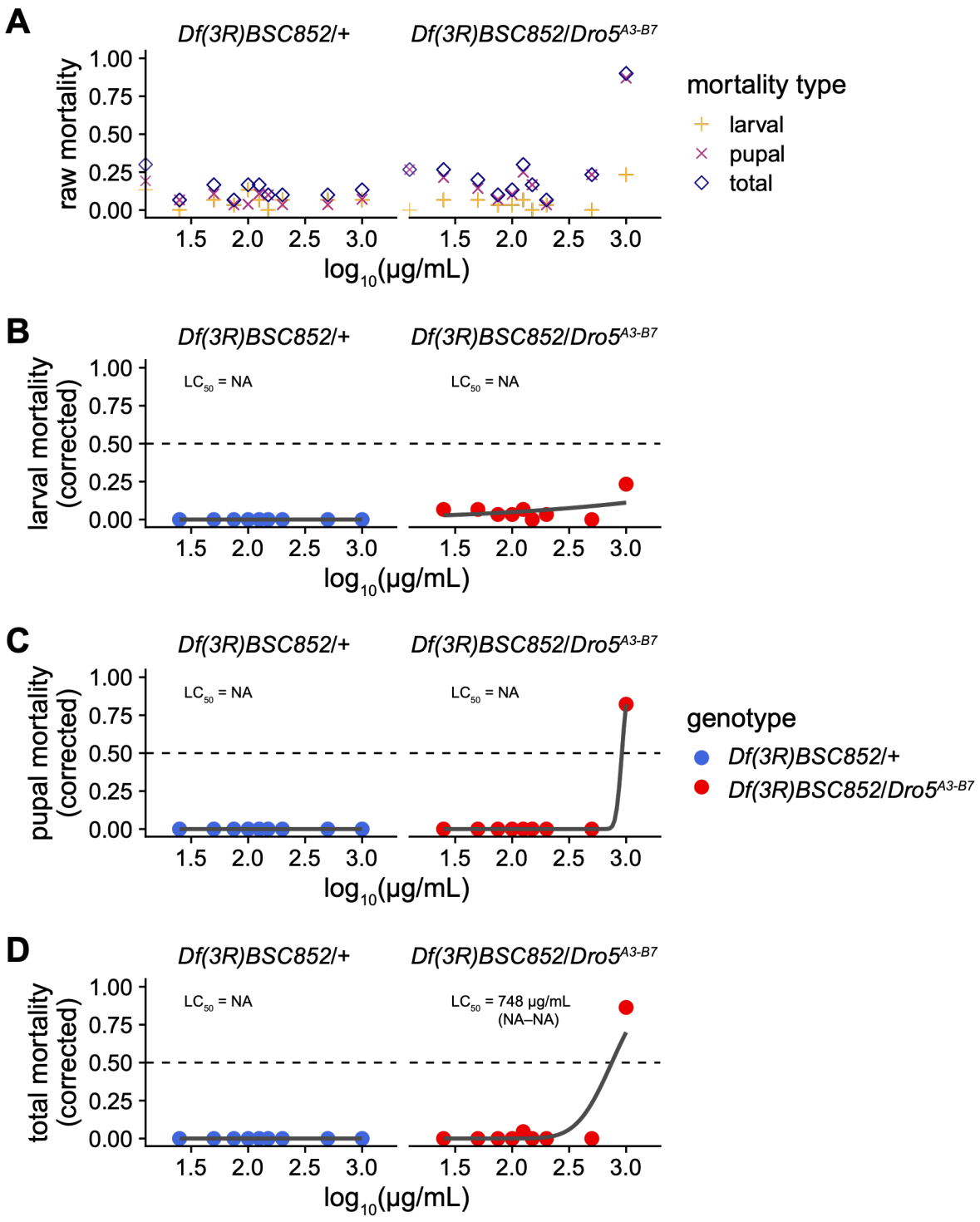


Figure 4.9. Developmental mortality of *Df(3R)BSC852*/+ and *Df(3R)BSC852/Dro5*$^{A3\text{-}B7}$ heterozygotes on media containing 25–1,000 µg/mL caffeine. (A) Raw proportional mortality, with control mortality (0 µg/mL caffeine) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; values are 'NA' when they could not be calculated from the data.

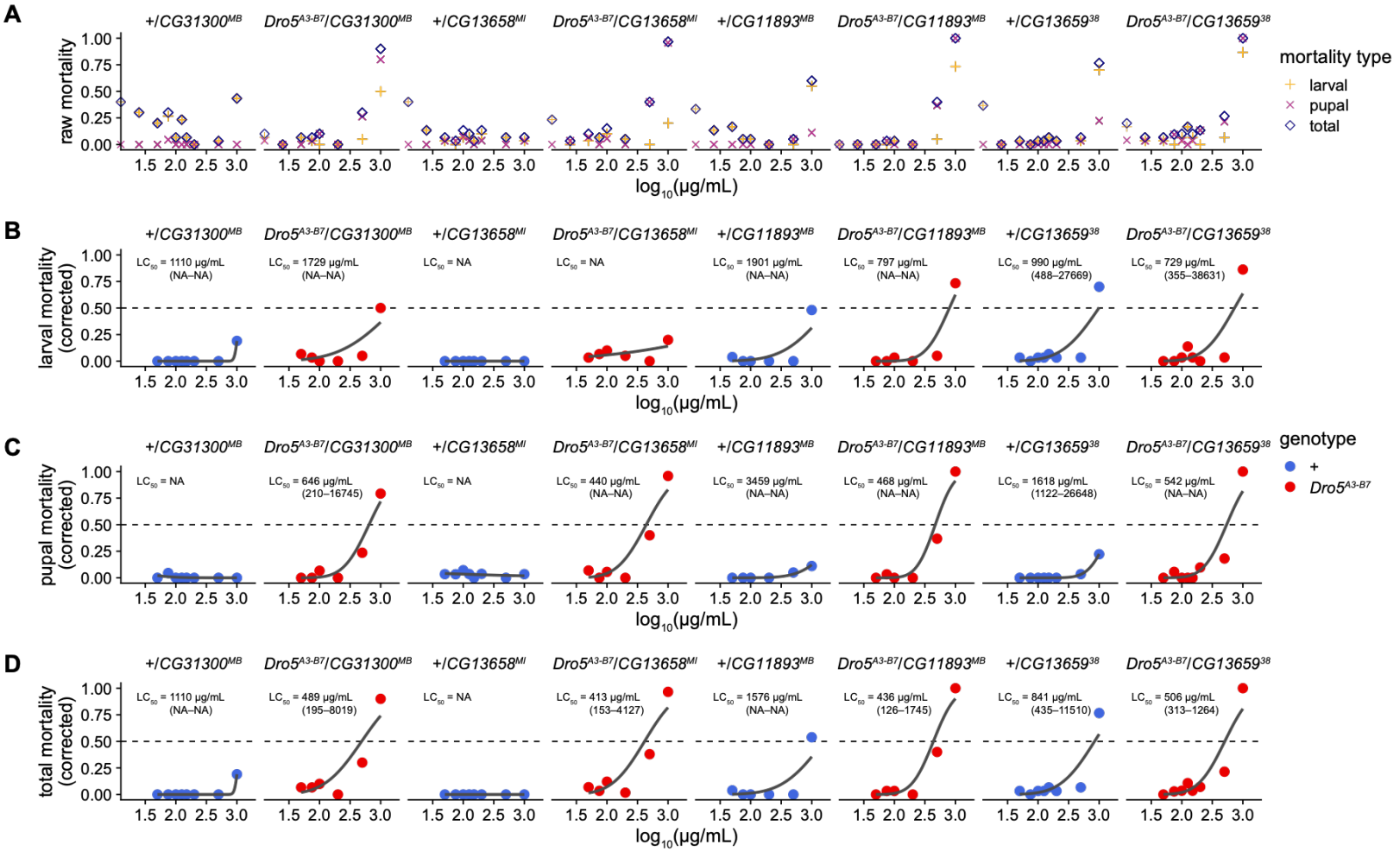


Figure 4.10. Developmental mortality of single-gene disruption alleles over either wild-type (+) or *Dro5*$^{A3\text{-}B7}$ alleles on media containing 25–1,000 µg/mL caffeine. (A) Raw proportional mortality, with control mortality (0 µg/mL caffeine) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; values are 'NA' when they could not be calculated from the data. Due to high mortality on the control doses, mortality was corrected using the raw 25 µg/mL dose (which was then not used in the probit regression).

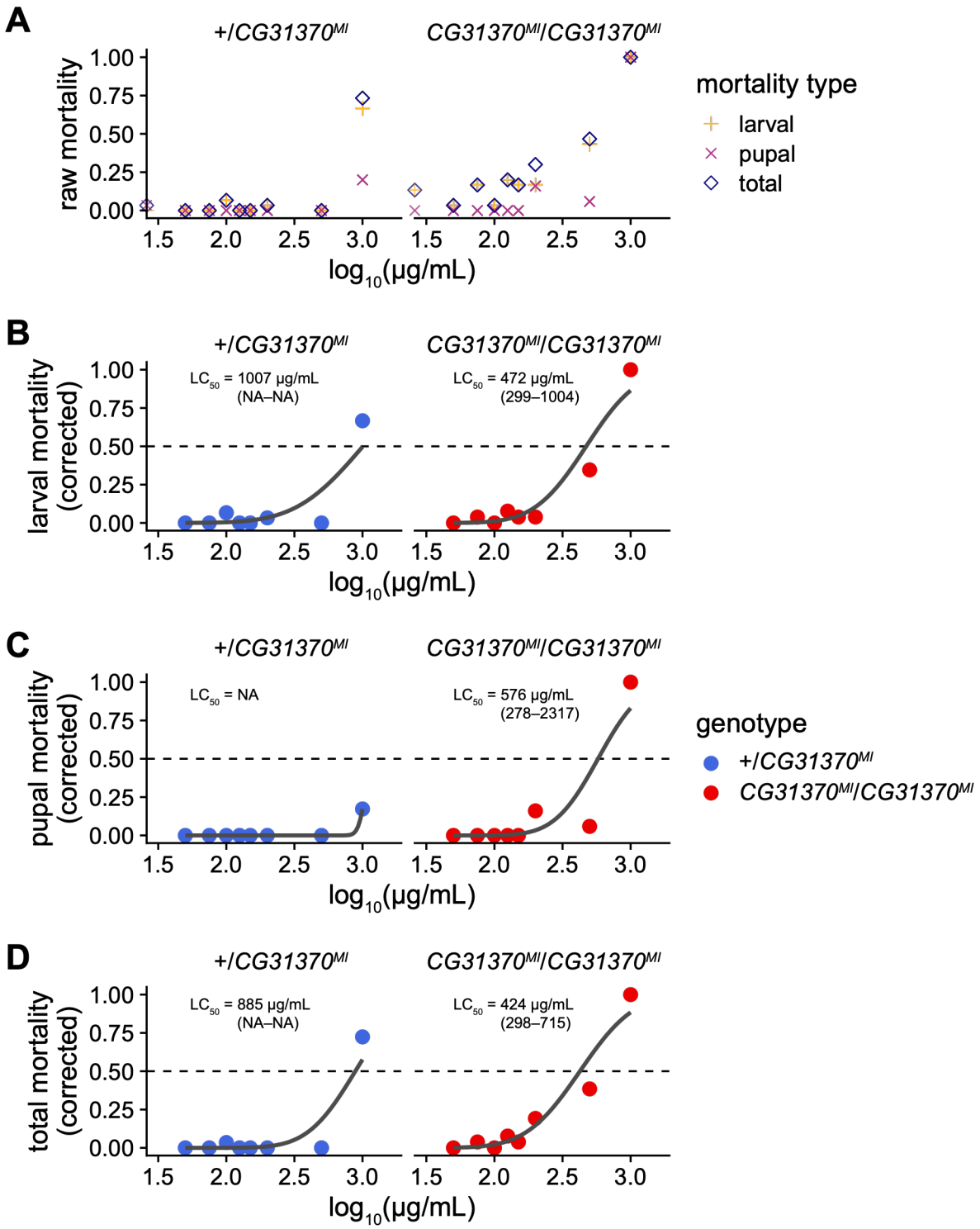


Figure 4.11. Developmental mortality of +/*CG31370*^{MI} heterozygotes and *CG31370*^{MI}/*CG31370*^{MI} homozygotes on media containing 50–1,000 µg/mL caffeine. (A) Raw proportional mortality, with control mortality (0 µg/mL caffeine) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; values are 'NA' when they could not be calculated from the data.

4.3.5. Misexpression of two Dro5 genes in detoxification tissues decreases developmental susceptibility to caffeine

Dro5 genes were transgenic misexpressed in the midgut, Malpighian tubules and fat body using the *6g1HR-GAL4-6c* (*HR-GAL4*) driver (Chung *et al.* 2007) and UAS-ORF lines generated in Chapter 4.2.2, which has previously been used to test the ability of genes to modulate toxicological phenotypes (Battlay *et al.* 2018; 2016; Chung *et al.* 2007; Daborn *et al.* 2007; Green *et al.* 2019). Unfortunately, due to stock loss, experiments with the UAS-*CG11893* and UAS-*CG31436* were unable to be conducted. A 95% CI was not able to be calculated for the larval LC_{50} for three genotypes, including the *HR>yw* control genotype (Fig. 4.12B), but pupal and total LC_{50}s (Fig. 4.12B–C) were able to be statistically compared. Misexpression of *CG31300* increased both the pupal LC_{50}, with a ratio of 1.48 (95% CI: 1.2, 1.82) and the total LC_{50}, with a ratio of 1.34 (95% CI: 1.12, 1.6), as did misexpression of *CG13659*, with a pupal LC_{50} ratio of 1.29 (95% CI: 1.09, 1.53) and a total LC_{50} ratio of 1.37 (95% CI: 1.19, 1.59). Curiously, misexpression of *CG31370* decreased the pupal LC_{50}, with a ratio of 0.76 (95% CI: 0.65, 0.90), and the total LC_{50}, with a ratio of 0.75 (95% CI: 0.65, 0.87). Neither misexpression of *CG31104* nor *CG13658* significantly changed pupal nor total LC_{50}s compared to the control. Larval LC_{50} comparisons between misexpression genotypes showed no significant differences (all LC_{50} ratio 95% CIs included 1).

4.3.6. *Dro5$^{A3\text{-}B7}$* homozygotes are developmentally susceptible to kojic acid

Dose-response assays were conducted on homozygous *Dro5$^{A3\text{-}B7}$* and wild-type (+/+) animals using media containing 2,000—6,000 μg/mL kojic acid, with a water-based control (0 μg/mL), and larval, pupal and total mortality was calculated (Fig. 4.13). Mutant:wild-type LC_{50} ratios were 0.77 (95% CI: 0.70, 0.86), 0.60 (95% CI: 0.52, 0.69) and 0.63 (95% CI: 0.58, 0.70) for larval, pupal and total mortalities, respectively. These data suggest that genes disrupted in the *Dro5$^{A3\text{-}B7}$* allele may contribute to kojic acid tolerance.

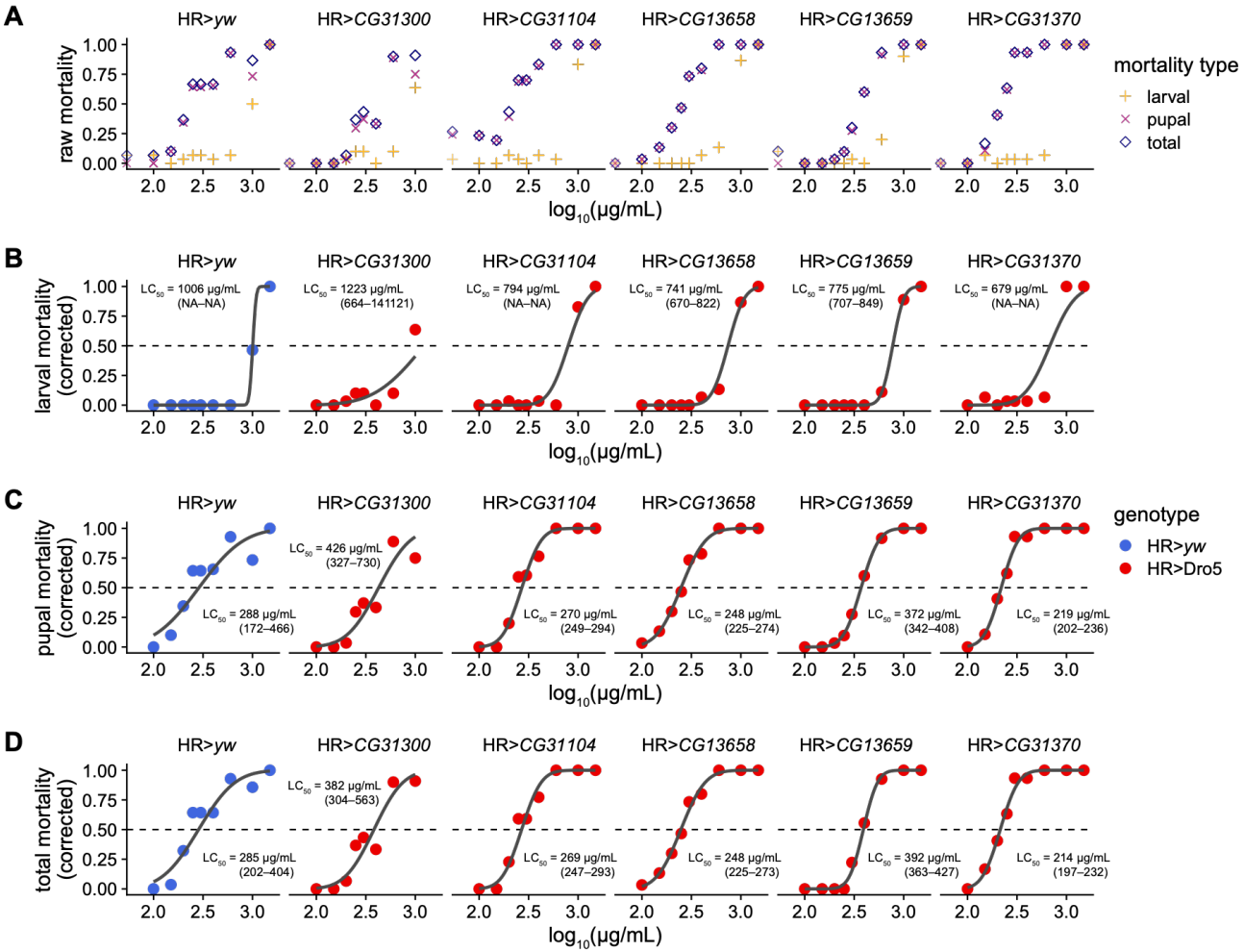


Figure 4.12. Developmental mortality of animals misexpressing one of five Dro5 ORFs (*CG31300*, *CG31104*, *CG13658*, *CG13659* or *CG31370*), or nothing (*yw* genetic background), with the *HR-GAL4* driver on media containing 100–1,500 µg/mL caffeine. (A) Raw proportional mortality, with control mortality (0 µg/mL caffeine) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; values are 'NA' when they could not be calculated from the data. A 1,500 µg/mL dose was not used with HR>*CG31300* animals.

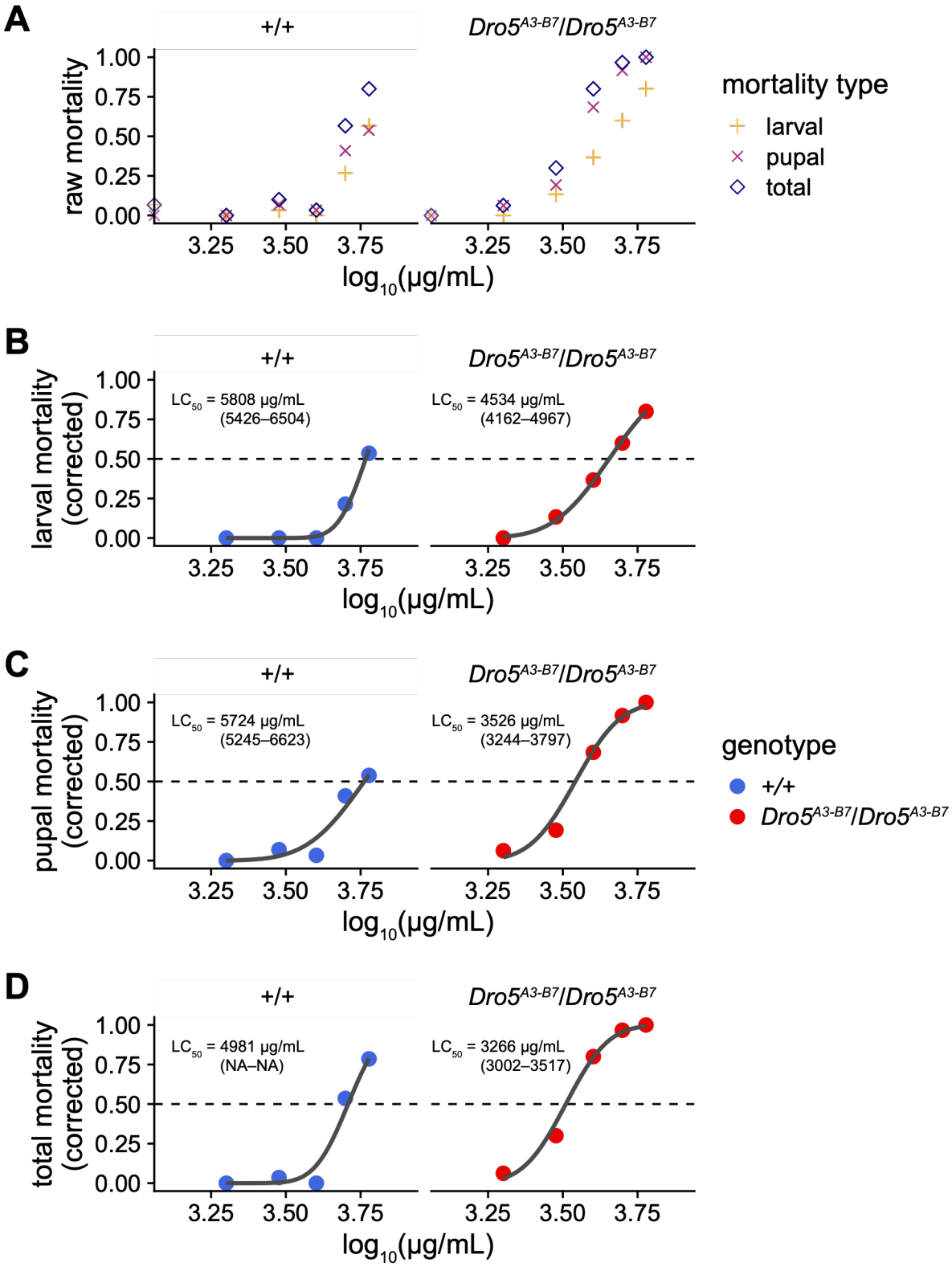


Figure 4.13. Developmental mortality of +/+ homozygotes and *Dro5*$^{A3\text{-}B7}$/*Dro5*$^{A3\text{-}B7}$ homozygotes on media containing 2,000–6,000 µg/mL kojic acid. (A) Raw proportional mortality, with control mortality (0 µg/mL kojic acid) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; values are 'NA' when they could not be calculated from the data.

4.3.7. No evidence that *Dro5$^{A3\text{-}B7}$* homozygotes have increased developmental susceptibility to salicin, curcumin, escin, esculin or quercetin

Quercetin, escin, esculin and curcumin are sparingly soluble in water, so stock solutions were made in ethanol, up to 10 mg/mL for escin, esculin and curcumin and 2 mg/mL for quercetin. Pilot experiments on the *w^{1118}* lab strain of *D. melanogaster* with 10-fold serial dilutions of these stock solutions suggested the highest concentrations possible at a maximum of 2% (v/v) ethanol in the media were not toxic to wild-type flies, and so single-dose experiments were conducted on homozygous *Dro5$^{A3\text{-}B7}$* and wild-type (*Dro5$^{A3\text{-}B7}$*/+) animals with media containing 200 μg/mL escin, 200 μg/mL esculin, 200 μg/mL curcumin, 40 μg/mL quercetin or 2% ethanol (control). Viability was scored for both successful pupation and successful eclosion (Fig. 4.14). No significant differences were found between genotypes for each medium and viability combination ($p > 0.05$, Welch's two-sided t-test with unequal variance), indicating homozygosity of the *Dro5$^{A3\text{-}B7}$* allele does not affect developmental viability on any of the four single-toxin substrates.

Salicin is very soluble in water, and so dose-response assays were conducted on homozygous *Dro5$^{A3\text{-}B7}$* and wild-type (*Dro5$^{A3\text{-}B7}$*/+) animals using media containing 62.5–8,000 μg/mL salicin, with a water-based control (0 μg/mL), and larval, pupal and total mortality was calculated (Fig. 4.15). No appreciable mortality dose-response could be detected for either genotype for any of the mortality metrics, meaning LC_{50} values could not be estimated.

These data show homozygosity of the *Dro5$^{A3\text{-}B7}$* allele does not alter tolerance to these five compounds at the doses used here, but they do not preclude changes in tolerance at higher doses.

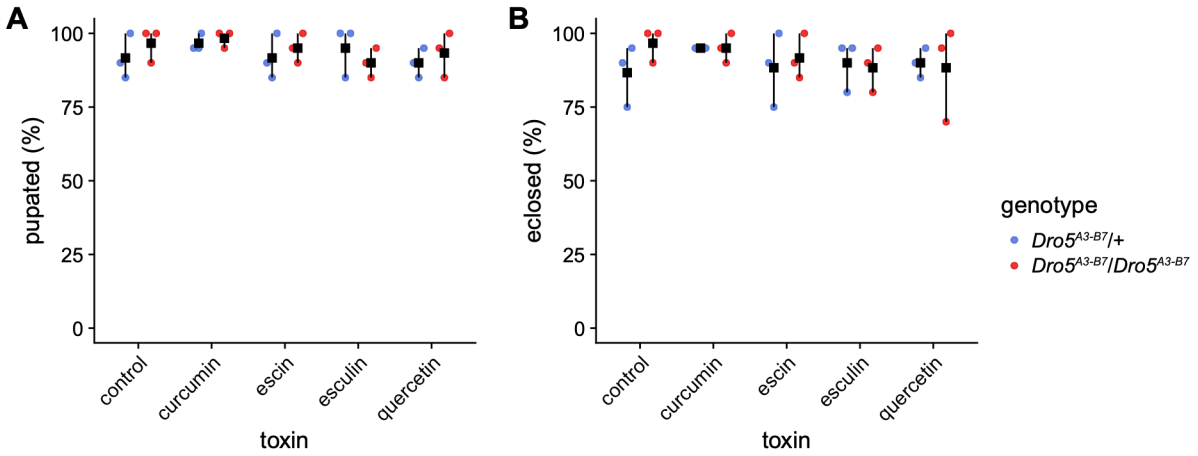


Figure 4.14. Developmental viability of *Dro5*$^{A3\text{-}B7}$/+ heterozygotes (blue) and *Dro5*$^{A3\text{-}B7}$/*Dro5*$^{A3\text{-}B7}$ homozygotes (red) on control media or media containing natural plant toxins—curcumin (200 μg/mL), escin (200 μg/mL), esculin (200 μg/mL) or quercetin (40 μg/mL)—scored as (A) survival to pupation or (B) survival to eclosion. n = 3 vials per genotype and medium, with 20 larvae per vial. Means are plotted as black squares with bootstrapped 95% confidence intervals.

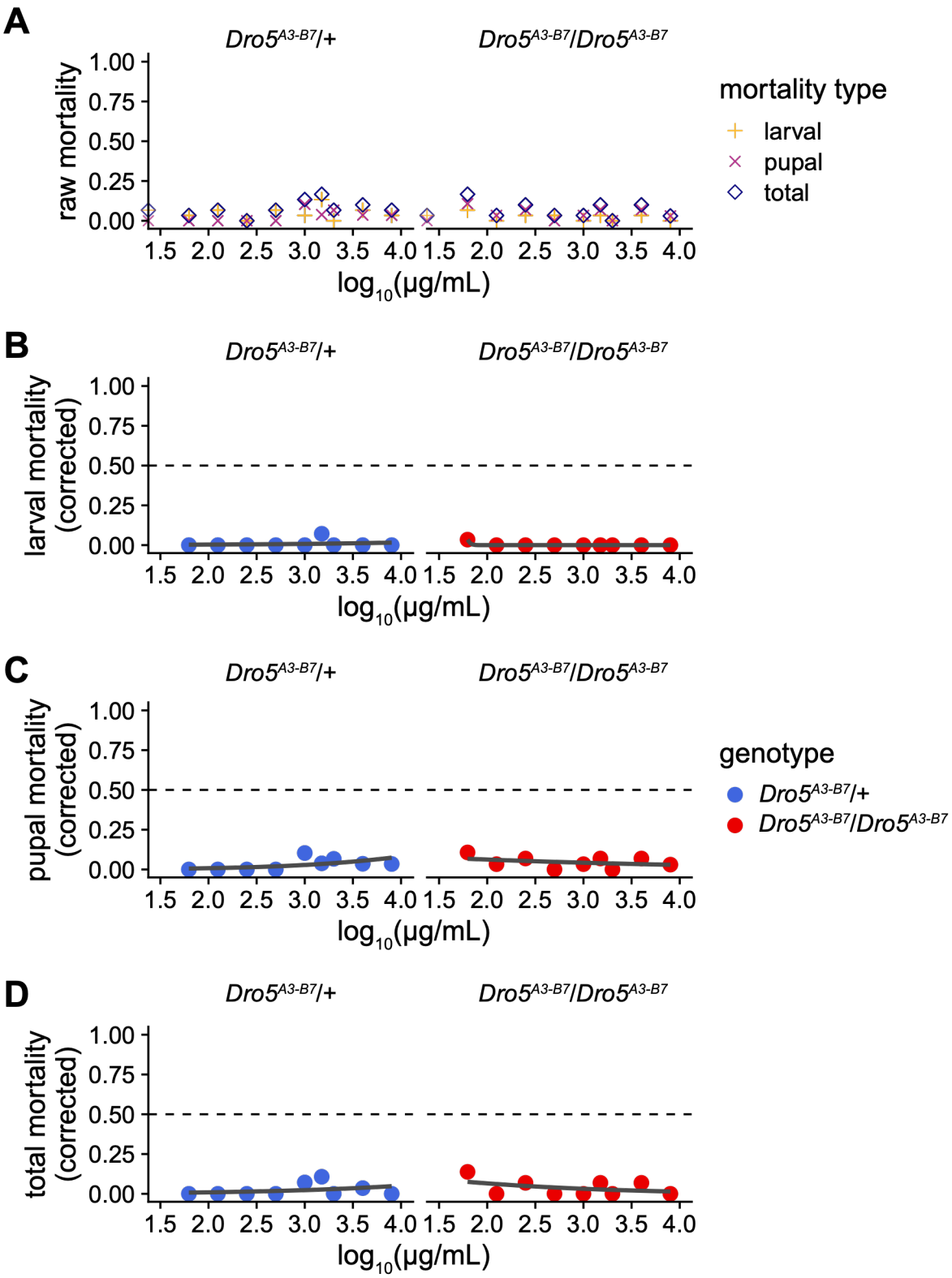


Figure 4.15. Developmental mortality of *Dro5*$^{A3\text{-}B7}$/+ heterozygotes and *Dro5*$^{A3\text{-}B7}$/*Dro5*$^{A3\text{-}B7}$ homozygotes on media containing 62.5–8,000 µg/mL salicin. (A) Raw proportional mortality, with control mortality (0 µg/mL salicin) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. No LC_{50} values could be calculated due to a lack of dose-response ($p > 0.05$, probit GLM).

4.3.8. *Dro5$^{A3\text{-}B7}$* homozygotes are developmentally viable on citrus-based media

To test whether Dro5 EcKLs are involved in tolerance to ecologically relevant mixtures of fruit-derived xenobiotic compounds, I conducted developmental viability assays of both homozygous *Dro5$^{A3\text{-}B7}$* and wild-type (*Dro5$^{A3\text{-}B7}$*/+) animals on semi-natural fruit media made from grapefruit, orange or mandarin juices; media containing lime juice was also made, but it failed to solidify adequately for use in experiments. Citrus fruits were chosen for their ease of juicing, as well as their status as preferred host fruits for *D. melanogaster* oviposition (Dweck *et al.* 2013). Viability was scored for both successful pupation and successful eclosion (Fig. 4.16). No significant differences were found between genotypes for each medium and viability combination ($p > 0.05$, Welch's two-sided t-test with unequal variance), indicating homozygosity of the *Dro5$^{A3\text{-}B7}$* allele does not affect developmental viability on any of the three fruit-based substrates.

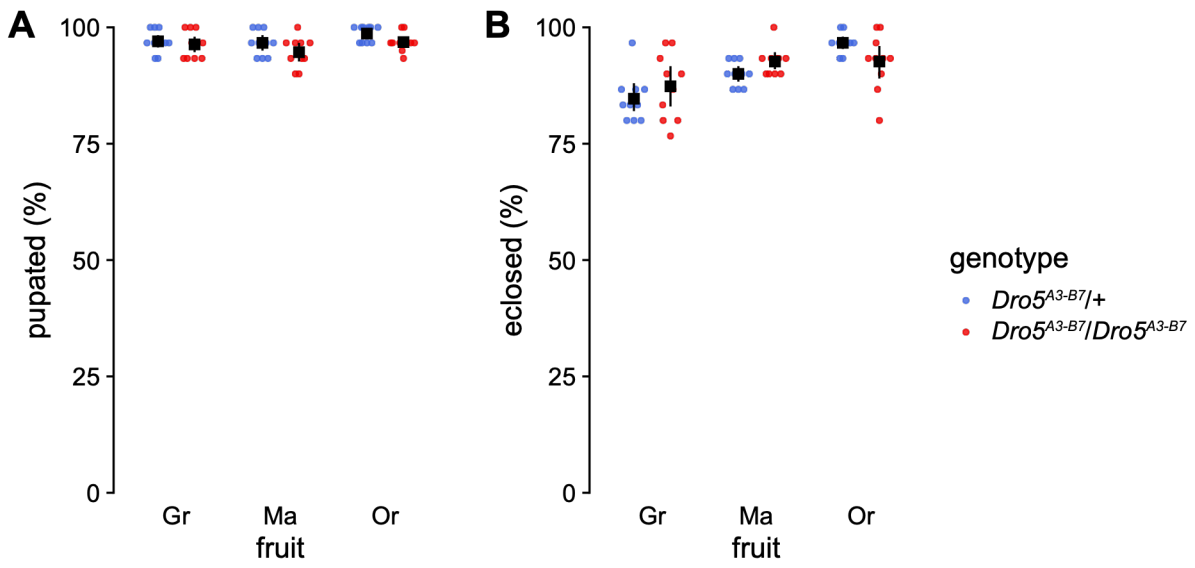


Figure 4.16. Developmental viability of *Dro5*$^{A3\text{-}B7}$/+ heterozygotes (blue) and *Dro5*$^{A3\text{-}B7}$/*Dro5*$^{A3\text{-}B7}$ homozygotes (red) on three semi-natural citrus fruit mediums—grapefruit (Gr), mandarin (Ma) and orange (Or)—scored as (A) survival to pupation or (B) survival to eclosion. n = 10 vials per genotype and medium, with 30 larvae per vial. Means are plotted as black squares with bootstrapped 95% confidence intervals.

4.4. Discussion

4.4.1. Which Dro5 EcKLs are most likely to be involved in caffeine detoxification?

In this chapter, I used a combination of gene knockout and gene misexpression experiments to test the hypothesis that Dro5 genes in *D. melanogaster* modulate developmental susceptibility to caffeine, which, if true, would be evidence for these genes playing a role in the detoxification of caffeine. Three types of mortality were scored: larval mortality—failure to successfully pupate, indicating death between the 1st instar and pupation; pupal mortality—failure to successfully eclose from the puparium given successful pupation, indicating death during metamorphosis; and total mortality—failure to successfully complete development, indicating death before the adult stage. *Dro5*$^{A3\text{-}B7}$ mutants had increased susceptibility to caffeine at both larval and pupal stages, the latter of which failed to be complemented by *CG11893* (Dro5-6) and *CG13659* (Dro5-7) mutant alleles, and possibly also *CG31300* (Dro5-1) and *CG13658* (Dro5-5) mutant alleles. Misexpression of *CG31300* (Dro5-1) and *CG13659* (Dro5-7) with the *HR-GAL4* driver decreased susceptibility to caffeine, consistent with the hypothesis that disruption of these genes specifically may be responsible for the increased susceptibility of *Dro5*$^{A3\text{-}B7}$ homozygotes to caffeine. Curiously, *CG31370* (Dro5-8) misexpression increased developmental susceptibility to caffeine.

A number of gene-experiment combinations were unfortunately not tested in this chapter, due to loss or absence of certain fly lines. Specifically, a *CG31104* (Dro5-2) single-gene disruption allele was not available, and the UAS-*CG11893* and UAS-*CG31436* constructs were not available to be misexpressed with *HR-GAL4*. In addition, a full Dro5-null allele was not generated by CRISPR-Cas9 mutagenesis as expected, meaning the caffeine susceptibility of *CG31370* (Dro5-8) and *CG31436* (Dro5-10) loss-of-function animals could not be appropriately tested. The analysis of some toxicological assay data also suffered from inappropriate dose ranges and high control mortality.

Despite these challenges, the data in this chapter support the hypothesis that Dro5 genes play a role in caffeine detoxification. Of the seven Dro5 genes in the *D. melanogaster* genome, *CG13659* (Dro5-7) is the best candidate for involvement in the

detoxification of caffeine, given that: its disruption, either with four other Dro5 genes in $Dro5^{A3\text{-}B7}$ or by itself ($CG13659^{38}$), increases caffeine susceptibility; misexpression of *CG13659* in detoxification tissues decreases caffeine susceptibility; and *CG13659* is transcriptionally induced by ingestion of caffeine in larvae (Table 4.3). It is also a good detoxification gene candidate more generally, as it is induced in six xenobiotic datasets and transcriptionally enriched in the midgut, Malpighian tubules, fat body and blood-brain barrier (Fig. 2.7; DeSalvo *et al.* 2014). *CG31300* (Dro5-1) is also a decent caffeine detoxification candidate: it is disrupted in the $Dro5^{A3\text{-}B7}$ allele, which a single-gene disruption allele of *CG31300* may fail to complement, and misexpression of *CG31300* decreases developmental susceptibility to caffeine. However, it is not induced by caffeine, and its expression is much less in larval detoxification tissues than other Dro5 genes (Leader *et al.* 2018), suggesting its role in caffeine detoxification in wild-type animals may be less important than *CG13659* (Dro5-7; Table 4.3).

In comparison, the role of *CG31370* (Dro5-8) in caffeine detoxification is much less clear. A naturally occurring loss-of-function allele, $CG31370^{del}$, is associated with reduced adult survival on caffeine in the DGRP (Fig. 4.7B), although it does not affect developmental survival on a lower concentration of caffeine (Fig. 4.7C). However, misexpression of *CG31370* increased—not decreased—developmental susceptibility to caffeine and there is not yet robust evidence that disruption of *CG31370* increases developmental susceptibility to caffeine. *CG31370* is also not transcriptionally induced by larval ingestion of caffeine (Table 4.3), and it is not a good detoxification candidate gene more generally (Chapter 2). If disruption of *CG31370* function is not the cause of $CG31370^{del}$'s association with caffeine susceptibility, the allele may affect the expression of another Dro5 gene, such as its upstream paralog *CG13659* (Dro5-7), by disrupting enhancers present in the first exon of *CG31370*. While basal expression of *CG13659* does not appear to be affected in homozygous $CG31370^{del}$ females (mean $\log_2$(FPKM) difference: 0.115, 95% CI: -0.153, 0.518) or males (mean $\log_2$(FPKM) difference: -0.065, 95% CI: -0.303, 0.337), the allele could conceivably affect *CG13659*'s transcriptional induction by caffeine. Alternatively, *CG31370* may be involved in an adult-specific caffeine detoxification pathway, a hypothesis that was not tested in this work due to a focus on developmental toxicology experiments.

Table 4.3. Collated evidence for the involvement of individual *Drosophila melanogaster* Dro5 EcKLs in caffeine detoxification.

Gene	Nom[a]	Induction (Z)[b]	Induction (K)[c]	DGRP[d]	Dro5^{A3-B7} KO[e]	Single KO[f]	ME[g]
CG31300	Dro5-1	no	no	-	yes	maybe	yes
CG31104	Dro5-2	strong	weak	-	yes	-	no
CG13658	Dro5-5	weak	no	-	yes	maybe	no
CG11893	Dro5-6	weak	no	-	yes	yes	-
CG13659	Dro5-7	strong	weak	-	yes	yes	yes
CG31370	Dro5-8	no	no	yes	-	maybe	no
CG31436	Dro5-10	weak	weak	-	-	-	-

[a] *Drosophila* EcKL nomenclature; see Chapter 2
[b] Induction by 1,553 μg/mL caffeine in Zhuo (2014) dataset; strong, > 3-fold; weak, ≤ 3-fold; see Chapter 4.1.4
[c] Induction by 1,500 μg/mL caffeine in Kee & Robin (unpublished) dataset; strong, > 3-fold; weak, ≤ 3-fold; see Chapter 4.1.4
[d] Association of genetic variation with caffeine-related phenotypes in the DGRP; -, no associated variation; see Chapter 4.3.3
[e] Disrupted in the *Dro5^{A3-B7}* allele, which is associated with increased developmental susceptibility to caffeine; -, not disrupted; see Chapter 4.3.4
[f] Single-gene disruption increases developmental caffeine susceptibility; -, not determined; see Chapter 4.3.4
[g] Transgenic misexpression (ME) of the gene with the *HR-GAL4* driver decreases developmental caffeine susceptibility; -, not determined; see Chapter 4.3.5

4.4.2. Caffeine toxicology and metabolism in *Drosophila*

As just discussed, the data in this chapter demonstrate that at least some Dro5 EcKLs—specifically *CG13659* and perhaps others—modulate the developmental susceptibility of *D. melanogaster* to caffeine. However, the specific molecular mechanism for this was not experimentally determined; indeed, there is much that is unknown about the toxicology of caffeine in *D. melanogaster* and in insects more broadly, making the possible mechanisms through which Dro5 genes are involved in caffeine detoxification somewhat open to speculation.

The molecular targets of caffeine have been comprehensively studied in humans and other vertebrates (Fredholm *et al.* 1999), but the same is not true in insects—while it is known that caffeine has acute effects on the insect nervous system (Mustard 2014), as well as chronic effects on insect development (Nathanson 1984; Nigsch *et al.* 1977), the molecular causes of caffeine toxicity in *D. melanogaster* and other insects are not well understood. Molecular targets of caffeine in the insect nervous system include the ryanodine receptor and phosphodiesterases, and possibly also adenosine receptors (the main neurological target in mammals) and dopamine receptors (Mustard 2014), some or all of which are likely responsible for caffeine's acute effects on behaviour and physiology (Nathanson 1984). Caffeine also inhibits proteins involved in DNA repair (Blasina *et al.* 1999; Tsabar *et al.* 2015; Zelensky *et al.* 2013) and increases the mutation rate *in vivo* (Kuhlmann *et al.* 1968), and *D. melanogaster* mutants for genes with functions in genome stability are highly developmentally sensitive to caffeine (Li *et al.* 2013), strongly suggesting exposure to caffeine indirectly causes DNA damage *in vivo*; this mechanism is likely partially responsible for the chronic developmental toxicity of caffeine. Feeding on food containing high concentrations of caffeine also causes death in *D. melanogaster* adults in 15–112 hours (Najarro *et al.* 2015), although the molecular causes of this have not been studied in detail, despite the validation of tolerance loci likely involved in detoxification (Najarro *et al.* 2015).

The metabolism of caffeine in insects is also poorly understood in comparison to that in mammals; the only detailed work so far comes from Coelho *et al.* (2015), who studied caffeine metabolism in adult *D. melanogaster* using a radiolabelled thin-layer chromatography (TLC) method. *D. melanogaster* metabolises caffeine to theobromine, paraxanthine and theophylline (all known mammalian metabolites), as well as five

unidentified metabolites (M1–5), one of which—M2—is the second-most abundant metabolite overall, accounting for a third of all ingested caffeine; 1,3,7-trimethyluric acid, a common metabolite in mammals and the only characterised caffeine metabolite with a hydroxyl group (in its enol form), does not appear to be produced in *D. melanogaster*. This work also implicated three P450 genes in caffeine metabolism based on their transcriptional induction by caffeine, as well as RNAi knockdown experiments that measured the effect of depleting each gene's transcript level with changes in metabolite abundance—*Cyp6d5* was implicated in both the production of theobromine and the inferred downstream metabolism of M2 to another unknown metabolite; *Cyp6a8* was implicated in the inferred downstream metabolism of theobromine, M2 and theophylline; and *Cyp12d1* was implicated in the production of M2 (Coelho *et al.* 2015). Importantly, *Cyp12d1* and *Cyp6d5* were found to be likely caffeine tolerance genes in adult flies by QTL mapping and RNAi knockdown in an independent study (Najarro *et al.* 2015).

The work by Coelho *et al.* (2015) was limited by the use of TLC and RNAi—which suffer from a number of drawbacks compared to more powerful methods like LC-MS, gene misexpression and CRISPR-Cas9 mutagenesis—and its focus on adult flies, and it leaves many questions unanswered that might be important for understanding how EcKLs—putative detoxicative kinases—might be involved in caffeine detoxification. First, it is unclear how the distinct molecular actions of caffeine and/or its metabolites exert toxicological effects at different developmental stages (i.e. larvae, pupae and adults). We also don't yet understand where caffeine metabolism occurs in the body, how it differs between life stages, nor which the metabolites are produced and their relative toxicities compared to caffeine. Given that some caffeine metabolites can produce similar physiological effects to caffeine *in vivo* (Benowitz *et al.* 1995; Geraets *et al.* 2006; Malki *et al.* 2006), we cannot assume that every step in the metabolism of caffeine is 'detoxification' *per se*, and it is possible that some caffeine metabolites might be as or more toxic in some contexts than caffeine (Geraets *et al.* 2006). Such hypothetical 'toxic caffeine metabolites' (TCMs) would also need to be detoxified and/or excreted rapidly to ameliorate their adverse effects and may contribute to the overall toxicity of caffeine at different life stages.

As the experiments in this chapter used developmental mortality as the toxicological endpoint in *D. melanogaster*, the effects of caffeine on development are of particular

interest to discuss. Genetic modulation of Dro5 function in this chapter seemed to mostly affect pupal mortality but not larval mortality—only $Dro5^{A3\text{-}B7}$ mutants appeared to be significantly more sensitive to caffeine during the larval stage compared to wild-type animals, while misexpression of *CG31300* (Dro5-1) and *CG13659* (Dro5-7), which decreased both pupal and total susceptibility to caffeine, did not appear to affect larval susceptibility compared to the misexpression of other genes. Similarly, putting the $CG11893^{MB}$ and $CG13659^{38}$ alleles in *trans* with the $Dro5^{A3\text{-}B7}$ allele also increased pupal and total susceptibility while failing to significantly affect larval susceptibility. Overall, this suggests that Dro5 genes are mostly affecting a mechanism of action for caffeine that disproportionately causes mortality during metamorphosis and not during larval development. Most of the pupal mortality observed at doses of caffeine close to the pupal LC_{50}—but below the larval LC_{50}—was at the pharate adult stage, with many animals either failing to initiate eclosion or dying stuck halfway out of the puparium. This appears to phenocopy *D. melanogaster* mutants for genes involved in genome stability (*Smc5*, *Smc6* and *MAGE*), which experience extensive mortality during metamorphosis after larval development on food containing 48.6–388.4 μg/mL (0.25–2 mM) caffeine, due to widespread apoptosis in the imaginal discs caused by DNA damage (Li *et al.* 2013). This suggests that Dro5 genes may interact with caffeine's effect on DNA repair pathways, possibly by lowering the active concentration of caffeine or TCMs available to reach the imaginal discs, through detoxicative phosphorylation.

While pupal mortality after caffeine ingestion is likely due to damage to imaginal discs during larval stages, the cause of larval mortality on higher doses of caffeine is unclear, although it is likely due to neurological effects through one of numerous receptors causing paralysis or the cessation of feeding (Mustard 2014). If this is true, the observation that genetic modulation of Dro5 function had little effect on larval mortality—except for the $Dro5^{A3\text{-}B7}$ allele—suggests these genes do little to protect the nervous system from caffeine and its metabolites. That damage to the imaginal discs—which likely causes the observed pupal mortality—is occurring during larval development as well suggests one of two hypotheses for why Dro5-mediated detoxification cannot adequately protect the CNS: one, because the genes are not expressed in the CNS; and two, because the compounds on which Dro5 enzymes act have less of an effect on the CNS than DNA repair mechanisms.

The first hypothesis does not appear likely. While the disparity in pupal-larval susceptibility between Dro5 misexpression and *Dro5*$^{A3\text{-}B7}$ genotypes could conceivably be explained by where the *HR-GAL4* driver is expressed—the midgut, Malpighian tubules and fat body (Chung *et al.* 2007)—misexpression of detoxification genes with the *HR-GAL4* driver can protect animals from toxic compounds with CNS-specific modes of action, such as imidacloprid and DDT (Daborn *et al.* 2007; Denecke *et al.* 2017b), suggesting that caffeine/TCM metabolism in one or more of these 'detox' tissues might be sufficient to protect the CNS. This hypothesis also fails to explain why *CG13659* (Dro5-7) mutants are not significantly more susceptible to caffeine as larvae compared to a control genotype—*CG13659*, the gene most likely to be involved in caffeine detoxification in this study (Chapter 4.4.1), is known to be expressed in the blood-brain barrier (DeSalvo *et al.* 2014), suggesting it has the capacity to protect the CNS directly, regardless of its expression in other detoxification tissues.

The second hypothesis for the disparity—that some caffeine metabolites affect DNA repair more than the nervous system—may be more plausible. As noted in Chapter 4.1.4, there is an important wrinkle in the hypothesis that Dro5 genes are involved in caffeine detoxification: it is unlikely that Dro5 enzymes, or any EcKL, can directly metabolise caffeine, as its chemical structure lacks the hydroxyl moieties required for phosphorylation (Fig. 4.2). However, some caffeine metabolites—potentially TCMs—may contain such moieties, as P450s are required for the production of as-yet-uncharacterised caffeine metabolites (Coelho *et al.* 2015), which often (although not always) act as monooxygenases (Guengerich 2001). As such, my proposed working hypothesis for the observation that modulation of caffeine susceptibility upon knockout or misexpression of Dro5 genes, such as *CG13659*, is that EcKL enzymes phosphorylate certain hydroxylated TCMs, reducing the amount of DNA damage they cause in the imaginal discs and increase developmental survival; these specific TCMs have a greater effect on DNA repair mechanisms than the nervous system, explaining the differential effects of Dro5 manipulation on pupal and larval mortality (Fig. 4.17).

Alternatively, high mortality during pupal, but not larval, stages might also be explained by the lack of excretion of caffeine or TCMs in sessile pupae, leaving them dependent on metabolic detoxification by EcKL enzymes. While the half-life of caffeine and its metabolites in *D. melanogaster* is not known, it is between 0.7–5 hours in mammals, with smaller animals generally having shorter half-lives (Nehlig 1999); it is

arguably unlikely that ingested caffeine appreciably persists in *D. melanogaster* in the wandering period between cessation of feeding and pupariation, which is approximately 12 hours at 25 ºC (Warren *et al.* 2006). However, as the pharmacokinetics of caffeine is unstudied in this species, I cannot rule out that accumulation of caffeine or its metabolites does not occur during metamorphosis.

In addition, the data presented in this chapter cannot exclude the possibility that Dro5 genes are not acting in the direct detoxification of caffeine or its metabolites, but are instead involved in a tolerance process that reduces the harmful effects of caffeine ingestion, in a similar manner as hypothesised for *CG33301* and *CG16898* and their association with methylmercury tolerance (Chapter 2.7.1).

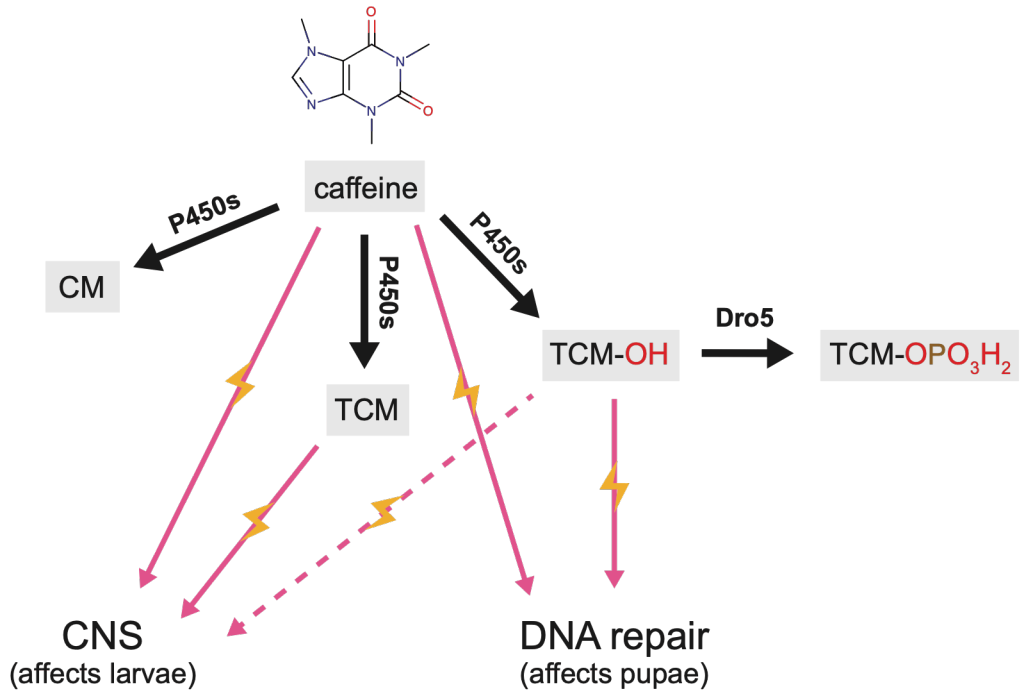


Figure 4.17. Working hypothesis for the mechanism behind modulation of caffeine susceptibility by Dro5 genes in *Drosophila melanogaster*. Caffeine is metabolised (black arrows) by P450s into multiple metabolites: CM (a non-toxic metabolite), TCM (a toxic metabolite) and TCM-OH (a hydroxylated toxic metabolite). Caffeine, TCM and TCM-OH all cause toxic effects (pink arrows with lightning bolts), but through slightly different mechanisms: caffeine affects both the CNS (causing larval mortality) and DNA repair (causing pupal mortality), while TCM affects only the CNS, and TCM-OH mostly affects DNA repair. Dro5 enzymes only phosphorylate (black arrow) and detoxify TCM-OH, which leads to a large reduction in pupal mortality but only a minor reduction in larval mortality.

4.4.3. Is caffeine detoxification in *Drosophila* ecologically relevant?

Caffeine is found in the leaves, fruits, seeds and/or flowers of a variety of plants, including species in the genera *Coffea*, *Camellia*, *Theobroma*, *Paullinia*, *Cola*, *Ilex* and *Citrus* (Anaya *et al.* 2006), and is a natural part of the diet of a number of insects, including coffee bean pests (Ceja-Navarro *et al.* 2015) and bees (Couvillon *et al.* 2015). *D. melanogaster* feeds not on fresh plant material but rotting fruit substrates (Markow & O'Grady 2007), which are unlikely to originate from the small, caffeine-rich fruits found in the *Coffea*, *Cola* and *Paullinia* genera. However, *Citrus* spp. are known to produce highly favourable host fruits for *D. melanogaster* (Dweck *et al.* 2013)—while caffeine is found in *Citrus* flowers, not fruits (Kretschmar & Baumann 1999), *Citrus* trees typically produce large numbers of flowers (Iglesias *et al.* 2007), raising the possibility that fruits and flowers decompose together, forming a developmental substrate for *D. melanogaster* larvae containing toxicologically relevant levels of caffeine. Whole *Citrus* flowers contain approximately 318 nmol/g (62 μg/g) caffeine (Kretschmar & Baumann 1999), meaning that a 1:1 flower to fruit ratio—a plausible upper limit for what might be found in nature—would produce a developmental substrate with only 31 μg/g caffeine, well below the developmental LC_{50}s determined in this study. However, such doses might produce adverse behavioural or developmental effects in natural environments, especially for non-adapted genotypes, producing a large enough fitness cost to select for efficient caffeine detoxification. A diverse collection of *Drosophila* species other than *D. melanogaster* use *Citrus* spp. fruits as developmental substrates in nature, including *D. simulans*, *D. pseudoobscura*, *D. persimilis*, *D. willistoni*, *D. immigrans* and *D. hydei* (Hoenigsberg *et al.* 1977), suggesting the ability to detoxify caffeine and related methylxanthine compounds may have been present in the ancestor of all or most *Drosophila* species.

Alternatively, the ability for *D. melanogaster* to detoxify caffeine might be a by-product of detoxification pathways for other, possibly unrelated compounds. *D. melanogaster* can detoxify a number of compounds that it was unlikely to encounter in nature before the arrival of human civilisation, including the plant alkaloid nicotine (Highfill *et al.* 2017), and a number of synthetic insecticides (Denecke *et al.* 2017b; Green *et al.* 2019; Joußen *et al.* 2008), indicative of a large detoxification breadth (DB) adaptive for its chemically unpredictable ecological niche (Chapter 4.1.2).

4.4.4. Dro5 EcKLs and other plant toxins

In this chapter, I did not find evidence that quercetin, escin, esculin, curcumin or salicin are developmentally toxic to *D. melanogaster* when ingested at the concentrations used, nor that any of the five genes disrupted by the $Dro5^{A3\text{-}B7}$ allele are required for survival on those doses. The work on the first four of these compounds was limited by their relatively low solubility in ethanol, and it is possible that some or all of these compounds are developmentally toxic to *D. melanogaster* when ingested at higher concentrations. These compounds were chosen because they contain hydroxyl groups, which made them plausible substrates for EcKL enzymes. There may be redundancy between all seven Dro5 genes or other EcKLs with respect to detoxifying these compounds, or they might be detoxified by other mechanisms, such as glycosylation or oxidation—previous work has shown that quercetin can be metabolised by insect prophenoloxidases *in vitro*, suggesting these enzymes may be responsible for its detoxification *in vivo* (Wu *et al.* 2015). Alternatively, these compounds simply may not need to be detoxified at the doses tested.

Salicin and other phenolic glycosides are good candidates for EcKL substrates, given they are detoxified by phosphorylation in Lepidoptera (Boeckler *et al.* 2016); however, I found in this chapter that concentrations of salicin up to 8,000 μg/mL (27.9 mM) were not developmentally toxic to wild-type or $Dro5^{A3\text{-}B7}$ *D. melanogaster* genotypes, which did not allow me to test whether knockout or misexpression of any Dro5 EcKLs modulates susceptibility to salicin. Based on these data, *D. melanogaster* appears broadly tolerant of salicin, although whether this is due to efficient metabolic detoxification or other tolerance mechanisms is currently unclear. Some *Drosophila* species have diets high in salicinoids, such as those in the virilis group (which includes *D. virilis*), which feed on decaying willow and poplar trees (Markow & O'Grady 2007; Spieth 1979). However, there are no conspicuously expanded EcKL clades in *D. virilis* (Chapter 2.4.1), meaning there are no clear candidates for salicinoid-detoxifying clades in this gene family.

Experiments in this chapter also failed to detect a difference in developmental viability between wild-type and $Dro5^{A3\text{-}B7}$ *D. melanogaster* genotypes when feeding on semi-natural fruit substrates made from citrus juices. As previously discussed (Chapter 4.4.3), decomposing citrus fruits are natural hosts for *D. melanogaster* and other *Drosophila*

spp. (Dweck *et al.* 2013; Hoenigsberg *et al.* 1977), suggesting they are well adapted to tolerate secondary metabolites present in these fruits; however, whether this tolerance is due to metabolic detoxification or simply that the secondary metabolites are not especially toxic at natural levels of exposure is unclear.

4.4.5. Dro5 EcKLs and kojic acid

An experiment in this chapter found that *Dro5*$^{A3\text{-}B7}$ mutant animals are more developmentally susceptible to kojic acid than wild-type controls, at both larval and pupal stages. This suggests that one or more Dro5 genes may be involved in the detoxification of kojic acid, a dihydroxylated compound that could hypothetically be phosphorylated by an EcKL enzyme (Fig. 4.2). However, neither single-gene complementation nor misexpression experiments were conducted to follow up on this phenotype.

Kojic acid was selected to use in experiments because it is both known to be toxic to *D. melanogaster* (Dobias *et al.* 1977) and produced as a secondary metabolite of known fungal competitors of *Drosophila* larvae (El-Kady *et al.* 2014). The concentrations of kojic acid used in experiments in this chapter (up to 8,000 μg/mL, 0.8% w/v, or 56.3 mM), while much higher than the concentrations of other toxins used, are likely to be ecologically relevant, given that many strains of *Aspergillus* spp. and *Penicillium* spp. regularly produce more than 0.5% w/v kojic acid in culture (Beard & Walton 1969; El-Kady *et al.* 2014). As such, it is highly plausible that *D. melanogaster* populations are regularly exposed to toxicologically relevant concentrations of kojic acid and that they have evolved metabolic detoxification mechanisms to increase their tolerance.

Kojic acid's mechanism of action is thought to be the inhibition of tyrosinases, including insect prophenoloxidases, which are enzymes required for development and immune function (Goudru *et al.* 2013; Zhang *et al.* 2019). Essentially nothing is known about the metabolism of kojic acid in insects, although kojic acid is substantially metabolised to sulfate and glucuronide conjugates in rats (Burnett *et al.* 2010), suggesting similarly conjugation-heavy metabolism could occur in insects. Phosphorylation—or indeed any conjugation reaction—would likely disrupt the binding of kojic acid to its target enzymes, rendering it less toxic and also facilitating its excretion. Whether or not this occurs *in vivo*—and whether Dro5 enzymes are indeed involved—remains to be determined.

4.4.6. Are all Dro5 EcKLs detoxification genes?

This chapter has presented multiple lines of genetic evidence that at least some genes in the Dro5 clade of EcKLs in *D. melanogaster* have detoxification—or X-class (Chapter 2.2; Gotoh 2012)—functions. But do all Dro5 genes to encode X-class enzymes? Or might some encode enzymes with E-class (endogenous substrate) or S-class (secondary metabolite substrate) functions?

Five of the seven Dro5 genes in *D. melanogaster* were predicted to be detoxification genes in Chapter 2, excluding *CG31370* (Dro5-8) and *CG31436* (Dro5-10), which were not induced by any of the xenobiotic datasets explored in that analysis. These latter two genes have broader patterns of tissue expression than other Dro5 genes (Fig. S2.4), but this is not strong evidence of a lack of an X-class function, and it is possible one or both of these genes are transcriptionally induced by as-yet-unidentified xenobiotic compounds. Additionally, *CG31370* may be implicated in adult caffeine tolerance, but this has yet to be independently validated outside of the DGRP. Despite this, neither gene has an obvious role in development, given that their loss-of-function TE-insertion alleles are fully complemented by the *Df(3R)BSC852* deficiency, which is a likely loss-of-function allele for the entire Dro5 clade in *D. melanogaster*. As such, the functions of these genes remain unknown.

Curiously, misexpression of *CG31104* (Dro5-2) and *CG13658* (Dro5-5) using the strong, ubiquitous *tub-GAL4* driver resulted in pharate adult lethality, with undifferentiated abdomens (Chapter 4.3.2); this appears to be at odds with the predicted detoxification functions of these genes (Chapter 2), although neither of these genes have strong lines of genetic evidence for a detoxification function like their paralog *CG13659* (Table 4.3). Misexpression of X-class P450s typically does not result in developmental arrest (Daborn *et al.* 2007), although it can sometimes occur, such as in the case of ubiquitous misexpression of *Cyp12a4*, which results in embryonic lethality (Bogwitz *et al.* 2005). In general, such misexpression-induced developmental arrest may be due to promiscuous action of a detoxification enzyme on an essential endogenous substrate that it typically does not encounter in its natural sites of expression, or the over-production of a metabolite that is toxic at high concentrations. Indeed, if detoxification enzymes have sufficiently broad substrate specificity, which may be adaptive to increase detoxification breadth (DB) for the organism, it is likely that they can act on some

endogenous substrates and high levels of ectopic expression may reduce fitness. *CG31104*, and to a lesser extent *CG13658*, are expressed in a relatively restricted number of tissues during larval and adult stages (Fig. S2.4)—of note, both genes show extremely low expression in the fat body, ecdysteroid signalling in which is required for the correct formation of the abdomen during metamorphosis (Bond *et al.* 2010); given that at least one EcKL phosphorylates ecdysteroids (Sonobe *et al.* 2006), it is possible that *CG31104* and *CG13658* misexpression in the fat body disrupts ecdysteroid signalling, producing the abdomen-specific morphological defects observed during ubiquitous misexpression. Consistent with this hypothesis, misexpression of *CG31104* (but not *CG13658*) using the *HR-GAL4* driver, which expresses GAL4 in the midgut, Malpighian tubules and fat body (Chung *et al.* 2007), resulted in a modest amount of pupal mortality on caffeine-free media (Fig. 4.12); the incomplete penetrance of this lethality, and the lack of lethality from *HR-GAL4* misexpression of *CG13658*, may be due to the difference in strength between the two drivers, or the incomplete persistence of *HR-GAL4*-driven expression in the fat body throughout metamorphosis. Another possible explanation for why misexpression of these two genes causes developmental defects is that they upset the dimeric stoichiometry of an essential EcKL by forming artificial heterodimers; discussion of the possible quaternary structure of EcKLs will be conducted in Chapter 6.6.

Dro5 EcKLs are found in all 12 *Drosophila* species studied in Chapter 2, and the broader order-ancestral clade they belong to—Dip1—is relatively well conserved across Diptera, with complete clade loss only found in three species of *Glossina*, haematophagous insects with very small predicted DB (Chapter 3.3.4). Within *Drosophila*, Dro5 genes are very evolutionarily dynamic, with extensive duplication and loss over time; they also have the largest difference in gene number between species, with nine genes in *D. simulans* and one gene in *D. mojavensis*. It is possible, however, that Dro5 genes might have similar functions to genes in the related clades Dro6, Dro7 and Dro39, all of which are also in Dip1—combining the total Dip1 gene count shows that *D. pseudoobscura*, *D. persimilis* and *D. grimshawi* have the fewest genes in this clade, with only three each (Fig. 3.12). *D. mojavensis* (seven Dip1 genes) is a cactophilic species that is exposed to many alkaloids in its developmental substrates (Markow & O'Grady 2007)—given that Dro5 genes have been linked to the detoxification of another alkaloid, caffeine, in this chapter, it is possible related EcKLs may be involved in detoxifying cactus-derived compounds; however, no associations between EcKLs and host

cactus use have been reported (Matzkin 2012). It is also interesting to note that *D. sechellia*—a *Morinda* fruit specialist (Yassin *et al.* 2016)—has pseudogenised its Dro5-5 (*CG13658*) and Dro5-6 (*CG11893*) orthologs (Table S2.1), raising the possibility that its diet may lack certain Dro5 substrates, leading to relaxed selection on certain genes in the clade.

Given that the Dip1 clade varies so widely in size between species, and is linked to detoxification in *D. melanogaster*, it is tempting to speculate that its members have detoxification functions across Diptera, as has been raised previously (Chapter 3.4.6). It is unclear if detoxification is the ancestral function of Dip1 genes, but the fact that the clade has been completely lost in some species, and that most of its constituent clades in *Drosophila*, specifically Dro6, Dro7 and Dro39, have undergone multiple complete loss events in this genus, means it is unlikely that the clade as a whole has a developmentally essential ancestral function that is conserved in at least one member per genome. However, this does not preclude the evolution of derived, essential functions for specific genes in certain taxa, which is likely to occur with some frequency in large gene families (Sezutsu *et al.* 2013).

4.4.7. Future research directions for Dro5 genes and detoxification

The obvious way to build on the work in this chapter is to complete the sets of experiments that were planned for the caffeine toxicology assays, including misexpressing *CG11893* and *CG31436* with *HR-GAL4*, and generating and characterising CRISPR-Cas9 knockout alleles for *CG31104*, *CG31370*, *CG31436* and/or the entire Dro5 clade with respect to caffeine susceptibility. Some experiments would ideally also be repeated with more appropriate (typically higher) dose ranges, including *Df(3R)BSC852* and single-gene complementation of the $Dro5^{A3\text{-}B7}$ allele, as well as *HR-GAL4*-driven misexpression with respect to larval mortality. Other experiments that would build on this work on caffeine detoxification would be misexpressing—and/or somatically disrupting (Port *et al.* 2020)—Dro5 genes with other GAL4 drivers, such as those expressing in just the midgut, Malpighian tubules, fat body or blood-brain barrier, to determine if Dro5 modulation of caffeine susceptibility is tissue-specific; it would also be worth exploring—with either qPCR or a gene trap reporter allele (Li-Kroeger *et al.* 2018)—if caffeine ingestion induces *CG13659* in a tissue-specific manner. Adult survival assays would also be important for properly testing the relationship between

CG31370^{del} and caffeine susceptibility, as well as testing if larval and adult metabolism of caffeine is similar. To test the hypothesis that Dro5 enzymes act on hydroxylated caffeine metabolites (Fig. 4.17), epistasis experiments could also be performed with P450 genes implicated in caffeine metabolism, such as *Cyp6d5* and *Cyp12d1*. Ultimately, radiolabelled or isotope-labelled caffeine metabolite tracing combined with Dro5 knockout or misexpression might be able to determine if phosphate conjugates of caffeine metabolites are indeed produced by Dro5 enzymes.

With respect to other xenobiotic compounds, the relationship between kojic acid and Dro5 genes would likely be a fruitful avenue for future research. RNA-seq or qPCR experiments could show whether any Dro5 genes are induced by ingestion of kojic acid, and more specific knockout and misexpression toxicological assays could narrow down the candidate genes as has been done with caffeine and *CG13659*. Salicin, and other alcoholic glycosides like helicin and arbutin, remain top candidates for Dro5 substrates, and similar experiments could be performed for these compounds as has been proposed for caffeine and kojic acid, if these compounds end up being toxic to *D. melanogaster* at ecologically plausible doses.

It is also tempting to explore the relationship between Dro5 genes and the secondary metabolism of filamentous fungal competitors, such as *Aspergillus* spp. Experiments were planned for this chapter—which ultimately didn't progress past pilot stages—that would have tested the susceptibility of Dro5 knockout and misexpression animals to feeding on wild-type or *ΔlaeA* knockout (secondary metabolite-deficient) strains of *Aspergillus nidulans* and *Aspergillus niger* (Kempken & Rohlfs 2010; Trienens *et al.* 2017), testing the hypothesis that these EcKLs are involved in the detoxification of insecticidal defence compounds synthesised by these fungi. Support for this hypothesis would demonstrate that Dro5 genes, and EcKLs in general, are required for ecologically relevant detoxification processes in *D. melanogaster*, which remains a poorly studied field.

4.5. Conclusion

This chapter has provided the first experimental evidence that insect EcKL genes are involved in detoxification in the model insect *Drosophila melanogaster*, a key hypothesis of this thesis. Multiple lines of evidence have linked the Dro5 genes—a large, dynamic clade containing many detoxification candidate genes identified in Chapter 2—to the detoxification of the plant secondary metabolite caffeine, and suggest an additional association with the fungal secondary metabolite kojic acid, both of which may be ecologically relevant toxins for *D. melanogaster*. This work lays the groundwork for future research into detoxicative kinases in *Drosophila* and may lead to a deeper understanding of caffeine toxicology in insects.

4.6. Supplementary Figures

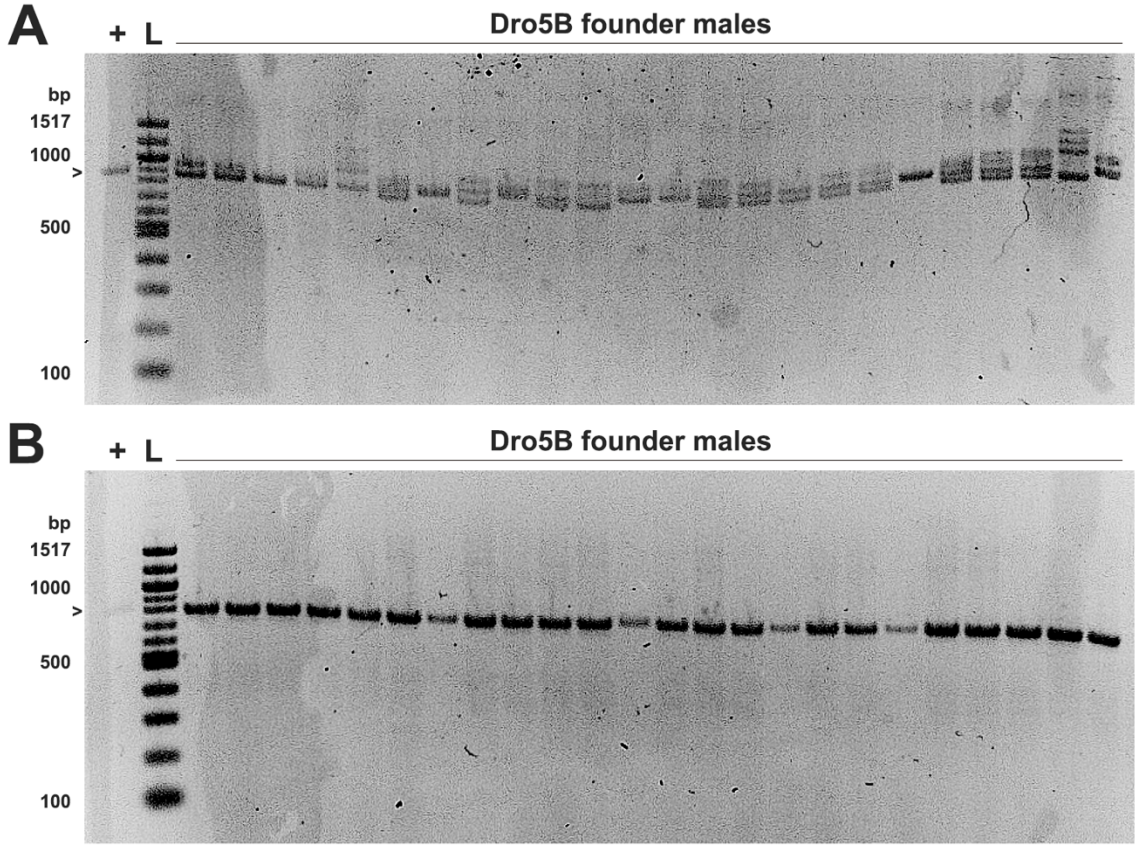


Figure S4.1. Representative agarose gel electrophoresis images of PCR products from Dro5B genotyping reactions using DNA from 24 founder males that are heterozygous for putatively mutagenised Dro5B loci, and a wild-type control (+). L, NEB 100 bp ladder (N3231). (A) PCR with the D5ΔB_1F/D5ΔB_1R primer pair, which flanks the first two gRNA target sites (Bg1 and Bg2) of pCFD6-Dro5B and produces an 859 bp amplicon (>) from wild-type chromosomes. (B) PCR with the D5ΔB_2F/D5ΔB_2R primer pair, which flanks the last two gRNA target sites (Bg3 and Bg4) of pCFD6-Dro5B and produces a 792 bp amplicon (>) from wild-type chromosomes. Successful CRISPR-Cas9 mutagenesis produces small indels, resulting in complex heteroduplex bands, which can be seen in (A) but not (B), indicating that Bg3 and Bg4 fail to cut.

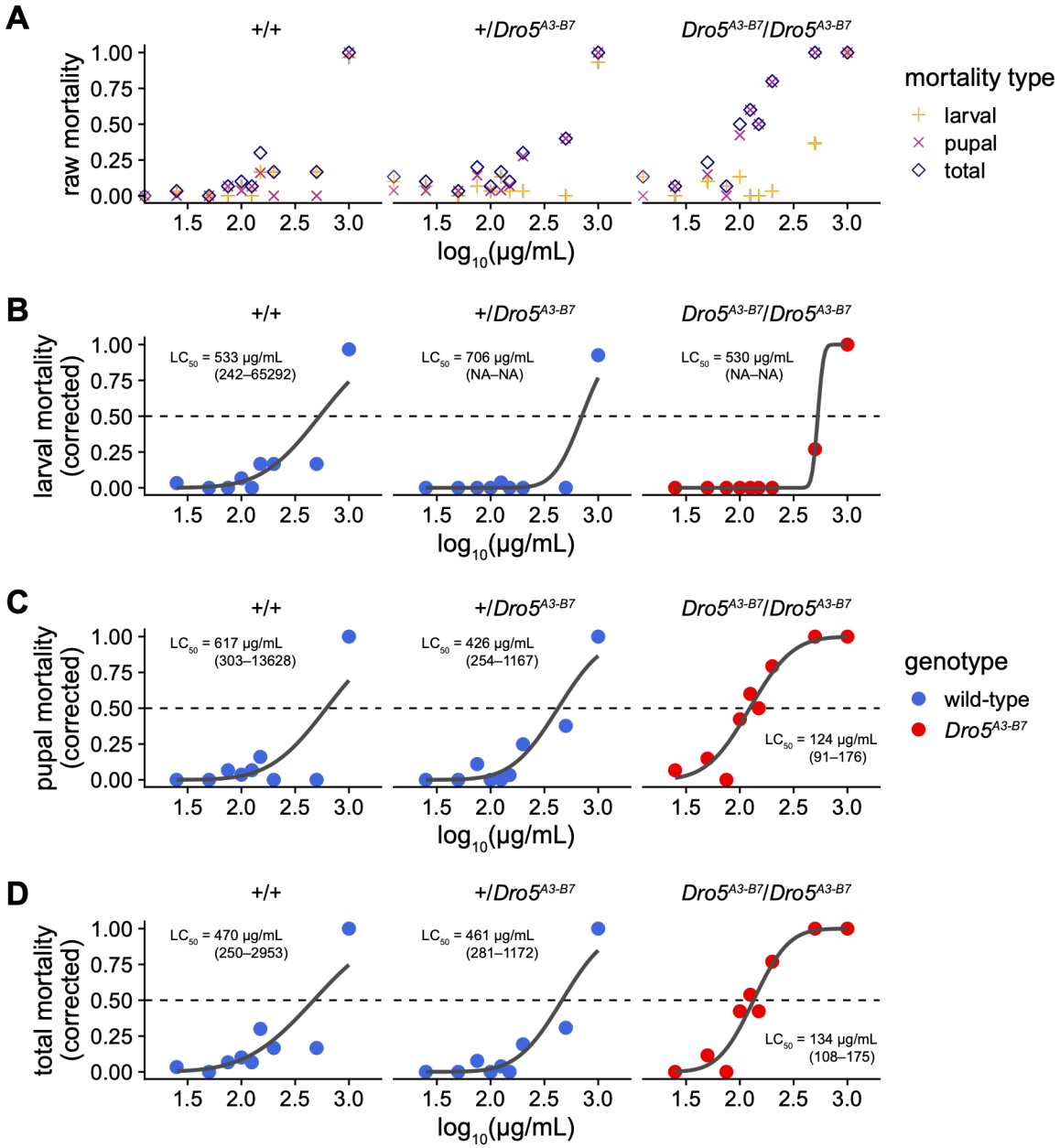


Figure S4.2. Developmental mortality of +/+ homozygotes, +/*Dro5*$^{A3\text{-}B7}$ heterozygotes and *Dro5*$^{A3\text{-}B7}$/*Dro5*$^{A3\text{-}B7}$ homozygotes on media containing 25–1,000 µg/mL caffeine. (A) Raw proportional mortality, with control mortality (0 µg/mL caffeine) on the left-most side of each plot (transparent). (B–D) Schneider-Orelli-corrected (B) larval, (C) pupal and (D) total mortality on non-zero doses, with probit regression curves for each genotype. LC_{50}s are written as point estimates with 95% CIs; values are 'NA' when they could not be calculated from the data.

Chapter 5

Functional genetic analyses of ecdysteroid kinase candidate EcKLs in *Drosophila melanogaster*

5.1. Introduction

5.1.1. Identifying and characterising ecdysteroid kinase-encoding genes

Ecdysteroid phosphorylation— catalysed by ecdysteroid kinases (EcKs)—inactivates ecdysteroids, leading to localised and/or global reductions in ecdysteroid signalling. However, as discussed in Chapters 1.3.4–7, the specific physiological functions of ecdysteroid phosphorylation are largely unknown, except for ovarian/embryonic ecdysteroid-phosphate (OEEP) recycling (Chapter 1.3.6) and possibly also phytoecdysteroid detoxification (Rharrabe *et al.* 2007). Given that ecdysteroids control many aspects of insect biology (Chapter 1.3.2), tissue-specific ecdysteroid phosphorylation could, in principle, regulate many physiological and developmental processes by producing localised decreases in free ecdysteroid concentration; this has been hypothesised as a reason that EcR/USP signalling is highly tissue-specific in *D. melanogaster* adults (Schwedes *et al.* 2011; Schwedes & Carney 2012). Understanding the functions of ecdysteroid-phosphate conjugates is likely to lead to further understanding about the regulation of ecdysteroid signalling in insects.

Very few EcKs have been biochemically characterised, and only one—BmEc22K, an ecdysteroid 22-kinase central to the OEEP system of *Bombyx mori* (Sonobe *et al.* 2006)—has been cloned and genetically identified. As discussed in Chapter 1.3.7, existing biochemical evidence is consistent with the hypothesis that insect EcKs are encoded by members of the EcKL gene family; EcKLs are therefore great candidates for reverse genetic approaches to explore the biology of ecdysteroid phosphorylation. Similarly to the characterisation of detoxification genes (Chapter 4.1), genetic characterisation of EcK-encoding genes complements biochemical characterisation of EcK enzymes, as phenotypes produced by genetic manipulation can shed light on the specific functions of ecdysteroid-phosphate conjugates that the enzymes produce. Also similarly to detoxification genes (Chapters 1.2.7 & 2), there may be ways to identify candidate EcK-encoding EcKLs using evolutionary and transcriptomic characteristics, including: being close homologs of known EcK genes; transcriptional induction by ecdysteroids, or being otherwise regulated by ecdysteroid-response pathways; and being conspicuously expressed in tissues related to ecdysteroidogenesis or ecdysteroid signalling at particular developmental timepoints. Once candidate EcK-encoding genes are found, they can be functionally validated by reverse genetic techniques, including gene

knockout, knockdown and/or misexpression, with the expectation that these may arrest or delay development, or produce defects in other ecdysteroid-related processes such as reproduction, behaviour or immunity. EcK mutants or misexpression animals might also phenocopy the genetic manipulation of other ecdysteroid catabolism genes, such as *Cyp18a1* (Guittard *et al.* 2011; Rewitz *et al.* 2010).

5.1.2. Ecdysteroid phosphorylation in *Drosophila melanogaster*

Drosophila melanogaster (Diptera: Drosophilidae) is the best-studied model insect in biology, but this is poorly reflected in our understanding of the biochemistry of ecdysteroid phosphorylation in this species (Chapter 1.3.6), at least compared to other insects such as *B. mori* (Lepidoptera: Bombycidae), *Manduca sexta* (Lepidoptera: Sphingidae) and *Schistocerca gregaria* (Orthoptera: Acrididae; Isaac & Rees 1984; 1985; Sonobe *et al.* 2006; Thompson *et al.* 1988; 1987a). Despite this, the powerful tools in *D. melanogaster* may be well positioned to enable the genetic identification of EcKs in this species.

Only two ecdysteroid-phosphate conjugates have been confidently identified in *D. melanogaster*: ecdysone (E) 22-phosphate in the follicle cells of the adult ovary, which is bound to an uncharacterised 50 kDa protein (Grau *et al.* 1995; Pis *et al.* 1995); and a 26-phosphate conjugate of 20,26-dihydroxyecdysone that can be formed in the S2 cell line (Guittard *et al.* 2011). Two other conjugates have been less-confidently identified in 3rd-instar larvae: 3-oxo-E 2-phosphate may be formed in response to ingestion of 20E (Hilton 2004), and 3-epi-20E 3-phosphate may be formed in response to injection of 20E (Sommé-Martin *et al.* 1988a). The higher-level biology of these ecdysteroid-phosphates has not been explored, and the EcKs responsible have not been identified; in addition, as noted in Chapter 1.3.6, it is unlikely that *D. melanogaster* has an OEEP system like other insects, and may instead use ecdysteroid-acyl conjugates as a maternal source of ecdysteroids during embryogenesis (Bownes *et al.* 1988).

The biological importance of ecdysteroid-phosphate conjugates in *D. melanogaster* is further suggested by the presence of a functional ecdysteroid-phosphate phosphatase (encoded by *CG13064*), which has highest activity towards 22- and 2-phosphates (Davies *et al.* 2007). This enzyme has also yet to be functionally characterised using genetic methods (Chapter 1.3.8).

5.1.3. *Wallflower* (*Wall*/*CG13813*) and *Pinkman* (*pkm*/*CG1561*) are *D. melanogaster* orthologs of *BmEc22K*

In Chapter 3, the EcKL gene family was categorised into 13 subfamily clades inferred to be present in the ancestor of all insects, and *BmEc22K* was found to be a member of subfamily A, which has been differentially retained across insect orders. As such, subfamily A genes are good candidates for encoding EcKs, although it is likely other subfamilies also contain EcK genes, as orthopterans lack the A subfamily, yet possess multiple ecdysteroid kinase activities (Chapter 3.3.2).

BmEc22K belongs to the Lep1 ancestral clade in Lepidoptera, one of 10 such clades (Lep1–8, Lep16–17) in subfamily A (Chapter 3.3.5), while Diptera contains four subfamily A ancestral clades: Dip8–11 (Chapter 3.3.4). Diptera subfamily A clades are monophyletic, and do not reliably group with any subfamily A clades in Lepidoptera (Fig. S5.1), suggesting Diptera does not possess direct orthologs of Lep1 genes. *D. melanogaster* possesses two subfamily A EcKLs, *CG13813*—named in this chapter *Wallflower* (*Wall*)—and *CG1561*—recently named *Pinkman* (*pkm*) by Santana *et al.* (2020)—which belong to Dip10 and Dip9, respectively. *Wall* and *pkm* are therefore many-to-many orthologs of *BmEc22K*, as well as many other EcKLs in Lepidoptera, including seven other subfamily A EcKLs in *B. mori*; despite this lack of direct orthology, these two EcKLs are good *a priori* EcK candidate genes in *D. melanogaster*.

Within Diptera, *pkm* orthologs in the Dip9 clade tend to be conserved in single-copy, but have experienced expansions in the Chironomoidea and has been independently lost at least three times, suggesting the clade is not essential in all dipteran species. *Wall* orthologs in the Dip10 clade also tend to be single-copy and have also experienced clade losses—at least once in the ancestor of the *Drosophila* subgenus Drosophila and at least once at the base of the suborder Nematocera—similarly suggesting this clade is not essential in all dipteran species (Fig. 3.12).

5.1.4. Published data on *pkm*

pkm is located on the X chromosome within a large intron of the kinesin-like gene *nod*, and encodes a 635 aa polypeptide with a single C-terminal EcKL domain and—uniquely among EcKLs in *D. melanogaster*—an intrinsically disordered N-terminal

region that spans the first 230 aa (Mészáros *et al.* 2018). *pkm* is strongly expressed in the adult eye, seen in both the FlyAtlas 2 dataset (Leader *et al.* 2018) and an older study that compared gene expression between the heads of wild-type and eyeless flies (Xu *et al.* 2004); it is also expressed a much lower levels in the CNS and fat body of pupae (Graveley *et al.* 2011).

High-throughput proteomic screening has suggested that the Pkm protein can physically interact with CG11200, Hsp23 and Hsp26 (Guruharsha *et al.* 2011), likely via its disordered N-terminal region, which are frequently involved in protein-protein interactions (Mészáros *et al.* 2018). CG11200 is a poorly characterised carbonyl reductase that is annotated as having NADP-retinol dehydrogenase activity in FlyBase (Thurmond *et al.* 2019), based on sequence similarity to the human retinol dehydrogenase *RDH11*, but its closest human ortholog appears to be *DHRSX*, which does not have a known substrate (Zhang *et al.* 2014a). Pkm's interactions with the small heat shock proteins Hsp23 and Hsp26 was noticed by Santana *et al.* (2020), who found that RNAi knockdown of *pkm* in motor neurons increases the mRNA and protein levels of these Hsps in adult heads, and decreases synapse number in larval neuromuscular junctions (NMJs). However, while qPCR and Western blots were performed on adult heads, knockdown experiments quantifying changes in NMJ synapse number were performed in 3rd-instar larvae—*pkm* is barely expressed in larvae (Graveley *et al.* 2011; Leader *et al.* 2018), suggesting it is unlikely to affect synapse development *in vivo* at this stage. As such, while *pkm* may regulate Hsp23 and Hsp26 levels in adults, this has not been conclusively linked to higher-level phenotypes.

Despite this, there are other lines of evidence that suggest *pkm* may function in the nervous system: *pkm* is down-regulated in pupae with loss-of-function alleles in *parkin*, a gene whose human homolog is implicated in Parkinson's disease (Greene *et al.* 2005); a TE-insertion near *pkm* can, in *trans*, suppress a *Smn* model of spinal muscular atrophy (Chang *et al.* 2008)—this TE (d10763) maps to a repetitive region 3–12 kb upstream of *pkm* that is also upstream of *rho-4*, meaning it cannot be definitely linked to *pkm*.

There may also be a connection between *pkm* and starvation pursued herein—Chatterjee *et al.* (2014) report a dataset wherein female adult *D. melanogaster* tissue dissections involving the oenocytes—clusters of cells associated with the fat body in larvae

and the epidermis in adults, which are involved in lipid metabolism and response to starvation (Makki *et al.* 2013)—were transcriptionally profiled under starvation and control conditions. Supplementary materials in the manuscript indicate that *pkm* is induced in oenocyte- and epidermis-containing abdominal cuticle samples, but not muscle, head, fat body or gut samples, under starvation conditions. In addition, *pkm* fails to be induced in abdominal cuticle samples in which a dominant negative isoform of the insulin receptor, InR^{DN}, is expressed specifically in oenocytes, suggesting insulin signalling is responsible for its induction (Chatterjee *et al.* 2014). As such, I hypothesise that *pkm* is involved in the starvation response in adult flies.

Overall, there are no clear links between *pkm* and ecdysteroid biology in published data. However, it is possible it has an important developmental function, perhaps related to the nervous system.

5.1.5. Published data on *Wall*

Wall is located on chromosome 3L, encodes a 430 aa polypeptide with a single EcKL domain, and is predominantly expressed in the midgut, being found near-exclusively in the midguts of 3rd-instar larvae, adult males and adult females, according to FlyAtlas 2 (Leader *et al.* 2018); it is also expressed in the embryonic midgut (Weiszmann *et al.* 2009), and has a high peak of expression 16–18 hr into embryogenesis (Graveley *et al.* 2011) at the same time as *Cyp18a1* (Fig. 5.1), which encodes an ecdysteroid 26-hydroxylase/carboxylase (Guittard *et al.* 2011; Rewitz *et al.* 2010). However, *Wall* is also enriched in the 3rd-instar larval ring gland (RG), with enrichment very high soon after the L2/L3 moult and decreasing over time (Ou *et al.* 2016); this may mean *Wall* is expressed in the prothoracic (ecdysteroidogenic) cells (PG), but it could also mean it is expressed in the corpus allatum cells and/or the corpus cardiacum cells.

Wall is transcriptionally induced 5 hr after the addition of 10^{-6} M 20E to culture media in 40 out of 41 surveyed cell lines (including S2 cells) and is the 8th-most consistently induced gene in the genome, along with known 20E primary-response genes (Stoiber *et al.* 2016); it was also independently found to be induced by 20E in the Kc167 cell line (Gauhar *et al.* 2009). Curiously, *Wall* doesn't appear to be induced by 20E in organ culture, but is induced by RNAi against *EcR* at—and 4 hr before and after—pupariation; this is in contrast with *Cyp18a1*, which is induced by 20E and repressed by *EcR* RNAi (Beckstead *et al.* 2005). *Wall* does not appear to be a primary-response gene to 20E in S2 cells (Mazina *et al.* 2015) and is therefore likely regulated by a 20E primary-response gene, likely *DHR3*, as *Wall* expression in *DHR3* mutant embryos is strongly repressed (-21-fold; Ruaud *et al.* 2009) and *DHR3* is rapidly induced by 20E in S2 cells (Mazina *et al.* 2015).

While there are no known phenotypes associated with *Wall* loss-of-function mutants, an upstream *EYpg2* insertion—*EY20330*—that likely allows for GAL4-mediated misexpression of *Wall* from its native locus was found to enhance the lethality of *D42-GAL4*-mediated motor neuron defects when present in single-copy, despite no lethality resulting from putative misexpression of *Wall* with *D42-GAL4* alone (Chang *et al.* 2013).

Overall, these data, with multiple, independent connections to ecdysteroid biology,

make *Wall* an outstanding EcK candidate gene. In addition, numerous aspects of these data led me to a hypothesis that *Wall* encodes the ecdysteroid 26-kinase responsible for the ecdysteroid 26-phosphorylation identified in S2 cells by Guittard *et al.* (2011), given that *Wall* is one of the only EcKLs appreciably expressed in S2 cells (Graveley *et al.* 2011; Stoiber *et al.* 2016).

5.1.6. Chapter Aim

In this chapter, I aim to test the hypothesis that *Wall* and *pkm*, two orthologs of the only known ecdysteroid kinase gene, themselves encode ecdysteroid kinases. To do this, a combination of gene disruption (CRISPR-Cas9 mutagenesis and RNAi knock-down) and gene misexpression (using existing and new UAS-ORF constructs, as well as a collection of GAL4 drivers) experiments were performed to test if either gene is essential for developmental progression, if misexpression of either gene phenocopies a loss of ecdysteroid signalling, and if *EPPase* co-misexpression can rescue developmental arrest phenotypes. In addition, I test the hypothesis that *Wall* encodes an ecdysteroid 26-kinase with epistasis experiments with *Cyp18a1*, and test the hypothesis that *pkm* is involved in the adult starvation response.

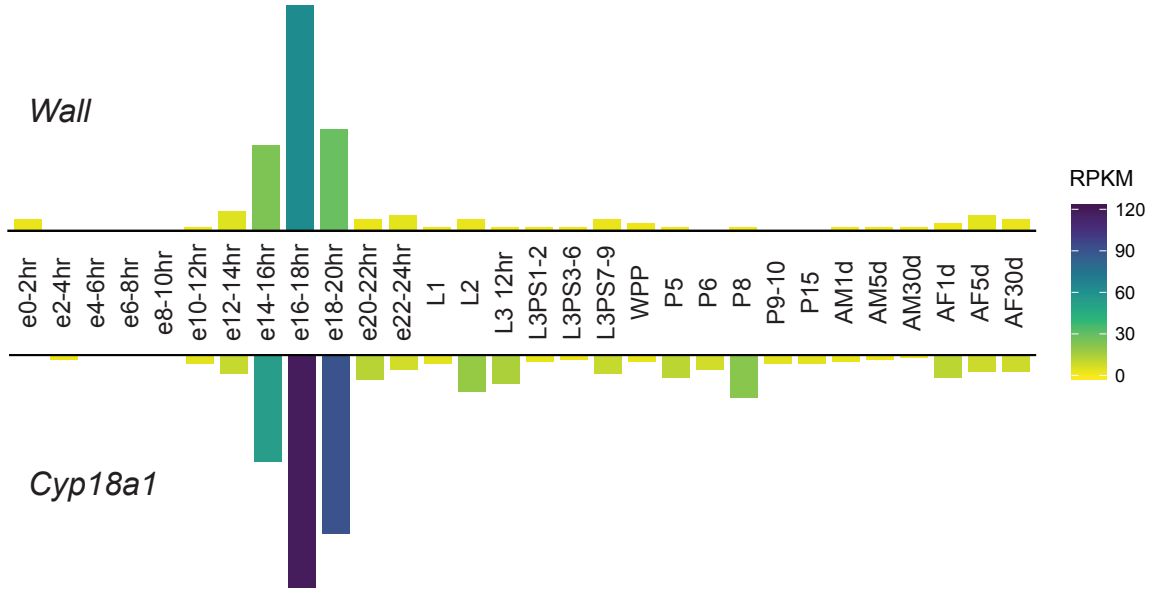


Figure 5.1. Expression of *Wall* and *Cyp18a1* throughout *Drosophila melanogaster* development, based on data from modENCODE (Graveley *et al.* 2011). Gene expression is given in RPKM (Reads Per Kilobase of transcript, per Million mapped reads) mapped onto the colour of each bar, while the height of the bars for each gene is normalised to the highest level of expression for that gene. e, embryo (in hrs after fertilisation); L1, 1st instar larva; L2, 2nd instar larva; L3, 3rd instar larva (12 hr after moulting and the salivary gland puff stages); WPP, white prepupa; P, pupa (stages from Bainbridge & Bownes 1981); AM, adult male (1–30 days post-eclosion); AF adult female (1–30 days post-eclosion).

5.2. Materials and Methods

5.2.1. Fly lines and husbandry

The following fly lines were obtained from the Bloomington Drosophila Stock Center (BDSC): *w**;; *Sb*^{1}/TM3, *actGFP*, *Ser*^{1} (BL4534), *Tango11*^{1}/CyO, *actGFP* (BL36320), FM7i, *actGFP*/C(1)DX, *y*^{1}, *f*^{1} (BL4559), FM7j (BL6418), *y*^{1}, *w*^{67c23};; *P{EPgy2}EY20330* (BL23106, called UAS-*Wall*^{EY} in this chapter), *w**; *nos-GAL4*; UAS-*Cas9.P2* (BL67083) and *elav-GAL4*; UAS-*Cas9.P2*/CyO (BL67073). The following UAS-dsRNA (and control genotype) lines were obtained from the Vienna Drosophila Resource Centre (VDRC): VL60100 ('KK control'), *w*^{1118} (VL60000; 'GD control'), *KK*^{Wall} (VL104249), *GD*^{Wall} (VL45409), *KK*^{pkm} (VL106503), *GD*^{pkm1} (VL32634) and *GD*^{pkm2} (VL32635). UAS-*Cyp18a1* (Guittard *et al.* 2011), UAS-*Dcr2*/CyO; *tub-GAL4*/TM6B and some GAL4 driver stocks (Table 5.1) were a kind gift of Philip Batterham (The University of Melbourne). *w**; *tub-GAL80*^{ts}; TM2/TM6B and *y*^{1}, *w**, *P{lacW}Fas2G0032*, *P{neoFRT}19A*/FM7c; *P{ey-FLP.N}5* were a kind gift of Michael Murray (The University of Melbourne). The strong loss-of-function *Cyp18a1*^{1} allele was described in Rewitz *et al.* (2010); a *Cyp18a1*^{1}/FM7i, *actGFP* line was a kind gift of Michael O'Connor (University of Minnesota). For GAL4 driver lines used in this chapter, see Table 5.1; other fly lines were made by routine crossing (see Chapter 5.6.1) or were previously described in Chapter 4.

For routine stock maintenance, flies were kept on yeast-cornmeal-molasses media ('lab media'; http://bdsc.indiana.edu/information/recipes/molassesfood.html) at 18 ºC, 21 ºC or 25 ºC in plastic vials sealed with cotton stoppers. Bioassays that were analysed together (each represented by a different figure or sub-figure in the results) were carried out as a group on the same batch of media at the same time to minimise intra-experiment batch effects, unless otherwise stated.

Table 5.1. GAL4 driver lines used in this chapter.

Name	Origin	Stock ID	Genotype	Expression
tub-GAL4	PB[a]	-	tub-GAL4/TM3, actGFP, Ser^{1}	ubiquitous (very high)
act-GAL4	PB	-	act-GAL4/CyO, actGFP	ubiquitous (high)
da-GAL4	PB	-	likely BL48489	ubiquitous (moderate)
elav-GAL4	PB	-	likely P{$w^{+mW.hs}$=GawB}$elav^{C155}$	pan-neuronal, imaginal discs, glia, trachea
Eip71CD-GAL4	Bloomington	BL6871	w^{1118}; P{w^{+mC}=MsrA-GAL4.657}TP1-1	mid-L3 larval brain, epidermis
HR-GAL4	PB	-	from Chung *et al.* 2007	midgut, fat body, Malpighian tubules
mex1-GAL4	PB	-	likely P{mex1-GAL4.2.1} construct	midgut (enterocytes)
ppl-GAL4	Bloomington	BL58768	w^{*}; P{w^{+mC}=ppl-GAL4.P}2	larval fat body, other tissues
UO-GAL4	PB	-	likely P{Uro-GAL4.T}	Malpighian tubule principle cells
e22c-GAL4	Bloomington	BL1973	y^{1}, w^{*}; P{$w^{+mW.hs}$=en2.4-GAL4}e22c/SM5	epidermis, embryo, imaginal discs, follicle cells
c204-GAL4	Bloomington	BL3751	w^{1118}; P{$w^{+mW.hs}$=GawB}c204/TM3, Ser^{1}	follicle cells, oocyte stage 8-14, probably other tissues
dan^{AC116}-GAL4	Bloomington	BL27591	y^{1}, w^{*}; P{$w^{+mW.hs}$=GawB}dan^{AC116}	eye, 3rd antennal segment regions of the eye disc
phm-GAL4	PB	-	phm-GAL4/CyO, actGFP	prothoracic gland, imaginal discs
btl-GAL4	Bloomington	BL8807	w^{*}; P{w^{+mC}=GAL4-btl.S}2, P{w^{+mC}=UASp-Act5C.T:GFP}2/CyO, P{w^{+m*}=lacZ.w^{+}}276	trachea
nSyb-GAL4	PB	-	likely BL51635	pan-neuronal
cg-GAL4	Bloomington	BL7011	w^{1118}; P{w^{+mC}=Cg-GAL4.A}2	fat body, haemocytes
Mef2-GAL4	Bloomington	BL27390	y^{1}, w^{*}; P{w^{+mC}=GAL4-Mef2.R}3	muscles
desat1-GAL4	Bloomington	BL65405	w^{*}; P{w^{+mC}=Desat1-GAL4.E800}4M/TM3, Sb^{1}	oenocytes, male accessory gland
Sgs3-GAL4	Bloomington	BL6870	w^{1118}; P{w^{+mC}=Sgs3-GAL4.PD}TP1	salivary glands

[a] PB, gifts from Philip Batterham (The University of Melbourne)

5.2.2. Generation of UAS-ORF lines

Cloning of ORFs into the pUASTattB vector and *D. melanogaster* transformation was as described in Chapter 4.2.2, with the following modifications. ORFs of *Wall* and *EPPase* (*CG13604*; short 312 aa isoform) were isolated by PCR from the DGRC cDNA clones IP11764 and GH09153, using the primer pairs CG13813_EagIF/CG13813_KpnIR and CG13604S_EagIF/CG13604_KpnIR, respectively (Table 5.2). ORFs of *BmEc22K* and two forms of *pkm*—the full 635 aa ORF ('pkm^F') and a truncated 430 aa ORF without the N-terminal intrinsically disordered region ('pkm^T')—were synthesised by Integrated DNA Technologies, with EagI and KpnI restriction sites (plus six additional nucleotides to allow for efficient digestion) at the N- and C-termini, respectively. Expected amplicon sizes from recombinant plasmids in colony PCR for *Wall*, *EPPase*, *BmEc22K*, pkm^F and pkm^T were 1,521 bp, 1,167 bp, 1,389 bp, 2,136 bp and 1,521 bp, respectively.

All correctly assembled plasmids were sent to TheBestGene Inc. (US) for microinjection and incorporation into the *D. melanogaster* genome at the attP40 site on chromosome 2 (BL25709), or in the case of *EPPase*, independently into both the attP40 site (BL25709) and the attP2 site (BL25710) on chromosome 3, to generate the UAS-$EPPase^2$ and UAS-$EPPase^3$ lines, respectively.

5.2.3. Generation of CRISPR-Cas9 mutant lines

The recombinant pCFD6 plasmids 'pCFD6-CG13813' and 'pCFD6-CG1561'—each of which express, under the control of a UAS promoter, four gRNAs that target either the *CG13813*/*Wall* locus (Fig. 5.3A) or the *CG1561*/*pkm* locus (Fig. 5.3B)—were designed *in silico* using Benchling (http://benchling.com), and cloned as described in Chapter 4.2.3, with the following modifications: the primer pairs for the amplification of the inserts for pCFD6-CG13813 were pCFD6_CG13813_1F/R, pCFD6_CG13813_2F/R and pCFD6_CG13813_3F/R; and the primer pairs for the amplification of the inserts for pCFD6-CG1561 were pCFD6_CG1561_1F/R, pCFD6_CG1561_2F/R and pCFD6_CG1561_3F/R (Table 5.2).

Correctly assembled plasmids were sent to TheBestGene Inc. for microinjection and incorporation into the *D. melanogaster* genome at the attP40 site on chromosome 2 (BL25709), to produce the homozygous fly lines UAS-$Wall^{pCFD6}$ and UAS-pkm^{pCFD6}.

Mutagenesis of the *Wall* locus on chromosome 3L was done using the previously described crossing scheme in Figure 4.2A. Mutagenesis of the *pkm* locus on the X chromosome was done using the crossing scheme in Figure 5.1. Single founder female flies during *pkm* mutagenesis were allowed to lay viable larvae before undergoing DNA extraction.

Wall PCR genotyping used the ΔD38_1F/1R and ΔD38_2F/2R primer pairs, which produce amplicons of size 698 bp and 845 bp, respectively, from wild-type chromosomes, and ΔD38_1F/2R, which produce 1,689 bp amplicons from wild-type chromosomes and amplicons between ~700–1,500 bp in the case of deletions between any two *Wall* gRNA pairs. *pkm* PCR genotyping used the ΔD37_1F/1R and ΔD37_2F/2R primer pairs, which produce amplicons of size 693 bp and 898 bp, respectively from wild-type chromosomes, and ΔD37_2F/1R, which produce 2,757 bp amplicons from wild-type chromosomes and amplicons between ~900–2,600 bp in the case of deletions between any two *pkm* gRNA pairs. PCR—2 min initial denaturation (95 ºC), then 2 min denaturation (95 ºC), 45 sec annealing (55 ºC) and 1.5 min extension (72 ºC) for 32 cycles, then a 5 min final extension (72 ºC)—was carried out with GoTaq Green Master Mix. Deletion loci from mutant flies (Figs. 4.2A & 5.2, blue boxes) were amplified using the genotyping primers above with four GoTaq Green PCR reactions per line, which were combined before gel-purification to allow for the detection of early-cycle polymerase-derived errors by close inspection of the sequencing chromatogram output. Gel-purified amplicons were sequenced using the appropriate genotyping primers at AGRF.

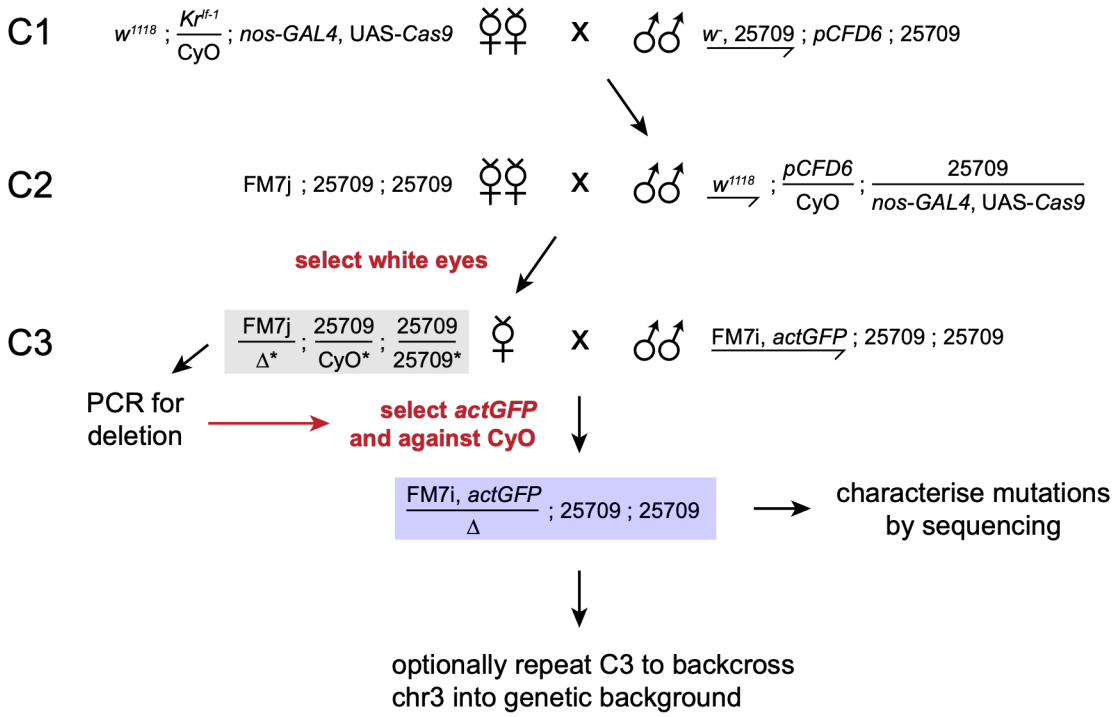


Figure 5.2. Crossing scheme for CRISPR-Cas9 mutagenesis on the X chromosome (chrX) using *pCFD6*-transformed (UAS-4xgRNA) flies, to produce a deletion allele (Δ) in a final line (blue box) that can be homozygosed if possible. The single females used in C3 (grey box) are 'founder females' of each potential mutant line. Asterisks indicate potentially on-target or off-target mutagenised chromosomes.

Table 5.2. Primer sequences used in Chapter 5. All oligonucleotides were synthesised by Integrated DNA Technologies (IDT).

Primer ID	Sequence (5' to 3')
pUASTattB_3F	CGCAGCTGAACAAGCTAAAC
pUASTattB_5R	TGTCACACCACAGAAGTAAGG
CG13813_EagIF	TAAGCACGGCCGATGAGCGATAACGAGCTTTC
CG13813_KpnIR	TAAGCAGGTACCTTAGCACTGCTCGGAGTC
CG13604S_EagIF	TAAGCACGGCCGATGTCTTCTTCCTCCACCTCC
CG13604_KpnIR	TAAGCAGGTACCCTAGGTGGCCGACAAGGC
pCFD6_CG13813_1F	CGGCCCGGGTTCGATTCCCGGCCGATGCAACCTATGGTCTCGC AACCAAGTTTCAGAGCTATGCTGGAAAC
pCFD6_CG13813_1R	AATTCACGCTGTATTATGCCTGCACCAGCCGGGAATCGAACC
pCFD6_CG13813_2F	GGCATAATACAGCGTGAATTGTTTCAGAGCTATGCTGGAAAC
pCFD6_CG13813_2R	CATGTCTATAGAGAATATTGTGCACCAGCCGGGAATCGAACC
pCFD6_CG13813_3F	CAATATTCTCTATAGACATGGTTTCAGAGCTATGCTGGAAAC
pCFD6_CG13813_3R	ATTTTAACTTGCTATTTCTAGCTCTAAAACTCGTGTTGTAGTAAG CATCCTGCACCAGCCGGGAATCGAACC
pCFD6_CG1561_1F	CGGCCCGGGTTCGATTCCCGGCCGATGCAATATAAACGATCCA CACTCGGTTTCAGAGCTATGCTGGAAAC
pCFD6_CG1561_1R	GATGTCGATCAGAAGTCACCTGCACCAGCCGGGAATCGAACC
pCFD6_CG1561_2F	GGTGACTTCTGATCGACATCGTTTCAGAGCTATGCTGGAAAC
pCFD6_CG1561_2R	GCGAGGCATATCGCATCAGTTGCACCAGCCGGGAATCGAACC
pCFD6_CG1561_3F	ACTGATGCGATATGCCTCGCGTTTCAGAGCTATGCTGGAAAC
pCFD6_CG1561_3R	ATTTTAACTTGCTATTTCTAGCTCTAAAACTGCATCGTTACAATG GGTAGTGCACCAGCCGGGAATCGAACC
ΔD38_1F	ACGCAATACAAACCCAGTACCAT
ΔD38_1R	GCTTTCAGGGAGCAGACCAC
ΔD38_2F	GCAAGGCGAGAATCATCCTG
ΔD38_2R	TTTTGCCTCGCTTACACGCA
ΔD37_1F	CTGCACTGCTTCGCTGTTATT
ΔD37_1R	TCGCTATGGAAACCCAACAGA
ΔD37_2F	TGCAGGGTAAGTGCTACGTT
ΔD37_2R	GCCTTTGGTCAGCGGATTCA
NC_Genomic_F	GCTGGCGAACTGTCAATCAC
NC_R1	GGAGCAGCGCTTATTTGCTC
pKC26_R	TGTAAAACGACGGCCAGT

5.2.4. Germline knockout, somatic knockout, RNAi knockdown and misexpression developmental viability assays

Egg-to-adult and larval-to-adult viability assays were conducted as described in Chapter 4.2.5, but at 25 ºC or 29 ºC depending on the experiment. The 'MultinomCI' function in the *DescTools* package in R was used to calculate confidence intervals for multinomial proportions (Chapter 5.3.2).

Misexpression phenotyping (Chapter 5.3.7) was conducted by crossing 3–5 UAS-responder females to *GAL4* males, letting them lay in lab media vials for 24 hours, with a minimum of three replicate vials per genotype; vials were left to develop at 25 ºC and were checked every 24 hours for developmental arrest phenotypes. Arrest phenotypes were fully penetrant unless otherwise noted.

5.2.5. Delayed-onset misexpression assay

Three groups of 10 *act-GAL4*/CyO, *actGFP* virgin females were crossed to *tub-GAL80^{ts}*; UAS-*Wall^{EY}* males and allowed to pre-mate in lab media vials at 18 ºC for 24 hours, then transferred to three fresh vials at 18 ºC (the GAL80^{ts} restrictive temperature) and allowed to lay for 24 hours. This transfer-and-lay process was repeated every 24 hours (with fresh females added after nine days) until 19 sets of three vials containing offspring were produced, and then all the vials were shifted to 29 ºC (the GAL80^{ts} permissive temperature) to complete development—each set of vials contained animals that started expressing *Wall^{EY}* at a different developmental stage. Eclosing adults were genotyped based on the presence or absence of the phenotypic markers on the CyO, *actGFP* chromosome. Adult genotype counts from each set of vials were analysed by the 'binom.test' function in R (with a Bonferroni correction for multiple tests) to determine if genotypic ratios were significantly different from the Mendelian expectation (1:1); if so, this was considered evidence that one genotype had less egg-to-adult viability than the other.

5.2.6. Food avoidance assay

tub-GAL80^{ts}; *da-GAL4* virgin females were crossed to *w^{1118}* or UAS-*Wall^{EY}* males and laid eggs on juice plates (Appendix 3.2) at 25 ºC for eight hours. 1st-instar larvae were transferred to lab media vials (20 larvae per vial, eight vials per genotype) and kept at

29 ºC. Two and three days post-hatching, the number of larvae located in the food substrate (mouthparts hidden in substrate; 'digging'), on top of the food substrate (mouthparts out of substrate; 'on food') or on the sides of the vials (body completely off substrate; 'side of vial') was scored for each genotype. Fisher's exact test was used to determine if there were significant differences in larval position between genotypes and/or timepoints, using the 'fisher.test' function in R.

5.2.7. Misexpression epistasis assays

Six *Cyp18a1*^{1}/FM7i, *actGFP*; UAS-*Wall*^{pU} virgin females or +; UAS-*Wall*^{pU} virgin females were crossed to *da-GAL4* or *Mef2-GAL4* males and allowed to lay in lab media vials for 24 hours, with 10 vials per genotype. Offspring were left to develop at 25 ºC for 14 days, and the number of individuals that reached pupariation, pupation (or pupariation/pupation for *da-GAL4* crosses), pharate adult differentiation and eclosion were scored. Fisher's exact test was used to determine if there were significant differences between the developmental outcomes of pairs of genotypes, using the 'fisher.test' function in R.

5.2.8. *EPPase/Wall* co-misexpression assays

For the *cg-GAL4* experiment, five *cg-GAL4* females were crossed to +; UAS-*EPPase*^{3}, UAS-*Wall*^{pU}; + or UAS-*Wall*^{pU}; UAS-*EPPase*^{3} males and allowed to lay in lab media vials for 24 hours, with 10 vials per genotype. Offspring were left to develop at 25 ºC for 14 days, and the number of individuals that reached pupariation, pupation, pharate adult differentiation and eclosion were scored. Fisher's exact test was used to determine if there were significant differences between the developmental outcomes of pairs of genotypes, using the 'fisher.test' function in R.

For the *btl-GAL4* experiment, five *btl-GAL4*/CyO females were crossed to +; UAS-*EPPase*^{3}, UAS-*Wall*^{pU}; + or UAS-*Wall*^{pU}; UAS-*EPPase*^{3} males and allowed to lay in lab media vials for 24 hours, with 10 vials per genotype. Offspring were left to develop at 25 ºC for 14 days, and eclosing adults were genotyped based on the presence or absence of the CyO balancer chromosome. Adult genotype counts were analysed by the 'binom.test' function in R (with a Bonferroni correction for multiple tests) to determine if genotypic ratios were significantly different from the Mendelian expectation (1:1); if so, this was considered evidence that one genotype had less egg-to-adult viability than

the other. Fisher's exact test—'fisher.test' in R—was used to compare the proportion of eclosing misexpression adults for pairs of genotypes.

5.2.9. Starvation assays

Starvation assays were conducted at 21–22 ºC using the 'wet starvation' method developed by Storelli *et al*. (2019). Adult flies were collected as virgins in the first 12 hours post-eclosion, separated by sex and kept in lab media vials for 2–3 days. Empty vials were half-filled with deionised water and cellulose acetate stoppers (Flystuff, 49-102) were pushed to the bottom and saturated. Excess water was thoroughly removed and 9–10 flies were placed into each vial (six vials per sex-genotype combination), which were closed with an additional dry stopper. Flies were moved to fresh vials twice a week (Mondays and Fridays). Survival was scored every 24 hours until all flies were dead. Male and female flies were analysed separately due to known sex differences in starvation resistance (Jang & Lee 2015). Survival analysis was conducted with the *survival* (v2.38) and *survminer* (v0.4.7) packages in R, using a log-rank test to compare survival curves; pairwise p-values were adjusted for multiple tests using the Benjamini-Hochberg procedure.

5.3. Results

5.3.1. CRISPR-Cas9 mutagenesis of *Wall* and *pkm*

Twenty-two putative deletions at the *Wall* locus were detected by PCR from 66 founder males. Two *Wall* deletion alleles were kept as homozygous-viable stocks and molecularly characterised: $Wall^{1}$ and $Wall^{20}$. $Wall^{1}$ comprises two deletions in the coding region of the genes—a 381 bp in-frame deletion in exon 1, and a 2 bp frameshift deletion in exon 2. $Wall^{20}$ comprises a single 605 bp frameshift deletion that spans exon 1 and 2 (Fig. 5.3A). Both alleles are predicted to significantly disrupt the function of the encoded protein, due to the deleted coding sequence (including the ATP-binding Brenner's motif) and frameshifts, and are likely strong loss-of-function alleles.

Eighteen putative deletions at the *pkm* locus were detected by PCR from 24 founder females. Two *pkm* composite deletion alleles were kept as homozygous-viable stocks and molecularly characterised: pkm^{2} and pkm^{4}. pkm^{2} comprises two deletions, one of 34 bp overlapping the transcription start site and another of 1,394 bp that induces a frameshift in exon 1, while pkm^{4} comprises three deletions, one of 34 bp that removes the transcription start site, one of 1,410 bp in-frame in exon 1, and another of 13 bp that induces a frameshift in exon 2 (Fig. 5.3B). Both alleles delete Brenner's motif, induce frameshifts and delete significant portions of the coding sequence, meaning it is likely both are strong loss-of-function alleles.

Requirements for *Wall* and *pkm* during development were tested in three ways: germline knockout (with single and double mutants; Chapter 5.3.2), somatic knockout (with the UAS-gRNA pCFD6 constructs and the ubiquitous *da-GAL4* driver; Chapter 5.3.3) and RNAi knockdown (with KK and GD UAS-*dsRNA* hairpin lines and the ubiquitous *tub-GAL4* driver; Chapter 5.3.4).

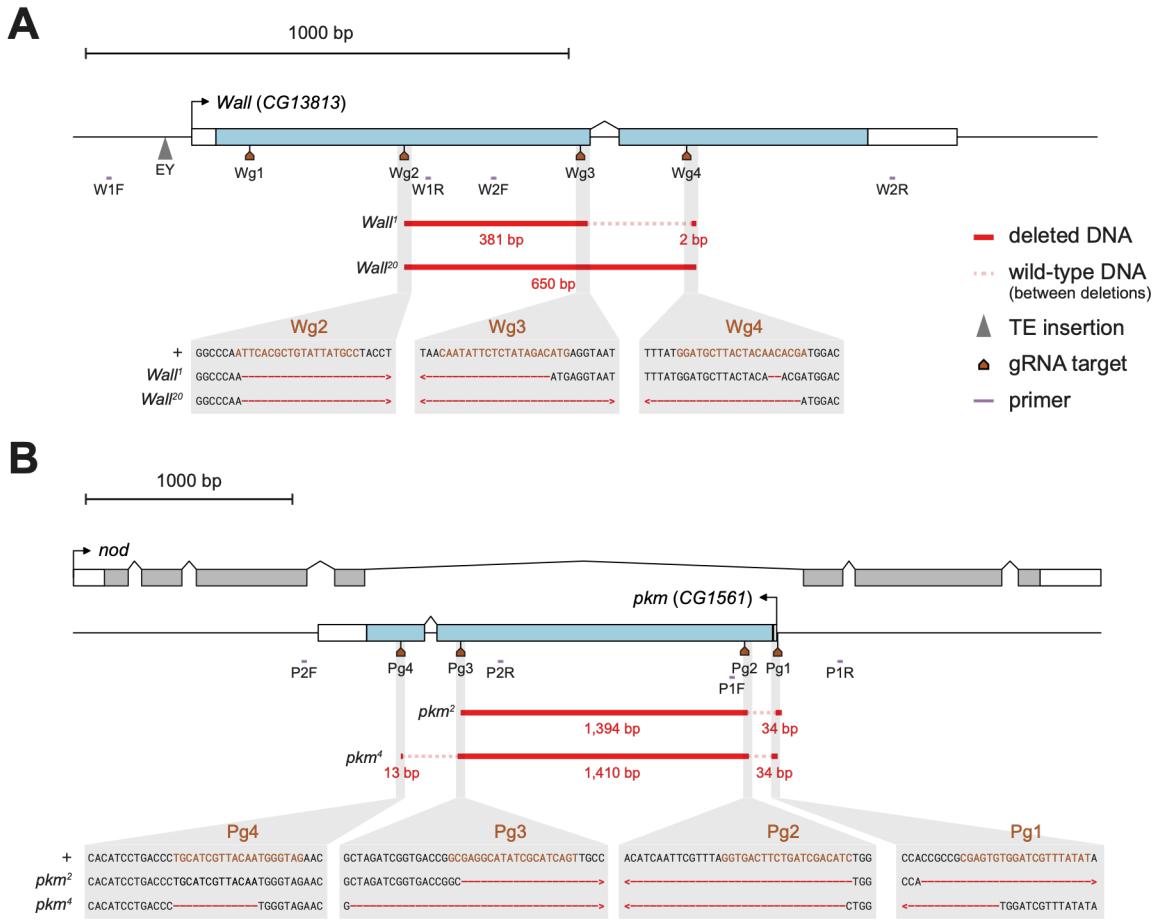


Figure 5.3. (A) The *Wall* (*CG13813*) locus on chromosome 3L of *Drosophila melanogaster*. Mapped onto the locus are the locations of gRNA target sites (Wg1–4; brown), genotyping primer-binding sites (purple), the *EYgp2* element *EY20330* (grey triangle), and two CRISPR-Cas9-induced deletion alleles (*Wall*^{1} and *Wall*^{20}; deletion sizes in red text). Neither deletion allele had lesions at the Wg1 target site. W1F, ΔD38_1F; W1R, ΔD38_1R; W2F, ΔD38_2F; W2R, ΔD38_2R. (B) The *pkm* (*CG1561*) locus on the X chromosome of *Drosophila melanogaster*, including the *pkm* gene (blue exon) and the *nod* gene (grey exons). Mapped onto the locus are the locations of gRNA target sites (Pg1–4; brown), genotyping primer-binding sites (purple), and two CRISPR-Cas9-induced deletion alleles (*pkm*^{2} and *pkm*^{4}; deletion sizes in red text). P1F, ΔD37_1F; P1R, ΔD37_1R; P2F, ΔD37_2F; P2R, ΔD37_2R. Wild-type DNA present in-between deletions is indicated with pale dashed lines. Grey boxes (bottom) are sequence-level detail of the deletions with respect to the wild-type genetic background (+). gRNA target sites are highlighted in brown, and deleted bases are red dashes; > and < symbols indicate that the deletion continues out of the frame of the highlighted sequence.

5.3.2. *Wall* and *pkm* single and double mutants have no obvious developmental defects

Although the *Wall* and *pkm* mutants generated above appeared to be homozygous viable—as balancer chromosomes could be crossed out from their stocks without obvious drops in viability—I sought to quantify this by tracking the larval-to-adult viability of *Wall* single mutants and the egg-to-adult viability of *pkm* single mutants and *Wall* and *pkm* double mutants.

There were no significant differences between the developmental outcomes of $Wall^{1}$/+ and $Wall^{1}$/$Wall^{20}$ heterozygotes ($p = 0.152$, Fisher's exact test; Fig. 5.4A), with the vast majority of all individuals successfully eclosing as viable adults. In crosses with pkm^{2} alleles, while FM7c hemizygotes appeared much less viable than other genotypes, there were no clear differences between the proportions of other genotypes, strongly suggesting that pkm^{2}/Y hemizygotes are developmentally viable (Fig. 5.4B). With respect to *Wall* and *pkm* double mutants, there was no significant difference in the genotypic ratios between $Wall^{1}$/$Wall^{1}$ and $Wall^{1}$/+ crosses ($p = 0.168$, Fisher's exact test; Fig. 5.4B), demonstrating double mutants (homozygous for $Wall^{1}$ and hemizygous for pkm^{2}) do not suffer any developmental viability loss compared to other genotypes, and that *Wall* and *pkm* do not act redundantly during development. The lower viability of FM7c hemizygotes is consistent with the known accumulation of fitness-affecting mutations on balancer chromosomes (Miller *et al.* 2016; 2018).

All generated *Wall* and *pkm* mutant stocks could be kept for multiple generations as homozygous stocks without obvious issues, suggesting neither gene is required for fertility, although this was not formally quantified.

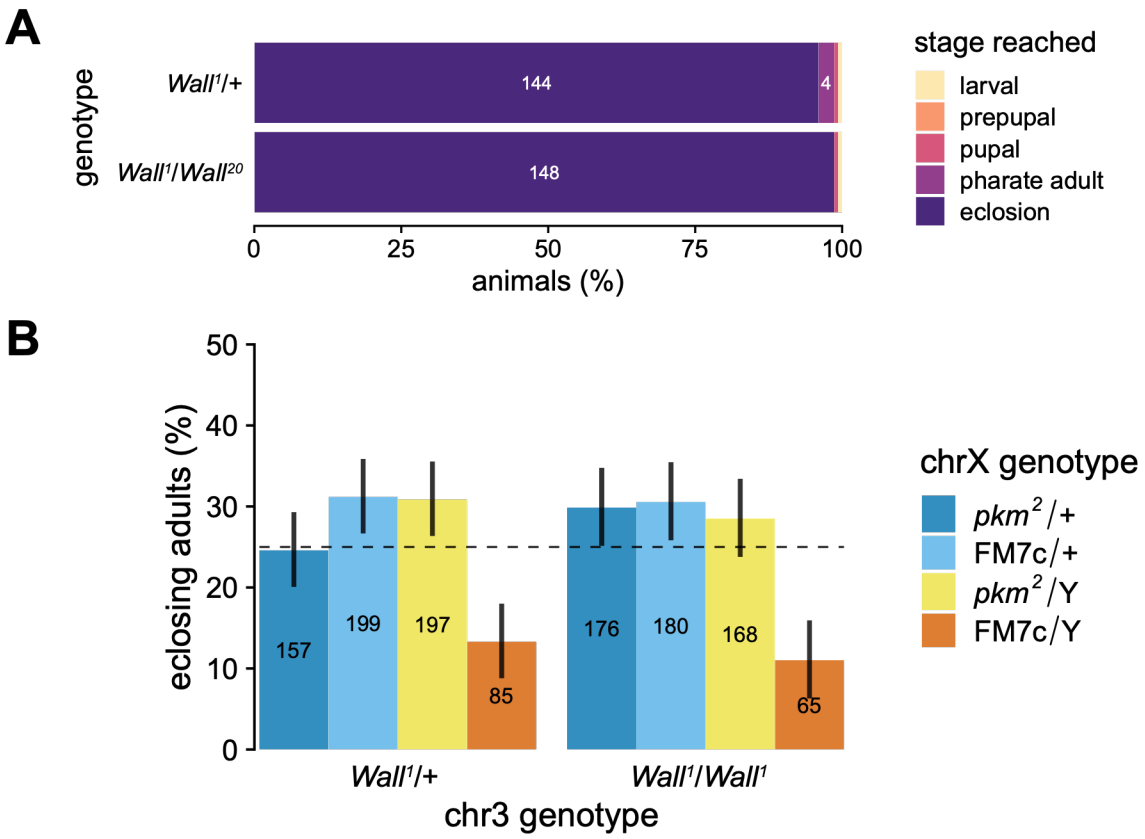


Figure 5.4. Developmental viability of *Wall* and *pkm* single and double mutant animals. (A) Larval-to-adult viability of offspring from crosses between $Wall^{1}$ females and wild-type (+) males (top) or $Wall^{20}$ males (bottom). Numbers on the bars are the number of individuals in each lethal phase category (for numbers greater than three). (B) Eclosing adult genotypic ratios of the offspring from crosses between pkm^{2}/FM7c;; $Wall^{1}$ females and $Wall^{1}$ (left) or w^{1118} (right) males. Bars are coloured by their chrX genotype, as determined by their sex and the presence or absence of the FM7c balancer chromosome; pkm^{2} hemizygotes are in yellow (single mutants on left, double mutants on right). The dashed line indicates the expected 1:1:1:1 genotypic ratio if all genotypes per cross are equally developmentally viable. Numbers on the bars are the number of adults of each genotype. Error bars are 97.5% confidence intervals (95% CI adjusted for two tests) for the proportion of each genotype within each cross.

5.3.3. Somatic knockout of *pkm* produces an immobility phenotype in adult flies

Despite the lack of obvious developmental viability defects of *Wall* or *pkm* null mutant animals, I decided to test if simultaneously driving a UAS-*Cas9* transgene and the gRNA-expressing UAS-*Wall*^{pCFD6} or UAS-*pkm*^{pCFD6} constructs in somatic tissues using a ubiquitous GAL4 driver would recapitulate the lack of phenotype. I used the moderately strong driver *da-GAL4*, rather than the stronger *act-GAL4* and *tub-GAL4* drivers, to induce ubiquitous somatic knockout of *Wall* and *pkm* in order to minimise some of the known fitness effects of expressing *Cas9* transgenes at high levels with the UAS/GAL4 system (Huynh *et al.* 2018).

Somatic knockout of *Wall* produced no detectable loss of egg-to-adult viability compared to non-gRNA-expressing animals (Fig. 5.5A), but knockout of *pkm* resulted in a substantial number of adults falling into the food media, being unable to remove themselves and 'drowning'; these adults were unable to be easily counted. Tipping vials upside-down before eclosion revealed that knockout adults eclosed in a 1:1 ratio with balancer adults, indicating no detectable pre-adult lethality (Fig. 5.5B), but had a seizure-like phenotype where their limbs were uncoordinated, effectively resulting in immobility (Fig. 5.5C), a phenotype not seen in *pkm* null mutants (Fig. 5.5D). This immobility phenotype was also observed when UAS-*pkm*^{pCFD6} and UAS-*Cas9* were driven with the (non-specific; Berger *et al.* 2007; Casas-Tintó *et al.* 2017) neuronal driver *elav-GAL4*, although this was not quantified.

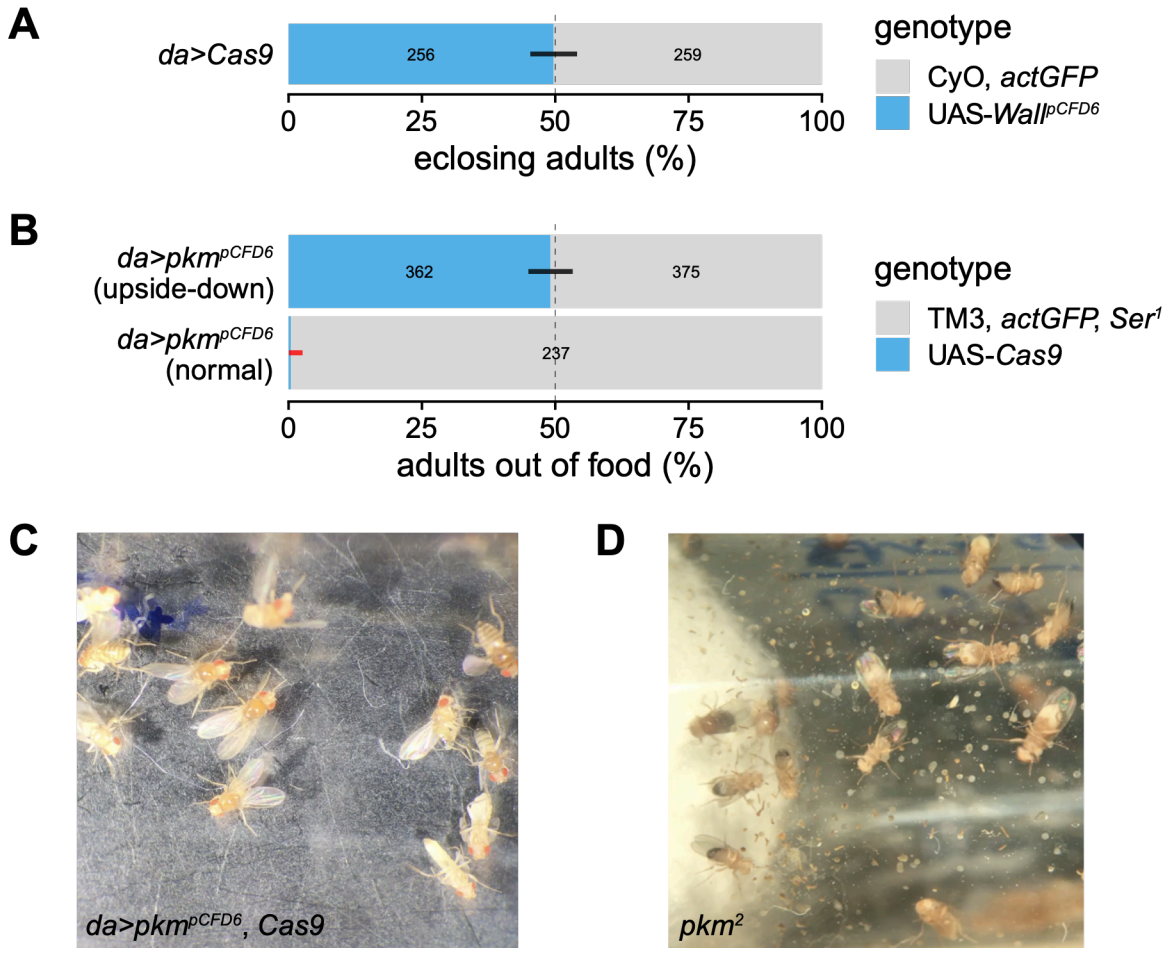


Figure 5.5. (A–B) Egg-to-adult viability of ubiquitous somatic CRISPR-Cas9 mutagenesis of (A) *Wall* and (B) *pkm* at 25 ºC, estimated from the adult genotypic ratios of offspring from crosses between (A) *da-GAL4* females and UAS-*Wall*^{pCFD6}/CyO, *actGFP*; UAS-*Cas9* males or (B) *da-GAL4* females and UAS-*pkm*^{pCFD6}; UAS-*Cas9*/TM3, *actGFP*, *Ser*^{1} males. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; black and red error bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than one). (A) Error bars are 95% confidence intervals for the proportion of UAS-*Wall*^{pCFD6}-containing (somatic knockout) heterozygotes. Vials were held in the normal orientation throughout development. (B) Error bars are 97.5% confidence intervals (95% CI adjusted for two tests) for the proportion of UAS-*Cas9*-containing (somatic knockout) heterozygotes. Vials were kept in their normal orientation after pupation ('normal') or tipped upside-down ('upside-down'), to check if pre-adult development is affected by somatic knockout. (C) Non-motile but otherwise viable *da>pkm*^{pCFD6}, *Cas9* adults, rescued from 'drowning' in food substrate by eclosing in vials tipped upside-down, viewed on the bottom side of a vial resting on its side; note the 'held-out' wing posture of many flies. (D) Motile, viable *pkm2* adults, viewed on the top side of a vial resting on its side.

5.3.4. Putative *Wall* and *pkm* RNAi knockdown causes developmental arrest, but may be due to off-target effects

Before the generation of *Wall* and *pkm* mutant alleles, complementary RNAi knockdown experiments were conducted to test if either gene is required for development. RNAi knockdown of *Wall* and *pkm* at 25 ºC had been attempted previously with two KK library UAS-*dsRNA* lines (Chapter 3.7), but I added to this experiment by also using independently generated GD lines that target different parts of each gene (Dietzl *et al.* 2007). I also performed knockdown at a higher temperature (29 ºC), which increases GAL4 activity and should result in stronger knockdown (Duffy 2002; Fortier & Belote 2000), as well as at 29 ºC with the addition of a UAS-*Dcr2* transgene, which can further increase knockdown efficiency (Dietzl *et al.* 2007).

For *Wall*, only the KK^{Wall} construct resulted in detectable developmental arrest before the adult stage at either 25 ºC (Fig. 5.6A) or 29 ºC (Fig. 5.6B), but both KK^{Wall} and GD^{Wall} had significant effects on developmental viability compared to controls (both $p < 2.2\times10^{-16}$, Fisher's exact test) at 29 ºC when paired with the UAS-*Dcr2* transgene, with KK^{Wall} resulting in mostly larval lethality, and GD^{Wall} resulting in a mix of pharate adult lethality and a 'drowning' phenotype whereby adults eclosed from the puparium but fell to the bottom of the vial and failed to escape from the semi-liquid food media (Fig. 5.7).

For *pkm*, none of the three constructs produced detectable developmental arrest at 25 ºC (Fig. 5.6A), but KK^{pkm} was semi-lethal before the adult stage at 29 ºC (Fig. 5.6B). KK^{pkm}, GD^{pkm1} and GD^{pkm2} all had significant effects on developmental viability compared to controls (all $p < 2.2\times10^{-16}$, Fisher's exact test) at 29 ºC when paired with the UAS-*Dcr2* transgene, with KK^{pkm} resulting in pupal and pharate adult lethality, and GD^{pkm1} and GD^{pkm2} producing similar phenotypes to GD^{Wall}, with pharate adult lethality and the 'drowning' phenotype (Fig. 5.7).

These knockdown results are suspicious, given that no developmental phenotypes were detected for *Wall* and *pkm* null mutant animals (Chapter 5.3.2), phenotypes for most of these constructs were only detectable with a very strong GAL4 driver and the UAS-*Dcr2* transgene at a high temperature, and that such extreme knockdown conditions are known to produce high levels of off-target knockdown (Dietzl *et al.* 2007).

However, the discrepancy between null mutant and RNAi knockdown animals has been noted before (El-Brolosy & Stainier 2017), and can be due to a phenomenon called 'genetic compensation', wherein gene knockout triggers a compensatory mechanism (such as the up-regulation of genes with similar functions) to restore a wild-type phenotype, while RNAi knockdown fails to trigger compensation, leading to a mutant phenotype (El-Brolosy *et al*. 2019; El-Brolosy & Stainier 2017).

To test the hypothesis that the phenotypic discrepancy between knockdown animals and null mutant animals was due to genetic compensation in mutants, I performed *pkm* knockdown in a pkm^2 null mutant background, with the expectation that if genetic compensation was occurring in mutants, knockdown would fail to produce a developmental arrest phenotype, while if the knockdown phenotype was due to off-target effects, the presence of the pkm^2 allele would have no effect on the phenotype. Females homozygous for the KK^{pkm} construct—and homozygous for either the pkm^2 allele (one of two such independently generated lines) or a wild-type allele (pkm^+)—were crossed to *tub-GAL4*/TM3, *actGFP*, Ser^1 males, and adult genotypic ratios of the offspring were scored, with half of the knockdown individuals (males) expected to be pkm^+ or pkm^2 hemizygotes (Fig. 5.6C). The proportion of adult knockdown individuals between knockdown crosses with pkm^+ and pkm^2 was not significantly different, with a pkm^2 (line 1) vs pkm^+ odds ratio of 1.67 (97.5% CI: 0.94, 3.0) and a pkm^2 (line 2) vs pkm^+ odds ratio of 1.02 (97.5% CI: 0.54, 1.91). This demonstrates that the presence or absence of *pkm* loss-of-function alleles does not affect the $tub{>}KK^{pkm}$ knockdown phenotype, strongly suggesting these phenotypes are due to off-target effects and not genetic compensation. A similar experiment to test the genetic compensation hypothesis for *Wall* knockdown phenotypes was planned but ultimately not conducted.

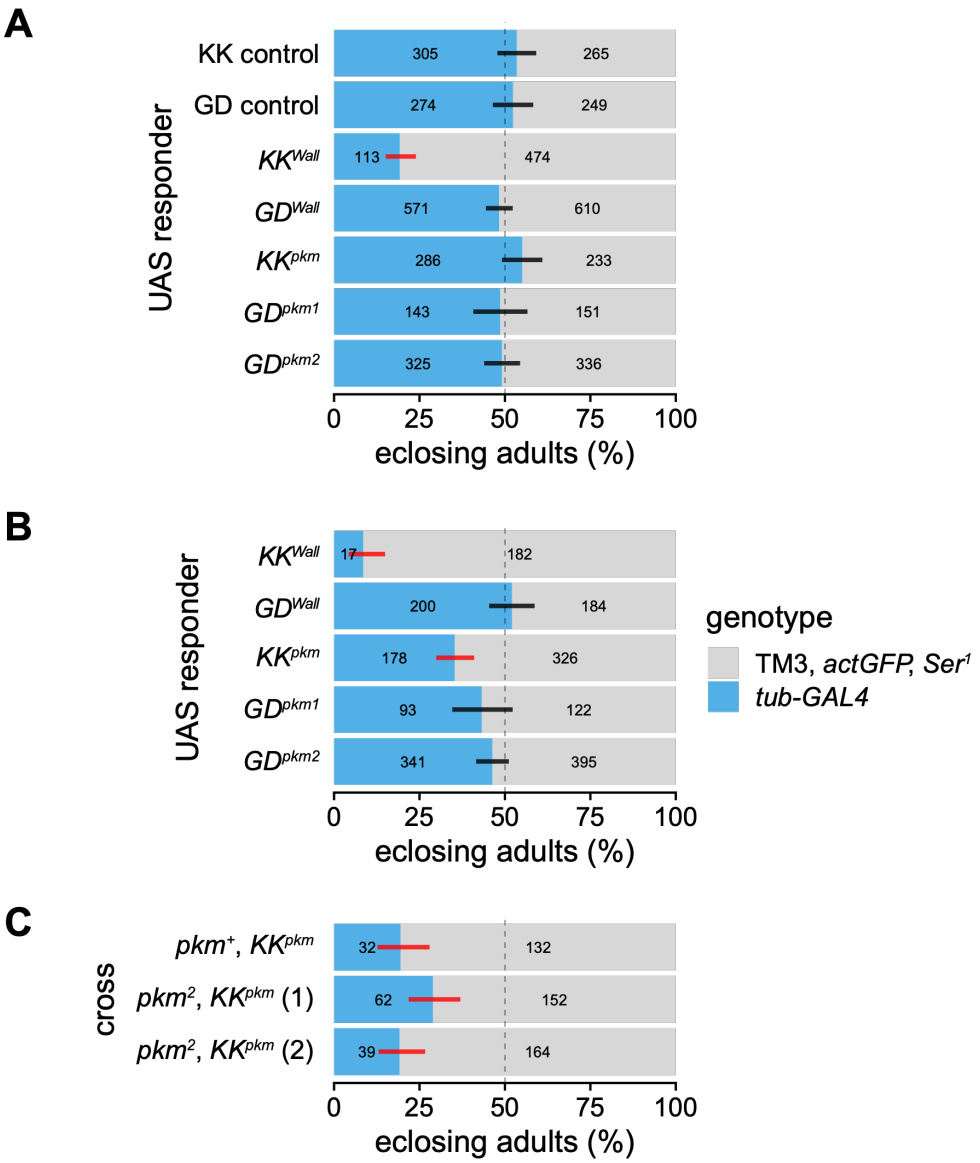


Figure 5.6. Egg-to-adult viability of RNAi knockdown (or control genotypes), estimated from the adult genotypic ratios of offspring from crosses between (A–B) *tub-GAL4*/TM3, *actGFP*, Ser^1 females and UAS-responder (or control genetic backgrounds) males, or (C) UAS-responder females and *tub-GAL4*/TM3, *actGFP*, Ser^1 males. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; black and red error bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than 0). KK control genotype is VL60100, GD control genotype is VL60000. (A) Knockdown crosses conducted at 25 ºC. Error bars are 99.3% confidence intervals (95% CI adjusted for seven tests) for the proportion of *tub-GAL4*-containing heterozygotes. Data from the KK control, GD control and KK^{pkm} crosses are from Fig. S2.8 and were not conducted at the same time as the other crosses. (B) Knockdown crosses conducted at 29 ºC. Error bars are 99% confidence intervals (95% CI adjusted for five tests) for the proportion of *tub-GAL4*-containing heterozygotes. (C) Knockdown crosses conducted at 29 ºC to test for genetic compensation (lack of knockdown phenotype in a null mutant background). Maternal genotypes were pkm^+; KK^{pkm} (VL106503; top), and two independently generated pkm^2; KK^{pkm} lines (middle and bottom). Error bars are 98.3% confidence intervals (95% CI adjusted for three tests) for the proportion of *tub-GAL4*-containing heterozygotes.

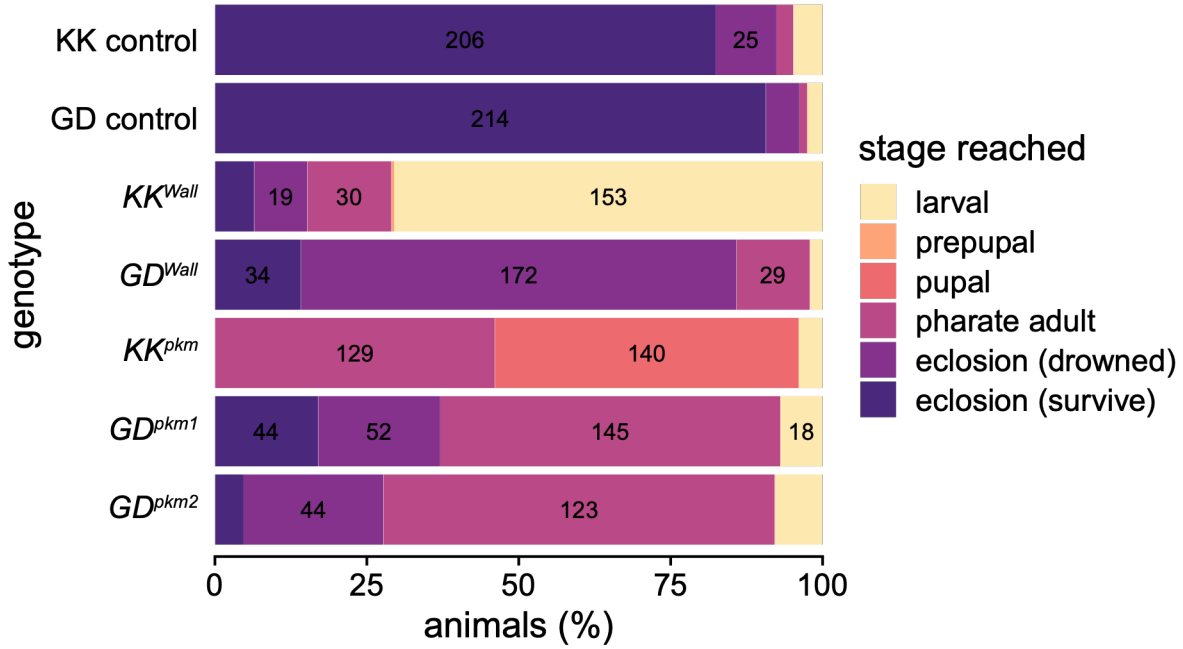


Figure 5.7. Larval-to-adult viability of GFP-negative offspring from crosses between UAS-*Dcr2*/CyO, *actGFP*; *tub-GAL4*/TM3, *actGFP*, *Ser*^{1} females and UAS-responder (or control genetic backgrounds) males. Numbers on the bars are the number of individuals in each lethal phase category (for numbers greater than 15).

5.3.5. Ubiquitous misexpression of *Wall*, but not *pkm*, *BmEc22K* or *EPPase*, causes developmental arrest

Without obvious requirements for *Wall* or *pkm* during development in *D. melanogaster*, I sought to test the hypothesis that these genes encode ecdysteroid kinases by misexpression, driving their UAS-ORF constructs (and an EPgy2 insertion upstream of the native *Wall* locus that contains a UAS and promoter region—UAS-$Wall^{EY}$) with the very strong, ubiquitous driver *tub-GAL4* at 25 ºC to see if ectopic expression arrested developmental progression. As a positive control for ecdysteroid 22-kinase activity, I also misexpressed *BmEc22K*, which encodes the *Bombyx mori* ecdysteroid 22-kinase (Sonobe *et al.* 2006), and as a positive control for ecdysteroid inactivation more generally, I used a UAS-*Cyp18a1* responder (Guittard *et al.* 2011) to misexpress the ecdysteroid 26-hydroxylase/carboxylase Cyp18a1, which phenocopies ecdysteroid biosynthesis mutants (Guittard *et al.* 2011; Rewitz *et al.* 2010). I was also curious about the phenotypic effects of misexpressing the short isoform of the *D. melanogaster EPPase* gene (also known as *CG13604*), which encodes an ecdysteroid-phosphate phosphatase (Davies *et al.* 2007).

Cyp18a1 misexpression completely arrested development, as expected (Guittard *et al.* 2011), as did misexpression of *Wall* with either the UAS-$Wall^{EY}$ line or the UAS-$Wall^{pU}$ line—examination of offspring vials under a fluorescent microscope (to detect the presence of the TM3, *actGFP*, Ser^{1} balancer chromosome) suggested that these misexpression genotypes failed to complete embryogenesis and hatch as larvae. Misexpression of the transgenes for *pkm* (both the full and truncated ORFs, to see if the presence or absence of the N-terminal disordered region affected developmental arrest), *BmEc22K* or *EPPase* failed to significantly alter adult genotypic ratios from Mendelian expectations, suggesting ubiquitous expression of these transgenes does not affect developmental progression (Fig. 5.8); qPCR was not conducted to check if the transgenes were being successfully expressed.

Misexpression of *Wall* using UAS-$Wall^{EY}$ and the strong, ubiquitous *act-GAL4* driver at 25 ºC also caused complete developmental arrest ($n = 158$ adults, $p < 2.2\times10^{-16}$, exact binomial test), although not all animals died as embryos, with some larvae hatching but dying before pupariation, with various morphological phenotypes, such as spiracle and rectal pad pigmentation, and tracheal and denticle belt defects (Fig. 5.9).

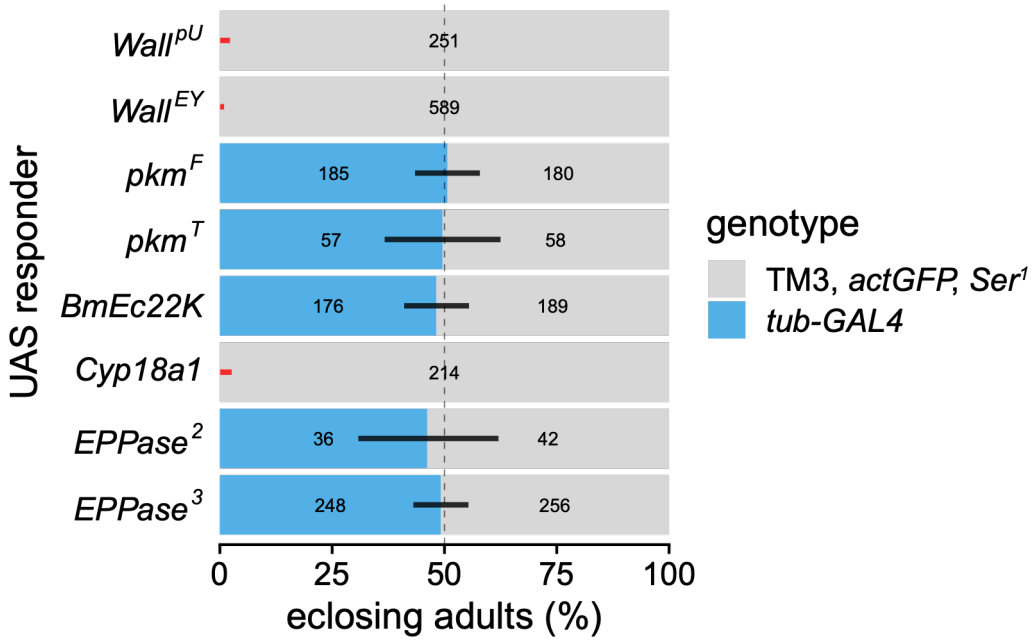


Figure 5.8. Egg-to-adult viability of strong, ubiquitous *Wall*, *pkm*, *BmEc22K*, *Cyp18a1* and *EPPase* misexpression, estimated from the adult genotypic ratios of offspring from crosses between *tub-GAL4*/TM3, *actGFP*, *Ser*^{1} females (genotypes in bold) and UAS-responder males at 25 ºC. *EPPase* misexpression was from two separate transgenic lines: UAS-*EPPase*^{2} is an integration on chr2, and UAS-*EPPase*^{3} is an integration on chr3. UAS-*pkm*^{F} expresses the full 635 aa *pkm* ORF, while UAS-*pkm*^{T} expresses a truncated 430 aa ORF that does not include the pkm N-terminal disordered region (Chapter 5.2.2). The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; error bars are 99.4% confidence intervals (95% CI adjusted for eight tests) for the proportion of *GAL4*-containing heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than 0).

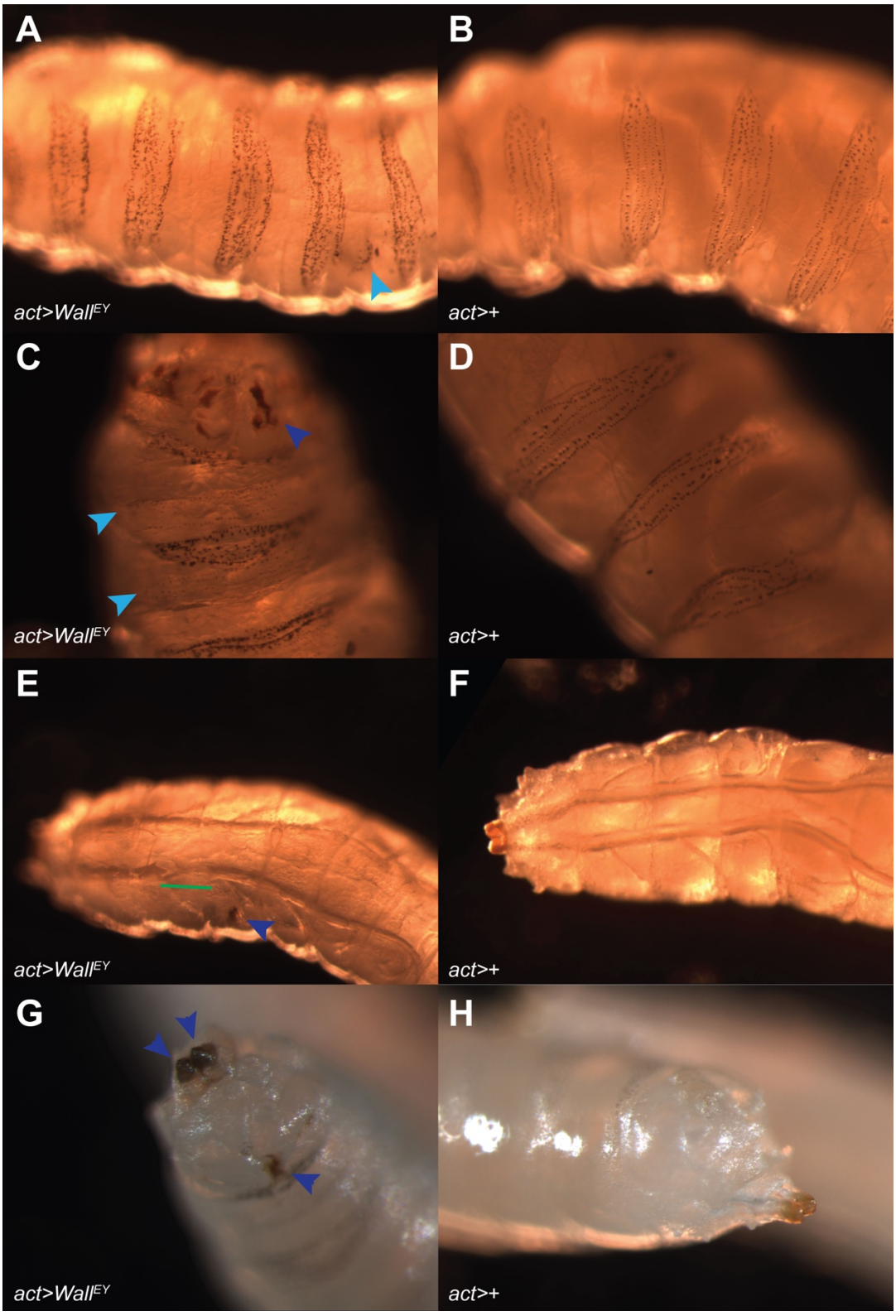


Figure 5.9. Phenotypes in larvae misexpressing UAS-*Wall*^{EY} with the *act-GAL4* driver (*act*>*Wall*^{EY}; left) compared with wild-type, non-misexpressing larvae (*act*>+; right), at 25 ºC. (A–D) Denticle belt disorganisation, including ectopic denticles in between belts (light blue arrowheads) and rectal pad melanisation (dark blue arrowheads). (E–F) Fluid-filled section of a tracheal dorsal trunk (green line segment), along with a melanised section of trachea (dark blue arrowhead). (G–H) Melanised posterior spiracles and rectal pad (dark blue arrowheads).

5.3.6. Ubiquitous misexpression of *Wall* causes developmental arrest until the middle of metamorphosis

While ubiquitous misexpression of *Wall* is embryonic lethal, it was unclear if misexpression affected later developmental stages. To investigate this, a delayed-onset misexpression experiment was conducted using the temperature-sensitive *tub-GAL80^{ts}* construct, which ubiquitously represses GAL4 activity at 18 ºC and permits GAL4 activity at 29 ºC, and the *act-GAL4* driver, which strongly expresses GAL4 in a ubiquitous pattern (*tub-GAL4* was not used because pilot experiments suggested *tub-GAL80^{ts}* was unable to fully repress its activity). Misexpression of *Wall* caused detectable drops in egg-to-adult viability when delayed 0–14 days into development, but did not affect viability from 15–18 days (Fig. 5.10A), suggesting animals become insensitive to *Wall* misexpression around the middle of metamorphosis (Fig. 5.10B).

5.3.7. Tissue-specific misexpression of *Wall* causes developmental arrest distinct from *Cyp18a1*

The developmental arrest phenotype observed upon ubiquitous misexpression of *Wall* is consistent with a possible ecdysteroid kinase function of its encoded protein. To further test this hypothesis, I conducted tissue-specific misexpression crosses to see if the patterns of developmental arrest produced by *Cyp18a1* misexpression (a positive control for ecdysteroid inactivation) matched that of *Wall* misexpression—a close match would be good evidence that Wall and Cyp18a1 act on the same or similar substrates.

Misexpression was conducted with two ubiquitous GAL4 drivers (*tub-GAL4*, used previously, and *da-GAL4*, a weaker driver) and 16 tissue-specific GAL4 drivers. The UAS-*Cyp18a1* construct produced developmental arrest when misexpressed with 16 out of 18 drivers (Fig. 5.11). The UAS-*Wall^{EY}* construct produced developmental arrest when misexpressed with 10 drivers, while the UAS-*Wall^{pU}* construct produced very similar results, but with no phenotype with the *ppl-GAL4* driver and generally caused developmental arrest at later life stages (Fig. 5.11); *btl-GAL4* and *cg-GAL4* misexpression also resulted in a small number of adult escapers. These differences are likely due to differences in expression strength: the *EPgy2* element contains 14 copies of UAS in its artificial enhancer region (Bellen *et al.* 2004), while pUASTattB contains only five (Bischof *et al.* 2007), meaning misexpression of *Wall* with UAS-*Wall^{EY}* is likely stronger

than that from UAS-*Wall*^{pU} using the same GAL4 driver. However, the broad consistency of the results from both an independent transgene construct and a native locus misexpression construct means these phenotypes can be confidently linked to the ectopic expression of the *Wall* ORF. Comparison of phenotypes between the three constructs can be found in Table S5.1.

Misexpressing UAS-*Wall*^{EY} with the *HR-GAL4* driver resulted in a dramatic phenotype: an extended wandering period during the 3rd larval instar, then defects during pupariation and pupation wherein the puparium was thin and elongated and anterior adult tissue differentiation appeared to occur without head eversion, resulting in dark thoracic (likely from wings) and red eye pigmentation developing internally (Fig. S5.2A–C). This phenotype suggests *Wall* misexpression with this driver can block epidermal and cuticular development during metamorphosis, but curiously allows for imaginal discs eversion to take place. The *HR-GAL4* driver expresses GAL4 along the whole length of the midgut, as well as in the Malpighian tubules and the fat body (Chung *et al.* 2007); however, the lack of phenotypes with the enterocyte-specific *mex1-GAL4* driver suggests that *Wall* misexpression in these midgut cells does not cause developmental defects (Phillips & Thomas 2006). Additionally, given that the fat body drivers *cg-GAL4* and *ppl-GAL4* failed to phenocopy *HR-GAL4*, the crucial tissues might be non-enterocyte midgut cells and/or the Malpighian tubules. Alternatively, *HR-GAL4* may drive at a low (yet sufficient) level in another tissue to produce this phenotype, such as the muscles.

Conspicuously, the *phm-GAL4* driver, which expresses GAL4 in the prothoracic cells of the ring gland (PG)—the ecdysteroidogenic tissue in larvae (Redfern 1983)—as well as some other tissues such as the wing disc (Casas-Tintó *et al.* 2017), caused embryonic lethality when misexpressing *Cyp18a1* as expected (Guittard *et al.* 2011), but produced no detectable phenotype when misexpressing either *Wall* construct (Fig. 5.11B). This result appears strongly inconsistent with *Wall* encoding a kinase that can act on ecdysteroids present in the PG, as well as *Wall*'s own native expression in the ring gland (Ou *et al.* 2016).

Misexpression of *Wall* with the *c204-GAL4* driver caused larval arrest with the UAS-*Wall*^{EY} construct and arrest during metamorphosis with the UAS-*Wall*^{pU} construct (Fig. 5.11A–B). The expression pattern of this driver is not well characterised—while it is

known to be expressed in the adult ovary (Manseau *et al.* 1997), it may be expressed in other tissues and at other life stages, given it causes embryonic arrest with UAS-*Cyp18a1* and the aforementioned arrest with UAS-*Wall* constructs. Curiously, *elav-GAL4* misexpression of *Wall* caused developmental arrest, while *nSyb-GAL4* misexpression did not, suggesting misexpression in neurons is not the cause of *elav-GAL4* arrest; the *elav-GAL4* driver, while often considered specific to neurons, expresses in other tissues, such as glia, trachea and the imaginal discs (Berger *et al.* 2007; Casas-Tintó *et al.* 2017), one or more of which is likely the cause of the developmental arrest observed here. Attempted misexpression of *Wall* and *Cyp18a1* with an *e22c-GAL4* line, which drives in the epidermis, did not result in obvious developmental defects (Fig. 5.11B), inconsistent with the defects seen when *Wall* or *Cyp18a1* were driven with *Eip71CD-GAL4* (Fig. 5.11A), suggesting the *e22c-GAL4*/SM5 line used may not contain a functional *GAL4* construct. A *UO-GAL4* line, which is meant to drive GAL4 in the principle cells of the Malpighian tubules, also failed to show a developmental arrest phenotype with either of the three responder constructs, despite the requirement of ecdysteroid signalling in this tissue during development (Gautam *et al.* 2015). Closer examination of the genotypes of these lines needs to be conducted.

Overall, these data demonstrate that *Wall* misexpression in the trachea, muscle, epidermis or fat body is sufficient to cause developmental arrest in *D. melanogaster*, while misexpression in the PG, imaginal discs, salivary gland, neurons and oenocytes does not cause any obvious phenotype. In addition, combined with the results of the ubiquitous delayed-onset misexpression experiment (Fig. 5.10), these data suggest that the pharate adult lethality observed through misexpression with some drivers (*cg-GAL4* and *btl-GAL4*) is likely due to misexpression of *Wall* during early metamorphosis or earlier, not during the pharate adult stage itself.

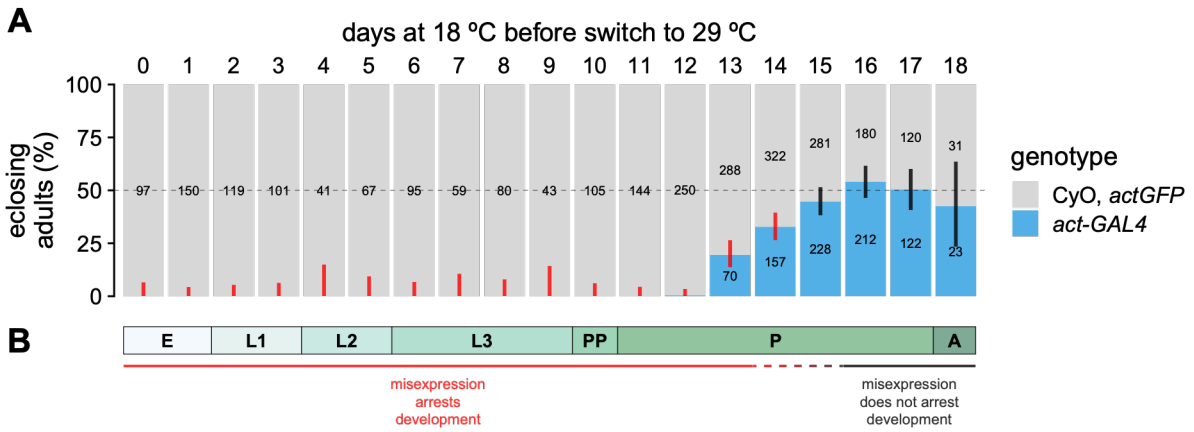


Figure 5.10. Delayed-onset of ubiquitous misexpression of *Wall* arrests development up until mid-metamorphosis. (A) Egg-to-adult viability of *Wall* misexpression with the ubiquitous driver *act-GAL4*, in concert with the UAS-*Wall*^{EY} element and the temperature-sensitive *tub-GAL80*^{ts} construct, which inhibits GAL4 at 18 ºC but not at 29 ºC; animals were moved from 18 ºC to 29 ºC at one of 19 timepoints (top), which began *Wall* misexpression. Viability is estimated from the adult genotypic ratios of offspring from crosses between *act-GAL4*/CyO, *actGFP* females and *tub-GAL80*^{ts}; UAS-*Wall*^{EY} males. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; error bars are 99.7% confidence intervals (95% CI adjusted for 19 tests) for the proportion of *act-GAL4*-containing heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than 10). (B) Approximate developmental life stages of *D. melanogaster* at 18 ºC, showing *Wall*-sensitive (red) and *Wall*-insensitive (black) developmental periods; dashed line indicates region of uncertainty. E, embryo; L1, 1st-instar larva; L3, 2nd-instar larva; L3, 3rd-instar larva; PP, prepupa; P, pupa; A, adult.

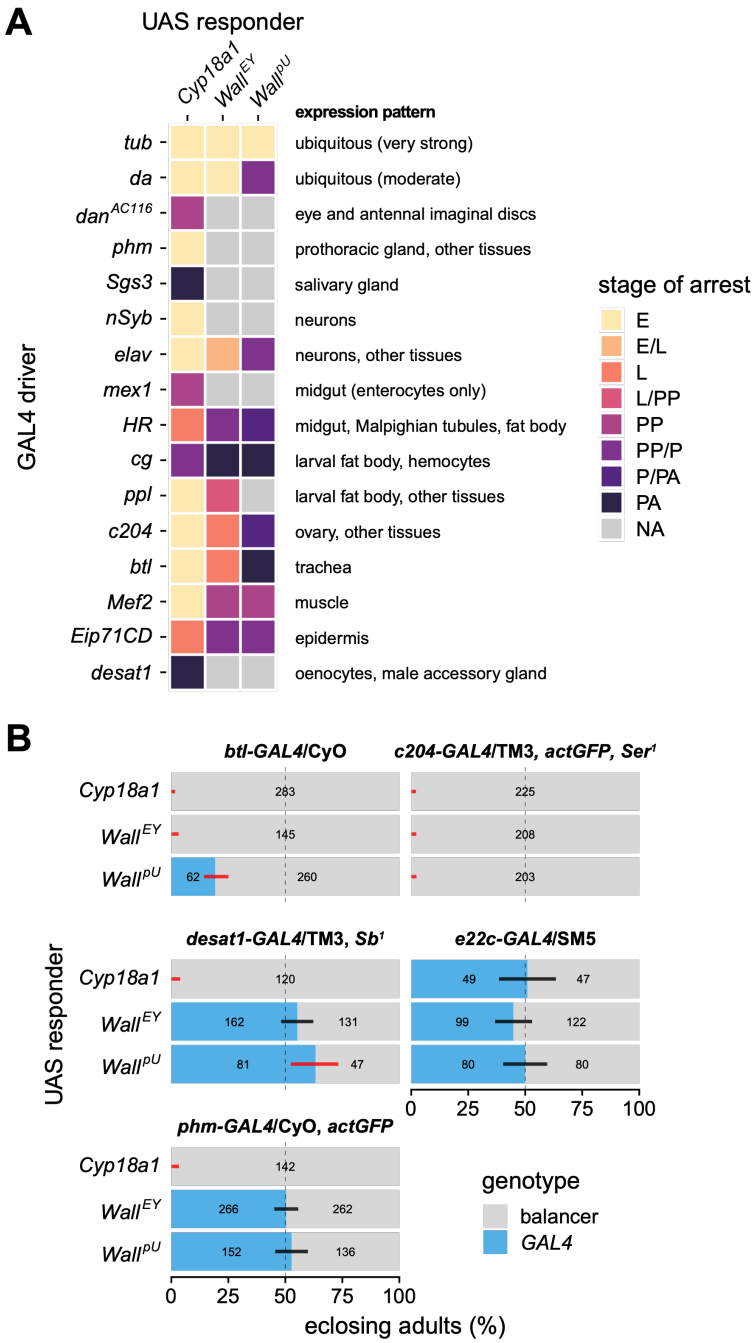


Figure 5.11. Tissue-specific misexpression of *Wall* and *Cyp18a1*, using the UAS-*Wall*^{EY}, UAS-*Wall*^{pU} and UAS-*Cyp18a1* constructs, at 25 °C. (A) Heat map of qualitative developmental arrest phenotypes for *Wall* and *Cyp18a1* misexpression using two ubiquitous GAL4 drivers and 14 tissue-specific GAL4 drivers. E, embryo; L, larva; PP, prepupa; P, pupa; PA, pharate adult; NA, no arrest (successful eclosion). When arrest occurred at more than one stage, two stages are indicated and separated with a slash. Misexpression of all three constructs with the *UO-GAL4* or *e22c-GAL4* drivers failed to cause developmental arrest, which is not shown in the heat map. For phenotypic descriptions, see Table S5.1. (B) Egg-to-adult viability of *Wall* and *Cyp18a1* misexpression, estimated from the adult genotypic ratios of offspring from crosses between *GAL4*/balancer females (genotypes in bold) and UAS-responder males. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; error bars are 98.3% confidence intervals (95% CI adjusted for three tests) for the proportion of *GAL4*-containing heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than 0).

5.3.8. Larval tracheal misexpression of *Wall* causes tracheal defects and food aversion

Misexpression of *Wall* with the trachea-specific *btl-GAL4* driver causes developmental arrest at the larval or pharate adult stages, depending on the UAS-responder construct and the temperature (Fig. 5.11A, Table S5.1). I noticed that *btl>Wall^{EY}* larvae (raised at 25 ºC) abandoned the food media to crawl up the sides of the vial, a phenotype that was shared with ubiquitously misexpressing *da>Wall^{EY}* larvae that were escaped from embryonic lethality using the *tub-GAL80^{ts}* construct and then moved to 29 ºC, as well as the *act>Wall^{EY}* larvae described previously (Chapter 5.3.6). Quantification of this food-aversion phenotype (Fig. 5.12A) showed *da>Wall^{EY}* larvae had significant differences in their positions within vials compared to non-misexpressing *da>w^{1118}* controls, at both two days and three days after hatching (both $p < 2.2\times10^{-16}$, Fisher's exact test); there was also a significant difference in positions between the two timepoints for misexpressing larvae ($p = 9.7\times10^{-14}$, Fisher's exact test), while there was no significant difference between non-misexpressing larvae ($p = 0.73$, Fisher's exact test), consistent with progression of the misexpression phenotype over time. Food aversion is a well-established consequence of hypoxia in both larvae and adult *D. melanogaster* (Wang:2015hz; Vigne & Frelin 2010; Wingen *et al.* 2017)—consistent with this, close examination of *btl>Wall^{EY}* and *da>Wall^{EY}* larvae revealed defects in tracheal filling (similar to those seen in *act>Wall^{EY}* and *c204>Wall^{EY}* larvae; Fig. 5.9E & Table S5.1), with parts of the dorsal trunks often filled with liquid (Fig. 5.12B), which can cause severe reductions in gas exchange (Parvy *et al.* 2012). Taken together, these data suggest that *Wall* misexpression in trachea causes hypoxia through defects in either tracheal development, maintenance or integrity. The food-aversion phenotype is also the basis for the gene name *Wallflower* (*Wall*), due to the tendency of the larvae to position themselves on the walls of the vial, rather than in the food with other larvae, an allusion to the 'wallflower' metaphor for social behaviour; *Wall* starts with a capital letter because the phenotype is due to misexpression, a dominant genotype that relates to the wild-type function of the Wall protein (see the FlyBase nomenclature rules; Thurmond *et al.* 2019).

Also of note, I noticed a curious pigmentation defect on the dorsal abdomen of both male and female *btl>Wall^{pU}* adult escapers (Fig. 5.12C); a hypothetical mechanistic basis of this phenotype will be proposed in Chapter 5.4.7.

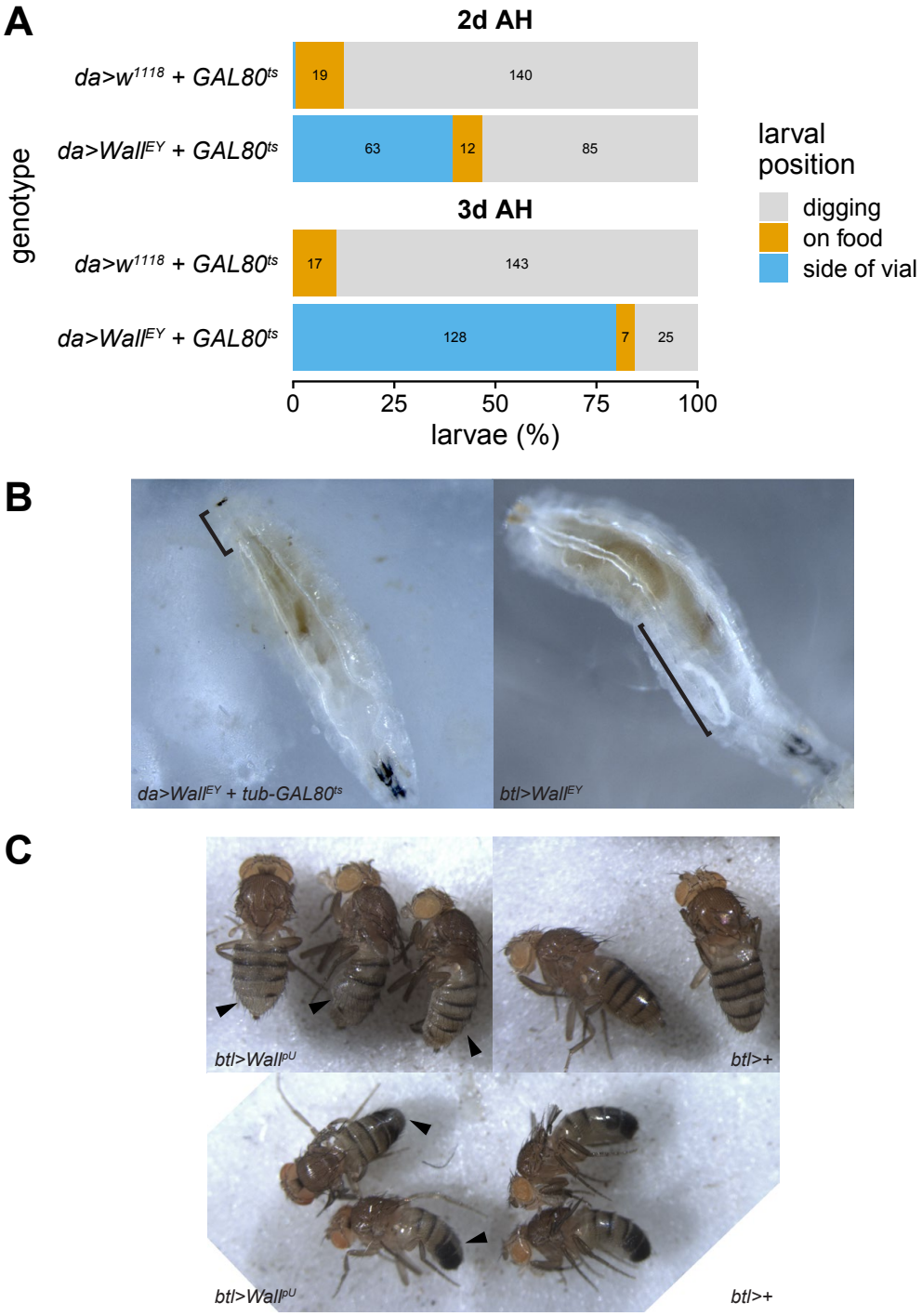


Figure 5.12. Misexpression of *Wall* affects tracheal integrity and adult pigmentation. (A) Food avoidance assay comparing *da>w*^{1118} *+ GAL80*^{ts} and *da>Wall*^{EY} *+ GAL80*^{ts} larvae 2 days (top) and 3 days (bottom) after hatching (AH). Numbers on the bars are the number of larvae in each position (for numbers greater than five). (B) Photos of larvae misexpressing *Wall* ubiquitously (*da>Wall*^{EY} *+ tub-GAL80*^{ts}, left) or specially in the trachea (*btl>Wall*^{EY}, right) retrieved from food media at 25 ºC, before food aversion occurred. Black segments indicate sections of tracheal dorsal trunks that are filled with liquid. (C) Adult flies (females, top; males, bottom) misexpressing *Wall* in the trachea (*btl>Wall*^{pU}, left) or with the *btl-GAL4* driver alone (*btl>+*, right). Arrowheads indicate dorsal abdomen pigmentation defects (areas of lightened pigment compared to the rest of the abdomen). Wings have been removed to expose the dorsal abdomen.

5.3.9. *Wall* misexpression phenotypes are not hypostatic to *Cyp18a1*

If *Wall* encodes an ecdysteroid 26-kinase, which acts on the products of the ecdysteroid 26-hydroxylase Cyp18a1, then *Wall* misexpression should be hypostatic to the wild-type function of *Cyp18a1*—i.e. misexpression of an ecdysteroid 26-kinase (Wall) should not cause developmental defects when the source of 26-hydroxyecdysteroids (Cyp18a1) is not functional. Evidence of such hypostasis would be strong genetic evidence of a biochemical interaction between the two genes and this hypothesis for Wall's function. In contrast to $Cyp18a1^{1}$ animals, which largely reach the pharate adult stage but fail to eclose (Rewitz *et al.* 2010), at 25 ºC, most *da*>$Wall^{pU}$ animals arrest development before the differentiation of adult structures (prepupal/pupal arrest), while most *Mef2*>$Wall^{pU}$ animals arrest development before pupation (prepupal arrest). I reasoned that if *Wall* misexpression phenotypes were hypostatic to the function of *Cyp18a1*, $Cyp18a1^{1}$ hemizygotes (25% of the offspring of these crosses) should reach the pharate adult stage instead of earlier developmental arrest due to *Wall* misexpression, increasing the proportion of pharate adult-lethal animals.

There were no significant differences in the developmental outcomes between $Cyp18a1^{1}$, *Mef2*>$Wall^{pU}$ and *Mef2*>$Wall^{pU}$ offspring (Fig. 5.13A; $p = 0.336$, Fisher's exact test), nor between $Cyp18a1^{1}$, *da*>$Wall^{pU}$ and *da*>$Wall^{pU}$ offspring (Fig. 5.13B; $p = 0.841$, Fisher's exact test). These data indicate that misexpression of *Wall* in $Cyp18a1^{1}$ hemizygotes produces the same developmental arrest phenotypes as misexpression of *Wall* in a wild-type background, strongly suggesting that *Wall* is not hypostatic to *Cyp18a1*, and that *Wall* is unlikely to encode an ecdysteroid 26-kinase (assuming Cyp18a1 is the only ecdysteroid 26-kinase in *D. melanogaster*).

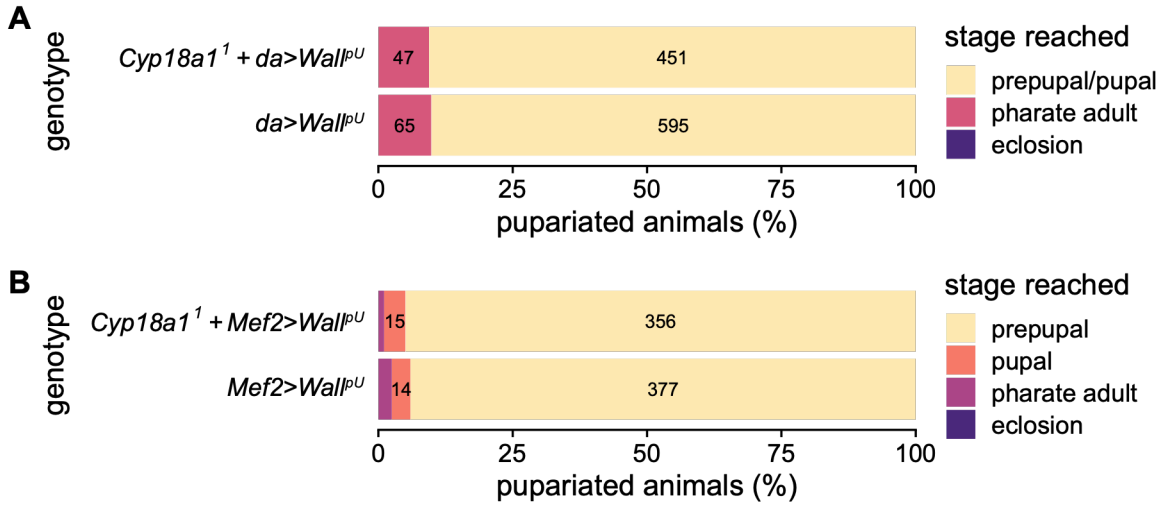


Figure 5.13. Tests of hypostasis between *Wall*^{pU} misexpression and *Cyp18a1*, using (A) the moderate, ubiquitous driver *da-GAL4* and (B) the muscle-specific driver *Mef2-GAL4*. Genotypes are pooled offspring genotypes from crosses with *Cyp18a1*^{1}/FM7i, *actGFP* chrX mothers (25% of offspring are *Cyp18a1*^{1} hemizygotes) or wild-type chrX mothers (100% of offspring have wild-type *Cyp18a1*). The number of individuals was not quantified before the prepupal stage, so proportions are given as the percentage of pupariated animals per genotype. Numbers on the bars are the number of individuals in each lethal phase category (for numbers greater than 10).

5.3.10. Co-misexpression of *EPPase* does not rescue developmental arrest caused by misexpression of *Wall*

If *Wall* misexpression results in the ectopic activity of an ecdysteroid kinase, resulting in developmental arrest due to the production of ecdysteroid-phosphate conjugates, I hypothesised that co-misexpression with a gene encoding an ecdysteroid-phosphate phosphatase, *EPPase* (*CG13604*), might rescue this developmental arrest by releasing free ecdysteroids from conjugates and increasing the ecdysteroid titre.

Co-misexpression experiments were conducted with the fat body/haemocyte-specific driver *cg-GAL4* and the trachea-specific driver *btl-GAL4*, both of which result in partially-penetrant developmental arrest at the pharate adult stage when driving UAS-*Wall*^{pU} at 25 ºC, with the expectation that if *EPPase* is able to rescue *Wall* misexpression, a higher proportion of adults will successfully eclose compared to *Wall* misexpression alone. Misexpression of UAS-*EPPase*^{3} alone did not cause detectable developmental arrest with either driver (Fig. 5.14). *cg-GAL4* co-misexpression significantly increased the proportion of eclosing adults compared to *Wall* misexpression alone, by a factor of 1.93 (95% CI: 1.16, 3.33; $p = 0.0081$, Fisher's exact test; Fig. 5.14A), but *btl-GAL4* co-misexpression significantly decreased the proportion of eclosing adults compared to *Wall* misexpression alone, by a factor of 7.04 (95% CI: 3.5, 15.7; $p = 7.3\times10^{-11}$, Fisher's exact test; Fig. 5.14B).

The opposing effect signs in these two experiments appear to contradict each other: while *EPPase* co-misexpression with *cg-GAL4* appears to partially rescue *Wall* misexpression, the opposite is true with *btl-GAL4*, with *EPPase* co-misexpression enhancing the developmental arrest penetrance of *Wall* misexpression. These differences may have been due to chance, with one or both effects due to insufficient sampling sizes; alternatively, *EPPase* misexpression may interact with *Wall* misexpression differently in different tissues—the likelihood of this explanation is unclear, given how much is unknown about the mechanism behind developmental arrest caused by *Wall* misexpression.

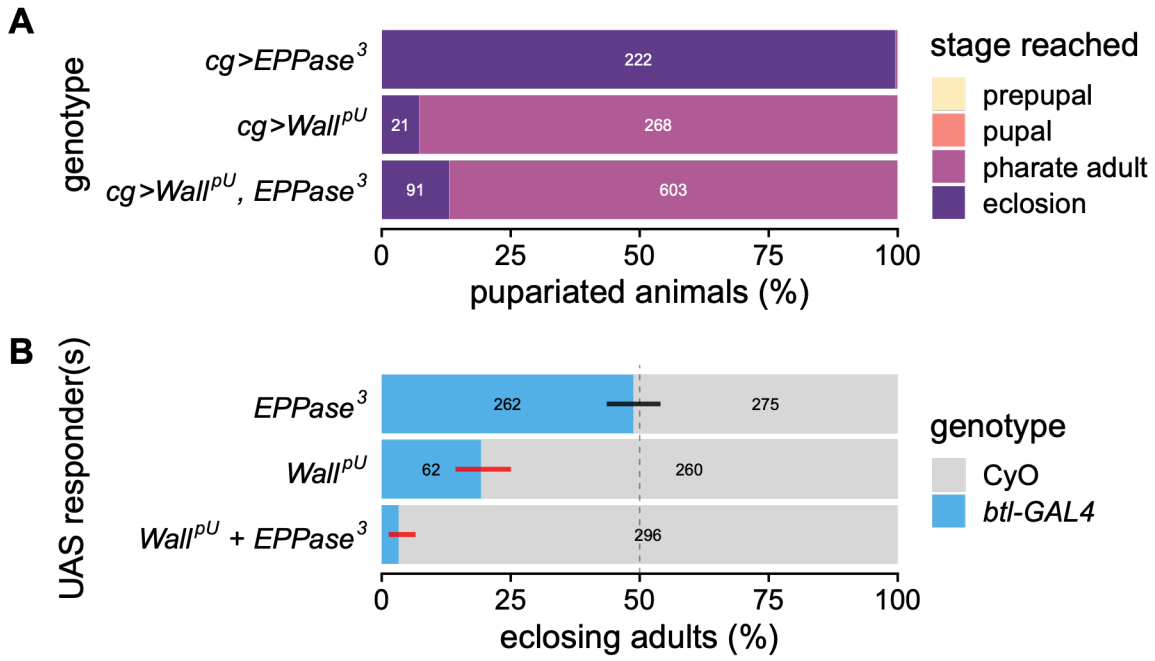


Figure 5.14. Tests of interaction between UAS-*Wall*^{pU} misexpression and UAS-*EPPase*^{3} misexpression using (A) the fat body/haemocyte-specific driver *cg-GAL4* and (B) the trachea-specific driver *btl-GAL4*. (A) Prepupal-to-adult viability of offspring from crosses between *cg-GAL4* females and UAS-responder males. The number of individuals was not quantified before the prepupal stage, so proportions are given as the percentage of pupariated animals per genotype. Numbers on the bars are the number of individuals in each lethal phase category (for numbers greater than 10). (B) Egg-to-adult viability estimated from the adult genotypic ratios of offspring from crosses between *btl-GAL4*/CyO females and UAS-responder males. The dashed line indicates the expected 1:1 genotypic ratio if both genotypes per cross are equally developmentally viable; error bars are 98.3% confidence intervals (95% CI adjusted for three tests) for the proportion of *btl-GAL4*-containing heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than 10).

5.3.11. *pkm* mutants are not more susceptible to starvation than wild-type controls

As previously stated (Chapter 5.1.4), *pkm* was found to be up-regulated in adult female oenocytes upon starvation (Chatterjee *et al.* 2014). This led to the hypothesis that *pkm* plays a role in the metabolic response to starvation, and that *pkm* loss-of-function mutants are more susceptible to starvation. Wet-starvation experiments (Storelli *et al.* 2019) were conducted with virgin male and virgin female flies, the offspring from three different crosses: pkm^4 females and w^{1118} males (yielding pkm^4/pkm^+ females and pkm^4/Y males); pkm^4 females and pkm^2 males (yielding pkm^4/pkm^2 females and pkm^4/Y males); and w^{1118} females and pkm^2 males (yielding pkm^2/pkm^+ females and pkm^+/Y males).

For heterozygous female genotypes (Fig. 5.15A), median survival in days were 9 (95% CI: 8, 9), 8 (95% CI: 8, 8) and 9 (95% CI: 8, 9) for pkm^4/pkm^2, pkm^4/pkm^+ and pkm^2/pkm^+, respectively, with no significant differences between any pairs of genotypes across entire survival curves (all $p > 0.6$, log-rank test). For hemizygous male genotypes (Fig. 5.15B), median survival in days were 7 (95% CI: 6, 7), 8 (95% CI: 7, 9) and 6 (95% CI: 6, 7) for pkm^4/Y (pkm^2 father), pkm^4/Y (w^{1118} father) and pkm^+/Y (pkm^2 father), respectively, with the only significant difference between the pkm^4/Y (pkm^2 father) and pkm^4/Y (w^{1118} father) genotypes ($p = 0.0037$, log-rank test; all other $p > 0.067$) across entire survival curves.

Given that there were no significant differences in survival between *pkm* loss-of-function animals and wild-type animals, these data do not support the hypothesis that *pkm* is involved in starvation resistance. The small significant difference between pkm^4/Y (pkm^2 father) and pkm^4/Y (w^{1118} father) males may have been due to differences in the autosomal genotype.

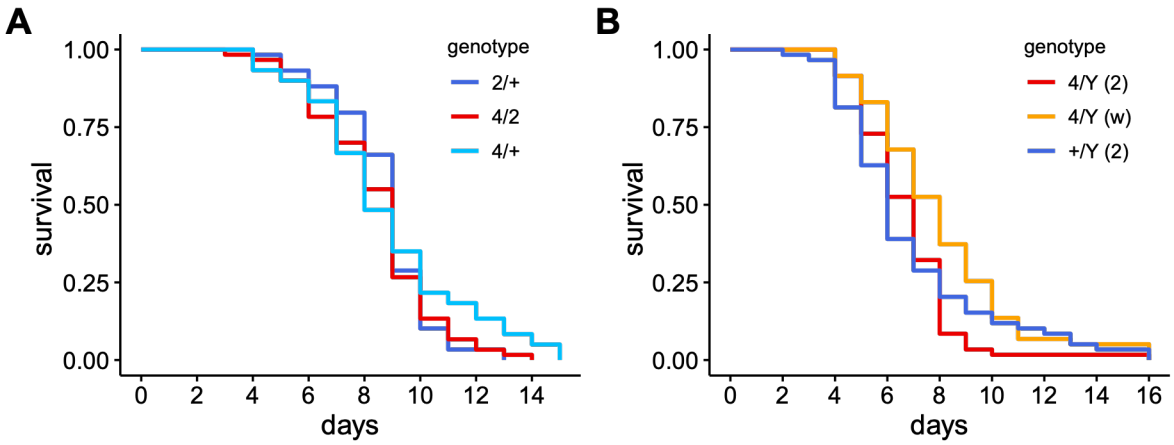


Figure 5.15. Wet-starvation survival experiments with *pkm* loss-of-function mutants (red and orange) and wild-type (dark blue and light blue) animals. (A) Virgin female survival curves. (B) Virgin male survival curves. Genotypes: 2/+, pkm^{2}/pkm^{+}; 4/2, pkm^{4}/pkm^{2}; 4/+, pkm^{4}/pkm^{+}; 4/Y (2), pkm^{4}/Y (pkm^{2} father); 4/Y (w), pkm^{4}/Y (w^{1118} father); +/Y (2), pkm^{+}/Y (pkm^{2} father). n = 59–60 adults per genotype.

5.4. Discussion

5.4.1. Discrepancies between gene disruption phenotypes

This chapter used three different gene disruption techniques to test the requirement of *Wall* and *pkm* function during development: null alleles, somatic knockout using Cas9, and RNAi knockdown. Single and double null mutant animals for *Wall* and *pkm* were completely viable, but somatic knockout of *pkm* either ubiquitously or with a pan-neuronal driver produced an immobility phenotype seemingly caused by a lack of motor coordination after eclosion, and RNAi knockdown of *Wall* and *pkm* under extreme conditions (high temperatures with a *Dcr2* transgene) produced developmental defects regardless of the construct used.

The typical explanation for phenotypic discrepancies between null alleles and RNAi is off-target knockdown, which may be a pervasive problem with RNAi technology (Seinen *et al.* 2011). However, another explanation is genetic compensation, wherein mutant alleles trigger a compensatory mechanism that restore the wild-type phenotype, but RNAi knockdown avoids this trigger, leading to a mutant phenotype (El-Brolosy & Stainier 2017). One proposed mechanism for genetic compensation is transcriptional adaptation, wherein mutant animals up-regulate genes that share a function with the disrupted gene; this process has recently been found to occur via the mRNA degradation pathway in zebrafish and mice (El-Brolosy *et al.* 2019), although it has not been studied in *D. melanogaster*. Of note, this mechanism requires the mutant allele be transcribed—the pkm^{2} allele characterised in this chapter (as well as the pkm^{4} allele) deletes the transcription start site of the gene, suggesting mutant mRNA may not be produced, although qPCR was not performed to check expression from the pkm^{2} allele. Despite this, I conducted a 'knockdown-in-mutant' test of the genetic compensation hypothesis, the results of which strongly suggested that the RNAi phenotypes are likely due to off-target effects, not a lack of genetic compensation.

The phenotypic discrepancy between *pkm* null alleles and somatic knockout is harder to explain, as there are very few published studies that have used somatic knockout in *D. melanogaster*, despite its potential in complementing or replacing tissue-specific RNAi knockdown (Huynh *et al.* 2018; Port *et al.* 2020; Port & Bullock 2016; Zirin *et al.*

2020). As such, it is unclear how widespread discrepancies between somatic knockout and null allele phenotypes are. Of course, one explanation could be that gRNA off-target mutagenesis is producing null alleles at other genomic loci, resulting in unintended phenotypes. While the gRNAs used in this study were designed to be as stringent as possible (following the approach of Gratz *et al.* 2014), occasional off-target mutagenesis may still have been occurring. The only issue with this explanation is that the somatic knockout immobility phenotype was extremely penetrant, with >99% of all adults being unable to move (Fig. 5.5B), which appears inconsistent with occasional off-target mutagenesis.

In theory, perfectly efficient somatic knockout of a gene—generating null alleles in every cell of the body—should completely phenocopy being homozygous for an inherited null allele. In practice, somatic knockout is likely not perfectly efficient, although efficiency increases substantially as the number of gRNAs targeting a locus also increases (Port & Bullock 2016)—this is one of the reasons the pCFD6 constructs used in this chapter were designed to express four gRNAs, and I would expect somatic knockout with the *pkm*^{pCFD6} construct to be highly efficient at inducing *pkm* null alleles in the vast majority of targeted cells. However, assuming somatic knockout of *pkm* was only moderately efficient, there may be a situation where such knockout might cause a phenotype not observed in null mutants: if genetic compensation occurs in a cell-autonomous fashion but needs to be coordinated (whereby some cells being wild-type and some cells being mutant can affect the global phenotype of the organism). In the case of *pkm* and the immobility phenotype, incomplete somatic knockout might disrupt neuronal networks by triggering a compensation pathway in only some neurons, altering how they interact or respond to stimuli.

Alternatively, somatic knockout might produce non-specific phenotypes through the ongoing DNA damage or chromatin remodeling that may occur due to the binding or cutting of DNA by Cas9 nucleases. This would likely be locus-dependent; indeed, no obvious phenotypes were visible in animals in which *Wall* was knocked out somatically (Fig. 5.5A).

5.4.2. Lack of *pkm*, *BmEc22K* and *EPPase* misexpression phenotypes

In this chapter, misexpression of ORFs using the UAS/GAL4 binary expression

system was used as a way to test the biochemical properties of enzymes. This has been used for *Cyp18a1*, which encodes an ecdysteroid 26-hydroxylase/carboxylase, in both *D. melanogaster* and *B. mori* (Guittard *et al.* 2011; Li *et al.* 2014; Rewitz *et al.* 2010), as well as an ecdysteroid 22-oxidase isolated from the entomopathogenic fungus *Nomuraea rileyi*, in *D. melanogaster* (Kamimura *et al.* 2012). In both cases, the enzyme's ecdysteroid catabolic function *in vitro* was recapitulated by phenotypes related to defects in ecdysteroid signalling *in vivo* upon misexpression of the ORF. One of the aims of this chapter was to test if misexpression of *Wall* and *pkm*, two candidate EcK-encoding genes, produced similar phenotypes, with the expectation that excess ecdysteroid phosphorylation *in vivo* would disrupt ecdysteroid signalling and developmental progression.

While the many phenotypes associated with misexpression of *Wall* will be discussed later (Chapter 5.4.7), it is also notable that misexpression of *pkm* (either full-length or without the N-terminal disordered region), *BmEc22K* or *EPPase* failed to arrest development when the very strong, ubiquitous driver *tub-GAL4* was used, which caused embryonic lethality with UAS-*Cyp18a1* and both UAS-*Wall* constructs. I chose to misexpress *BmEc22K* as a positive control for ectopic ecdysteroid 22-phosphorylation, as the phenotypic consequences of this are unknown; since 22-phosphates are thought to be physiologically inactive (Makka *et al.* 2002), the expectation was that *BmEc22K* misexpression would phenocopy *Cyp18a1* misexpression. *EPPase* misexpression was also tested to see if ectopic ecdysteroid dephosphorylation produced developmental defects, which if observed, would be evidence that ecdysteroid-phosphate conjugates play an important role during *D. melanogaster* development.

The simplest explanation for a lack of phenotype for the misexpression of all three genes could be that the misexpression was not taking place at all—qPCR validation of expression from the integrated pUASTattB constructs was not performed. Expression of an ORF from an integrated pUASTattB vector was clearly demonstrated here in the case of UAS-*Wall*^{pU}, which near-perfectly matched that of the native locus UAS-*Wall*^{EY} construct, albeit with less severe phenotypes (see Chapter 5.3.7), suggesting that the experimental procedure wasn't flawed in principle. It is possible undetected mutations in some fraction of the pUASTattB vector molecules prevented the expression of the ORFs, affecting some constructs but not others.

Another explanation for the lack of phenotypes is that misexpression of these ORFs simply does not affect ecdysteroid signalling *in vivo*. In the case of *pkm*, the substrates of its encoded enzyme may not be ecdysteroids at all, which is perfectly plausible. However, since misexpression of *BmEc22K* was expected to produce functional ecdysteroid 22-kinase, it is possible that misexpression of *BmEc22K* in *D. melanogaster* is insufficient to produce a functional enzyme—perhaps BmEc22K requires post-translational modification *in vivo* that does not occur in *D. melanogaster*. Sonobe *et al.* (2006) report the transfection of Sf9 cells (derived from the lepidopteran *Spodoptera frugiperda*) with a *BmEc22K*-expressing vector, resulting in ecdysteroid 22-phosphorylation *in vitro* with a cell extract; it is possible Sf9 cells, which are relatively closely related to *B. mori*, have the capacity to activate *BmEc22K* sufficiently to produce a detectable reaction. In the case of *EPPase*, ecdysteroid-phosphate conjugates may not be required for *D. melanogaster* development as hoped; alternatively, the short isoform, which was misexpressed in this chapter, may be inactive or otherwise less active than the long isoform—of note, the ecdysteroid-phosphate phosphatase activity reported by Davies *et al.* (2007) was found using the long isoform, and the short isoform was not tested for activity in their study. I did not misexpress the long EPPase isoform here due to problems with cloning.

5.4.3. Possible functions of *pkm*

Despite the inconsistent gene disruption phenotypes observed in this chapter (discussed in Chapter 5.4.1), the total weight of evidence may suggest that *pkm* has a function in the nervous system: knockdown with GD UAS-*dsRNA* constructs causes a 'drowning' phenotype, which is consistent with the seizure-like phenotype observed upon somatic knockout in the nervous system; knockdown of *pkm* in the nervous system affects the expression of *Hsp* genes, which may affect neuronal development and activity (Santana *et al.* 2020; although see Chapter 5.1.4); a TE-insertion near *pkm* affects a motor neuron phenotype (Chang *et al.* 2008); and *pkm* has also been transcriptionally linked to the *parkin* gene, which is associated with neuromuscular disorders in humans (Greene *et al.* 2005).

If *pkm* does function in the nervous system, there are no obvious neurological phenotypes in null mutants, suggesting the gene is dispensable for development, the phenotype is subtle, or mutant alleles can be compensated for. However, as previously

discussed (Chapter 5.4.1), both mutant alleles likely prevent normal transcription at the *pkm* locus, making transcriptional adaptation as described by El-Brolosy *et al.* (2019) an unlikely mechanism for hypothetical genetic compensation. Of note, *pkm* mutant animals also appeared to have morphologically normal eyes—this is significant given *pkm*'s extremely high and specific expression in adult eyes (Leader *et al.* 2018). While it is possible *pkm* mutants have visual defects or visual neurodegeneration phenotypes, this was not tested in this study.

I also proposed a hypothesis in Chapter 5.1.4 that *pkm* is involved in the starvation response in adult flies, based on transcriptomic data published by Chatterjee *et al.* (2014)—starvation experiments in this chapter with *pkm* null mutants failed to show either tolerance or susceptibility to starvation compared to wild-type flies, suggesting either *pkm* plays no role in starvation or this phenotype is also genetically compensated.

Another possible clue as to *pkm*'s function relates to *CG11200*, whose protein product may physically interact with Pkm (Guruharsha *et al.* 2011) and also has a TE-insertion associated with suppressing a motor neuron phenotype in the same manner as *pkm* (Chang *et al.* 2008). *CG11200* is predicted to encode a carbonyl reductase/dehydrogenase, but it does not have a known substrate. Its closest mammalian homolog is *DHRSX*, a predicted reductase/dehydrogenase in *H. sapiens* that has been linked to starvation-induced autophagy (Zhang *et al.* 2014a) but also has an unknown substrate. Curiously, FlyBase predicts that CG11200 is a retinol dehydrogenase based on sequence similarity to the human retinol dehydrogenase *RDH11* (Thurmond *et al.* 2019), and a zebrafish pigmentation mutation called *pye*, which produces reduced numbers of xanthophores and disrupts melanocyte patterning, has been mapped to *dhrsx*, the *Danio rerio* ortholog of *CG11200* (Mellgren & Johnson 2006). Given that *pkm* is strongly expressed in the *D. melanogaster* eye, if CG11200 and Pkm physically interact, this suggests that Pkm may be involved in retinoid or carotenoid metabolism, and these compounds might be good candidate Pkm substrates.

5.4.4. Future experiments on *pkm*

The outstanding questions regarding *pkm* relate to its substrate and any potential functions it may have. qPCR-validated misexpression of *pkm* should be performed to see

if misexpression indeed produces no developmental defects. Ideally, biochemical assays with heterologously expressed Pkm would be conducted to test its substrate specificity *in vitro*. From a gene disruption perspective, *pkm* mutants should be carefully examined for phenotypes related to visual function, given *pkm*'s high expression in the adult eye. The work of Santana *et al.* (2020) should also be replicated with somatic knockout experiments to confirm the relationship between *pkm*, *Hsp23* and *Hsp26*, as well as the synapse development phenotype they report. In order to fully resolve the issues with inconsistent phenotypes between somatic knockout and null alleles, independent gRNA-containing constructs should be produced; a transgenic CRISPR knockout line (which contains a single sgRNA, targeted to ~250 bp downstream of the translation start site; BL84057) from the 'TRiP-KO' library (Zirin *et al.* 2020) has recently been made available and could be used for this purpose. A wider array of neuron-specific and non-neuron-specific GAL4 drivers should also be used to pinpoint the specific tissues and/or neurons with which the immobility phenotype is associated.

5.4.5. Possible functions of *Wall*

Like *pkm*, I found in this chapter that null mutants for *Wall* did not have any obvious developmental or morphological defects. This suggests that *Wall* is not required for development, which is consistent with the recent loss of the *Wall*/Dro38 clade in the *Drosophila* subgenus Drosophila, which includes the annotated species *D. mojavenesis*, *D. virilis* and *D. grimshawi* (Chapter 2), and the more ancient loss of the Dip10 clade in the suborder Nematocera (Chapter 3). Given that *Wall* orthologs are otherwise relatively well conserved in Diptera, it is possible *Wall*'s function may have become redundant or inessential in *Drosophila* and Nematocera, either due to replacement by another gene or changes in the developmental or physiological importance of its substrate(s). In this respect, it is interesting to compare *Wall* with *Cyp18a1*, which shares a strong peak of expression with *Wall* during embryogenesis: despite a requirement for *Cyp18a1* function during metamorphosis, embryogenesis proceeds normally in $Cyp18a1^{1}$ mutants, suggesting the gene's absence can be compensated for (Rewitz *et al.* 2010); curiously, *Cyp18a1* has also been lost in the *Anopheles gambiae* species complex (Diptera: Nematocera; Neafsey *et al.* 2015). *Wall* mutants may be genetically compensated too; of note, neither *Wall* null allele characterised in this chapter likely prevents the transcription of the *Wall* locus, raising the possibility of transcriptional adaptation

through the mRNA degradation pathway, whereby expression of frameshifted mutant transcripts containing early stop codons can trigger the up-regulation of related genes (El-Brolosy *et al.* 2019).

It is possible *Wall* mutants do have a phenotype different from wild-type but it is too subtle to observe without a specific experimental setup or assay, such as a small defect in developmental timing or an adverse reaction to changing environmental conditions. If *Wall* does have such a function, it is likely related to ecdysteroids, due to the multiple lines of evidence presented in Chapter 5.1.5. *Wall* is natively expressed in the embryonic, larval and adult midgut (Leader *et al.* 2018; Weiszmann *et al.* 2009), as well as the ring gland (RG; Ou *et al.* 2016). *Wall*'s expression in the adult midgut appears to be mainly in the enteroendocrine cells and visceral muscle (Buchon *et al.* 2013; Dutta *et al.* 2013), which could be relevant to its function if it interacts with hormone signalling in the gut. However, it is unclear which part of the RG *Wall* is expressed in: the RG in higher Diptera is a composite gland comprised of three endocrine cell types that exist as three separate glands in other insects: the prothoracic gland (PG; which synthesises ecdysteroids), the corpus cardiacum (CC; which synthesises adipokinetic hormone) and the corpus allatum (CA; which synthesises juvenile hormones; Dai & Gilbert 1991). While one might assume the strong induction of *Wall* in *D. melanogaster* cell lines by 20E suggests that it might be expressed in the PG cells, PG ecdysteroidogenesis produces only E in *D. melanogaster*, with conversion to 20E occurring in peripheral tissues (Petryk *et al.* 2003), meaning there is unlikely to be a higher concentration of 20E in the PG than anywhere else in the body. There is also no evidence that *Wall* is induced by E *in vivo* (Beckstead *et al.* 2007). Additionally, the lack of phenotype upon overexpression of *Wall* in the PG strongly suggests that its substrate is not found in this tissue, making it unlikely that *Wall* would be natively expressed in the PG.

5.4.6. What is Wall's substrate?

Given the many reasons to hypothesise that Wall is an ecdysteroid kinase (Chapter 5.1.5), two logical substrates to first consider are 20E and E. The former is unlikely for multiple reasons: 20E is an essential signalling molecule in virtually all tissues during development (Gilbert 2011), yet developmental arrest through *Wall* misexpression is highly tissue-dependent; *Wall* misexpression in the embryo probably does not recapitulate the Halloween phenotype like *Cyp18a1* misexpression (see Chapter 5.4.7;

Rewitz *et al.* 2010); and while 20E induces *Wall* expression in cell lines, other 20E-induced genes rise and plateau in expression over time rather than showing a 'pulse-like' pattern as would be expected from rapid inactivation of 20E within the cell (Stoiber *et al.* 2016). E as a substrate is also unlikely, given that PG-specific misexpression of *Wall* causes no phenotype; the same logic also excludes ecdysteroid precursors, such as 2-deoxy-E or 2,22-dideoxy-E, as likely substrate candidates.

Other possible substrates are 26-hydroxylated ecdysteroids, produced by Cyp18a1 acting on E and 20E; this hypothesis was proposed in Chapter 5.1.5 to explain the co-expression of *Wall* and *Cyp18a1*, as well as the presence of ecdysteroid 26-kinase activity in the S2 cell line (Guittard *et al.* 2011). While the tissue-specificity of *Wall* misexpression might be explained by this hypothesis as the tissue-specific production of 26-hydroxyecdysteroids by *Cyp18a1*, the *Cyp18a1* epistasis experiments in this chapter seem to undermine the hypothesis, with no apparent rescue of *Wall*-misexpression developmental arrest in a $Cyp18a1^{1}$ background. This hypothesis may still be true if: Cyp18a1 is not the only ecdysteroid 26-hydroxylase in *D. melanogaster*; the $Cyp18a1^{1}$ allele is not a complete loss-of-function allele (Rewitz *et al.* 2010); or Wall can utilise other substrates besides 26-hydroxylated ecdysteroids. The plausibility of these postulates is hard to judge. Follow-up experiments that examine the ability of *Cyp18a1*-transfected S2 cells to produce ecdysteroid 26-phosphate conjugates when *Wall* is knocked out or knocked down would be a good way to further test this hypothesis.

If *Wall* does not encode an ecdysteroid 26-kinase, another EcKL likely does so. While, as previously noted (Chapter 5.1.5), *Wall* is the only EcKL transcriptionally induced in S2 cells ('S2-DGRC' isolate) by the presence of 20E (Stoiber *et al.* 2016), basal levels of expression of some EcKLs are relatively high in S2 cell isolates (Fig. S5.3). It is not clear which S2 cell isolate was used by Guittard *et al.* (2011), but genes with high expression in both isolates that are well conserved in *Drosophila* and Diptera are *CG2004* (Dro41-0/Dip12/subfamily B), *CG31975* (Dro44-0/Dip17/subfamily E) and *JhI-26* (Dro46-0/Dip16/subfamily F); any of these genes may encode the ecdysteroid 26-kinase, although they lack some of the other characteristics that initially made *Wall* such an attractive EcK candidate (Chapter 5.1.5).

3-oxoecdysteroids (Fig. 1.5A) are known to be produced in *D. melanogaster* larval fat body, epidermis and hindgut (Sommé-Martin *et al.* 1988a; 1988b) and may be active

hormones *in vivo* (Baker *et al.* 2003; Sommé-Martin *et al.* 1990); the putative 3-oxo-E 2-phosphate found by Hilton (2004) also suggests that 3-oxoecdysteroids may be phosphorylated *in vivo*. 3-epiecdysteroids may also be active hormones in larvae (Baker *et al.* 2003; Sommé-Martin *et al.* 1988a); of note, 3-epi-20E 3-phosphate might be formed in larvae despite the parallel formation of 3-epi-Eoic acids (Sommé-Martin *et al.* 1988a), raising the possibility that 3-phosphorylation could be a recycling reaction that protects 3-epiecdysteroids from irreversible 26-hydroxylation and carboxylation (Fig. 1.5).

As there is clearly much more to discover about ecdysteroid metabolism in *D. melanogaster* and other insects, unknown ecdysteroid metabolites may act as Wall's substrates instead. As described below, three genes encoding enzymes that may be associated with ecdysteroid metabolism—*Cyp301a1*, *Cyp303a1* and *spidey*—could belong to biochemical pathways that produce Wall's substrate.

Cyp301a1 is a P450 highly conserved across insects whose disruption, either by TE-insertion or RNAi knockdown, results in partially penetrant abdominal epidermal defects during metamorphosis, which can be rescued by feeding on 20E during larval development; this has led to speculation that the substrate of Cyp301a1 is an ecdysteroid (Sztal *et al.* 2012). As complete loss-of-function alleles of *Cyp301a1* have yet to be isolated and characterised, it is unclear what other developmental processes it may be involved in. However, *Cyp301a1* is expressed strongly in the larval trachea and the epidermis (Leader *et al.* 2018; Sztal *et al.* 2012), both of which appear to be tissues particularly sensitive to *Wall* misexpression, consistent with a hypothesis that Wall can act on an enzymatic product of Cyp301a1.

Cyp303a1 (also called *nompH*) is also a highly conserved P450 related to Halloween P450s that is required for embryogenesis, adult eclosion and the development of adult sensory organs (Willingham & Keil 2004; Wu *et al.* 2019b), and it has been speculated that it is involved in the metabolism of an ecdysteroid-like molecule (Wu *et al.* 2019b). *Cyp303a1* is expressed mostly during metamorphosis in pharate adult wings, but also has a peak of expression in the RG earlier in metamorphosis (Wu *et al.* 2019b); curiously, misexpression of *Cyp303a1* with the *tub-GAL4* driver can rescue null mutant phenotypes, but does not produce any developmental lethality on its own, suggesting it does not catabolise an active hormone (Wu *et al.* 2019b). While it is possible Wall

may act on the product of Cyp303a1, the tissue-specific expression profiles do not match well, unless the Cyp303a1 protein is produced locally but moves to other tissues to exert some of its functions.

spidey (also called *Kar*) encodes a putative lipid reductase/dehydrogenase that is required for normal ecdysteroid signalling in oenocytes; overexpression of *spidey* in oenocytes reduces whole-body levels of 20E, suggesting it may convert 20E to another, uncharacterised ecdysteroid (Chiang *et al.* 2016). While Wall could conceivably act on the enzymatic product of spidey, one piece of evidence against this hypothesis is that *Wall* misexpression in oenocytes fails to arrest development.

Of course, it is possible Wall does not act on ecdysteroids, which would suggest subfamily A EcKLs have a variety of endogenous substrates—a not-so-unlikely hypothesis given the phylogenomic evidence that subfamily A EcKLs may have transitioned from endogenous substrates to xenobiotic substrates multiple times in the Lepidoptera (Chapter 3). One possibility is that Wall can phosphorylate non-ecdysteroidal lipids, such as sphingolipids and glycerolipids; loss-of-function alleles of the lipid-phosphate phosphatase gene *wun* result in embryonic tracheal defects (Ile *et al.* 2012), suggesting that hyperphosphorylation of its products (by a kinase such as Wall) may produce a similar phenotype. Other substrates could be juvenile hormones or juvenile hormone-like compounds that also control aspects of insect development (Riddiford *et al.* 2010), although juvenile hormone diol kinases have already been identified in insects (Maxwell *et al.* 2002a; 2002b).

Rather than being due to tissue-specific substrates, an alternative explanation for the tissue-specific misexpression phenotypes of *Wall* could be tissue-specific post-translational modifications (PTMs). Wall has no predicted PTMs according to iProteinDB (Hu *et al.* 2019c), although this does not prove they do not occur *in vivo*. Another explanation for the tissue-dependent misexpression phenotypes is that an ecdysteroid-phosphate phosphatase (probably not EPPase/CG13604) is selectively expressed in certain tissues and protects them from the accumulation of ecdysteroid-phosphate conjugates.

5.4.7. How does *Wall* misexpression produce developmental defects?

Given that Wall's substrate is unknown, it is hard to say with any certainty what the misexpression of *Wall* is doing at a direct molecular level. However, the tissue-specific phenotypes catalogued in this chapter allow some insights into the possible signalling pathways and biological processes that can be influenced by Wall; once Wall's substrate is known, this information should be very useful in determining the compound's role in *D. melanogaster* development, and when and where it acts in the animal. *Wall* misexpression produces a complex collection of phenotypes depending on the tissue, but they can be distilled down to the following: tracheal defects; cuticle and denticle belt defects; ectopic melanisation or pigmentation defects; head eversion failure; and pharate adult lethality.

Wall misexpression with the *btl-GAL4*, *da-GAL4*, *act-GAL4* and *c204-GAL4* drivers produced defects in tracheal integrity, which may be due to loss of waterproofing, as seen in hydrocarbon biosynthesis mutants (Parvy *et al.* 2012); lack of tracheal integrity can also be due to defects in tracheal moulting (Kim *et al.* 2018a). Ecdysteroid signalling through EcR is required in the trachea for normal development (Buhler *et al.* 2018) and the 20E primary-response nuclear receptor DHR3, which controls the native expression of *Wall* in embryos, is also required for tracheal development during embryonic and larval stages (Lam *et al.* 1999; Ruaud *et al.* 2009). Null alleles in DHR78—a nuclear receptor involved in ecdysteroid signalling (Fisk & Thummel 1998) that can also inhibit EcR/Usp signalling (Zelhof *et al.* 1995)—also result in larval tracheal defects similar to those seen here, with fluid-filled regions (Astle *et al.* 2003; Fisk & Thummel 1998). While 20E is unlikely to be a substrate of Wall (Chapter 5.4.6), there may be other ecdysteroids that signal in the trachea, and it is possible the trachea have a deeply conserved endocrinological function in arthropods, as they share a developmental origin with the ring gland and are specified by similar genetic hierarchies (Cheng *et al.* 2014; Sánchez-Higueras *et al.* 2014). That pharate adult lethality was observed in putatively more moderate *btl>Wall*^{pU} misexpression animals also suggests that Wall's substrate affects the development of the trachea during metamorphosis as well as in larvae.

Wall misexpression with *act-GAL4* produced disordered and even ectopic denticle belts, which are parallel groups of trichomes on the larval cuticle that function in

mobility. Ecdysteroid signalling via EcR/Usp is required for trichome formation, as local inactivation of ecdysteroids prevents the local formation of trichomes in embryos (Chanut-Delalande *et al.* 2014), and hypomorphic alleles of *EcR* result in larval denticle belt defects (Bender *et al.* 1997). Denticle belt patterning is also controlled by the EGFR and Wingless (Wg) signalling pathways (Szuts *et al.* 1997), which are integrated through the transcription factor Shavenbaby (Svb; Chanut-Delalande *et al.* 2006). *polished rice* (*pri*; also known as *tarsal-less*/*tar*)—a polycistronic locus that encodes four peptides that control the maturation of Svb (Kondo *et al.* 2010)—is a direct target of EcR/Usp and is strongly co-regulated with DHR3 (Chanut-Delalande *et al.* 2014).

Wall misexpression with *act-GAL4* (or the poorly characterised *c204-GAL4*) produced dark melanisation of the posterior spiracles and rectal pad. Adult pigmentation defects were also seen in successfully eclosing *btl>Wall*^{pU} individuals, and localised melanisation defects were sometimes seen in tracheal branches when misexpressing with *act-GAL4*, *da-GAL4* or *btl-GAL4*. Melanisation in insects is part of the innate immune response to microbial pathogens and endoparasitoids, and dysregulation of immune pathways can lead to dramatic tracheal melanisation and localised 'melanotic pseudotumors' (Minakhina & Steward 2006; Tang *et al.* 2008). However, to the best of my knowledge, melanised spiracles and rectal pads are not frequently observed in immune system mutants, suggesting these phenotypes may be due to other, unknown processes.

Pharate adult lethality can be caused by a number of issues during metamorphosis, including defects in fat body remodelling (Bond *et al.* 2010), βFTZ-F1 inactivation (Rewitz *et al.* 2010), loss of oenocytes and hydrocarbon biosynthesis (Chiang *et al.* 2016; Gutierrez *et al.* 2007), and neuropeptide signalling defects required for eclosion (Loveall & Deitcher 2010; Park *et al.* 2003). Given that misexpression of *Wall* in the fat body caused pharate adult lethality, this suggests that Wall's substrate may affect fat body remodelling, which is already known to be under the control of ecdysteroid signalling (Bond *et al.* 2010).

Strong, ubiquitous misexpression of *Wall* with the *tub-GAL4* driver produced embryonic lethality. While this was not investigated in more detail, under a light microscope, misexpression embryos appeared to have secreted larval cuticle and may have been pharate larvae (Table S5.1); this is contrast with misexpression of *Cyp18a1*, which

phenocopies 'Halloween' ecdysteroidogenic mutants (Rewitz *et al.* 2010), so named for their lack of cuticle (Gilbert 2004). This strongly suggests that Wall does not inactivate 20E but does affect an essential process during embryogenesis; this also may explain why native *Wall* expression appears so strongly restricted to the midgut in the embryo (Weiszmann *et al.* 2009).

The *HR>Wall^{EY}* phenotype (Fig. S5.2) is reminiscent of phenotypes associated with muscular and neuronal defects during metamorphosis, resulting in head eversion failure, an elongated puparium and red eye pigmentation developing within the thorax (Cauchi *et al.* 2008; Diao *et al.* 2015; Zhao *et al.* 2008). Given that the muscle-specific *Mef2-GAL4* driver also produces arrest during metamorphosis, albeit before the differentiation of adult structures, this suggests that Wall can act within muscles to prevent normal developmental progression. Ecdysteroid signalling through EcR-B1 is required for the apoptosis of abdominal muscles after head eversion (Zirin *et al.* 2013), which may be linked to the *HR>Wall^{EY}* phenotype, although it is possible other ecdysteroid signalling pathways also affect muscle development.

While neuron-specific misexpression (via *nSyb-GAL4*) produced no obvious developmental phenotypes, the fact that the UAS-*Wall^{EY}* (*EY20330*) construct enhances the lethality of the *D42>DNGlued* phenotype in *trans* (Chang *et al.* 2013) suggests that ectopic Wall can interact with neuromuscular junctions (NMJs); given that ecdysteroid signalling is involved in the development of NMJs (Liu *et al.* 2010), this may be what Wall is affecting.

Some phenotypes are linked by multiple possible molecular mechanisms, which represent strong candidates for pathways that are affected by Wall. Tracheal defects and adult pigmentation defects seen with *btl-GAL4* could be explained by an interaction between the substrate of Wall and the scavenger receptor Dsb, which is involved in both tracheal integrity and adult pigmentation (Dembeck *et al.* 2015; Wingen *et al.* 2017). In addition, dysregulation either upstream or downstream of Svb and *pri* may explain both the denticle belt and tracheal phenotypes. Indeed, loss-of-function alleles in *Megf8*, the *D. melanogaster* homolog of the human gene *MEGF8*, cause suspiciously similar phenotypes to misexpression of *Wall*, including larval food aversion and disordered denticle belts; trachea-specific knockdown of *Megf8* with *btl-GAL4* causes developmental lethality, suggesting it has a specific role in the trachea (Lloyd *et al.* 2018).

While the exact signalling roles of Mefg8 are unknown, it likely acts upstream of EGFR, which regulates Svb (Szuts *et al.* 1997)—this further suggests that Wall's substrate may interact with Svb. Curiously, Svb and *pri* also act in the adult midgut, where they coordinate the behaviour of stem cells under the control of EcR signalling (Hayek *et al.* 2019). Given that *Wall* is natively expressed in the midgut (Leader *et al.* 2018), this raises the possibility that it functions here too, although perhaps in a redundant fashion given the lack of abnormal phenotypes observed in *Wall* null mutants.

5.4.8. Future experiments on *Wall*

Given the lack of null allele phenotypes for *Wall*, future work could focus—as with *pkm*—on more subtle phenotypic assays, such as those focusing on developmental timing; if null mutant animals show a phenotype, the UAS-*Wall*^{pCFD6} construct could be used to explore the tissue-specific nature of the phenotype, if any. A genetic compensation test similar to that done for *pkm* with RNAi knockdown in a mutant background should also be conducted.

However, the most fruitful line of future work on *Wall* is likely to be related to its misexpression phenotypes and determining Wall's substrate. 20E feeding experiments—which are common in the literature (Niwa *et al.* 2010; Parvy *et al.* 2014; Talamillo *et al.* 2008; Yoshiyama *et al.* 2006; Zhou *et al.* 2004) and were planned for this chapter but not carried out—could be used to test whether ectopic Wall is disrupting 20E signalling specifically; assaying the levels of E and 20E in misexpression animals would also be helpful in testing this hypothesis. Close inspection of the embryonic-lethal *Wall* misexpression phenotype should also be conducted to confirm that it does not match that of Halloween mutants. A larger collection of GAL4 drivers (including CA- and CC-specific drivers) could also be used to more precisely pinpoint the exact tissues where Wall has an effect, and drivers with potentially ambiguous expression patterns used in this study—such as *HR-GAL4*, *c204-GAL4*, *elav-GAL4* and *ppl-GAL4*—should be examined with fluorescent reporters to discover exactly where they express GAL4. Epistasis experiments, like those in this chapter with *Cyp18a1*, could also be used to test the possible biochemical associations between *Wall* and *Cyp301a1*, *Cyp303a1* and *spidey*; independent loss-of-function alleles of *Cyp18a1* could also be generated to replicate the experiments performed here. *Wall*-knockout or -knockdown experiments with S2 cells could also be performed to test whether this abolishes the

ecdysteroid 26-kinase activity observed by Guittard *et al.* (2011). Examination of changes in some of the signalling pathways mentioned in Chapter 5.4.7 should also be done to test those mechanistic hypotheses. Ultimately, biochemical assays on recombinant Wall protein would be a powerful way to determine its substrate. Comparative structural analyses of both Wall and BmEc22K would also be useful in determining how these enzymes function and what determines their substrate specificity.

5.5. Conclusion

This chapter has provided the first in-depth functional genetic characterisation of candidate ecdysteroid kinase-encoding genes in insects, focusing on the *BmEc22K* orthologs *Wall* and *pkm* in *Drosophila melanogaster*. While neither gene appears essential for development or reproduction, striking developmental arrest phenotypes due to the misexpression of *Wall* are suggestive of hidden complexity in tissue-specific ecdysteroid signalling and metabolism in *D. melanogaster*, particularly in the trachea, fat body, muscles and epidermis. This chapter also serves to highlight the possible phenotypic inconsistencies that may arise between functional genetic methodologies, and reinforces the importance of careful consideration, design and validation of RNAi knockdown, somatic knockout and misexpression experiments when exploring the functions of uncharacterised genes.

5.6. Supplementary Materials

5.6.1. Fly crossing for genotype creation

w^{1118}; *Kr*$^{If-1}$/CyO, *actGFP*; *Sb*^{1}/TM3, *actGFP*, *Ser*^{1} (also known as *w*^{1118}-DB^{GFP}) was made by routine crosses, starting with separately crossing BL4534 and BL36320 to *w*^{1118}-DB (Chapter 4.2.1), until the desired genotype was reached. FM7j; 25709; 25709 flies were made by routine crosses, starting with separately crossing BL6418 and BL25709 with *w*^{1118}-DB, until the desired genotype was reached. FM7i, *actGFP*/C(1)DX, *y*^{1}, *f*^{1}; 25709; 25709 flies were made by routine crosses, starting with separately crossing BL4559 and BL25709 with *w*^{1118}-DB, until the desired genotype was reached. FM7j;; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1} flies were made by routine crosses, starting with BL6418 females and *w*^{1118}-DB males, until the desired genotype was reached. The FM7c balancer was crossed out from *y*^{1}, *w**, *P{lacW}Fas2G0032*, *P{neoFRT}19A*/FM7c; *P{ey-FLP.N}5* using *w*^{1118}-DB and then crossed to *w*^{1118}. *pkm*^{2}/FM7c;; *Wall*^{1} flies were made by routine crosses, starting with crossing *pkm*^{2}/FM7i, *actGFP* females and FM7c/Y males, *w*^{1118}-DB females and FM7c/Y males, and *w*^{1118}-DB females and *Wall*^{1} males; the presence of each deletion allele in the final stock was confirmed by PCR as described in Chapter 5.2.3. The chr3 UAS-*Cas9* construct was crossed out of BL67083 by routine crosses involving *w*^{1118}-DB. UAS-*Wall*^{pCFD6}/CyO, *actGFP*; UAS-*Cas9* flies were made by routine crosses, starting with separately crossing UAS-*Wall*^{pCFD6} and UAS-*Cas9* to *w*^{1118}-DB^{GFP}, until the desired genotype was reached. UAS-*pkm*^{pCFD6}; UAS-*Cas9*/TM3, *actGFP*, *Ser*^{1} flies were made by routine crosses, starting with separately crossing UAS-*pkm*^{pCFD6} and UAS-*Cas9* to *w*^{1118}-DB^{GFP}, until the desired genotype was reached. UAS-*Dcr2*/CyO, *actGFP*; *tub-GAL4*/TM3, *actGFP*, *Ser*^{1} was made by routine crosses, starting with UAS-*Dcr2*/CyO; *tub-GAL4*/TM6B and *w*^{1118}-DB^{GFP}, and selecting the appropriate genotypes. *tub-GAL80*^{ts}; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1} was made by routine crosses, starting with *w**; *tub-GAL80*^{ts}; TM2/TM6B males and *w*^{1118}-DB females, until the desired genotype was reached. *tub-GAL80*^{ts}; UAS-*Wall*^{EY} flies were made by routine crosses, starting with BL23106 males and *w*^{1118}-DB females, and then crossing to *tub-GAL80*^{ts}; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1}, until the desired genotype was reached. *tub-GAL80*^{ts}; *da-GAL4* flies were made by routine crosses, starting with *da-GAL4* males and *w*^{1118}-DB females, and then crossing to *tub-GAL80*^{ts}; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1}, until the desired genotype was reached. *Cyp18a1*^{1}/FM7i, *actGFP*; UAS-*Wall*^{pU} flies were made by routine crosses, starting with crossing FM7j;; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1} females to UAS-*Wall*^{pU} males, and

Cyp18a1^{1}/FM7i, *actGFP* females to FM7j/Y;; *Sb*^{1}/TM6B, *Antp*^{Hu}, *Tb*^{1} males, until the desired genotype was reached. UAS-*Wall*^{pUI}; UAS-*EPPase*^{3} flies were made by routine crosses, starting with separately crossing UAS-*Wall*^{pUI} and UAS-*EPPase*^{3} to *w*^{1118}-DB, until the desired genotype was reached; homozygosity of each construct was confirmed by eye colour. *c204-GAL4*/TM3, *actGFP*, *Ser*^{1} flies were made by crossing BL3751 and BL4534 and selecting the appropriate genotypes. *w*^{1118}; *Kr*$^{If-1}$/CyO; *nos-GAL4*, UAS-*Cas9* was made as previously described (Chapter 4.2.1).

pkm^{2}, *w*$^{-}$; KK^{pkm}, *w*$^{+mC}$; + lines were made by crossing *pkm*^{2}, *w*$^{-}$; +; + females to *w*$^{-}$; KK^{pkm}, *w*$^{+mC}$; + males (VL106503), then crossing F1 males (*pkm*^{2}, *w*$^{-}$/Y; KK^{pkm}, *w*$^{+mC}$/+; +) back to *pkm*^{2}, *w*$^{-}$; +; + females. Red-eyed male and female F2 individuals (*pkm*^{2}, *w*$^{-}$; KK^{pkm}, *w*$^{+mC}$/+) were crossed together and pairs of single male and female red-eyed F3 offspring (either *pkm*^{2}, *w*$^{-}$; KK^{pkm}, *w*$^{+mC}$/+; + or *pkm*^{2}, *w*$^{-}$; KK^{pkm}, *w*$^{+mC}$; + which are virtually phenotypically identical as a single copy of *w*$^{+mC}$ results in a near-wild-type eye colour) were crossed to establish separate lines that were either homozygous for KK^{pkm}, *w*$^{+mC}$ or segregating a wild-type 2nd chromosome. DNA from the parents of each line was individually prepared using the single-fly squish-prep protocol (Appendix 3.1) and a genomic reverse primer (NC_R1; Table 5.2) was used, along with the NC_Genomic_F and pKC26_R primers from Green *et al.* (2014), to genotype the KK-RNAi library insertion site (VIE-260B-2) on the 2nd chromosome in a three-primer PCR using GoTaq Green Master Mix—cycling conditions: 3 min 95 ºC initial denaturation, followed by 35 cycles of 30 sec 95 ºC denaturation, 30 sec 60 ºC annealing and 45 sec 72 ºC extension, then 5 min 72 ºC final extension. NC_Genomic_F and NC_R1 produce a 724 bp amplicon from wild-type chromosomes, while NC_Genomic_F and pKC26_R produce a ~600 bp amplicon from KK^{pkm}, *w*$^{+mC}$ chromosomes (the large NC_Genomic_F and NC_R1 amplicon from KK^{pkm}, *w*$^{+mC}$ chromosomes fails to amplify with a 45 sec extension). PCR reactions were run on a 1.5% agarose gel for 20 min at 300V, and the presence of a single 724 bp band (and absence of a ~600 bp band) from both parents indicated the resultant line was completely homozygous for KK^{pkm}, *w*$^{+mC}$. DNA from VL106503 flies was used as a homozygous KK^{pkm}, *w*$^{+mC}$ control, and DNA from a *Wall*^{1} fly was used as a homozygous wild-type 2nd chromosome control.

5.6.2. Supplementary Figures

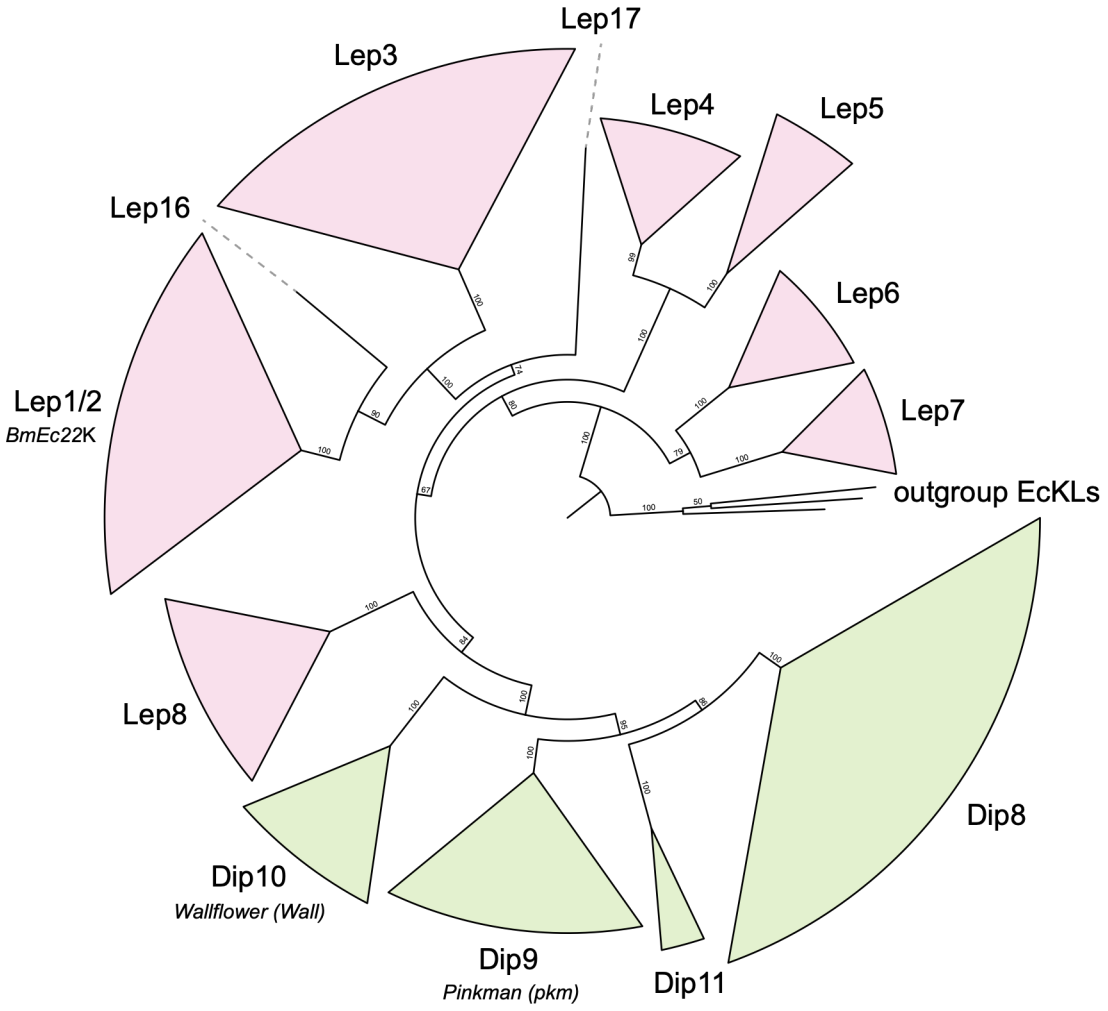


Figure S5.1. Phylogenetic tree of EcKLs in subfamily A from the orders Diptera (green) and Lepidoptera (pink), collapsed to the order-ancestral clades defined in Chapter 3, with the locations of the *BmEc22K*, *Wall* and *pkm* genes noted. The Lep1 and Lep2 clades are not well resolved on this tree and have been collapsed together. The tree was generated as described in Chapter 3.2.2. Branch numbers are ultra-fast bootstrap support values from UFBoot2, where values of 95 or above are considered reliable (Hoang *et al.* 2018). Tree is rooted with EcKLs from outside of the A subfamily.

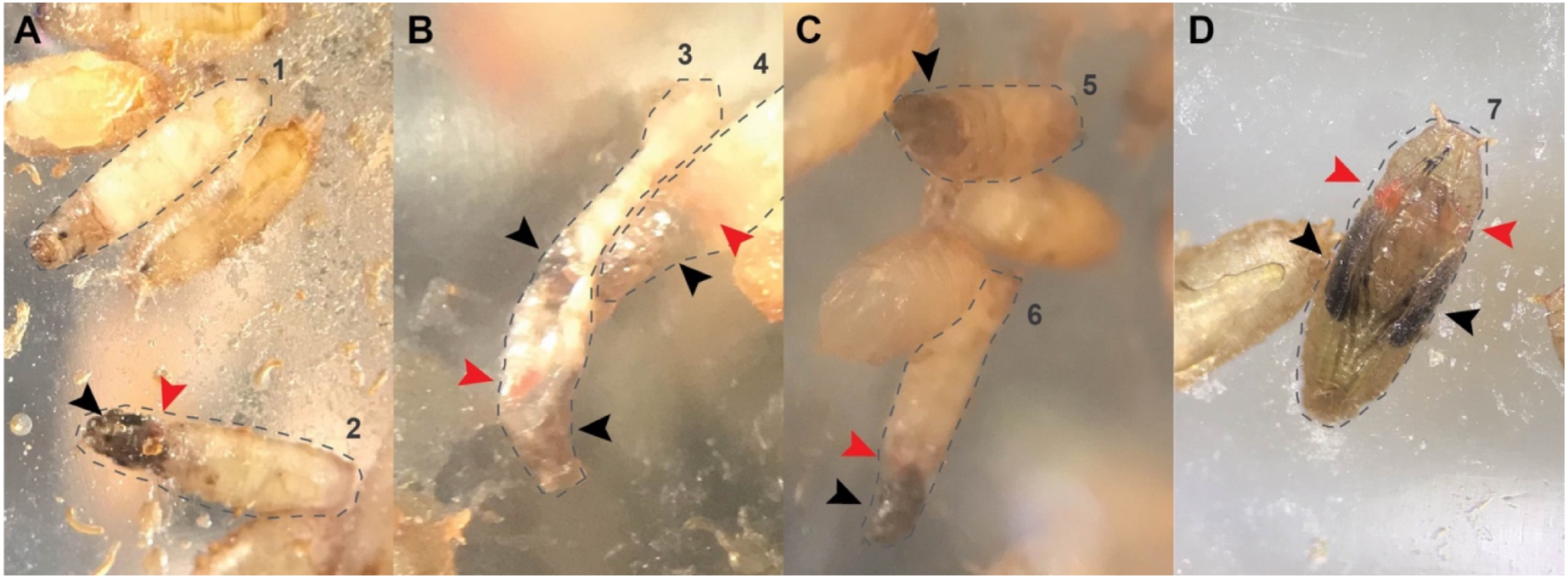


Figure S5.2. (A–C) Phenotype of *HR>Wall*^{EY} animals at 25 ºC (photos also contain some non-misexpression animals). Grey dashed lines separate the visible regions of each misexpression animal of note, which are also numbered (1–6). Regions of dark (thoracic/wing) pigmentation are marked with black arrowheads, regions of red (eye) pigmentation are marked with red arrowheads. (A) Animal 1 shows an early stage of the phenotype, before the differentiation of adult structures within the puparium. Animal 2 shows a later stage, where the adult thorax and head have differentiated without head eversion, resulting in red eye pigment posterior to the dark pigment of the wings. (B) Similar phenotypes to animal 2 in animals 3 and 4, although 3 has two regions of dark pigmentation. Note that animal 3 appears 'deflated', which was common in older animals, suggestive of cuticular defects. (C) Animal 5 is viewed mostly from a superior perspective, obscuring the red pigment, while animal 6 is viewed along the length of the body, again revealing red pigment posterior to the dark pigment. (D) Wild-type animal (7) showing normal pharate adult development and the positions of the eyes and the wings.

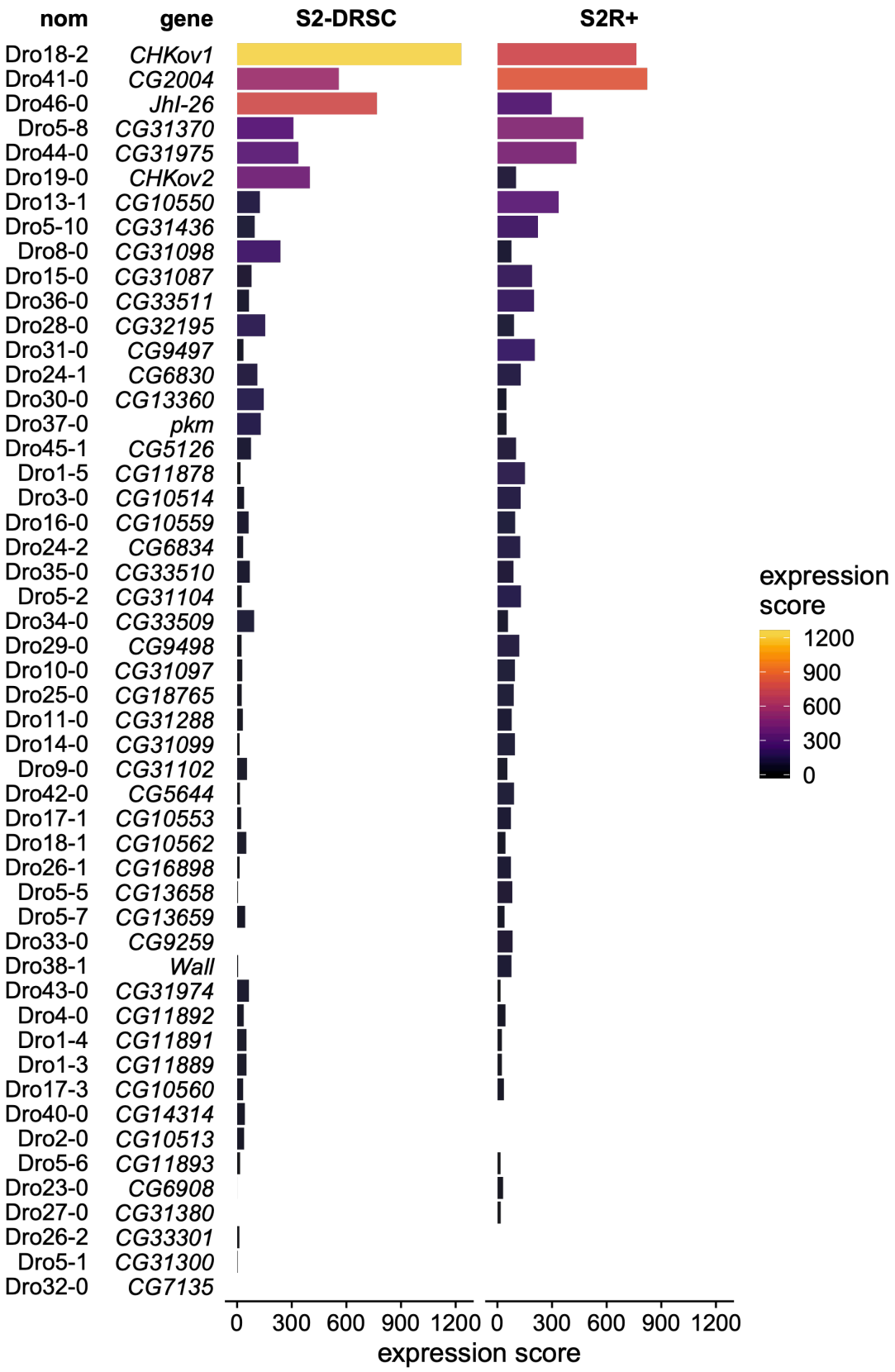


Figure S5.3. Basal expression ('expression scores' from microarrays) of *D. melanogaster* EcKL genes in two isolates of the S2 cell line—S2-DGRC and S2R+—as reported by Cherbas *et al.* (2011). Genes are ordered by combined total expression score. nom, *Drosophila* genus EcKL nomenclature (see Chapter 2).

5.6.3. Supplementary Tables

Table S5.1. Phenotypic descriptions of misexpression of *Cyp18a1* and *Wall* with various GAL4 drivers.

Driver	Expression	UAS-Cyp18a1	UAS-Wall^{EY}	UAS-Wall^{pU}
tub-GAL4	ubiquitous (very high)	Embryonic lethal	Embryonic lethal, appears to produce pharate larvae	Embryonic lethal, appears to produce pharate larvae
act-GAL4	ubiquitous (high)	Embryonic lethal	Embryonic and larval lethal, larvae have dark posterior spiracles and brown patches of epidermis, particularly in the posterior (rectal pad). Also denticle belts may be disrupted	ND[a]
da-GAL4	ubiquitous (moderate)	Embryonic lethal	Embryonic and larval lethal, larvae have dark posterior spiracles and brown patches of epidermis, particularly in the posterior (rectal pad). Also denticle belts may be disrupted	Not embryonic lethal, but some larvae stop growing and wander early (short and 'tubby' like phenotype). Majority of animals pupariate but gas bubble movement is impaired and pupation does not occur (no head eversion). No tracheal defects observed.
dan^{AC116}-GAL4	eye, 3rd antennal segment regions of the eye disc	Prepupal lethal, gas bubble defects, no head eversion	NOP[b]	NOP
phm-GAL4	prothoracic gland, imaginal discs	Embryonic lethal	NOP	NOP
Sgs3-GAL4	salivary glands	Pharate adult semi-lethal, most adults eclose but some do not. Those that do not eclose have defects in mouthparts.	NOP	NOP
nSyb-GAL4	pan-neuronal	Embryonic lethal	NOP	NOP
elav-GAL4	pan-neuronal, imaginal discs, glia, trachea	Embryonic lethal	Embryonic and larval lethal. Larvae may have similar phenotype to act-GAL4 but this was not checked in detail.	Prepupal/pupal lethal

Driver	Expression	UAS-Cyp18a1	UAS-Wall^{EY}	UAS-Wall^{pU}
mex1-GAL4	midgut (enterocytes only)	Prepupal lethal, gas bubble defects and puparium defects	NOP	NOP
HR-GAL4	midgut, fat body, Malpighian tubules	Larval lethal, larvae never pupariate, most die wandering as very small L1/L2, some grow to terminal size L3 but continue to feed	Extended wandering period as larvae, and ultimately prepupal/pupal lethal—hard to tell because puparia never shorten and tan properly. However, some 'pupae' start to form adult structures, such as eyes and wings (can see red and black pigment) but internalised—head eversion does not occur	Pupal/pharate adult lethal. If pharate adult lethal, abdominal development is impaired and individuals never successfully eclose.
cg-GAL4	fat body, haemocytes	Prepupal/pupal lethal. no adult differentiation, some puparium defects (flattened)	Pharate adult lethal, some pharate lethal adults have defects in abdomen	Pharate adult semi-lethal, some pharate lethal adults have defects in abdomen
ppl-GAL4	larval fat body, other tissues	Embryonic lethal	Prepupal lethal, but also partially larval lethal. Affected larvae have trouble moulting and sometimes get their mouthhooks caught in their cuticle. Appears to be some sort of double mouthhook phenotype	NOP
c204-GAL4	follicle cells, oocyte stage 8-14, probably other tissues	Embryonic lethal	Tracheal defects (discontinuous gas), dark posterior spiracles, some brown patches of epidermis (rectal pad). Some larvae die at early instars, some become large 3rd instar but die on the side of vials, some failing to pupariate and forming mishapen pseudo-puparia	Pupal/pharate adult lethal. If pupal lethal, development is arrested after head eversion and adult differentiation does not appear to happen (no pigmentation of eyes or bristles)
btl-GAL4	trachea	Embryonic lethal	Larval lethal, larvae die on the sides of vials probably as early 3rd instar (medium sized larvae). Some evidence of tracheal melanisation, but no dark spiracles or epidermis	Pharate adult semi-lethal, vast majority of escaping adults have patches of light pigmentation on their dorsal abdomens

Driver	Expression	UAS-Cyp18a1	UAS-Wall^{EY}	UAS-Wall^{pU}
Mef2-GAL4	muscles	Embryonic lethal	Prepupal lethal. Gas bubble always misplaced (middle of puparium or posterior) and head eversion never occurs. Apolysis never occurs	Prepupal lethal. Gas bubble always misplaced (middle of puparium or posterior) and head eversion never occurs. Apolysis never occurs
Eip71CD-GAL4	mid-L3 larval brain, epidermis	Larvae grow normally until wandering, wherein they wander for a considerably long time, eventually die on side of vials without pupariating. Some larvae died buried deep in food media	Prepupal and pupal lethal, with some pharate adult lethal. Some animals fail to evert the head, others evert but have misplaced gas bubbles. Some appear relatively normal pharate adults but never eclose	Prepupal and pupal lethal, with some pharate adult lethal. Some animals fail to evert the head, others evert but have misplaced gas bubbles. Some appear relatively normal pharate adults but never eclose
desat1-GAL4	oenocytes, male accessory gland	Pharate adults fail to eclose. Some have defects in dorsal abdomen, some appear morphologically normal	NOP	NOP
UO-GAL4	Malpighian tubule principle cells	NOP	NOP	NOP
e22c-GAL4	epidermis, embryo, imaginal discs, follicle cells	NOP	NOP	NOP

[a] NOP, no obvious phenotype; [b] ND, not determined

Chapter 6

General Discussion

6.1. Introduction

The motivating question at the heart of this project was: what are the functions of the ecdysteroid kinase-like (EcKL) gene family? In Chapter 1.5, I outlined two broad hypotheses with respect to this question, that some EcKLs:

1. mediate the catabolism and recycling of ecdysteroid hormones, and therefore play important roles in developmental endocrinology, and
2. are responsible for the phosphorylation of dietary toxins seen across insects.

Hypothesis 1 has already been supported by previous work by Haruyuki Sonobe and colleagues (Ito *et al.* 2008; Ito & Sonobe 2009; Sonobe *et al.* 2006; Sonobe & Ito 2009), who identified *BmEc22K*, the only biochemically and genetically characterised ecdysteroid kinase (EcK), but I aimed to discover the first EcK genes in *Drosophila melanogaster* using reverse genetic methods. This work was partially successful: while I did not find unequivocal evidence of another EcK in the EcKL family, the *BmEc22K* ortholog *Wallflower* (*Wall*) has many properties consistent with encoding an EcK enzyme, and given the interesting phenotypes produced by its misexpression, identifying its substrate is likely to lead to a deeper understanding of ecdysteroid signalling and its role in the development of *D. melanogaster* and other insects (Chapter 5). Two other well-conserved EcKLs in *D. melanogaster*—*CG5644* (Appendix 1) and *CG31098* (Appendix 2)—may also encode EcKs, although they have yet to be characterised in detail.

Hypothesis 2 was much more strongly supported by this work, with multiple lines of functional and phylogenomic evidence, including:

- a 'detoxification score' shows nearly half of *D. melanogaster* EcKLs are detoxification candidate genes, a metric benchmarked against the cytochrome P450s, a known detoxification gene family (Chapter 2);
- phenotypic variation related to toxic stress is associated with genomic and transcriptomic variation in a number of EcKLs in *D. melanogaster*, including susceptibility to natural (caffeine, methylmercury and ethanol) and synthetic (imidacloprid and chlorantraniliprole) toxins (Chapter 2);
- EcKL gene family size across insects is associated with the sizes of known

detoxification gene families, diet and the 'detoxification breadth' (DB; Chapter 1.2.1) metric, and EcKLs have reduced rates of duplication in taxa with small DB (Chapter 3);

- EcKL gene family size and the size of two ancestral clades are strongly associated with host plant diversity in herbivorous Lepidoptera (Chapter 3); and
- the rapidly evolving Dro5 genes in *Drosophila* modulate developmental susceptibility to two natural toxins, caffeine and kojic acid, in *D. melanogaster* (Chapter 4).

I believe there is now sufficient evidence to claim that the EcKL gene family functions in both ecdysteroid metabolism and detoxification processes, and it should be considered a multi-functional gene family in a similar manner to the P450s. However, this raises many further questions about the EcKLs, which I will discuss in the rest of this chapter.

6.2. EcKLs as ecdysteroid kinases

6.2.1. How many EcKLs encode ecdysteroid kinases?

Despite the initial annotation of the EcKL gene family as solely encoding ecdysteroid kinases (EcKs), data in this chapter demonstrate that at least some EcKLs function in detoxification and instead encode xenobiotic kinases. As such, it is unclear how many EcKLs truly encode EcKs and whether this is the function of an appreciable fraction of the gene family as a whole.

There are four known positions on the ecdysteroid nucleus that can be phosphorylated: C-2, C-3, C-22 and—after hydroxylation—C-26 (Fig. 1.6). How one might estimate the number of required EcKs in an insect genome depends on the answers to a number of questions:

1. How many unique ecdysteroid-phosphate species are present in each insect?
2. How specific is each EcK with respect to hydroxyl position? (i.e. Can one enzyme phosphorylate multiple different hydroxyl groups?)
3. How specific is each EcK with respect to its substrate? (i.e. Can one enzyme phosphorylate multiple different ecdysteroids?)
4. How tissue-specific are individual ecdysteroid kinases? (i.e. Are there two or more EcK isozymes that catalyse the same reaction but are spatially or temporally separate and encoded by separate genes?)

Question 1 is hard to answer, given that detailed biochemical data on the composition of the 'ecdysteroidome' are only really available for a handful of species, and even then the data are likely patchy, having been generated many decades ago. The species with the largest number of known ecdysteroid-phosphate species is *Bombyx mori*, with 13—three 2-phosphates, four 3-phosphates and six 22-phosphates—based on 13 unique free ecdysteroids (The Ecdysone Handbook, 3rd edition; http://ecdybase.org), however, it may possess more that have yet to be detected, such as 26-phosphates found in the related species *Manduca sexta* (Feldlaufer *et al.* 1987). Answers to questions 2 and 3 are unclear based on the limited amount of biochemical data available on EcK substrate specificity. BmEc22K is thought to only phosphorylate C-22, based on inhibition assays (Sonobe *et al.* 2006), but that it cannot phosphorylate other

hydroxyl positions has not been robustly demonstrated; it can, however, phosphorylate at least four ecdysteroids at the C-22 position (Sonobe *et al.* 2006), suggesting it could be responsible for all 22-phosphates found in this species. Question 4 is also very hard to answer. The existence of the paralogous Halloween genes *spook* and *spookier* in *D. melanogaster*, as well as many other Cyp307 paralogs (Rewitz *et al.* 2007; Sztal *et al.* 2007; Caitlyn Perry, unpublished data), suggest that ecdysteroid biosynthetic enzymes can overlap in their substrate specificity if they are spatially and/or temporally separated, although the reason why this has occurred repeatedly to Cyp307 enzymes is currently unclear. It is possible that similar arrangements have evolved for EcKs as well, perhaps in insects where EcKLs generally conserved as single-copy orthologs are present as a small number of paralogs, such as subfamily C in *Acyrthosiphon pisum* (Hemiptera: Aphidoidea) or subfamily D in *Neodiprion lecontei* and *Athalia rosae* (Hymenoptera: Tenthredinoidea; Chapter 3).

Overall, if a species' EcK activities are position-promiscuous, substrate-promiscuous and are not split between isoenzymes, only one EcK gene might be needed in its genome, but if a species' EcK activities are position-specific, substrate-specific and many are split between multiple isozymes, many dozens of EcK genes might be needed in its genome. In all likelihood, a realistic scenario probably falls somewhere in-between, with no more than 10 or so EcK genes per genome, depending on the number of other ecdysteroid catabolism genes also present. However, this is just a speculative guess, and future research on the biochemistry of ecdysteroid phosphorylation should aim to shed light on the matter.

6.2.2. Conservation of ecdysteroid kinases in insects

In Chapter 1.3.4, I hypothesised that genes encoding ecdysteroid catabolic enzymes may have lower conservation than genes encoding ecdysteroid biosynthetic enzymes, based on the wide variety of seemingly independent ecdysteroid inactivation pathways observed across insects (Rharrabe *et al.* 2007). The phylogenomic data described in Chapter 3 tend to support this 'poor conservation' hypothesis with respect to the EcKL family, which likely encodes all or most of the ecdysteroid kinases in insects (see Chapter 1.3.7): no single EcKL subfamily is retained in all insect taxa, although some clades are highly conserved as single-copy orthologs within orders.

While there is currently insufficient data to estimate how often transitions in ecdysteroid-phosphate functionality occur, or how much redundancy exists in ecdysteroid phosphorylation pathways, it appears likely that many types of EcK activity are poorly conserved between taxa, with ovarian/embryonic ecdysteroid-phosphate (OEEP) recycling systems (Chapter 1.3.6) in some taxa (Lepidoptera and Orthoptera) but not others (*Drosophila*), and even OEEP systems in relatively closely related species can differ in the particular ecdysteroid-phosphate conjugates used—*Bombyx mori* uses 22-phosphates and 3-phosphates (Sonobe & Ito 2009) and *Manduca sexta* uses 26- and 2-phosphates (Feldlaufer *et al.* 1987), yet they are both members of the superfamily Bombycoidea.

Why transitions between ecdysteroid-phosphate conjugates should occur with any regularity is up for debate. In the case of simple catabolism, any number of ecdysteroid modification or conjugation reactions should be able to substitute for each other, allowing essentially neutral changes in biochemistry without changing the required endocrinological effect—but ecdysteroid recycling is a more complex system. For example, with respect to the *B. mori* and *M. sexta* OEEP systems, the inferred transition between 22- and 26-phosphate conjugates (or vice versa) may have also required a transition between 22- and 26-phosphatases (or vice versa), unless the ancestral enzyme possessed both activities; transitions between broader conjugation types—such as phosphorylation and acylation, which may have happened in an ancestor of Diptera or of Lepidoptera—may have been even harder to evolve, with a guarantee of two enzymes being replaced simultaneously.

It is also hard to imagine such changes in biochemistry being adaptive, unless a species has multiple ecdysteroid recycling systems that overlap spatially or temporally that need to avoid interacting with each other biochemically, resulting for selection on changes in substrates or reactions. Along these lines, it is tempting to speculate that the presence of ecdysteroid-acyl conjugates as a maternal ecdysteroid contribution in the eggs of *D. melanogaster* (Bownes *et al.* 1988)—rather than ecdysteroid-phosphates like in other insects—might have something to do with the presence of ovarian ecdysteroid 22-phosphates (Grau *et al.* 1995; Pis *et al.* 1995): it may be that the latter conjugates control a distinct endocrinological process that needs to be kept separate from the maternal ecdysteroid contribution, requiring a different biochemistry. Such interactions (or lack thereof) should become more obvious as more is discovered about

ecdysteroid conjugation and its physiological functions.

6.2.3. Candidate ecdysteroid kinase EcKLs

Many ecdysteroid-phosphate conjugates are not yet associated with an EcK gene, and 'deorphaning' these conjugates requires the identification and functional validation of candidate genes. Despite my hypothesis in Chapter 5 that *Wallflower* encodes the ecdysteroid 26-kinase proposed by Guittard *et al.* (2011), experimental data did not support the hypothesis, and no additional EcK–conjugate pairs have been discovered in this thesis. As such, future research will need to identify and validate other candidate genes.

Wall remains a strong EcK candidate, although as discussed in Chapter 5.4.6, there are no clear candidates for its substrate. In Lepidoptera, other genes in the EcKL A subfamily paralogous to *BmEc22K* and the Lep1 clade, perhaps in the Lep2–3 and Lep16 (to a lesser extent, Lep4–8) clades, may be EcKs; in particular, the EcKs responsible for the 2- and 3-phosphates in *B. mori* and the 2- and 26-phosphates in *M. sexta* are likely to belong to these clades. This is in contrast to Orthoptera, which lacks subfamily A genes, meaning their EcKs—most notably the biochemically characterised 2-kinase and 22-kinase (Kabbouh & Rees 1993; 1991)—are likely not encoded by direct orthologs of *BmEc22K* (Chapter 3.3.2), and must be found in other subfamilies.

CG31098 (Dro8-0/Dip5/subfamily H), which likely plays a role in choriogenesis in *D. melanogaster* (Appendix 2), may also encode an EcK, specifically the ecdysteroid 22-kinase hypothesised to exist in the adult ovary of this species (Grau *et al.* 1995; Pis *et al.* 1995); this or another EcK could also be encoded by *CG5644* (Dip42-0/Dip14/subfamily D), misexpression of which (either ubiquitously or in the prothoracic cells of the ring gland) causes developmental arrest (Appendix 1). Both of these genes are relatively well conserved as single-copy orthologs in Diptera (Chapter 3), suggesting at least one EcK is also well conserved in this order.

Subfamily C EcKLs are well conserved in most insect orders (Fig. 3.6) and the *D. melanogaster* ortholog *CG14314* (Dro40-0/Dip13) has enriched expression in the nervous system (Leader *et al.* 2018)—given that DopEcR functions in the nervous system (Evans *et al.* 2014; Ishimoto *et al.* 2012; Petruccelli *et al.* 2020; 2016), this raises the

possibility that *CG14314* has a conserved role interacting with DopEcR by phosphorylating its ecdysteroid ligands. Interestingly, CG14314 is also predicted to physically interact with *D. melanogaster* EPPase (CG13604; Murali *et al.* 2011), which is also predominantly expressed in the nervous system (Leader *et al.* 2018) and has been previously hypothesised to be functionally linked to DopEcR (Chapter 1.3.8), raising the possibility that all three proteins interact both biochemically and physically and providing a strong basis for the hypothesis that CG14314's substrate is an ecdysteroid. Given that misexpression of *CG14314* may not cause developmental arrest (Appendix 1)—somewhat inconsistent with a role in catabolising active ecdysteroids—it is possible CG14314's ligand is not required for development and instead functions only in behaviour (Petruccelli *et al.* 2020); alternatively, its activity could be tightly regulated post-translationally, explaining why transcriptional mis- or over-expression fails to cause ectopic ecdysteroid kinase activity and developmental arrest.

Based on the above reasoning, it is likely there are at least a few EcKs present in the A+B+C+D+E+F+G+H+L+M insect superclade (Fig. 3.10), but whether or not the subfamily I and J+K clades also contain EcKs remains to be determined. EcKLs in the insect subfamily I are highly retained as single-copy orthologs across insects, with few absences in annotated genomes (including *D. melanogaster*; Chapter 3), suggesting its substrate and function relate to a biological process essential in most insects. It is tempting to speculate that subfamily I genes could encode EcKs, although it currently does not seem that certain ecdysteroid-phosphate conjugates appear consistently across all or most insects (Rharrabe *et al.* 2007).

6.3. EcKLs as detoxification enzymes

6.3.1. Phosphorylation and the classical model of detoxification

The biochemistry of detoxification has classically been divided into three 'phases' (Williams 1951), but the rigid distinctions between these phases don't appear to hold in nature (see Chapter 1.2.2 & Fig. 1.2). As an understudied detoxification reaction, phosphorylation—which would typically be thought of as a 'Phase II' reaction that necessarily acts on the products of Phase I enzymes—also resists this classification system. Existing evidence, limited as it is, suggests that detoxicative kinases can act on three broad classes of substrates: already-hydroxylated (AH) compounds, Phase I products (PIPs) and Phase II products (PIIPs): known AH substrates are phytoecdysteroids, phenolic compounds and harmol; known PIPs are terfenadine metabolites; and known PIIPs are salicinoid glycosides and midazolam glycosides (see Table 2.1). Whether detoxicative kinases (assumed to be EcKLs) typically act on one class of substrate more than the others is yet to be determined, but it is likely all three classes are substrates for EcKLs in most or all insects.

Why detoxicative kinases act on PIIPs (particularly glycosides) is unclear, as such metabolites are presumably already detoxified. One explanation is that phosphorylation prevents the reactivation of glycoside conjugates by glycosidases, as proposed by Boeckler *et al.* (2016). Aglycones of phenolic glycosides may be detoxified by prophenoloxidases in the gut (Wu *et al.* 2015), but glycosides may be absorbed into the hemolymph before this can occur; insects are already known to modify glycosides in other ways before hydrolysis to limit their toxicity (Pentzold *et al.* 2014). Another related explanation is that many PIIP-phosphate metabolites detected *in vivo* are the products of side reactions that are not adaptive *per se* but occur because some detoxicative kinases function to directly phosphorylate ingested glycosides, which are common plant secondary metabolites (Boeckler *et al.* 2011; Dobler *et al.* 2011). Boeckler *et al.* (2016) also suggest that phosphorylation could act as a 'metabolic tag' to selectively remove glycosides from particular tissues with phosphate-specific efflux transporters, but this explanation may not be plausible, as glycoside-specific transporters are already known to exist in insects (Strauss *et al.* 2013).

Due to a lack of appropriate hydroxyl groups, it is unlikely that detoxicative kinases act on glutathione moieties, and PIIP substrates may be limited to the products of UDP-glycosyltransferases (UGTs), unless detoxicative kinases can also act on the products of more exotic—and poorly understood—forms of conjugation, such as serine and threonine conjugation (Novoselov *et al.* 2015).

6.3.2. Physiological considerations of detoxicative phosphorylation

Mitchell (2015) notes that, since phosphate is vital for primary metabolism (e.g. its use in the synthesis of ATP and other nucleotides), detoxicative phosphorylation could be limited by the dietary intake of phosphorus in the diet—however, he notes that in humans ~40% of ingested phosphate is not absorbed by the gastrointestinal tract, so it may not be in short supply. In insects, phosphorous deficiency unsurprisingly limits growth (Bergwitz *et al.* 2013; Perkins *et al.* 2004; Visanuvimol & Bertram 2011), but insect diets vary in their phosphate content, and so using a proportion of available phosphate in order to utilise a toxin-containing food resource may be a viable feeding strategy for some insects. Boeckler *et al.* (Boeckler *et al.* 2016) note that gypsy moths appear to use approximately 25% of their phosphorus intake on salicinoid phosphorylation. There is some evidence that specialist caterpillars grow faster than generalist caterpillars on natural diets (Coley *et al.* 2006), which is consistent with the hypothesis that generalist herbivores with large detoxification breadth (DB) expend a non-trivial amount of their energy and nutrient budgets on metabolic detoxification, and nutritional supply and demand may explain variation in detoxification enzymes used between generalist insects. Conversely, excess dietary phosphate reduces growth and survival in grasshoppers (Cease *et al.* 2016), which excrete excess phosphate in frass (Cease *et al.* 2016; Zhang *et al.* 2014b); hypothetically, such excretion could be due to the excretion of phosphate conjugates formed by EcKLs, although this has not been determined. Curiously, a number of EcKLs in *D. melanogaster* are very highly enriched for expression in the Malpighian tubules—the main osmoregulatory and excretory organs of insects (Beyenbach *et al.* 2010)—such as *CG9259* (Dro33-0/Dip20/subfamily J), *CG10513* (Dro2-0/Dip3/subfamily H), *CG10514* (Dro3-0/Dip3/subfamily H) and *CG11892* (Dro4-0/Dip3/subfamily H). These genes are not otherwise good detoxification candidate genes (Chapter 2), suggesting they could have a role in the excretion of other compounds by the Malpighian tubules—these genes may be good candidates for regulating phosphate levels via excretion.

6.3.3. Natural X-class EcKL substrates

The question of which genes are good candidates for encoding X-class ('xenobiotic function'; Chapter 2.2) EcKLs has previously been discussed in Chapters 2 and 3, but the identification of possible naturally occurring substrates for these enzymes deserves further attention. Of course, all previously identified phosphorylated xenobiotic compounds collated in Table 2.1 are strongly hypothesed to be EcKL substrates, but apart from some phenolics, phenolic glycosides, harmol and ecdysteroids, there are few naturally occurring compounds that have been demonstrated to be detoxified by phosphorylation.

Phytoecdysteroids are good candidates for EcKL substrates in herbivorous insects, given their wide distribution in plants (Dinan 2001) and the known functions of at least some E-class EcKLs in ecdysteroid metabolism (Sonobe *et al.* 2006). Consistent with this, *Spodoptera littoralis* (Lepidoptera: Noctuidae) and *Manduca sexta* (Lepidoptera: Sphingidae) larval midgut homogenates can convert ecdysteroids to 2- and 22-phosphates (Webb *et al.* 1996; 1995) and 2- and 3-phosphates (Weirich *et al.* 1986; Williams *et al.* 1997), respectively, and acetylphosphate conjugates are dominant metabolites of ingested ecdysteroids in *Locusta migratoria* (Orthoptera: Acrididae; Modde *et al.* 1984). Curiously, however, phytoecdysteroids are primarily converted to acyl esters in many noctuid moths (see references in Duan *et al.* 2020), suggesting that phytoecdysteroid phosphorylation is not the dominant detoxification pathway in all insects. In addition, that noctuid moths have large expansions of the detoxification candidate Lep1 and Lep8 EcKL clades (Chapter 3) suggests that these clades may not be primarily utilising phytoecdysteroids as substrates in most species. However, although no ecdysteroid-phosphate conjugates have been identified in the larval midgut of *Helicoverpa armigera* (Lepidoptera: Noctuidae) after ingesting ecdysteroids (Duan *et al.* 2020; Robinson *et al.* 1987), nine EcKLs are up-regulated in the larval midgut 3 hours after ingestion of 20E (Duan *et al.* 2020). Of note, all of these genes are in the Lep1 and Lep8 clades—while these nine EcKLs may not be involved in phytoecdysteroid inactivation *per se*, they may be induced by a xenobiotic-response pathway that is activated by the ingestion of exogenous ecdysteroids; in support of this hypothesis, Duan *et al.* (2020) also find numerous P450 genes up-regulated in the same dataset, which may be induced by the same hypothetical pathway.

Other steroidal secondary metabolites of plants are also good EcKL substrate candidates, including withanolides, cardenolides and cucurbitacins (Agrawal *et al.* 2012; Dinan *et al.* 1997; Glotter 1991), as are glycosides, either formed by the action of UGTs (Chapter 6.3.1) or those directly ingested, such as flavonoid glycosides or iridoid glycosides, which are widespread in plants (Ibanez *et al.* 2012; Peterson *et al.* 2006; Wang *et al.* 2017). Also, as hypothesised in Chapter 4, hydroxylated metabolites of toxins such as methylxanthine alkaloids may also be EcKL substrates. Ecdysteroids present in fungi (mycoecdysteroids; Kovganko 1999) might also be EcKL substrates for fungivorous insects.

Filamentous fungal competitors of insects such as *Drosophila* spp. likely also synthesise hydroxylated toxins that could be EcKL substrates, such as kojic acid (see Chapter 3), patulin and citrinin (Beard & Walton 1969; Paterson *et al.* 1987).

6.4. Other EcKL functions

6.4.1. Juvenile hormone and EcKLs

Ipi10G08, a juvenile hormone (JH)-inducible EcKL in *Ips pini* (Coleoptera: Scolytidae) belongs to subfamily E (Bearfield *et al.* 2008); *JhI-26* from *Drosophila melanogaster* (Diptera: Drosophilidae) is another JH-inducible gene (Dubrovsky *et al.* 2000) but belongs to subfamily F—this implies that multiple EcKLs may have independently evolved to be regulated by JH. There is compelling genetic evidence that *JhI-26* functions in reproduction and modulates the expression of sperm proteins (Liu *et al.* 2014), while it has been hypothesised that *Ipi10G08* is involved in the regulation of aggregation pheromone biosynthesis in the midgut of *I. pini* (Bearfield *et al.* 2008). The substrates of these enzymes have not been identified but are likely small-molecule downstream effectors of JH signalling that link global changes in JH titre to specific physiological processes; as such, identifying the substrates of these EcKLs would likely shine substantial light on the molecular mechanisms underpinning JH signalling in insects.

Subfamily E EcKLs are numerous in coleopteran genomes (Chapter 3), raising the question of how many of them have similar functions to *Ipi10G08*; the same can be said of subfamily F EcKLs, which have expanded substantially in some lineages of Hymenoptera and Hemiptera (Chapter 3). In particular, the aphid *Acyrthosiphon pisum* (Hemiptera: Aphididae) has 13 subfamily F EcKLs—it is tempting to speculate that at least some of these genes are involved in regulating aphid reproduction or polyphenism, given the role of JH in these processes (Brisson 2010; Ishikawa *et al.* 2012).

6.4.2. Non-ecdysteroidal E-class EcKL substrates

The cytochrome P450s contain most of the Halloween genes, which encode ecdysteroid biosynthetic enzymes (Niwa & Niwa 2014), but many other P450s encode E-class ('endogenous function'; Chapter 2.2) enzymes with a variety of substrates and/or products, from hydrocarbons to juvenile hormones (Table 2.2). In a similar way, while it is clear that some E-class EcKL genes encode ecdysteroid kinases, it is very likely that other E-class genes in the family encode enzymes with non-ecdysteroidal substrates.

As suggested in Chapter 5, some EcKLs could be biochemically associated with E-class P450s that catalyse unknown—yet seemingly developmentally essential—reactions, such as Cyp301a1 (Sztal *et al.* 2012) and Cyp303a1 (Wu *et al.* 2019b). While these P450s might be involved in ecdysteroid metabolism (Sztal *et al.* 2012; Wu *et al.* 2019b), they might also be involved in the biosynthesis or degradation of other endogenous hormones that act as ligands for currently-orphaned nuclear receptors (Fahrbach *et al.* 2012), the activity of which some EcKLs could also regulate. In addition, prostaglandins and other eicosanoids—some of which have hydroxyl groups—modulate reproduction and immunity in insects (Stanley 2006; Tootle 2013) and might be phosphorylated by EcKLs. Eicosanoids are known to be conjugated in mammals (Dalli *et al.* 2014), although—to my knowledge—eicosanoid conjugates have not yet been detected in insects, possibly because their metabolism is currently poorly studied in this taxon, despite a couple of decades of progress (Stanley & Kim 2018). Tyrosine-O-phosphate is also found in *D. melanogaster* and other *Drosophila* spp. and appears to act as a storage form of tyrosine before it is used in cuticular tanning during metamorphosis (Kramer & Hopkins 1987; Mitchell *et al.* 1960); the kinase responsible for its formation has yet to be described (to my knowledge) and could be encoded by an EcKL gene.

Despite the assumption in this thesis that the relatively close homology between EcKLs and the choline/ethanolamine kinases and aminoglycoside phosphotransferases (and the known small-molecule substrate of BmEc22K; Sonobe *et al.* 2006) implies that all EcKLs must also have small-molecule substrates, it is of course also possible that some EcKLs use peptides or proteins as substrates, given that the gene family belongs to the PKinase clan (CL0016), which includes many protein kinase families (El-Gebali *et al.* 2018). Little is known about the structure of the EcKL active site that confirms or disconfirms this possibility—although Rebecca Gledhill-Smith, another PhD student in the Robin Lab, is currently working in this area.

6.4.3. S-class EcKL functions

The focus of this thesis has been on the hypothetical E- and X-class functions of EcKLs, but it is also possible that some EcKLs may have S-class ('secondary function'; Chapter 2.2) functions in the metabolism of secondary metabolites, such as defensive compounds, pheromones or pigments, again in a similar manner as the P450s (Beran *et al.* 2019; Wojtasek & Leal 1999).

A natural fit for S-class EcKLs could be the storage and/or degradation of steroidal secondary metabolites in insects, such as lucibufagins, defensive compounds produced by fireflies (Coleoptera: Lampyridae; Eisner *et al.* 1978); it would be illuminating (no pun intended) to explore the recently sequenced genomes of fireflies to see if they possess conspicuously expanded clades of EcKLs, in a similar manner to the large expansion of the Cyp303 clade, hypothesised to be involved in lucibufagin biosynthesis (Fallon *et al.* 2018). Defensive compounds sequestered from an insect's diet or otherwise from their environment might also be stored in an inactive form through EcKL-mediated phosphorylation, which could be dephosphorylated in response to predation, in a similar manner to the 'mustard oil bomb' involving glucosinolates (Kazana *et al.* 2007), although such phosphorylated defensive compounds have, to my knowledge, yet to be detected in insects. S-class EcKLs could also be involved in the degradation of odorants in the olfactory system, perhaps acting on the catalytic products of carboxylcholinesterases (Leal 2013).

Even though phosphorylation might naturally be thought of more as an 'inactivating' reaction, S-class EcKLs could also hypothetically be involved in the biosynthesis of secondary metabolites, such as in the biosynthesis of the psychoactive fungal secondary metabolite psylocibin, which involves a phosphotransferase similar to insect EcKLs (Fricke *et al.* 2017).

6.5. The evolution of EcKL functions

6.5.1. Evolutionary transitions between EcKL functions

Transitions between functions is thought to be relatively common in large, dynamic gene families, particularly for those encoding enzymes (Podar *et al.* 2005; van Loo *et al.* 2018), such as the P450s (Kawashima & Satta 2014; Sezutsu *et al.* 2013). As such, it is likely this has occurred at least a few times in the EcKL gene family in insects. For example, if lepidopteran EcKLs in the Lep1 and Lep8 clades at least partially encode detoxicative kinases, as suggested by their association with host plant diversity (Chapter 3.3.10), this indicates that multiple transitions may have occurred between ecdysteroid kinase activity and detoxicative kinase activity in EcKL subfamily A in Lepidoptera. *B. mori*'s sole Lep1 gene—*BmEc22K*—encodes an ecdysteroid kinase (Sonobe *et al.* 2006), yet *Manduca sexta*, which is also a member of the superfamily Bombycoidea, possesses eight Lep1 genes, all of which appear orthologous to *BmEc22K* (Fig. S3.4). Similarly, expansions of the Lep1 clades in Noctuoidea, Papilionoidea, Pyraloidea, Geometroidea and Tortricoidea all appear to have occurred independently from Bombycoidea (Fig. S3.4), with seemingly no fixed gene duplications occurring before the divergence of these Lepidopteran superfamilies—suggestive that ecdysteroid-to-xenobiotic substrate transitions may have occurred at least six times in Lepidoptera. Alternatively, it is possible the ancestral Lep1 gene was a bi-functional enzyme that acted on ecdysteroid and xenobiotic substrates—perhaps phytoecdysteroids.

Further evidence for functional transitions of EcKLs comes from Diptera. In this order, subfamily H has greatly expanded (Fig. 3.10) and contains 88% (21/24) of the detoxification candidate genes in *D. melanogaster* identified in Chapter 2, including the Dro5 genes experimentally linked to detoxification in Chapter 4; and it is likely many other subfamily H genes are detoxification genes in other species. However, this subfamily also contains a number of genes that may have roles in development or reproduction, such as *CG31098* (Dro8-0/Dip5; Appendix 2); even within the hyper-diverse subfamily H clade Dip4, which contains 11 detoxification candidates in *D. melanogaster*, there are a number of genes that likely have no function in detoxification, such as *CG31099* (Dro14-0) and *CG31087* (Dro15-0). In addition, Dip4 genes in *Glossina* are largely

conserved (Fig. 3.24), suggesting at least some of them have important E-class functions. Overall, this suggests multiple transitions have occurred between E-class and X-class (and perhaps S-class) functions in the Dip4 clade alone and are likely common in the H subfamily in general.

As discussed in Chapter 3.4.9, it appears possible that certain subfamilies (such as A and H) may have independently expanded in multiple insect orders to fulfil X-class functions while remaining stable at very low copy-number in other orders—if this is the case, it may also imply evolutionary transitions between an ancestral E- or S-class function and a derived X-class function; whether or not these independent subfamily expansions involve convergent neofunctionalisation towards similar X-class substrates remains to be determined. Given the size of some of these subfamilies in certain taxa—e.g. 76 subfamily H EcKLs in *Blattella germanica* (Blattodea: Ectobiidae) and 32 subfamily A EcKLs in *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae)—it seems likely that there is at least some functional diversity within these large expansions in order to increase the detoxification breadth (DB) of these lineages, given that DB and EcKL family size are positively associated (Chapter 3).

On a similar note, there is currently not enough data to confidently claim that certain EcKL subfamilies in insects contain strictly E-, X- or even S-class genes. Although it seems unlikely that many of the subfamilies that are conserved as single-copy orthologs in most genomes (such as subfamilies C, D, G and I; Fig. 3.7) have undergone many—or any—functional transitions between insect lineages, such an interpretation (the 'ortholog conjecture'; Koonin 2005) remains controversial and may not be valid (Stamboulian *et al.* 2020; Studer & Robinson-Rechavi 2009). Ultimately, functional data from multiple insect lineages will be required to make any claims about the functions of entire EcKL subfamilies.

6.5.2. What was the ancestral EcKL function in Tetraconata?

My limited phylogenetic analysis of the EcKL family across the Tetraconata (Hexapoda + Crustacea; Richter 2002) in Chapter 3 (Fig. 3.10) strongly suggested that insect EcKLs are not monophyletic, with insect subfamilies able to be grouped into higher-level clades (A+B+C+D+E+F+G+H+L+M, I and J+K). Given this, and the apparent non-monophyly of other hexapod and crustacean EcKLs, it is possible that there may

have been multiple EcKL genes in the most recent common ancestor of the Tetraconata ($MRCA_{Tet}$), although there was poor signal in many parts of the phylogenetic tree, making the assignment of orthology between insect EcKL subfamilies and EcKLs in Diplura, Malacostraca and Copepoda impossible except in one case, where it appears the malacostracan *Hyallela azteca* has a direct ortholog of the insect I subfamily. However, given that some crustacean lineages were not sampled for this phylogeny—such as Branchiura, Ostracoda and Mystacocarida—I ultimately cannot reject the hypothesis that all EcKLs in the Tetraconata are derived from a single gene present in the $MRCA_{Tet}$.

If $MRCA_{Tet}$ did possess only one EcKL (or all extant Tetraconata EcKLs are derived from a single gene), an interesting question becomes: what was the function of this gene? The presence of ecdysteroid-phosphate conjugates in crustaceans (Subramoniam 2000), as well as ecdysteroid-phosphate phosphatases (Asada *et al*. 2014), suggests that at least some crustacean EcKLs encode ecdysteroid kinases (EcKs), but given the largely unknown relationships between insect and crustacean EcKLs, as well as the unknown genetic identity of crustacean EcKs, there is no evidence yet for or against ecdysteroid phosphorylation being an ancestral function of the EcKL family. Likewise, it is also unclear if arthropods outside of insects are capable of synthesising phosphate metabolites of xenobiotic compounds—to my knowledge, they have yet to be detected; an ancestral detoxification function—while possible—currently appears to be less likely than an ancestral ecdysteroid metabolism function for the EcKLs.

6.5.3. What was the origin of the EcKL family?

If all extant Tetraconata EcKLs are derived from multiple genes, the issue of the ancestral function of EcKLs may shift to the origin of the gene family, which remains unclear. BLAST searches conducted in Chapter 3 failed to find EcKL genes in the genomes or transcriptomes of panarthropods outside of the Tetraconata—Myriapoda, Chelicerata, Onychophora and Tardigrada (Fig. 3.2)—suggesting that either EcKLs have been independently lost in all four of these lineages ('independent loss' hypothesis), or the gene family was derived from a horizontal gene transfer event in an ancestor of the Tetraconata ('HGT' hypothesis).

Evidence for independent loss is the presence of the DUF1679 gene family (PF07914;

(El-Gebali *et al.* 2018) in nematodes (Nematoda: Ecdysozoa), which appears closely related to the EcKLs and was previously considered part of the DUF227 gene family (McElwee *et al.* 2004; Patel *et al.* 2008; Taubert *et al.* 2008; Thomas 2006), the old name for the EcKLs. As mentioned in Chapter 1.4, DUF1679 may have some transcriptional characteristics of detoxification genes (McElwee *et al.* 2004; Taubert *et al.* 2008), despite phosphate conjugates of xenobiotic compounds—to my knowledge—not being detected in nematodes; DUF1679 genes are otherwise very poorly functionally characterised. Evidence (albeit weak evidence) for HGT is the presence of genes annotated as EcKLs in bacteria and fungi (El-Gebali *et al.* 2018), which suggests that either Tetraconata EcKLs were derived from bacteria or fungi, or bacterial or fungal EcKLs were derived from Tetraconata EcKLs. Curiously, a bacterial EcKL of unknown function was likely horizontally transferred from marine bacteria to fish in the Clupeocephala (Teleostei: Actinopterygii) and retained as a single-copy ortholog in at least 19 fish genomes; the gene is present on genomic scaffolds with vertically inherited genes, and it contains an intron that is supported with RNA-seq data, strongly suggesting the presence of the gene is not due to contamination of the assembly with bacterial DNA (Sun *et al.* 2015). It is possible that some or all bacterial and/or fungal EcKLs are misannotated due to an overly wide definition of the EcKL family in Pfam (El-Gebali *et al.* 2018), but it is still likely they are homologous to Tetraconata EcKLs. Ultimately, 3D structural information may be needed to determine the deep phylogenetic relationships between Tetraconata EcKLs, bacterial EcKLs, fungal EcKLs and nematode DUF1679s, due to their low sequence identity (Lundin *et al.* 2012; Rost 1999).

6.6. Structural properties of EcKL proteins

As noted in Chapter 1.4, EcKLs are a fairly uniform gene family with respect to the overall tertiary structure of their encoded polypeptides, both in the Pfam database (El-Gebali *et al.* 2018)—where 87% of proteins contain a single EcKL domain, 8% contain two domains and 5% contain three or more domains—and in the arthropod EcKL annotations I conducted in Chapter 3—where 98.6% of EcKLs possess a single domain and the remainder possess either two EcKL domains or additional N- or C-terminal intrinsically disordered regions (Chapter 3.3.7).

The disordered regions—found in Dip9 and Dip10 (subfamily A) proteins, including the *D. melanogaster* gene *pkm* characterised in Chapter 5—may have protein-protein interaction functions (Dunker *et al.* 2008), suggesting proteins with these features may either bind to other proteins to regulate one or both partner's activity, but they can also respond to redox potential (Jakob *et al.* 2014), meaning the catalytic activity of these EcKLs might be particularly sensitive to their chemical environment.

The quaternary structure of EcKLs is currently unknown. However, there are a number of hints that EcKLs may form homomeric complexes *in vivo*. The subfamily E EcKL Ipi10G08 appears to exist as monomers, dimers and unidentified oligomers (pentamers or higher) on immunoblots of whole body samples of the pine engraver beetle *Ips pini* (Bearfield *et al.* 2008). As mentioned, a number of EcKL genes in subfamilies H and J also possess two EcKL domains (Chapter 3.3.7)—such multidomain structures can be indicative of contiguous multimerisation, as seen in other small-molecule kinases (Compaan & Ellington 2003; Hoffman *et al.* 2008). Hilton (2004) found that ecdysteroid kinase activity in the midgut cytosol of *Spodoptera littoralis*—likely due to one or more EcKL enzymes—was inhibited by the presence of either the reducing agent dithiothreitol (DTT) or the non-ionic detergent Nonidet P40; this is suggestive of the active enzymes (probably EcKLs) being composed of multiple subunits, given similar inhibition is observed with juvenile hormone diol kinase, which is active as a homodimer (Maxwell *et al.* 2002a). In addition, choline kinases—a fellow member of the PKinase clan to which the EcKLs belong—exist as dimers and tetramers *in vivo* (Aoyama *et al.* 2004; Peisach *et al.* 2003).

If EcKLs do form multimers *in vivo*, this may be important for increasing enzyme-substrate encounter rates, increasing enzyme stability, allosteric regulation, and protein activation more generally (see references in Hagner *et al.* 2018). Alternatively, multimerisation could be effectively non-adaptive and may have evolved through neutral processes (Lynch 2012). However, if EcKLs frequently form heteromers, this may result in lower-than-expected genetic redundancy throughout the gene family, due to functional dependency between paralogs (Dandage & Landry 2019). Broad-based determination of EcKL crystal structures and other biochemical analyses will be needed in order to see how many EcKLs form multimers and the nature of these multimers.

6.7. Future research directions

Ways to extend the work in this thesis have previously been discussed in Chapters 3, 4.4.7, 5.4.4 and 5.4.8. But given that so little is known about the EcKLs, there are many other future research directions available for this gene family.

6.7.1. Outstanding questions in *Drosophila melanogaster*

Drosophila melanogaster, whose EcKLs I started to functionally characterise in this study, will likely remain a key focus of future work on EcKL function, due to its powerful genetic toolkit and the collation of evolutionary, genomic and transcriptomic data on its EcKLs in this thesis (see Chapter 2).

With respect to ecdysteroid metabolism, the two to four ecdysteroid-phosphate conjugates known to be present in this species have yet to be linked to ecdysteroid kinases (EcKs). *CG31098* (Dro8-0), *CG5644* (Dro42-0) and *CG14314* (Dro40-0) are good candidates for encoding some of these EcKs (Chapter 6.2.3) and targeted gene disruption techniques such as germline and somatic CRISPR-Cas9 mutagenesis should be used to uncover their possible functions in reproduction and development. Past these three genes, a concerted effort to discover and characterise all ecdysteroid-phosphate species in *D. melanogaster*—ideally over the life cycle of the insect—would provide important data; in particular, it might shed light on whether or not ecdysteroid-phosphates are involved in supplying the active ecdysteroid titre during the pupal-adult transition in this species (Chapter 1.3.5).

With respect to detoxification, there are many intriguing phenotypes and genes that came out of the work in Chapter 2 (Scanlan *et al.* 2020) that deserve to be followed up, including the links between developmental methylmercury tolerance and *CG16898* (Dro26-1) and *CG33301* (Dro26-2), and the EcKLs with detoxification scores of 4, such as *CG31288* (Dro11-0) and *CG6908* (Dro23-0), which have been linked to a curious meconium colour phenotype (Appendix 1) and insecticide tolerance phenotypes (Chapter 2), respectively. Other research should aim to further characterise the possible roles of EcKLs in the detoxification of ecologically appropriate toxins, such as the secondary metabolites produced by *Aspergillus* spp. in their interactions with *D. melanogaster*

larvae and other saprophagous insects (Chapter 4.4.7).

Other EcKLs with high levels of conservation in other species but no clear possible functions should also be a focus of study in *D. melanogaster*, such as *CG2004* (Dro41-0/subfamily B) and the subfamily J EcKLs (*CG9259*, *CG33509*, *CG33510* and *CG33511*).

6.7.2. Functional genetics of EcKLs in other insects

Other insects present many additional opportunities to explore the function of EcKLs, for both E- and X-class functions, particularly given the rapid implementation of CRISPR-Cas9 mutagenesis and gene misexpression techniques in non-model species (Bi *et al.* 2016; Fandino *et al.* 2019; Wang *et al.* 2019; Zeng *et al.* 2016). Obvious examples are known or candidate ecdysteroid kinase genes in Lepidoptera, particularly *BmEc22K*, for which gene disruption phenotypes have yet to be produced; knocking out *BmEc22K* would go a long way to explore the functions of ecdysteroid 22-phosphate conjugates in not just reproduction, but in the rest of *Bombyx mori* development. Likewise, there are many orthologs of *BmEc22K* in *Manduca sexta* that should be characterised for their role in reproduction and development, both to see if they encode the predicted ecdysteroid 2- or 26-kinases involved in embryonic development in this species (Thompson *et al.* 1987b; 1985), but also if they play an important ecdysteroid-recycling function during metamorphosis (Warren & Gilbert 1986). A very exciting research program could be designed around exploring the functions of Lep1 and Lep8 EcKLs in detoxification in generalist moth species, such as *M. sexta*, *Helicoverpa* spp. and *Lymantria dispar*, the latter of which I hypothesise uses subfamily A EcKLs to phosphorylate and detoxify dietary salicinoids and other phenolic glycosides (Boeckler *et al.* 2016; Chapter 3.4.7).

Nasonia vitripennis (Hymenoptera: Pteromalidae) is arguably the most genetically tractable hymenopteran insect (Chaverra Rodriguez *et al.* 2020; Li *et al.* 2017a; Lynch 2015) and presents an opportunity to study EcKLs conserved throughout Holometabola (alongside parallel work in *D. melanogaster*), Hymenoptera-specific EcKLs, as well as EcKLs that might be involved in parasitoidism, given the large expansion of subfamily E and J genes in this and related species (Chapter 3). Locusts and other orthopterans could also be the focus of searches for both the X-class EcKLs responsible for the phosphorylation of terfenadine and midazolam metabolites (Olsen *et al.* 2015; 2014) and

the ecdysteroid kinases known to be important for reproduction and embryonic development in this taxon (Isaac & Rees 1984; 1985; Isaac *et al.* 1982). Functional genetics in non-dipteran insects would also allow for the characterisation of the otherwise highly conserved EcKL subfamily I, which is not present in *D. melanogaster*.

6.7.3. EcKLs in non-insects

As essentially nothing is known about the functions of EcKLs outside of insects, functional genetics approaches, as well as more targeted phylogenomic analyses, could go a long way in determining if this gene family is also involved in development, reproduction and/or detoxification in non-insect hexapods and crustaceans. More broad-sampling phylogenetic analyses, incorporating bacterial and fungal EcKLs, as well as potentially DUF1679 sequences, might also go some way towards resolving the evolutionary relationships between these groups of genes, although as mentioned in Chapter 6.5.3, structural information might be required (Chapter 6.7.4).

Of note, a recent paper by Will *et al.* (2020) identified an EcKL (*Ophcf2|00159*) in the entomopathogenic fungus *Ophiocordyceps camponoti-floridani*, which infects the ant *Camponotus floridanus*. *Ophcf2|00159*'s encoded protein has a signal peptide domain and is up-regulated during host infection compared to fungal culture, suggesting it may be secreted by the fungi into its host. Will *et al.* (2020) speculate that this EcKL may be involved in changing host foraging behaviour via the regulation of the host ecdysteroid titre, in a similar manner to the baculovirus ecdysteroid 22-glycosyltransferase EGT, which has been linked to the behavioural manipulation of caterpillars via a reduction in active ecdysteroid titre (Evans & O'Reilly 1998; Hoover *et al.* 2011; O'Reilly & Miller 1989). However, unlike EGT, which is thought to be derived from a lepidopteran UGT (Hughes 2013), BLAST searches with the Ophcf2|00159 protein sequence show high sequence identity to other fungal EcKLs, and not insect EcKLs, suggesting the hypothetical ecdysteroid kinase activity of this enzyme may be due to convergent evolution rather than functional conservation of a horizontally acquired gene. Regardless of its origin, the function of *Ophcf2|00159* in ant-fungus interactions deserves much greater study.

6.7.4. EcKL biochemistry

As a complement to the evolutionary and functional genetic work that has taken place

through this project, understanding the biochemistry of EcKL proteins—including their 3D structures, catalytic mechanisms and substrate specificities—will be an important part of piecing together the overall biology of the EcKL gene family. Rebecca Gledhill-Smith, another PhD student in the Robin Lab, has been working in this area, focusing on the heterologous expression of EcKL proteins in bacterial and insect cell systems, with the ultimate aims of producing the first EcKL crystal structure and conducting *in vitro* biochemical assays using purified EcKL proteins.

Once these aims are achieved, many experimental avenues will open up, including the ability to conduct *in vitro* screening of EcKL enzymes—either E-, X- or S-class—with potential substrates, as is routinely done for other enzyme families (Snoeck *et al.* 2019; Wang *et al.* 2018), as well as explore the functional consequences of the structural similarities and differences between EcKLs and other related kinases, such as aminoglycoside phosphotransferases and choline/ethanolamine kinases. Substrate screening will be a particularly powerful tool to determine the likely *in vivo* substrates of *D. melanogaster* EcKLs characterised in this thesis, such as *Wall* (Chapter 5) and the Dro5 EcKLs (Chapter 4).

6.7.5. Practical applications of EcKL research

While much of the research around the EcKL gene family is likely to remain in the category of 'basic science', there are a couple of ways understanding the functions and properties of EcKLs could have practical applications. One such application relates to insecticide resistance. Despite the published association between azinphos-methyl and *CHKov1* in *D. melanogaster* (Aminetzach *et al.* 2005) and unvalidated TWAS associations between *CG6908* and malathion and chlorantraniliprole (Green *et al.* 2019; Chapter 2), there has yet to be—to my knowledge—any validated links between insecticide resistance in the field and any EcKL genes. However, given that the EcKL family is very likely a detoxification gene family across insects, it remains a possibility that some future insecticide resistance genotype may involve an EcKL—in such a situation, the present work, as well as any future work, characterising this gene family will suddenly become very useful for understanding the molecular basis of this resistance.

Another application also relates to insecticides, albeit from a different angle. A major

problem with insecticide use in agriculture, as well as in disease-vector control, is adverse effects on non-target organisms, particularly non-target insects such as bees and other pollinators (Arena & Sgolastra 2014; Siviter *et al.* 2018; Tsvetkov *et al.* 2017). As such, the development of highly specific, narrow spectrum insecticides that kill only targeted pest species are a major goal of much work surrounding insecticide biology. EcKL research could be applied to solve this problem by identifying poorly conserved, yet essential (within a particular taxonomic group), EcKLs that could be inhibited by rationally designed small molecules—as kinases, EcKLs may be particularly 'druggable' in a similar manner to protein kinases, which are attractive drug targets in biomedicine (Paul *et al.* 2020; Sioud & Leirdal 2007; Yueh *et al.* 2019). My phylogenomic analyses in Chapter 3 identify a number of EcKL clades that appear highly conserved within either Diptera, Lepidoptera or Hymenoptera but are poorly conserved in other insect taxa—these genes could be candidates for insecticides that only target one of these orders.

6.8. Conclusion

In this thesis, I have presented the first detailed analysis of the poorly characterised ecdysteroid kinase-like (EcKL) gene family across insects. My hypothesis that the EcKLs are a novel detoxification family, responsible for a neglected aspect of insect toxicology—'detoxification-by-phosphorylation'—was well supported by both phylogenomic and functional evidence. I also provided the first detailed functional characterisation of two ecdysteroid kinase candidate genes, *Wallflower* and *Pinkman*, in *Drosophila melanogaster*, which suggested that ecdysteroid metabolism in this species may be much more complex than previously understood. This work provides a solid model for the simultaneous evolutionary and functional characterisation of uncharacterised gene families, particularly those that encode enzymes with unknown substrates, as well as methods for integrating evolutionary, genomic and transcriptomic data to predict the functions of genes in detoxification processes in insects and potentially other taxa.

References

Abbott, W.S., 1925. A Method of Computing the Effectiveness of an Insecticide. *J Econ Entomol* 18, 265–267. doi:10.1093/jee/18.2.265a

Abdou, M.A., He, Q., Wen, D., Zyaan, O., *et al.*, 2011. Drosophila Met and Gce are partially redundant in transducing juvenile hormone action. *Insect biochemistry and molecular Biology* 41, 938–945. doi:10.1016/j.ibmb.2011.09.003

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., *et al.*, 2000. The genome sequence of Drosophila melanogaster. *Science* 287, 2185–2195. doi:10.1126/science.287.5461.2185

Afschar, S., Toivonen, J.M., Hoffmann, J.M., Tain, L.S., *et al.*, 2016. Nuclear hormone receptor DHR96 mediates the resistance to xenobiotics but not the increased lifespan of insulin-mutant Drosophila. *PNAS* 113, 1321–1326. doi:10.1073/pnas.1515137113

Agrawal, A.A., Petschenka, G., Bingham, R.A., Weber, M.G., *et al.*, 2012. Toxic cardenolides: chemical ecology and coevolution of specialized plant-herbivore interactions. *New Phytologist* 194, 28–45. doi:10.1111/j.1469-8137.2011.04049.x

Ahn, S.-J., Badenes Pérez, F.R., Reichelt, M., Svatos, A., *et al.*, 2011. Metabolic detoxification of capsaicin by UDP-glycosyltransferase in three Helicoverpa species. *Archives of Insect Biochemistry and Physiology* 78, 104–118. doi:10.1002/arch.20444

Ahn, S.-J., Vogel, H., Heckel, D.G., 2012. Comparative analysis of the UDP-glycosyltransferase multigene family in insects. *Insect biochemistry and molecular Biology* 42, 133–147. doi:10.1016/j.ibmb.2011.11.006

Akagi, K., Sarhan, M., Sultan, A.-R.S., Nishida, H., *et al.*, 2016. A biological timer in the fat body comprising Blimp-1, βFtz-f1 and Shade regulates pupation timing in Drosophila melanogaster. *Development* 143, 2410–2416. doi:10.1242/dev.133595

Alexander, J.L., Beagan, K., Orr-Weaver, T.L., McVey, M., 2016. Multiple mechanisms contribute to double-strand break repair at rereplication forks in Drosophila follicle cells. *PNAS* 113, 13809–13814. doi:10.1073/pnas.1617110113

Ali, J.G., Agrawal, A.A., 2012. Specialist versus generalist insect herbivores and plant defense. *Trends in Plant Science* 17, 293–302. doi:10.1016/j.tplants.2012.02.006

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., *et al.*, 1990. Basic local alignment search tool. *Journal of molecular biology* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Alyokhin, A., Chen, Y.H., 2017. Adaptation to Toxic Hosts as a Factor in the Evolution of Insecticide Resistance. *Current Opinion in Insect Science*. doi:10.1016/j.cois.2017.04.006

Ambrosino, L., Chiusano, M.L., 2017. Transcriptologs: A Transcriptome-Based Approach to Predict Orthology Relationships:. *Bioinformatics and Biology Insights* 11, 117793221769013. doi:10.1177/1177932217690136

Ameku, T., Yoshinari, Y., Fukuda, R., Niwa, R., 2017. Ovarian ecdysteroid biosynthesis and female germline stem cells. *Fly* 127, 1–9. doi:10.1080/19336934.2017.1291472

Aminetzach, Y.T., Macpherson, J.M., Petrov, D.A., 2005. Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. *Science* 309, 764–767. doi:10.1126/science.1112699

Anaya, A.L., Cruz-Ortega, R., Waller, G.R., 2006. Metabolism and ecology of purine alkaloids. *Frontiers in Bioscience-Landmark* 11, 2354–2370. doi:10.2741/1975

Anderson, C.J., Oakeshott, J.G., Tay, W.T., Gordon, K.H.J., *et al.*, 2018. Hybridization and gene flow in the mega-pest lineage of moth, Helicoverpa. *Proc Natl Acad Sci U S A* 115, 5034–5039. doi:10.1073/pnas.1718831115

Andrew Pitoniak, D.B., 2015. Mechanisms and Functions of Nrf2 Signaling in Drosophila. *Free Radical Biology and Medicine* 88, 302–313. doi:10.1016/j.freeradbiomed.2015.06.020

Aoyama, C., Liao, H., Ishidate, K., 2004. Structure and function of choline kinase isoforms in mammalian cells. *Prog. Lipid Res.* 43, 266–281. doi:10.1016/j.plipres.2003.12.001

Arbuckle, K., Rodríguez de la Vega, R.C., Casewell, N.R., 2017. Coevolution takes the sting out of it: Evolutionary biology and mechanisms of toxin resistance in animals. *Toxicon* 140, 118–131. doi:10.1016/j.toxicon.2017.10.026

Arena, M., Sgolastra, F., 2014. A meta-analysis comparing the sensitivity of bees to pesticides. *Ecotoxicology* 23, 324–334. doi:10.1007/s10646-014-1190-1

Armstrong, R.N., 2002. Glutathione S-transferases: reaction mechanism, structure, and function. *Chem. Res. Toxicol.* 4, 131–140. doi:10.1021/tx00020a001

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., *et al.*, 2012. ExPASy: SIB bioinformatics resource portal. *Nucl Acids Res* 40, W597–W603. doi:10.1093/nar/gks400

Asada, M., Kato, Y., Matsuura, T., Watanabe, H., 2014. Early Embryonic Expression of a Putative Ecdysteroid-Phosphate Phosphatase in the Water Flea, Daphnia magna (Cladocera: Daphniidae). *J Insect Sci* 14, 181–181. doi:10.1093/jisesa/ieu043

Ashburner, M., 1990. Puffs, genes, and hormones revisited. *Cell* 61, 1–3. doi:10.1016/0092-8674(90)90205-S

Astle, J., Kozlova, T., Thummel, C.S., 2003. Essential roles for the Dhr78 orphan nuclear receptor during molting of the Drosophila tracheal system. *Insect biochemistry and molecular Biology* 33, 1201–1209. doi:10.1016/j.ibmb.2003.06.011

Atkins, W.M., 2020. Mechanisms of promiscuity among drug metabolizing enzymes and drug transporters. *FEBS Journal* 287, 1306–1322. doi:10.1111/febs.15116

Awmack, C.S., Leather, S.R., 2002. Host plant quality and fecundity in herbivorous insects. *Annu Rev Entomol* 47, 817–844. doi:10.1146/annurev.ento.47.091201.145300

Baars, A.J., Zijlstra, J.A., Jansen, M., Vogel, E., *et al.*, 1980. Biotransformation and Spectral Interaction of Xenobiotics with Subcellular Fractions from Drosophila Melanogaster, in: *Further Studies in the Assessment of Toxic Actions*, Archives of Toxicology. Springer, Berlin, Heidelberg, Berlin, Heidelberg, pp. 54–58. doi:10.1007/978-3-642-67729-8_11

Bainbridge, S.P., Bownes, M., 1981. Staging the metamorphosis of Drosophila melanogaster. *J Embryol Exp Morphol* 66, 57–80.

Baker, K.D., Shewchuk, L.M., Kozlova, T., Makishima, M., *et al.*, 2003. The Drosophila Orphan Nuclear Receptor DHR38 Mediates an Atypical Ecdysteroid Signaling Pathway. *Cell* 113, 731–742. doi:10.1016/S0092-8674(03)00420-3

Baker, K.D., Warren, J.T., Thummel, C.S., Gilbert, L.I., *et al.*, 2000. Transcriptional activation of the Drosophila ecdysone receptor by insect and plant ecdysteroids. *Insect biochemistry and molecular Biology* 30, 1037–1043. doi:10.1016/S0965-1748(00)00075-8

Balabanidou, V., Grigoraki, L., Vontas, J., 2018. Insect cuticle: a critical determinant of insecticide resistance. *Current Opinion in Insect Science* 27, 68–74. doi:10.1016/j.cois.2018.03.001

Balieira, K.V.B., Mazzo, M., Bizerra, P.F.V., de Jesus Santos Guimarães, A.R., *et al.*, 2018. Imidacloprid-induced oxidative stress in honey bees and the antioxidant action of caffeine. *Apidologie* 49, 562–572. doi:10.1007/s13592-018-0583-1

Barbosa, P., Martinat, P., Waldvogel, M., 1986. Development, fecundity and survival of the herbivore Lymantria dispar and the number of plant species in its diet. *Ecological Entomology* 11, 1–6. doi:10.1111/j.1365-2311.1986.tb00273.x

Baric, R.S., Crosson, S., Damania, B., Miller, S.I., *et al.*, 2016. Next-Generation High-Throughput Functional Annotation of Microbial Genomes. *mBio* 7, aad6253. doi:10.1128/mBio.01245-16

Bassett, M.H., McCarthy, J.L., Waterman, M.R., Sliter, T.J., 1997. Sequence and developmental expression of Cyp18, a member of a new cytochrome P450 family from Drosophila. *Molecular and Cellular Endocrinology* 131, 39–49. doi:10.1016/S0303-7207(97)00093-2

Bateman, A., Coggill, P., Finn, R.D., 2010. DUFs: families in search of function. *Acta Crystallogr F Struct Biol Cryst Commun* 66, 1148–1152. doi:10.1107/S1744309110001685

Battlay, P., Leblanc, P.B., Green, L., Garud, N.R., *et al.*, 2018. Structural Variants and Selective Sweep Foci Contribute to Insecticide Resistance in the Drosophila Genetic Reference Panel. *G3 (Bethesda)* 8, 3489–3497. doi:10.1534/g3.118.200619

Battlay, P., Schmidt, J.M., Fournier-Level, A., Robin, C., 2016. Genomic and Transcriptomic Associations Identify a New Insecticide Resistance Phenotype for the Selective Sweep at the Cyp6g1 Locus of Drosophila melanogaster. *G3 (Bethesda)* g3.116.031054. doi:10.1534/g3.116.031054

Beard, R.L., Walton, G.S., 1969. Kojic acid as an insecticidal mycotoxin. *J Invertebr Pathol* 14, 53–59. doi:10.1016/0022-2011(69)90010-X

Bearfield, J.C., Box, C.D., Keeling, C.I., Young, S., *et al.*, 2008. Isolation, endocrine regulation and transcript distribution of a putative primary JH-responsive gene from the pine engraver, Ips pini (Coleoptera: Scolytidae). *Insect biochemistry and molecular Biology* 38, 256–267. doi:10.1016/j.ibmb.2007.11.009

Beckstead, R.B., Lam, G., Thummel, C.S., 2007. Specific transcriptional responses to juvenile hormone and ecdysone in Drosophila. *Insect biochemistry and molecular Biology* 37, 570–578. doi:10.1016/j.ibmb.2007.03.001

Beckstead, R.B., Lam, G., Thummel, C.S., 2005. The genomic response to 20-hydroxyecdysone at the onset of Drosophila metamorphosis. *Genome Biol* 6, R99. doi:10.1186/gb-2005-6-12-r99

Bedell, V.M., Person, A.D., Larson, J.D., McLoon, A., *et al.*, 2012. The lineage-specific gene ponzr1 is essential for zebrafish pronephric and pharyngeal arch development. *Development* 139, 793–804. doi:10.1242/dev.071720

Beedessee, G., Hisata, K., Roy, M.C., Van Dolah, F.M., *et al.*, 2019. Diversified secondary metabolite biosynthesis gene repertoire revealed in symbiotic dinoflagellates. *Sci. Rep.* 9, 1–12. doi:10.1038/s41598-018-37792-0

Bellen, H.J., Levis, R.W., He, Y., Carlson, J.W., *et al.*, 2011. The Drosophila gene disruption project: progress using transposons with distinctive site specificities. *Genetics* 188, 731–743. doi:10.1534/genetics.111.126995

Bellen, H.J., Levis, R.W., Liao, G., He, Y., *et al.*, 2004. The BDGP Gene Disruption Project. *Genetics* 167, 761–781. doi:10.1534/genetics.104.026427

Bender, M., Imam, F.B., Talbot, W.S., Ganetzky, B., *et al.*, 1997. Drosophila Ecdysone Receptor Mutations Reveal Functional Differences among Receptor Isoforms. *Cell* 91, 777–788. doi:10.1016/S0092-8674(00)80466-3

Benowitz, N.L., Jacob, P., Mayan, H., Denaro, C., 1995. Sympathomimetic effects of paraxanthine and caffeine in humans. *Clin Pharmacol Ther* 58, 684–691. doi:10.1016/0009-9236(95)90025-X

Beran, F., Köllner, T.G., Gershenzon, J., Tholl, D., 2019. Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. *New Phytologist* 223, 52–67. doi:10.1111/nph.15718

Berenbaum, M.R., Johnson, R.M., 2015. Xenobiotic detoxification pathways in honey bees. *Current Opinion in Insect Science* 10, 51–58. doi:10.1016/j.cois.2015.03.005

Berger, C., Renner, S., Lüer, K., Technau, G.M., 2007. The commonly used marker ELAV is transiently expressed in neuroblasts and glial cells in the Drosophila embryonic CNS. *Developmental Dynamics* 236, 3562–3568. doi:10.1002/dvdy.21372

Bergsten, J., 2005. A review of long-branch attraction. *Cladistics* 21, 163–193. doi:10.1111/j.1096-0031.2005.00059.x

Bergwitz, C., Wee, M.J., Sinha, S., Huang, J., *et al.*, 2013. Genetic Determinants of Phosphate Response in Drosophila. *PLoS ONE* 8, e56753. doi:10.1371/journal.pone.0056753

Bernard, G., Pathmanathan, J.S., Lannes, R., Lopez, P., *et al.*, 2018. Microbial Dark Matter Investigations: How Microbial Studies Transform Biological Knowledge and Empirically Sketch a Logic of Scientific Discovery. *Genome Biology and Evolution* 10, 707–715. doi:10.1093/gbe/evy031

Bernhardt, R., 2006. Cytochromes P450 as versatile biocatalysts. *Journal of biotechnology* 124, 128–145. doi:10.1016/j.jbiotec.2006.01.026

Beydon, P., Claret, J., Porcheron, P., Lafont, R., 1981. Biosynthesis and inactivation of ecdysone during the pupal-adult development of the cabbage butterfly, Pieris brassicae L. *Steroids* 38, 633–650. doi:10.1016/0039-128X(81)90083-0

Beydon, P., Girault, J.-P., Lafont, R., 1987. Ecdysone metabolism in Pieris brassicae during the feeding last larval instar. *Archives of Insect Biochemistry and Physiology* 4, 139–149. doi:10.1002/arch.940040207

Beyenbach, K.W., Skaer, H., Dow, J.A.T., 2010. The developmental, molecular, and transport biology of Malpighian tubules. *Annual review of entomology* 55, 351–374. doi:10.1146/annurev-ento-112408-085512

Bi, H.L., Xu, J., Tan, A.J., Huang, Y.P., 2016. CRISPR/Cas9-mediated targeted gene mutagenesis in Spodoptera litura. *Insect Sci.* 23, 469–477. doi:10.1111/1744-7917.12341

Bialecki, M., Shilton, A., Fichtenberg, C., Segraves, W.A., *et al.*, 2002. Loss of the Ecdysteroid-Inducible E75A Orphan Nuclear Receptor Uncouples Molting from Metamorphosis in Drosophila. *Developmental Cell* 3, 209–220. doi:10.1016/S1534-5807(02)00204-6

Binning, A., Darby, F.J., Heenan, M.P., Smith, J.N., 1967. The conjugation of phenols with phosphate in grass grubs and flies. *Biochemical Journal* 103, 42–48. doi:10.1042/bj1030042

Birnbaum, S.S.L., Abbot, P., 2020. Gene Expression and Diet Breadth in Plant-Feeding Insects: Summarizing Trends. *Trends in Ecology & Evolution* 35, 259–277. doi:10.1016/j.tree.2019.10.014

Birnbaum, S.S.L., Rinker, D.C., Gerardo, N.M., Abbot, P., 2017. Transcriptional profile and differential fitness in a specialist milkweed insect across host plants varying in toxicity. *Mol Ecol*. doi:10.1111/mec.14401

Bischof, J., Bjorklund, M., Furger, E., Schertel, C., *et al.*, 2013. A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. *Development* 140, 2434–2442. doi:10.1242/dev.088757

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., *et al.*, 2007. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* 104, 3312–3317. doi:10.1073/pnas.0611511104

Blasina, A., Price, B.D., Turenne, G.A., McGowan, C.H., 1999. Caffeine inhibits the checkpoint kinase ATM. *Current Biology* 9, 1135–1138. doi:10.1016/S0960-9822(99)80486-2

Blumenthal, T., 1998. Gene clusters and polycistronic transcription in eukaryotes. *BioEssays* 20, 480–487. doi:10.1002/(SICI)1521-1878(199806)20:6<480::AID-BIES6>3.0.CO;2-Q

Boeckler, G.A., Gershenzon, J., Unsicker, S.B., 2011. Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry* 72, 1497–1509. doi:10.1016/j.phytochem.2011.01.038

Boeckler, G.A., Paetz, C., Feibicke, P., Gershenzon, J., *et al.*, 2016. Metabolism of poplar salicinoids by the generalist herbivore Lymantria dispar (Lepidoptera). *Insect biochemistry and molecular Biology* 78, 39–49. doi:10.1016/j.ibmb.2016.08.001

Bogwitz, M.R., Chung, H., Magoc, L., Rigby, S., *et al.*, 2005. Cyp12a4 confers lufenuron resistance in a natural population of Drosophila melanogaster. *Proc Natl Acad Sci U S A* 102, 12807–12812. doi:10.1073/pnas.0503709102

Bond, N.D., Hoshizaki, D.K., Gibbs, A.G., 2010. The role of 20-hydroxyecdysone signaling in Drosophila pupal metabolism. *Comp Biochem Physiol A Mol Integr Physiol* 157, 398–404. doi:10.1016/j.cbpa.2010.08.025

Borges, R., Khan, I., Johnson, W.E., Gilbert, M.T.P., *et al.*, 2015. Gene loss, adaptive evolution and the co-evolution of plumage coloration genes with opsins in birds. *BMC Genomics* 16, 675. doi:10.1186/s12864-015-1924-3

Bownes, M., 1992. Why is there sequence similarity between insect yolk proteins and vertebrate lipases? *J Lipid Res* 33, 777–790.

Bownes, M., Shirras, A., Blair, M., Collins, J., *et al.*, 1988. Evidence that insect embryogenesis is regulated by ecdysteroids released from yolk proteins. *Proc Natl Acad Sci U S A* 85, 1554–1557. doi:10.2307/31268

Brady, J.P., Richmond, R.C., 1992. An evolutionary model for the duplication and divergence of esterase genes in Drosophila. *J Mol Evol* 34, 506–521. doi:10.1007/BF00160464

Brand, C.L., Larracuente, A.M., Presgraves, D.C., 2015. Origin, evolution, and population genetics of the selfish Segregation Distorter gene duplication in European and African populations of Drosophila melanogaster. *Evolution* 69, 1271–1283. doi:10.1111/evo.12658

Brattsten, L.B., 1988. Enzymic adaptations in leaf-feeding insects to host-plant allelochemicals. *Journal of chemical ecology* 14, 1919–1939. doi:10.1007/BF01013486

Breitling, R., Laubner, D., Adamski, J., 2001. Structure-based Phylogenetic Analysis of Short-chain Alcohol Dehydrogenases and Reclassification of the 17beta-Hydroxysteroid Dehydrogenase Family. *Mol. Biol. Evol.* 18, 2154–2161. doi:10.1093/oxfordjournals.molbev.a003761

Brenner, S., 1987. Phosphotransferase sequence homology. *Nature* 329, 21. doi:10.1038/329021a0

Bretschneider, A., Heckel, D.G., Vogel, H., 2016. Know your ABCs: Characterization and gene expression dynamics of ABC transporters in the polyphagous herbivore Helicoverpa armigera. *Insect biochemistry and molecular Biology*. doi:10.1016/j.ibmb.2016.03.001

Brisson, J.A., 2010. Aphid wing dimorphisms: linking environmental and genetic control of trait variation. *Phil. Trans. R. Soc. B* 365, 605–616. doi:10.1098/rstb.2009.0255

Broadus, J., McCabe, J.R., Endrizzi, B., Thummel, C.S., *et al.*, 1999. The Drosophila βFTZ-F1 Orphan Nuclear Receptor Provides Competence for Stage-Specific Responses to the Steroid Hormone Ecdysone. *Molecular cell* 3, 143–149. doi:10.1016/S1097-2765(00)80305-6

Brock, A.R., Seto, M., Smith-Bolton, R.K., 2017. Cap-n-Collar Promotes Tissue Regeneration by Regulating ROS and JNK Signaling in the Drosophila melanogaster Wing Imaginal Disc. *Genetics* 206, 1505–1520. doi:10.1534/genetics.116.196832

Brown, J.J., Reed-Larsen, D., 1991. Ecdysteroids and insect host/parasitoid interactions. *Biological control* 1, 136–143. doi:10.1016/1049-9644(91)90112-D

Buchon, N., Broderick, N.A., Poidevin, M., Pradervand, S., *et al.*, 2009. Drosophila Intestinal Response to Bacterial Infection: Activation of Host Defense and Stem Cell Proliferation. *Cell Host and Microbe* 5, 200–211. doi:10.1016/j.chom.2009.01.003

Buchon, N., Osman, D., David, F.P.A., Yu Fang, H., *et al.*, 2013. Morphological and Molecular Characterization of Adult Midgut Compartmentalization in Drosophila. *Cell Reports* 3, 1725–1738. doi:10.1016/j.celrep.2013.04.001

Buergin, C.G., Connat, J.L., 1989. Ecdysone metabolism in the host-parasitoid-system Trichoplusia ni/Chelonus sp. *Archives of Insect Biochemistry and Physiology* 11, 79–92. doi:10.1002/arch.940110203

Buhler, K., Clements, J., Winant, M., Bolckmans, L., *et al.*, 2018. Growth control through regulation of insulin-signaling by nutrition-activated steroid hormone in Drosophila. *Development* dev.165654. doi:10.1242/dev.165654

Bujnicki, J.M., 2000. Phylogeny of the Restriction Endonuclease-Like Superfamily Inferred from Comparison of Protein Structures. *J Mol Evol* 50, 39–44. doi:10.1007/s002399910005

Bujold, M., Gopalakrishnan, A., Nally, E., King-Jones, K., 2010. Nuclear Receptor DHR96 Acts as a Sentinel for Low Cholesterol Concentrations in Drosophila melanogaster. *Molecular and cellular biology* 30, 793–805. doi:10.1128/MCB.01327-09

Burnett, C.L., Bergfeld, W.F., Belsito, D.V., Hill, R.A., *et al.*, 2010. Final Report of the Safety Assessment of Kojic Acid as Used in Cosmetics:. *Int J Toxicol* 29, 244S–273S. doi:10.1177/1091581810385956

Burnham, K.P., Anderson, D.R. (Eds.), 2004. Model Selection and Multimodel Inference. Springer New York, New York, NY. doi:10.1007/b97636

Buszczak, M., Freeman, M.R., Carlson, J.R., Bender, M., *et al.*, 1999. Ecdysone response genes govern egg chamber development during mid-oogenesis in Drosophila. *Development* 126, 4581–4589. doi:10.1016/S0012-1606(74)80016-3

Calabrese, E.J., Mattson, M.P., 2017. How does hormesis impact biology, toxicology, and medicine? *Aging Mech Dis* 3, 1–8. doi:10.1038/s41514-017-0013-z

Calla, B., Noble, K., Johnson, R.M., Walden, K.K.O., *et al.*, 2017. Cytochrome P450 diversification and hostplant utilization patterns in specialist and generalist moths: Birth, death and adaptation. *Mol Ecol* 26, 6021–6035. doi:10.1111/mec.14348

Carlisle, D.B., Jenkin, P.M., 1959. Terminology of Hormones. *Nature* 183, 336–337. doi:10.1038/183336a0

Carney, G.E., Wade, A.A., Sapra, R., Goldstein, E.S., *et al.*, 1997. DHR3, an ecdysone-inducible early-late gene encoding a Drosophila nuclear receptor, is required for embryogenesis. *Proc Natl Acad Sci U S A* 94, 12024–12029. doi:10.1073/pnas.94.22.12024

Carvalho, M., Schwudke, D., Sampaio, J.L., Palm, W., *et al.*, 2010. Survival strategies of a sterol auxotroph. *Development* 137, 3675–3685. doi:10.1242/dev.044560

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., *et al.*, 2012. Artemis - an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469. doi:10.1093/bioinformatics/btr703

Casas-Tintó, S., Arnés, M., Ferrús, A., 2017. Drosophila enhancer-Gal4 lines show ectopic expression during development. *Open Science* 4, 170039. doi:10.1098/rsos.170039

Casida, J.E., 1970. Mixed-function oxidase involvement in the biochemistry of insecticide synergists. *J. Agric. Food Chem.* 18, 753–772. doi:10.1021/jf60171a013

Cauchi, R.J., Davies, K.E., Liu, J.-L., 2008. A Motor Function for the DEAD-Box RNA Helicase, Gemin3, in Drosophila. *PLOS Genetics* 4, e1000265. doi:10.1371/journal.pgen.1000265

Cease, A.J., Fay, M., Elser, J.J., Harrison, J.F., 2016. Dietary phosphate affects food selection, post-ingestive phosphorus fate, and performance of a polyphagous herbivore. *Journal of Experimental Biology* 219, 64–72. doi:10.1242/jeb.126847

Ceja-Navarro, J.A., Vega, F.E., Karaoz, U., Hao, Z., *et al.*, 2015. Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nature Communications* 6, 1–9. doi:10.1038/ncomms8618

Chahine, S., O'Donnell, M.J., 2011. Interactions between detoxification mechanisms and excretion in Malpighian tubules of Drosophila melanogaster. *Journal of Experimental Biology* 214, 462–468. doi:10.1242/jeb.048884

Chahine, S., O'Donnell, M.J., 2009. Physiological and molecular characterization of methotrexate transport by Malpighian tubules of adult Drosophila melanogaster. *Journal of Insect Physiology* 55, 927–935. doi:10.1016/j.jinsphys.2009.06.005

Chakraborty, M., Emerson, J.J., Macdonald, S.J., Long, A.D., 2019. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nature Communications* 10, 1–11. doi:10.1038/s41467-019-12884-1

Chakraborty, M., VanKuren, N.W., Zhao, R., Zhang, X., *et al.*, 2018. Hidden genetic variation shapes the structure of functional elements in Drosophila. *Nat Genet* 50, 20–25. doi:10.1038/s41588-017-0010-y

Challi, R.J., Kumar, S., Dasmahapatra, K.K., Jiggins, C.D., *et al.*, 2016. Lepbase: the Lepidopteran genome database *bioRxiv* 5, 4737. doi:10.1101/056994

Chan, Y.-B., Robin, C., 2019. Reconciliation of a gene network and species tree. *Journal of Theoretical biology* 472, 54–66. doi:10.1016/j.jtbi.2019.04.001

Chang, H.C.-H., Dimlich, D.N., Yokokura, T., Mukherjee, A., *et al.*, 2008. Modeling Spinal Muscular Atrophy in Drosophila. *PLoS ONE* 3, e3209. doi:10.1371/journal.pone.0003209

Chang, L., Kreko, T., Davison, H., Cusmano, T., *et al.*, 2013. Normal dynactin complex function during synapse growth in Drosophila requires membrane binding by Arfaptin. *Mol Biol Cell* 24, 1749–1764. doi:10.1091/mbc.E12-09-0697

Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F., *et al.*, 2006. Shavenbaby Couples Patterning to Epidermal Cell Shape Control. *PLoS Biol* 4, e290. doi:10.1371/journal.pbio.0040290

Chanut-Delalande, H., Hashimoto, Y., Pelissier-Monier, A., Spokony, R., *et al.*, 2014. Pri peptides are mediators of ecdysone for the temporal control of development. *Nat Cell Biol* 127, 4115. doi:10.1038/ncb3052

Chatterjee, D., Katewa, S.D., Qi, Y., Jackson, S.A., *et al.*, 2014. Control of metabolic adaptation to fasting by dILP6-induced insulin signaling in Drosophila oenocytes. *PNAS* 111, 17959–17964. doi:10.1073/pnas.1409241111

Chaverra Rodriguez, D., Dalla Benetta, E., Heu, C.C., Rasgon, J.L., *et al.*, 2020. Germline mutagenesis of Nasonia vitripennisthrough ovarian delivery of CRISPR-Cas9 ribonucleoprotein. *Insect Mol Biol* imb.12663. doi:10.1111/imb.12663

Chavez, V.M., Marques, G., Delbecque, J.P., Kobayashi, K., *et al.*, 2000. The Drosophila disembodied gene controls late embryonic morphogenesis and codes for a cytochrome P450 enzyme that regulates embryonic ecdysone levels. *Development* 127, 4115–4126.

Chen, J.-H., Webb, T.J., Powls, R., Rees, H.H., 1996. Purification and Characterisation of Haemolymph 3-Dehydroecdysone 3β-Reductase in Relation to Ecdysteroid Biosynthesis in the Cotton Leafworm Spodoptera littoralis. *European Journal of Biochemistry* 242, 394–401. doi:10.1111/j.1432-1033.1996.0394r.x

Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C., *et al.*, 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet* 8, 762–775. doi:10.1038/nrg2193

Chen, J.H., Kabbouh, M., Fisher, M.J., Rees, H.H., 1994. Induction of an inactivation pathway for ecdysteroids in larvae of the cotton leafworm, Spodoptera littoralis. *Biochemical Journal* 301, 89–95. doi:10.1042/bj3010089

Chen, Y., Jakoncic, J., Wang, J., Zheng, X., *et al.*, 2008. Structural and Functional Characterization of the C-Terminal Domain of the Ecdysteroid Phosphate Phosphatase from Bombyx mori Reveals a New Enzymatic Activity†. *Biochemistry* 47, 12135–12145. doi:10.1021/bi801318w

Cheng, C., Ko, A., Chaieb, L., Koyama, T., *et al.*, 2014. The POU Factor Ventral Veins Lacking/Drifter Directs the Timing of Metamorphosis through Ecdysteroid and Juvenile Hormone Signaling. *PLOS Genetics* 10, e1004425. doi:10.1371/journal.pgen.1004425

Cherbas, L., Hu, X., Zhimulev, I., Belyaeva, E., *et al.*, 2003. EcR isoforms in Drosophila: testing tissue-specific requirements by targeted blockade and rescue. *Development* 130, 271–284. doi:10.1242/dev.00205

Cherbas, L., Willingham, A., Zhang, D., Yang, L., *et al.*, 2011. The transcriptional diversity of 25 Drosophila cell lines. *Genome Res.* 21, 301–314. doi:10.1101/gr.112961.110

Chiang, Y.N., Tan, K.J., Chung, H., Lavrynenko, O., *et al.*, 2016. Steroid Hormone Signaling Is Essential for Pheromone Production and Oenocyte Survival. *PLOS Genetics* 12, e1006126. doi:10.1371/journal.pgen.1006126

Chow, C.Y., Wolfner, M.F., Clark, A.G., 2013. Using natural variation in Drosophila to discover previously unknown endoplasmic reticulum stress genes. *PNAS* 110, 9013–9018. doi:10.1073/pnas.1307125110

Christesen, D., Yang, Y.T., Somers, J., Robin, C., *et al.*, 2017. Transcriptome Analysis of Drosophila melanogaster Third Instar Larval Ring Glands Points to Novel Functions and Uncovers a Cytochrome p450 Required for Development. *G3 (Bethesda)* 7, 467–479. doi:10.1534/g3.116.037333

Chung, H., Bogwitz, M.R., McCart, C., Andrianopoulos, A., *et al.*, 2007. Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the Drosophila melanogaster insecticide resistance gene Cyp6g1. *Genetics* 175, 1071–1077. doi:10.1534/genetics.106.066597

Chung, H., Sztal, T., Pasricha, S., Sridhar, M., *et al.*, 2009. Characterization of Drosophila melanogaster cytochrome P450 genes. *Proc Natl Acad Sci U S A* 106, 5731–5736. doi:10.1073/pnas.0812141106

Cinnamon, E., Makki, R., Sawala, A., Wickenberg, L.P., *et al.*, 2016. Drosophila Spidey/Kar Regulates Oenocyte Growth via PI3-Kinase Signaling. *PLOS Genetics* 12, e1006154. doi:10.1371/journal.pgen.1006154

Claudianos, C.C., Ranson, H.H., Johnson, R.M.R., Biswas, S.S., *et al.*, 2006. A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee. *Insect Mol Biol* 15, 615–636. doi:10.1111/j.1365-2583.2006.00672.x

Clay, N.A., Lehrter, R.J., Kaspari, M., 2017. Towards a geography of omnivory: Omnivores increase carnivory when sodium is limiting. *J Anim Ecol* 86, 1523–1531. doi:10.1111/1365-2656.12754

Clifton, B.D., Librado, P., Yeh, S.-D., Solares, E., *et al.*, 2016. Rapid Functional And Sequence Differentiation of a Tandemly-Repeated Species-Specific Multigene Family in Drosophila. *Mol. Biol. Evol.* 34, 51–65. doi:10.1093/molbev/msw212

Coelho, A., Fraichard, S., Le Goff, G., Faure, P., *et al.*, 2015. Cytochrome P450-Dependent Metabolism of Caffeine in Drosophila melanogaster. *PLoS ONE* 10, e0117328. doi:10.1371/journal.pone.0117328

Coley, P.D., Bateman, M.L., Kursar, T.A., 2006. The effects of plant quality on caterpillar growth and defense against natural enemies. *Oikos* 115, 219–228. doi:10.1111/j.2006.0030-1299.14928.x

Compaan, D.M., Ellington, W.R., 2003. Functional consequences of a gene duplication and fusion event in an arginine kinase. *Journal of Experimental Biology* 206, 1545–1556. doi:10.1242/jeb.00299

Cong, Q., Shen, J., Borek, D., Robbins, R.K., *et al.*, 2016. Complete genomes of Hairstreak butterflies, their speciation and nucleo-mitochondrial incongruence. *Sci. Rep.* 6, 1–15. doi:10.1038/srep24863

Cooper, G.M., Brown, C.D., 2008. Qualifying the relationship between sequence conservation and molecular function. *Genome Res.* 18, 201–205. doi:10.1101/gr.7205808

Coordinators, N.R., 2016. Database resources of the National Center for Biotechnology Information. *Nucl Acids Res* 44, D7–D19. doi:10.1093/nar/gkv1290

Cornell, H.V., Hawkins, B.A., 2003. Herbivore responses to plant secondary compounds: a test of phytochemical coevolution theory. *American Naturalist* 161, 507–522. doi:10.1086/368346

Cornwell, W., Nakagawa, S., 2017. Phylogenetic comparative methods. *Current Biology* 27, R333–R336. doi:10.1016/j.cub.2017.03.049

Couvillon, M.J., Toufailia, Al, H., Butterfield, T.M., Schrell, F., *et al.*, 2015. Caffeinated Forage Tricks Honeybees into Increasing Foraging and Recruitment Behaviors. *Current Biology* 25, 2815–2818. doi:10.1016/j.cub.2015.08.052

Cranston, P.S., Hardy, N.B., Morse, G.E., 2011. A dated molecular phylogeny for the Chironomidae (Diptera). *Systematic Entomology* 37, 172–188. doi:10.1111/j.1365-3113.2011.00603.x

Cruz, J., Martín, D., Franch-Marro, X., 2020. Egfr Signaling Is a Major Regulator of Ecdysone Biosynthesis in the Drosophila Prothoracic Gland. *Current Biology* 30, 1547–1554.e4. doi:10.1016/j.cub.2020.01.092

Cvicek, V., Goddard, W.A., Abrol, R., 2016. Structure-Based Sequence Alignment of the Transmembrane Domains of All Human GPCRs: Phylogenetic, Structural and Functional Implications. *PLoS Comp Biol* 12, e1004805. doi:10.1371/journal.pcbi.1004805

Daborn, P.J., Yen, J.L., Bogwitz, M.R., Le Goff, G., *et al.*, 2002. A Single P450 Allele Associated with Insecticide Resistance in Drosophila. *Science* 297, 2253–2256. doi:10.1126/science.1074170

Daborn, P.J.P., Lumb, C.C., Boey, A.A., Wong, W.W., *et al.*, 2007. Evaluating the insecticide resistance potential of eight Drosophila melanogaster cytochrome P450 genes by transgenic over-expression. *Insect biochemistry and molecular Biology* 37, 8–8. doi:10.1016/j.ibmb.2007.02.008

Dai, J.-D., Gilbert, L.I., 1991. Metamorphosis of the corpus allatum and degeneration of the prothoracic glands during the larval-pupal-adult transformation of Drosophila melanogaster: A cytophysiological analysis of the ring gland. *Developmental biology* 144, 309–326. doi:10.1016/0012-1606(91)90424-2

Dalli, J., Chiang, N., Serhan, C.N., 2014. Identification of 14-series sulfido-conjugated mediators that promote resolution of infection and organ protection. *Proc Natl Acad Sci U S A* 111, E4753–E4761. doi:10.1073/pnas.1415006111

Danchin, A., Fang, G., 2016. Unknown unknowns: essential genes in quest for function. *Microbial Biotechnology* 9, 530–540. doi:10.1111/1751-7915.12384

Danchin, E.G.J., Gouret, P., Pontarotti, P., 2006. Eleven ancestral gene families lost in mammals and vertebrates while otherwise universally conserved in animals. *BMC Evolutionary Biology* 6, 5. doi:10.1186/1471-2148-6-5

Dandage, R., Landry, C.R., 2019. Paralog dependency indirectly affects the robustness of human cells. *Mol Syst Biol* 15, 1531. doi:10.15252/msb.20198871

Danielsen, E.T., Moeller, M.E., Yamanaka, N., Ou, Q., *et al.*, 2016. A Drosophila Genome-Wide Screen Identifies Regulators of Steroid Hormone Production and Developmental Timing. *Developmental Cell* 37, 558–570. doi:10.1016/j.devcel.2016.05.015

Danneels, E.L., Rivers, D.B., de Graaf, D.C., 2010. Venom proteins of the parasitoid wasp Nasonia vitripennis: recent discovery of an untapped pharmacopee. *Toxins (Basel)* 2, 494–516. doi:10.3390/toxins2040494

Darby, F.J., Heenan, M.P., Smith, J.N., 1966. The absence of glucuronide conjugates from 1-naphthol dosed flies and grass grubs; Detection of 1-naphthylphosphate. *Life Sciences* 5, 1499–1502. doi:10.1016/0024-3205(66)90225-6

Date, P., Dweck, H.K.M., Stensmyr, M.C., Shann, J., *et al.*, 2013. Divergence in Olfactory Host Plant Preference in D. mojavensis in Response to Cactus Host Use. *PLoS ONE* 8, e70027. doi:10.1371/journal.pone.0070027

Davies, L., Anderson, I.P., Turner, P.C., Shirras, A.D., *et al.*, 2007. An unsuspected ecdysteroid/steroid phosphatase activity in the key T-cell regulator, Sts-1: Surprising relationship to insect ecdysteroid phosphate phosphatase. *Proteins* 67, 720–731. doi:10.1002/prot.21357

de Castro, É.C.P., Zagrobelny, M., Cardoso, M.Z., Bak, S., 2017. The arms race between heliconiine butterflies and Passiflora plants – new insights on an ancient subject. *Biological Reviews* 93, 555–573. doi:10.1111/brv.12357

De Fine Licht, H.H., Boomsma, J.J., 2010. Forage collection, substrate preparation, and diet composition in fungus-growing ants. *Ecological Entomology* 35, 259–269. doi:10.1111/j.1365-2311.2010.01193.x

De Fine Licht, H.H., Schiøtt, M., Rogowska-Wrzesinska, A., Nygaard, S., *et al.*, 2013. Laccase detoxification mediates the nutritional alliance between leaf-cutting ants and fungus-garden symbionts. *PNAS* 110, 583–587. doi:10.1073/pnas.1212709110

De Geyter, E., Swevers, L., Soin, T., Geelen, D., *et al.*, 2012. Saponins do not affect the ecdysteroid receptor complex but cause membrane permeation in insect culture cell lines. *Journal of Insect Physiology* 58, 18–23. doi:10.1016/j.jinsphys.2011.09.005

Delaux, P.-M., Varala, K., Edger, P.P., Coruzzi, G.M., *et al.*, 2014. Comparative Phylogenomics Uncovers the Impact of Symbiotic Associations on Host Genome Evolution. *PLOS Genetics* 10, e1004487. doi:10.1371/journal.pgen.1004487

Delbecque, J.-P., Weidner, K., Hoffmann, K.H., 1990. Alternative sites for ecdysteroid production in insects. *Invertebrate Reproduction & Development* 18, 29–42. doi:10.1080/07924259.1990.9672126

Dembeck, L.M., Huang, W., Magwire, M.M., Lawrence, F., *et al.*, 2015. Genetic Architecture of Abdominal Pigmentation in Drosophila melanogaster. *PLOS Genetics* 11, e1005163. doi:10.1371/journal.pgen.1005163

Demuth, J.P., Bie, T.D., Stajich, J.E., Cristianini, N., *et al.*, 2006. The Evolution of Mammalian Gene Families. *PLoS ONE* 1, e85. doi:10.1371/journal.pone.0000085

Demuth, J.P., Hahn, M.W., 2009. The life and death of gene families. *BioEssays* 31, 29–39. doi:10.1002/bies.080085

Denecke, S., Fusetto, R., Batterham, P., 2017a. Describing the role of Drosophila melanogaster ABC transporters in insecticide biology using CRISPR-Cas9 knockouts. *Insect biochemistry and molecular Biology*. doi:10.1016/j.ibmb.2017.09.017

Denecke, S., Fusetto, R., Martelli, F., Giang, A., *et al.*, 2017b. Multiple P450s and Variation in Neuronal Genes Underpins the Response to the Insecticide Imidacloprid in a Population of Drosophila melanogaster. *Sci. Rep.* 7, 1–11. doi:10.1038/s41598-017-11092-5

Deng, H., 2014. Multiple roles of Nrf2-Keap1 signaling: Regulation of development and xenobiotic response using distinct mechanisms. *Fly* 8, 0–7. doi:10.4161/fly.27007

Deng, H., Kerppola, T.K., 2014. Visualization of the Drosophila dKeap1-CncC interaction on chromatin illumines cooperative, xenobiotic-specific gene activation. *Development* 141, 3277–3288. doi:10.1242/dev.110528

Deng, H., Kerppola, T.K., 2013. Regulation of Drosophila metamorphosis by xenobiotic response regulators. *PLOS Genetics* 9, e1003263–e1003263. doi:10.1371/journal.pgen.1003263

Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., *et al.*, 2010. PheWAS - demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26, 1205–1210. doi:10.1093/bioinformatics/btq126

Denton, J.F., Lugo-Martinez, J., Tucker, A.E., Schrider, D.R., *et al.*, 2014. Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Comp Biol* 10, e1003998. doi:10.1371/journal.pcbi.1003998

Dermauw, W., Van Leeuwen, T., 2014. The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. *Insect biochemistry and molecular Biology* 45, 89–110. doi:10.1016/j.ibmb.2013.11.001

DeSalvo, M.K., Hindle, S.J., Rusan, Z.M., Orng, S., *et al.*, 2014. The Drosophila surface glia transcriptome: evolutionary conserved blood-brain barrier processes. *Front. Neurosci.* 8, 1727. doi:10.3389/fnins.2014.00346

Dessimoz, C., Gil, M., 2010. Phylogenetic assessment of alignments reveals neglected tree signal in gaps. *Genome Biology* 11, R37. doi:10.1186/gb-2010-11-4-r37

Diao, F., Mena, W., Shi, J., Park, D., *et al.*, 2015. The Splice Isoforms of the Drosophila Ecdysis Triggering Hormone Receptor Have Developmentally Distinct Roles. *Genetics* 202, 175–189. doi:10.1534/genetics.115.182121

Dierick, H.A., Greenspan, R.J., 2006. Molecular analysis of flies selected for aggressive behavior. *Nat Genet* 38, 1023–1031. doi:10.1038/ng1864

Dietzl, G., Chen, D., Schnorrer, F., Su, K.-C., *et al.*, 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature* 448, 151–156. doi:doi:10.1038/nature05954

Dinan, L., Whiting, P., Girault, J.P., Lafont, R., *et al.*, 1997. Cucurbitacins are insect steroid hormone antagonists acting at the ecdysteroid receptor. *Biochemical Journal* 327 (Pt 3), 643–650. doi:10.1042/bj3270643

Dinan, L.L., 2001. Phytoecdysteroids: biological aspects. *Phytochemistry* 57, 325–339. doi:10.1016/S0031-9422(01)00078-4

Dobias, J., Nemec, P., Brtko, J., 1977. The inhibitory effect of kojic acid and its two derivatives on the development of Drosophila melanogaster. *Biológica* 32, 417–421.

Dobler, S., Petschenka, G., Pankoke, H., 2011. Coping with toxic plant compounds – The insect's perspective on iridoid glycosides and cardenolides. *Phytochemistry* 72, 1593–1604. doi:10.1016/j.phytochem.2011.04.015

Dolinski, K., Botstein, D., 2007. Orthology and Functional Conservation in Eukaryotes. *Annu. Rev. Genet.* 41, 465–507. doi:10.1146/annurev.genet.40.110405.090439

Domanitskaya, E., Anllo, L., Schüpbach, T., 2014. Phantom, a cytochrome P450 enzyme essential for ecdysone biosynthesis, plays a critical role in the control of border cell migration in Drosophila. *Developmental biology* 386, 408–418. doi:10.1016/j.ydbio.2013.12.013

Dow, J.A.T., Pandit, A., Davies, S.A., 2018. New views on the Malpighian tubule from post-genomic technologies. *Current Opinion in Insect Science*. doi:10.1016/j.cois.2018.05.010

Drosophila 12 Genomes Consortium, 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature* 450, 203–218. doi:10.1038/nature06341

Duan, H., Yang, X., Bu, Z., Li, X., *et al.*, 2020. Identification and Characterization of Genes Involved in Ecdysteroid Esterification Pathway Contributing to the High 20-Hydroxyecdysone Resistance of Helicoverpa armigera. *Front. Physiol.* 11, 253. doi:10.3389/fphys.2020.00508

Dubrovsky, E.B., Dubrovskaya, V.A., Bilderback, A.L., Berger, E.M., 2000. The Isolation of Two Juvenile Hormone-Inducible Genes in Drosophila melanogaster. *Developmental biology* 224, 10–10. doi:10.1006/dbio.2000.9800

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., *et al.*, 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* eaal3327. doi:10.1126/science.aal3327

Duffy, J.B., 2002. GAL4 system in Drosophila: A fly geneticist's swiss army knife. *Genesis* 34, 1–15. doi:10.1002/gene.10150

Duneau, D., Sun, H., Revah, J., Miguel, K.S., *et al.*, 2018. Signatures of Insecticide Selection in the Genome of Drosophila melanogaster. *G3 (Bethesda)* 8, 3469–3480. doi:10.1534/g3.118.200537

Dunker, A.K., Silman, I., Uversky, V.N., Sussman, J.L., 2008. Function and structure of inherently disordered proteins. *Current Opinion in Structural Biology* 18, 756–764. doi:10.1016/j.sbi.2008.10.002

Dunkov, B.C., Guzov, V.M., Mocelin, G., Shotkoski, F., *et al.*, 1997. The Drosophila cytochrome P450 gene Cyp6a2: structure, localization, heterologous expression, and induction by phenobarbital. *DNA Cell Biol.* 16, 1345–1356. doi:10.1089/dna.1997.16.1345

Durand, N., Carot-Sans, G., Bozzolan, F., Rosell, G., *et al.*, 2011. Degradation of Pheromone and Plant Volatile Components by a Same Odorant-Degrading Enzyme in the Cotton Leafworm, Spodoptera littoralis. *PLoS ONE* 6, e29147. doi:10.1371/journal.pone.0029147

Dutilh, B.E., Backus, L., Edwards, R.A., Wels, M., *et al.*, 2013. Explaining microbial phenotypes on a genomic scale: GWAS for microbes. *Briefings in Functional Genomics* 12, 366–380. doi:10.1093/bfgp/elt008

Dutta, D., Xiang, J., Edgar, B.A., 2013. RNA Expression Profiling from FACS-Isolated Cells of the Drosophila Intestine. *Current Protocols in Stem Cell Biology* 27, 401. doi:10.1002/9780470151808.sc02f02s27

Dweck, H.K.M., Ebrahim, S.A.M., Kromann, S., Bown, D., *et al.*, 2013. Olfactory Preference for Egg Laying on Citrus Substrates in Drosophila. *Current Biology* 23, 2472–2480. doi:10.1016/j.cub.2013.10.047

Dworkin, I., Jones, C.D., 2008. Genetic Changes Accompanying the Evolution of Host Specialization in Drosophila sechellia. *Genetics* 181, 721–736. doi:10.1534/genetics.108.093419

Eckhardt, M., Haider, M., Dorn, S., Müller, A., 2014. Pollen mixing in pollen generalist solitary bees: a possible strategy to complement or mitigate unfavourable pollen properties? *Journal of Animal Ecology* 83, 588–597. doi:10.1111/1365-2656.12168

Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., *et al.*, 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci U S A* 112, 8362–8366. doi:10.1073/pnas.1503926112

Editors, 2007. Name that gene! *Nat Struct Mol Biol* 14, 681–681. doi:10.1038/nsmb0807-681

Ehrlich, P.R., Raven, P.H., 1964. Butterflies and plants: A study in coevolution. *Evolution* 18, 586–608. doi:10.1111/j.1558-5646.1964.tb01674.x

Eickbush, T.H., Eickbush, D.G., 2007. Finely Orchestrated Movements: Evolution of the Ribosomal RNA Genes. *Genetics* 175, 477–485. doi:10.1534/genetics.107.071399

Eisner, T., Wiemer, D.F., Haynes, L.W., Meinwald, J., 1978. Lucibufagins: Defensive steroids from the fireflies Photinus ignitus and P. marginellus (Coleoptera: Lampyridae). *Proc. Natl Acad. Sci. USA* 75, 905–908. doi:10.1073/pnas.75.2.905

El-Brolosy, M.A., Kontarakis, Z., Rossi, A., Kuenne, C., *et al.*, 2019. Genetic compensation triggered by mutant mRNA degradation. *Nature* 568, 193–197. doi:10.1038/s41586-019-1064-z

El-Brolosy, M.A., Stainier, D.Y.R., 2017. Genetic compensation: A phenomenon in search of mechanisms. *PLOS Genetics* 13, e1006780. doi:10.1371/journal.pgen.1006780

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., *et al.*, 2018. The Pfam protein families database in 2019. *Nucl Acids Res* 47, D427–D432. doi:10.1093/nar/gky995

El-Kady, I.A., Zohri, A.N.A., Hamed, S.R., 2014. Kojic Acid Production from Agro-Industrial By-Products Using Fungi. *Biotechnology Research International* 2014, 1–10. doi:10.1155/2014/642385

Ellens, K.W., Christian, N., Singh, C., Satagopam, V.P., *et al.*, 2017. Confronting the catalytic dark matter encoded by sequenced genomes. *Nucl Acids Res* 45, 11495–11514. doi:10.1093/nar/gkx937

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20, 425. doi:10.1186/s13059-019-1832-y

Enayati, A.A., Ranson, H., Hemingway, J., 2005. Insect glutathione transferases and insecticide resistance. *Insect Mol Biol* 14, 3–8. doi:10.1111/j.1365-2583.2004.00529.x

Enya, S., Ameku, T., Igarashi, F., Iga, M., *et al.*, 2014. A Halloween gene noppera-bo encodes a glutathione S-transferase essential for ecdysteroid biosynthesis via regulating the behaviour of cholesterol in Drosophila. *Sci. Rep.* 4, 6586. doi:10.1038/srep06586

Enya, S., Daimon, T., Igarashi, F., Kataoka, H., *et al.*, 2015. The silkworm glutathione S-transferase gene noppera-bo is required for ecdysteroid biosynthesis and larval development. *Insect biochemistry and molecular Biology* 61, 1–7. doi:10.1016/j.ibmb.2015.04.001

Espeland, M., Breinholt, J., Willmott, K.R., Warren, A.D., *et al.*, 2018. A Comprehensive and Dated Phylogenomic Analysis of Butterflies. *Current Biology*. doi:10.1016/j.cub.2018.01.061

Evans, O.P., O'Reilly, D.R., 1998. Purification and kinetic analysis of a baculovirus ecdysteroid UDP-glucosyltransferase. *Biochemical Journal* 330 (Pt 3), 1265–1270. doi:10.1042/bj3301265

Evans, P.D., Bayliss, A., Reale, V., 2014. GPCR-mediated rapid, non-genomic actions of steroids: Comparisons between DmDopEcR and GPER1 (GPR30). *Gen Comp Endocrinol* 195, 157–163. doi:10.1016/j.ygcen.2013.10.015

Everett, L.J., Huang, W., Zhou, S., Carbone, M.A., *et al.*, 2020. Gene expression networks in the Drosophila Genetic Reference Panel. *Genome Res.* 30, 485–496. doi:10.1101/gr.257592.119

Fahrbach, S.E., Smagghe, G., Velarde, R.A., 2012. Insect nuclear receptors. *Annual review of entomology* 57, 83–106. doi:10.1146/annurev-ento-120710-100607

Fallon, T.R., Lower, S.E., Chang, C.-H., Bessho-Uehara, M., *et al.*, 2018. Firefly genomes illuminate parallel origins of bioluminescence in beetles. *eLife* 7, 236. doi:10.7554/eLife.36495

Fandino, R.A., Haverkamp, A., Bisch-Knaden, S., Zhang, J., *et al.*, 2019. Mutagenesis of odorant coreceptor Orco fully disrupts foraging but not oviposition behaviors in the hawkmoth Manduca sexta. *Proc Natl Acad Sci U S A* 116, 15677–15685. doi:10.1073/pnas.1902089116

Feldlaufer, M.F., Svoboda, J.A., Thompson, M.J., Wilzer, K.R., 1987. Fate of maternally-acquired ecdysteroids in unfertilized eggs of Manduca sexta. *Insect Biochemistry* 18, 219–221. doi:10.1016/0020-1790(88)90027-3

Feldlaufer, M.F., Weirich, G.F., Imberski, R.B., Svoboda, J.A., 1995. Ecdysteroid production in Drosophila melanogaster reared on defined diets. *Insect biochemistry and molecular Biology* 25, 709–712. doi:10.1016/0965-1748(95)00009-K

Felsenstein, J., 1978. Cases in which Parsimony or Compatibility Methods will be Positively Misleading. *Systematic Biol.* 27, 401–410. doi:10.1093/sysbio/27.4.401

Fernández, R., Gabaldón, T., 2020. Gene gain and loss across the metazoan tree of life. *Nature Ecology & Evolution 2017 1:4* 9, 1–10. doi:10.1038/s41559-019-1069-x

Feyereisen, R., 2020. Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. *Molecular Phylogenetics and Evolution* 143, 106695. doi:10.1016/j.ympev.2019.106695

Feyereisen, R., 2011. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. *Biochimica et Biophysica Acta* 1814, 19–28. doi:10.1016/j.bbapap.2010.06.012

Feyereisen, R., 2006. Evolution of insect P450. *Biochem. Soc. Trans.* 34, 1252–1255. doi:10.1042/BST0341252

Feyereisen, R., 1995. Molecular biology of insecticide resistance. *Toxicol. Lett.* 82-83, 83–90. doi:10.1016/0378-4274(95)03470-6

Field, L.M., Blackman, R.L., Tyler-Smith, C., Devonshire, A.L., 1999. Relationship between amount of esterase and gene copy number in insecticide-resistant Myzus persicae (Sulzer). *Biochemical Journal* 339 (Pt 3), 737–742. doi:10.1042/bj3390737

Field, L.M., Devonshire, A.L., Forde, B.G., 1988. Molecular evidence that insecticide resistance in peach-potato aphids (Myzus persicae Sulz.) results from amplification of an esterase gene. *Biochemical Journal* 251, 309–312. doi:10.1042/bj2510309

Field, L.M., Williamson, M.S., Moores, G.D., Devonshire, A.L., 1993. Cloning and analysis of the esterase genes conferring insecticide resistance in the peach-potato aphid, Myzus persicae (Sulzer). *Biochemical Journal* 294 (Pt 2), 569–574. doi:10.1042/bj2940569

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. *Nucl Acids Res* 39, W29–W37. doi:10.1093/nar/gkr367

Fiore, M.F., de Lima, S.T., Carmichael, W.W., McKinnie, S.M.K., *et al.*, 2020. Guanitoxin, renaming a cyanobacterial organophosphate toxin. *Harmful Algae* 92, 101737. doi:10.1016/j.hal.2019.101737

Fischer, H.M., Wheat, C.W., Heckel, D.G., Vogel, H., 2008. Evolutionary Origins of a Novel Host Plant Detoxification Gene in Butterflies. *Mol. Biol. Evol.* 25, 809–820. doi:10.1093/molbev/msn014

Fisk, G.J., Thummel, C.S., 1998. The DHR78 Nuclear Receptor Is Required for Ecdysteroid Signaling during the Onset of Drosophila Metamorphosis. *Cell* 93, 543–555. doi:10.1016/S0092-8674(00)81184-8

Fletcher, J.C., Thummel, C.S., 1995. The Drosophila E74 gene is required for the proper stage- and tissue-specific transcription of ecdysone-regulated genes at the onset of metamorphosis. *Development* 121, 1411–1421. doi:10.1006/dbio.1993.1315

Fortier, E., Belote, J.M., 2000. Temperature-dependent gene silencing by an expressed inverted repeat in Drosophila. *Genesis* 26, 240–244. doi:10.1002/(SICI)1526-968X(200004)26:4<240::AID-GENE40>3.0.CO;2-P

Fortin, J.-P., Tan, J., Gascoigne, K.E., Haverty, P.M., *et al.*, 2019. Multiple-gene targeting and mismatch tolerance can confound analysis of genome-wide pooled CRISPR screens. *Genome Biol* 20, 564. doi:10.1186/s13059-019-1621-7

Fournier, D., Maturano, M., Gagnoux, L., Ziliani, P., *et al.*, 1995. Insect Acetylcholinesterase and Resistance to Insecticides, in: *Enzymes of the Cholinesterase Family*. Springer, Boston, MA, Boston, MA, pp. 149–153. doi:10.1007/978-1-4899-1051-6_32

Francino, M.P., 2005. An adaptive radiation model for the origin of new gene functions. *Nat Genet* 37, 573–577. doi:10.1038/ng1579

Francis, J.S., Acevedo, C.R., Muth, F., Leonard, A.S., 2019. Nectar quality changes the ecological costs of chemically defended pollen. *Current Biology* 29, R679–R680. doi:10.1016/j.cub.2019.05.046

Fredholm, B.B., Bättig, K., Holmén, J., Nehlig, A., *et al.*, 1999. Actions of Caffeine in the Brain with Special Reference to Factors That Contribute to Its Widespread Use. *Pharmacol Rev* 51, 83–133. doi:10.1002/ardp.19913240502

Fricke, J., Blei, F., Hoffmeister, D., 2017. Enzymatic Synthesis of Psilocybin. *Angew. Chem. Int. Ed.* 56, 12352–12355. doi:10.1002/anie.201705489

Friedman, R., 2011. Genomic organization of the glutathione S-transferase family in insects. *Molecular Phylogenetics and Evolution* 61, 924–932. doi:10.1016/j.ympev.2011.08.027

Fujii, S., Toyama, A., Amrein, H., 2008. A Male-Specific Fatty Acid ω-Hydroxylase, SXE1, Is Necessary for Efficient Male Mating in Drosophila melanogaster. *Genetics* 180, 179–190. doi:10.1534/genetics.108.089177

Fusetto, R., Denecke, S., Perry, T., O'Hair, R.A.J., *et al.*, 2017. Partitioning the roles of CYP6G1 and gut microbes in the metabolism of the insecticide imidacloprid in Drosophila melanogaster. *Sci. Rep.* 7, 11339. doi:10.1038/s41598-017-09800-2

Galperin, M.Y., Koonin, E.V., 2010. From complete genome sequence to "complete" understanding? *Trends in biotechnology* 28, 398–406. doi:10.1016/j.tibtech.2010.05.006

Garau, G., Di Guilmi, A.M., Hall, B.G., 2005. Structure-Based Phylogeny of the Metallo-β-Lactamases 49, 2778–2784. doi:10.1128/AAC.49.7.2778-2784.2005

Garen, A., Kauvar, L., Lepesant, J.-A., 1977. Roles of ecdysone in Drosophila development. *Proc Natl Acad Sci U S A* 74, 5099–5103. doi:10.1073/pnas.74.11.5099

Garud, N.R., Petrov, D.A., 2016. Elevated Linkage Disequilibrium and Signatures of Soft Sweeps Are Common in Drosophila melanogaster. *Genetics* 203, 863–880. doi:10.1534/genetics.115.184002

Gauhar, Z., Sun, L.V., Hua, S., Mason, C.E., *et al.*, 2009. Genomic mapping of binding regions for the Ecdysone receptor protein complex. 19, 1006–1013. doi:10.1101/gr.081349.108

Gautam, N.-K., Tapadia, M., 2010. Ecdysone signaling is required for proper organization and fluid secretion of stellate cells in the Malpighian tubules of Drosophila melanogaster. *Int J Dev Biol* 54, 635–642. doi:10.1387/ijdb.092910ng

Gautam, N.-K., Verma, P., Tapadia, M.G., 2015. Ecdysone regulates morphogenesis and function of Malpighian tubules in Drosophila melanogaster through EcR-B2 isoform. *Developmental biology* 398, 163–176. doi:10.1016/j.ydbio.2014.11.003

Gay, P., 1978. Drosophila genes which intervene in multiplication of sigma virus. *Molec. Gen. Genet.* 159, 269–283. doi:10.1007/BF00268263

Gazara, R.K., Cardoso, C., Bellieny-Rabelo, D., Ferreira, C., *et al.*, 2017. De novo transcriptome sequencing and comparative analysis of midgut tissues of four non-model insects pertaining to Hemiptera, Coleoptera, Diptera and Lepidoptera. *Gene* 627, 85–93. doi:10.1016/j.gene.2017.06.008

Gaziova, I., Bonnette, P.C., Henrich, V.C., Jindra, M., 2004. Cell-autonomous roles of the ecdysoneless gene in Drosophila development and oogenesis. *Development* 131, 2715–2725. doi:10.1242/dev.01143

Ge, W., Deng, Q., Guo, T., Hong, X., *et al.*, 2015. Regulation of Pattern Formation and Gene Amplification During Drosophila Oogenesis by the miR-318 microRNA. *Genetics* 200, 255–265. doi:10.1534/genetics.115.174748

Gegear, R.J., Manson, J.S., Thomson, J.D., 2007. Ecological context influences pollinator deterrence by alkaloids in floral nectar. *Ecol Lett* 10, 375–382. doi:10.1111/j.1461-0248.2007.01027.x

Geraets, L., Moonen, H.J.J., Wouters, E.F.M., Bast, A., *et al.*, 2006. Caffeine metabolites are inhibitors of the nuclear enzyme poly(ADP-ribose)polymerase-1 at physiological concentrations. *Biochem Pharmacol* 72, 902–910. doi:10.1016/j.bcp.2006.06.023

Gharib, B., Legay, J.-M., de Reggi, M., 1981. Potentiation of developmental abilities of diapausing eggs of Bombyx mori by 20-hydroxyecdysone. *Journal of Insect Physiology* 27, 711–713. doi:10.1016/0022-1910(81)90008-1

Gibson, J.M., Isaac, R.E., Dinan, L.N., Rees, H.H., 1984. Metabolism of [3H]-ecdysone in Schistocerca gregaria; formation of ecdysteroid acids together with free and phosphorylated ecdysteroid acetates. *Archives of Insect Biochemistry and Physiology* 1, 385–407. doi:10.1002/arch.940010409

Giguère, V., 2015. Editorial: What's in a Name, or the Impact of Misnomers in Endocrine Research. *Molecular Endocrinology* 29, 789–790. doi:10.1210/me.2015-1119

Gil, D.L., Rose, H.A., Yang, R.S., Young, R.G., *et al.*, 1974. Enzyme induction by phenobarbital in the Madagascar cockroach, Gromphadorhina portentosa. *Comp Biochem Physiol B* 47, 657–662. doi:10.1016/0305-0491(74)90013-3

Gilbert, L.I., 2011. Insect Endocrinology. Academic Press. doi:10.1016/C2009-0-62089-4

Gilbert, L.I., 2004. Halloween genes encode P450 enzymes that mediate steroid hormone biosynthesis in Drosophila melanogaster. *Molecular and Cellular Endocrinology* 215, 1–10. doi:10.1016/j.mce.2003.11.003

Giraldo-Calderón, G.I., Emrich, S.J., MacCallum, R.M., Maslen, G., *et al.*, 2015. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucl Acids Res* 43, D707–D713. doi:10.1093/nar/gku1117

Giraudo, M., Unnithan, G.C., Le Goff, G., Feyereisen, R., 2010. Regulation of cytochrome P450 expression in Drosophila: Genomic insights. *Pestic Biochem Physiol* 97, 115–122. doi:10.1016/j.pestbp.2009.06.009

Giribet, G., Edgecombe, G.D., 2017. Current Understanding of Ecdysozoa and its Internal Phylogenetic Relationships. *Integrative and Comparative Biology* 57, 455–466. doi:10.1093/icb/icx072

Glotter, E., 1991. Withanolides and related ergostane-type steroids. *Nat Prod Rep* 8, 415–440. doi:10.1039/np9910800415

Gonzalez, D., Fraichard, S., Grassein, P., Delarue, P., *et al.*, 2018. Characterization of a Drosophila glutathione transferase involved in isothiocyanate detoxification. *Insect biochemistry and molecular Biology* 95, 33–43. doi:10.1016/j.ibmb.2018.03.004

Good, R.T., Gramzow, L., Battlay, P., Sztal, T., *et al.*, 2014. The Molecular Evolution of Cytochrome P450 Genes within and between Drosophila Species. *Genome Biology and Evolution* 6, 1118–1134. doi:10.1093/gbe/evu083

Goodacre, N.F., Gerloff, D.L., Uetz, P., Fraser, C.M., 2014. Protein Domains of Unknown Function Are Essential in Bacteria. *mBio* 5, e00744–13. doi:10.1128/mBio.00744-13

Gotoh, O., 2012. Evolution of Cytochrome P450 Genes from the Viewpoint of Genome Informatics. *Biological & pharmaceutical bulletin* 35, 812–817. doi:10.1248/bpb.35.812

Goudru, H.G., Kumar, S., Jayalakshmi, S.K., Ballal, C.R., *et al.*, 2013. Purification and characterization of prophenoloxidase from cotton bollworm, Helicoverpa armigera. *Entomological Research* 43, 55–62. doi:10.1111/1748-5967.12002

Gout, J.F., Lynch, M., 2015. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol. Biol. Evol.* doi:10.1093/molbev/msv095

Gratz, S.J., Ukken, F.P., Rubinstein, C.D., Thiede, G., *et al.*, 2014. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in Drosophila. *Genetics* 196, 961–971. doi:10.1534/genetics.113.160713

Grau, V., Gutzeit, H.O., 1990. Asymmetrically distributed ecdysteroid-related antigens in follicles and young embryos of Drosophila melanogaster. *Roux's Arch Dev Biol* 198, 295–302. doi:10.1007/BF00377396

Grau, V., Lafont, R., 1994. The Distribution of Ecdysone Metabolites Within the Body of Adult Drosophila-Melanogaster Females and Their Sites of Production. *Journal of Insect Physiology* 40, 87–96. doi:10.1016/0022-1910(94)90115-5

Grau, V., Pis, J., Lafont, R., 1995. Ovary-specific interaction of ecdysone 22-phosphate with proteins in adult Drosophila melanogaster (Diptera: Drosophilidae). *EJE* 92, 189–189.

Graur, D., Zheng, Y., Azevedo, R.B.R., 2015. An evolutionary classification of genomic function. *Genome Biology and Evolution* 7, 642–645. doi:10.1093/gbe/evv021

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., *et al.*, 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* 471, 473–479. doi:10.1038/nature09715

Green, E.W., Fedele, G., Giorgini, F., Kyriacou, C.P., 2014. A Drosophila RNAi collection is subject to dominant phenotypic effects. *Nat Meth* 11, 222–223. doi:10.1038/nmeth.2856

Green, L., Battlay, P., Fournier-Level, A., Good, R.T., *et al.*, 2019. Cis- and trans-acting variants contribute to survivorship in a naïve Drosophila melanogaster population exposed to ryanoid insecticides. *Proc Natl Acad Sci U S A* 116, 10424–10429. doi:10.1073/pnas.1821713116

Greene, J.C., Whitworth, A.J., Andrews, L.A., Parker, T.J., *et al.*, 2005. Genetic and genomic studies of Drosophila parkin mutants implicate oxidative stress and innate immune responses in pathogenesis. *Hum Mol Genet* 14, 799–811. doi:10.1093/hmg/ddi074

Gu, X.-C., Zhang, Y.-N., Kang, K., Dong, S.-L., *et al.*, 2015. Antennal Transcriptome Analysis of Odorant Reception Genes in the Red Turpentine Beetle (RTB), Dendroctonus valens. *PLoS ONE* 10, e0125159. doi:10.1371/journal.pone.0125159

Guengerich, F.P., 2001. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* 14, 611–650. doi:10.1021/tx0002583

Guittard, E., Blais, C., Maria, A., Parvy, J.-P., *et al.*, 2011. CYP18A1, a key enzyme of Drosophila steroid hormone inactivation, is essential for metamorphosis. *Developmental biology* 349, 35–45. doi:10.1016/j.ydbio.2010.09.023

Guruharsha, K.G., Obar, R.A., Mintseris, J., Aishwarya, K., *et al.*, 2014. Drosophila Protein interaction Map (DPiM). *Fly* 6, 246–253. doi:10.4161/fly.22108

Guruharsha, K.G., Rual, J.-F., Zhai, B., Mintseris, J., *et al.*, 2011. A Protein Complex Network of Drosophila melanogaster. *Cell* 147, 690–703. doi:10.1016/j.cell.2011.08.047

Gutierrez, E., Wiggins, D., Fielding, B., Gould, A.P., 2007. Specialized hepatocyte-like cells regulate Drosophila lipid metabolism. *Nature* 445, 275–280. doi:10.1038/nature05382

Hagner, K., Setayeshgar, S., Lynch, M., 2018. Stochastic protein multimerization, activity, and fitness. *Physical review. E* 98, 2889. doi:10.1103/PhysRevE.98.062401

Hahn, M.W., Han, M.V., Han, S.-G., 2007. Gene Family Evolution across 12 Drosophila Genomes. *PLOS Genetics* 3, e197. doi:10.1371/journal.pgen.0030197

Hakim, R.S., Baldwin, K., Smagghe, G., 2010. Regulation of midgut growth, development, and metamorphosis. *Annual review of entomology* 55, 593–608. doi:10.1146/annurev-ento-112408-085450

Hall, J.P.W., Willmott, K.R., 2000. Patterns of feeding behaviour in adult male riodinid butterflies and their relationship to morphology and ecology. *Biological journal of the Linnean society* 69, 1–23. doi:10.1111/j.1095-8312.2000.tb01666.x

Hall, R.M., Schwarz, S., 2016. Resistance gene naming and numbering: is it a new gene or not? *J Antimicrob Chemother* 71, 569–571. doi:10.1093/jac/dkv351

Hamby, K.A., Becher, P.G., 2016. Current knowledge of interactions between Drosophila suzukii and microbes, and their potential utility for pest management. *J Pest Sci* 89, 621–630. doi:10.1007/s10340-016-0768-1

Hammer, T.J., Bowers, M.D., 2015. Gut microbes may facilitate insect herbivory of chemically defended plants. *Oecologia* 179, 1–14. doi:10.1007/s00442-015-3327-1

Hammer, T.J., Janzen, D.H., Hallwachs, W., Jaffe, S.P., *et al.*, 2017. Caterpillars lack a resident gut microbiome. *Proc Natl Acad Sci U S A* 114, 9641–9646. doi:10.1073/pnas.1707186114

Hardy, N.B., Peterson, D.A., Ross, L., Rosenheim, J.A., 2018. Does a plant-eating insect's diet govern the evolution of insecticide resistance? Comparative tests of the pre-adaptation hypothesis. *Evolutionary Applications* 11, 739–747. doi:10.1111/eva.12579

Hardy, S., Legagneux, V., Audic, Y., Paillard, L., 2010. Reverse genetics in eukaryotes. *Biology of the Cell* 102, 561–580. doi:10.1042/BC20100038

Harmatha, J., Dinan, L., 1997. Biological activity of natural and synthetic ecdysteroids in the B11 bioassay. *Archives of Insect Biochemistry and Physiology* 35, 219–225. doi:10.1002/(SICI)1520-6327(1997)35:1/2<219::AID-ARCH20>3.0.CO;2-D

Harris, T.W., Arnaboldi, V., Cain, S., Chan, J., *et al.*, 2020. WormBase: a modern Model Organism Information Resource. *Nucl Acids Res* 48, D762–D767. doi:10.1093/nar/gkz920

Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., *et al.*, 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution 2017 1:4* 2, 557–566. doi:10.1038/s41559-017-0459-1

Harrop, T.W., Denecke, S., Yang, Y.T., Chan, J., *et al.*, 2018. Evidence for activation of nitenpyram by a mitochondrial cytochrome P450 in Drosophila melanogaster. *Pest. Manag. Sci.* 74, 1616–1622. doi:10.1002/ps.4852

Harrop, T.W.R., Pearce, S.L., Daborn, P.J., Batterham, P., 2014. Whole-genome expression analysis in the third instar larval midgut of Drosophila melanogaster. *G3 (Bethesda)* 4, 2197–2205. doi:10.1534/g3.114.013870

Hasić, D., Tannier, E., 2019. Gene tree species tree reconciliation with gene conversion. *J. Math. Biol.* 78, 1981–2014. doi:10.1007/s00285-019-01331-w

Hawkins, J.S., Proulx, S.R., Rapp, R.A., Wendel, J.F., 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proc Natl Acad Sci U S A* 106, 17811–17816. doi:10.1073/pnas.0904339106

Hayek, Al, S., Alsawadi, A., Kambris, Z., Boquete, J.-P., *et al.*, 2019. Shavenbaby protein isoforms orchestrate the self-renewal versus differentiation of Drosophila intestinal stem cells. *bioRxiv* 127, 627554. doi:10.1101/627554

Heenan, M.P., Smith, J.N., 1974. Water-soluble metabolites of p-nitrophenol and 1-naphthyl N-methylcarbamate in flies and grass grubs. Formation of glucose phosphate and phosphate conjugates. *Biochemical Journal* 144, 303–310. doi:10.1042/bj1440303

Heidel-Fischer, H., Vogel, H., 2015. Molecular mechanisms of insect adaptation to plant secondary compounds. *Current Opinion in Insect Science*. doi:10.1016/j.cois.2015.02.004

Heidel-Fischer, H.M., Kirsch, R., Reichelt, M., Ahn, S.-J., *et al.*, 2019. An Insect Counteradaptation against Host Plant Defenses Evolved through Concerted Neofunctionalization. *Mol. Biol. Evol.* 36, 930–941. doi:10.1093/molbev/msz019

Heigwer, F., Port, F., Boutros, M., 2018. RNA Interference (RNAi) Screening in Drosophila. *Genetics* 208, 853–874. doi:10.1534/genetics.117.300077

Hellman, K., Nielsen, P.A., Ek, F., Olsson, R., 2016. An ex Vivo Model for Evaluating Blood-Brain Barrier Permeability, Efflux, and Drug Metabolism. *ACS Chem Neurosci* 7, 668–680. doi:10.1021/acschemneuro.6b00024

Helvig, C., Koener, J.F., Unnithan, G.C., Feyereisen, R., 2004a. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. *Proc Natl Acad Sci U S A* 101, 4024–4029. doi:10.1073/pnas.0306980101

Helvig, C., Tijet, N., Feyereisen, R., Walker, F.A., *et al.*, 2004b. Drosophila melanogaster CYP6A8, an insect P450 that catalyzes lauric acid (ω-1)-hydroxylation. *Biochem Biophys Res Commun* 325, 1495–1502. doi:10.1016/j.bbrc.2004.10.194

Hentze, J.L., Moeller, M.E., Jørgensen, A.F., Bengtsson, M.S., *et al.*, 2013. Accessory Gland as a Site for Prothoracicotropic Hormone Controlled Ecdysone Synthesis in Adult Male Insects. *PLoS ONE* 8, e55131. doi:10.1371/journal.pone.0055131.s002

Highfill, C.A., Tran, J.H., Nguyen, S.K.T., Moldenhauer, T.R., *et al.*, 2017. Naturally Segregating Variation at Ugt86Dd Contributes to Nicotine Resistance in Drosophila melanogaster. *Genetics* 207, 311–325. doi:10.1534/genetics.117.300058

Hilton, J.P., 2004. Isolation and characterization of ecdysteroid phosphotransferases. University of Liverpool, Liverpool.

Hindle, S.J., Bainton, R.J., 2014. Barrier mechanisms in the Drosophila blood-brain barrier. *Front. Neurosci.* 8, 5. doi:10.3389/fnins.2014.00414

Hindle, S.J., Munji, R.N., Dolghih, E., Gaskins, G., *et al.*, 2017. Evolutionarily Conserved Roles for Blood-Brain Barrier Xenobiotic Transporters in Endogenous Steroid Partitioning and Behavior. *Cell Reports* 21, 1304–1316. doi:10.1016/j.celrep.2017.10.026

Ho, J., Tumkaya, T., Aryal, S., Choi, H., *et al.*, 2019. Moving beyond P values: data analysis with estimation graphics. *Nat Meth* 16, 565–566. doi:10.1038/s41592-019-0470-3

Hoang, D.T., Chernomor, O., Haeseler, von, A., Minh, B.Q., *et al.*, 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* 35, 518–522. doi:10.1093/molbev/msx281

Hoang, K., Matzkin, L.M., Bono, J.M., 2015. Transcriptional variation associated with cactus host plant adaptation in Drosophila mettleri populations. *Mol Ecol* 24, 5186–5199. doi:10.1111/mec.13388

Hobson, R.P., 1935a. On a fat-soluble growth factor required by blow-fly larvae: Distribution and properties. *The Biochemical Journal* 29, 1292–1296. doi:10.1042/bj0291292

Hobson, R.P., 1935b. On a fat-soluble growth factor required by blow-fly larvae: Identity of the growth factor with cholesterol. *The Biochemical Journal* 29, 2023–2026. doi:10.1042/bj0292023

Hock, T., Cottrill, T., Keegan, J., Garza, D., 2000. The E23 early gene of Drosophila encodes an ecdysone-inducible ATP-binding cassette transporter capable of repressing ecdysone-mediated gene activation. *Proc Natl Acad Sci U S A* 97, 9519–9524. doi:10.1073/pnas.160271797

Hoenigsberg, H.F., Palomino, J.J., Chiappe, C., Rojas, G.G., *et al.*, 1977. Population genetics in the American tropics : XI. Seasonal and temporal variations in relative frequencies of species belonging to the Willistoni group in Colombia. *Oecologia* 27, 295–304. doi:10.1007/BF00345562

Hoffman, G.G., Davulcu, O., Sona, S., Ellington, W.R., 2008. Contributions to catalysis and potential interactions of the three catalytic domains in a contiguous trimeric creatine kinase. *FEBS Journal* 275, 646–654. doi:10.1111/j.1742-4658.2007.06226.x

Hoffmann, K.H., Bulenda, D., Thiry, E., Schmid, E., 1985. Apolar ecdysteroid esters in adult female crickets, Gryllus bimaculatus. *Life Sciences* 37, 185–192. doi:10.1016/0024-3205(85)90422-9

Holliday, R., 1964. A mechanism for gene conversion in fungi. *Genet. Res.* 5, 282–304. doi:10.1017/S0016672300001233

Holmes, R.S., Wright, M.W., Laulederkind, S.J.F., Cox, L.A., *et al.*, 2010. Recommended nomenclature for five mammalian carboxylesterase gene families: human, mouse, and rat genes and proteins. *Mamm. Genome* 21, 427–441. doi:10.1007/s00335-010-9284-4

Honjo, K., Mauthner, S.E., Wang, Y., Skene, J.H.P., *et al.*, 2016. Nociceptor-Enriched Genes Required for Normal Thermal Nociception. *Cell Reports* 16, 295–303. doi:10.1016/j.celrep.2016.06.003

Hoover, K., Grove, M., Gardner, M., Hughes, D.P., *et al.*, 2011. A gene for an extended phenotype. *Science* 333, 1401. doi:10.1126/science.1209199

Hopkins, D.H., Fraser, N.J., Mabbitt, P.D., Carr, P.D., *et al.*, 2017. Structure of an Insecticide Sequestering Carboxylesterase from the Disease Vector Culex quinquefasciatus: What Makes an Enzyme a Good Insecticide Sponge? *Biochemistry* 56, 5512–5525. doi:10.1021/acs.biochem.7b00774

Horner, M.A., Pardee, K., Liu, S., King-Jones, K., *et al.*, 2009. The Drosophila DHR96 nuclear receptor binds cholesterol and regulates cholesterol homeostasis. *Genes & Development* 23, 2711–2716. doi:10.1101/gad.1833609

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50, 346–363. doi:10.1002/bimj.200810425

Howard, J.J., 1990. Infidelity of leafcutting ants to host plants: resource heterogeneity or defense induction? *Oecologia* 82, 394–401. doi:10.1007/BF00317488

Howard, J.J., 1988. Leafcutting and Diet Selection: Relative Influence of Leaf Chemistry and Physical Features. *Ecology* 69, 250–260. doi:10.2307/1943180

Howard, J.J., 1987. Leafcutting Ant Diet Selection: The Role of Nutrients, Water, and Secondary Chemistry. *Ecology* 68, 503–515. doi:10.2307/1938455

Howard, J.J., Green, T.P., Wiemer, D.F., 1989. Comparative deterrency of two terpenoids to two genera of attine ants. *Journal of chemical ecology* 15, 2279–2288. doi:10.1007/BF01012081

Hu, B., Hu, S., Huang, H., Wei, Q., *et al.*, 2019a. Insecticides induce the co-expression of glutathione S-transferases through ROS/CncC pathway in Spodoptera exigua. *Pestic Biochem Physiol* 155, 58–71. doi:10.1016/j.pestbp.2019.01.008

Hu, B., Huang, H., Wei, Q., Ren, M., *et al.*, 2019b. Transcription factors CncC/Maf and AhR/ARNT coordinately regulate the expression of multiple GSTs conferring resistance to chlorpyrifos and cypermethrin in Spodoptera exigua. *Pest. Manag. Sci.* 75, 2009–2019. doi:10.1002/ps.5316

Hu, X., Cherbas, L., Cherbas, P., 2003. Transcription Activation by the Ecdysone Receptor (EcR/USP): Identification of Activation Functions. *Molecular Endocrinology* 17, 716–731. doi:10.1210/me.2002-0287

Hu, Y., Sopko, R., Chung, V., Foos, M., *et al.*, 2019c. iProteinDB: An Integrative Database of Drosophila Post-translational Modifications. *G3 (Bethesda)* 9, 1–11. doi:10.1534/g3.118.200637

Hubbell, S.P., Howard, J.J., Wiemer, D.F., 1984. Chemical Leaf Repellency to an Attine Ant: Seasonal Distribution Among Potential Host Plant Species. *Ecology* 65, 1067–1076. doi:10.2307/1938314

Hughes, A.L., 2013. Origin of ecdysteroid UDP-glycosyltransferases of baculoviruses through horizontal gene transfer from Lepidoptera. *Coevolution* 1, 1–7. doi:10.1080/23256214.2013.858497

Huson, D.H., Bryant, D., 2005. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* 23, 254–267. doi:10.1093/molbev/msj030

Huynh, N., Zeng, J., Liu, W., King-Jones, K., 2018. A Drosophila CRISPR/Cas9 Toolkit for Conditionally Manipulating Gene Expression in the Prothoracic Gland as a Test Case for Polytene Tissues. *G3 (Bethesda)* g3.200539.2018. doi:10.1534/g3.118.200539

Ibanez, S., Gallet, C., Després, L., 2012. Plant insecticidal toxins in ecological networks. *Toxins (Basel)* 4, 228–243. doi:10.3390/toxins4040228

Iglesias, D.J., Cercós, M., Colmenero-Flores, J.M., Naranjo, M.A., *et al.*, 2007. Physiology of citrus fruiting. *Brazilian Journal of Plant Physiology* 19, 333–362. doi:10.1590/S1677-04202007000400006

Ile, K.E., Tripathy, R., Goldfinger, V., Renault, A.D., 2012. Wunen, a Drosophila lipid phosphate phosphatase, is required for septate junction-mediated barrier function. *Development* 139, 2535–2546. doi:10.1242/dev.077289

Innan, H., Kondrashov, F., 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 11, 97–108. doi:10.1038/nrg2689

Isaac, R.E., Rees, H.H., 1984. Isolation and identification of ecdysteroid phosphates and acetylecdysteroid phosphates from developing eggs of the locust, Schistocerca gregaria. *Biochemical Journal* 221, 459–464. doi:10.1042/bj2210459

Isaac, E.R., Rees, H.H., 1985. Metabolism of maternal ecdysteroid-22-phosphates in developing embryos of the desert locust, Schistocerca gregaria. *Insect Biochemistry* 15, 65–72. doi:10.1016/0020-1790(85)90045-9

Isaac, R.E., Rose, M.E., Rees, H.H., Goodwin, T.W., 1983a. Identification of the 22-phosphate esters of ecdysone, 2-deoxyecdysone, 20-hydroxyecdysone and 2-deoxy-20-hydroxyecdysone from newly laid eggs of the desert locust, Schistocerca gregaria. *Biochemical Journal* 213, 533–541. doi:10.1042/bj2130533

Isaac, R.E., Rose, M.E., Rees, H.H., Goodwin, T.W., 1982. Identification of ecdysone-22-phosphate and 2-deoxyecysone-22-phosphate in eggs of the desert locust, Schistocerca gregaris, by fast atom bombardment mass spectrometry and n.m.r. spectroscopy. *J. Chem. Soc., Chem. Commun.* 0, 249–251. doi:10.1039/C39820000249

Isaac, R.E., Sweeney, F.P., Rees, H.H., 1983b. Enzymic hydrolysis of ecdysteroid phosphate during embryogenesis in the desert locust (Schistocerca gregaria). *Biochim. Biophys. Acta* 50, 340. doi: 10.1042/bst0110379

Ishikawa, A., Ogawa, K., Gotoh, H., Walsh, T.K., *et al.*, 2012. Juvenile hormone titre and related gene expression during the change of reproductive modes in the pea aphid. *Insect Mol Biol* 21, 49–60. doi:10.1111/j.1365-2583.2011.01111.x

Ishimoto, H., Sakai, T., Kitamoto, T., 2009. Ecdysone signaling regulates the formation of long-term courtship memory in adult Drosophila melanogaster. *PNAS* 106, 6381–6386. doi:10.1073/pnas.0810213106

Ishimoto, H., Wang, Z., Rao, Y., Wu, C.-F., *et al.*, 2012. A novel role for ecdysone in Drosophila conditioned behavior: linking GPCR-mediated non-canonical steroid action to cAMP signaling in the adult brain. *PLOS Genetics* 9, e1003843–e1003843. doi:10.1371/journal.pgen.1003843

Isoe, J., Koch, L.E., Isoe, Y.E., Rascón, A.A., *et al.*, 2019. Identification and characterization of a mosquito-specific eggshell organizing factor in Aedes aegypti mosquitoes. *PLoS Biol* 17, e3000068. doi:10.1371/journal.pbio.3000068

Ito, Y., Sonobe, H., 2009. The Role of Ecdysteroid 22-Kinase in the Accumulation of Ecdysteroids in Ovary of Silkworm Bombyx mori. *Annals of the New York Academy of Sciences* 1163, 421–424. doi:10.1111/j.1749-6632.2009.04465.x

Ito, Y., Yasuda, A., Sonobe, H., 2008. Synthesis and Phosphorylation of Ecdysteroids During Ovarian Development in the Silkworm, Bombyx mori. *Zoological Science* 25, 721–727. doi:10.2108/zsj.25.721

Jacobsen, D.J., Raguso, R.A., 2018. Lingering Effects of Herbivory and Plant Defenses on Pollinators. *Current Biology* 28, R1164–R1169. doi:10.1016/j.cub.2018.08.010

Jakob, U., Kriwacki, R., Uversky, V.N., 2014. Conditionally and Transiently Disordered Proteins: Awakening Cryptic Disorder To Regulate Protein Function. *Chem. Rev.* 114, 6779–6805. doi:10.1021/cr400459c

Jang, T., Lee, K.P., 2015. The genetic basis for mating-induced sex differences in starvation resistance in Drosophila melanogaster. *Journal of Insect Physiology* 82, 56–65. doi:10.1016/j.jinsphys.2015.09.002

Jaroszewski, L., Li, Z., Krishna, S.S., Bakolitsa, C., *et al.*, 2009. Exploration of Uncharted Regions of the Protein Universe. *PLoS Biol* 7, e1000205. doi:10.1371/journal.pbio.1000205

Jindra, M., Uhlirova, M., Charles, J.-P., Smykal, V., *et al.*, 2015. Genetic Evidence for Function of the bHLH-PAS Protein Gce/Met As a Juvenile Hormone Receptor. *PLOS Genetics* 11, e1005394. doi:10.1371/journal.pgen.1005394

Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., *et al.*, 2018a. Phylogenomics and the evolution of hemipteroid insects. *Proc Natl Acad Sci U S A* 19, 201815820. doi:10.1073/pnas.1815820115

Johnson, R.M., Harpur, B.A., Dogantzis, K.A., Zayed, A., *et al.*, 2018b. Genomic footprint of evolution of eusociality in bees: floral food use and CYPome "blooms." *Insect. Soc.* 65, 445–454. doi:10.1007/s00040-018-0631-x

Johnson, R.N., O'Meally, D., Chen, Z., Etherington, G.J., *et al.*, 2018c. Adaptation and conservation insights from the koala genome. *Nat Genet* 65, 1. doi:10.1038/s41588-018-0153-5

Joußen, N., Agnolet, S., Lorenz, S., Schöne, S.E., *et al.*, 2012. Resistance of Australian Helicoverpa armigera to fenvalerate is due to the chimeric P450 enzyme CYP337B3. *Proc Natl Acad Sci U S A* 109, 15206–15211. doi:10.1073/pnas.1202047109

Joußen, N., Heckel, D.G., Haas, M., Schuphan, I., *et al.*, 2008. Metabolism of imidacloprid and DDT by P450 CYP6G1 expressed in cell cultures of Nicotiana tabacum suggests detoxification of these insecticides in Cyp6g1-overexpressing strains of Drosophila melanogaster, leading to resistance. *Pest. Manag. Sci.* 64, 65–73. doi:10.1002/ps.1472

Junqueira, A.C.M., Azeredo-Espin, A.M.L., Paulo, D.F., Marinho, M.A.T., *et al.*, 2016. Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity. *Sci. Rep.* 6, 1–13. doi:10.1038/srep21762

Jurgens, G., Wieschaus, E., Nüsslein-Volhard, C., Kluding, H., 1984. Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. *Wilhelm Roux' Archiv* 193, 283–295. doi:10.1007/BF00848157

Júnior, F.E.B., Macedo, G.E., Zemolin, A.P., da Silva, G.F., *et al.*, 2016. Oxidant effects and toxicity of Croton campestris in Drosophila melanogaster. *Pharmaceutical Biology* 54, 3068–3077. doi:10.1080/13880209.2016.1207089

Kabbouh, M., Rees, H.H., 1993. Characterization of ATP: Ecdysone 3-acetate 2-phosphotransferase (ecdysone 3-acetate 2-kinase) in Schistocerca gregaria larvae. *Insect biochemistry and molecular Biology* 23, 73–79. doi:10.1016/0965-1748(93)90084-6

Kabbouh, M., Rees, H.H., 1991. Characterization of the ATP:2-deoxyecdysone 22-phosphotransferase (2-deoxyecdysone 22-kinase) in the follicle cells of Schistocerca gregaria. *Insect Biochemistry* 21, 57–64. doi:10.1016/0020-1790(91)90064-L

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., Haeseler, von, A., *et al.*, 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Meth* 8, 163. doi:10.1038/nmeth.4285

Kamba, M., Sonobe, H., Mamiya, Y., Hara, N., 1995. Isolation and identification of 3-epi-ecdysteroids from diapause eggs of the silkworm, Bombyx mori. *The Journal of Sericultural Science of Japan*. doi:10.11416/kontyushigen1930.64.333

Kamba, M., Sonobe, H., Mamiya, Y., Hara, N., *et al.*, 2000. 3-Epi-22-Deoxy-20,26-Dihydroxyecdysone and 3-Epi-22-Deoxy-16β,20-Dihydroxyecdysone, and their 2-Phosphates from Eggs of the Silkworm, Bombyx mori. *Natural Product Letters* 14, 469–476. doi:10.1080/10575630008043787

Kamimura, M., Saito, H., Niwa, R., Niimi, T., *et al.*, 2012. Fungal ecdysteroid-22-oxidase, a new tool for manipulating ecdysteroid signaling and insect development. *J. Biol. Chem.* 287, 16488–16498. doi:10.1074/jbc.M112.341180

Kamita, S.G., Hammock, B.D., 2010. Juvenile hormone esterase: biochemistry and structure. *Journal of pesticide science* 35, 265–274. doi:10.1584/jpestics.R10-09

Kang, J., Kim, J., Choi, K.-W., 2011. Novel Cytochrome P450, cyp6a17, Is Required for Temperature Preference Behavior in Drosophila. *PLoS ONE* 6, e29800. doi:10.1371/journal.pone.0029800

Kanost, M.R., Arrese, E.L., Cao, X., Chen, Y.-R., *et al.*, 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, Manduca sexta. *Insect biochemistry and molecular Biology*. doi:10.1016/j.ibmb.2016.07.005

Kaplanis, J.N., Robbins, W.E., Thompson, M.J., Dutky, S.R., 1973. 26-Hydroxyecdysone: New Insect Molting Hormone from the Egg of the Tobacco Hornworm. *Science* 180, 307–308. doi:10.1126/science.180.4083.307

Kaplanis, J.N., Thompson, M.J., Dutky, S.R., Robbins, W.E., 1980. The ecdysteroids from young embryonated eggs of the tobacco hornworm. *Steroids* 36, 321–336. doi:10.1016/0039-128X(80)90006-9

Kaplanoglu, E., Chapman, P., Scott, I.M., Donly, C., 2017. Overexpression of a cytochrome P450 and a UDP-glycosyltransferase is associated with imidacloprid resistance in the Colorado potato beetle, Leptinotarsa decemlineata. *Sci. Rep.* 7, 1762. doi:10.1038/s41598-017-01961-4

Karginov, T.A., Pastor, D.P.H., Semler, B.L., Gomez, C.M., 2017. Mammalian Polycistronic mRNAs and Disease. *Trends in genetics: TIG* 33, 129–142. doi:10.1016/j.tig.2016.11.007

Karimi, K., Fortriede, J.D., Lotay, V.S., Burns, K.A., *et al.*, 2017. Xenbase: a genomic, epigenomic and transcriptomic model organism database. *Nucl Acids Res* 46, D861–D868. doi:10.1093/nar/gkx936

Katju, V., Bergthorsson, U., 2013. Copy-number changes in evolution: rates, fitness effects and adaptive significance. *Front. Genet.* 4. doi:10.3389/fgene.2013.00273

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010

Katsube, K.I., Sakamoto, K., Tamamura, Y., Yamaguchi, A., 2009. Role of CCN, a vertebrate specific gene family, in development. *Development, Growth & Differentiation* 51, 55–67. doi:10.1111/j.1440-169X.2009.01077.x

Kawahara, A.Y., Plotkin, D., Espeland, M., Meusemann, K., *et al.*, 2019. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl Acad Sci U S A* vol. 3148, 201907847. doi:10.1073/pnas.1907847116

Kawashima, A., Satta, Y., 2014. Substrate-Dependent Evolution of Cytochrome P450: Rapid Turnover of the Detoxification-Type and Conservation of the Biosynthesis-Type. *PLoS ONE* 9, –e100059. doi:10.1371/journal.pone.0100059

Kayser, H., Wtnkler, T., Spindler-Barth, M., 1997. 26-Hydroxylation of Ecdysteroids is Catalyzed by a Typical Cytochrome P-450-Dependent Oxidase and Related to Ecdysteroid Resistance in an Insect Cell Line. *European Journal of Biochemistry* 248, 707–716. doi:10.1111/j.1432-1033.1997.00707.x

Kazana, E., Pope, T.W., Tibbles, L., Bridges, M., *et al.*, 2007. The cabbage aphid: a walking mustard oil bomb. *Proceedings of the Royal Society B: Biological Sciences* 274, 2271–2277. doi:10.1098/rspb.2007.0237

Keck, F., Rimet, F., Bouchez, A., Franc, A., 2016. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecology and Evolution* 6, 2774–2780. doi:10.1002/ece3.2051

Keeling, C.I., Bearfield, J.C., Young, S., Blomquist, G.J., *et al.*, 2006. Effects of juvenile hormone on gene expression in the pheromone-producing midgut of the pine engraver beetle, Ips pini. *Insect Mol Biol* 15, 207–216. doi:10.1111/j.1365-2583.2006.00629.x

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., *et al.*, 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC bioinformatics* 19, 511. doi:10.1186/s12859-018-2203-5

Kempken, F., Rohlfs, M., 2010. Fungal secondary metabolite biosynthesis – a chemical defence strategy against antagonistic animals? *Fungal Ecology* 3, 107–114. doi:10.1016/j.funeco.2009.08.001

Kendall, R.O., 1959. More larval foodplants from Texas. *Journal of the Lepidopterists Society*.

Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., *et al.*, 2012. Symbiont-mediated insecticide resistance. *Proc Natl Acad Sci U S A* 109, 8618–8622. doi:10.1073/pnas.1200231109

Kim, D.-H., Kim, Y.-J., Adams, M.E., 2018a. Endocrine regulation of airway clearance in Drosophila. *Proc Natl Acad Sci U S A* 115, 201717257–1540. doi:10.1073/pnas.1717257115

Kim, J., Suh, H., Kim, S., Kim, K., *et al.*, 2006. Identification and characteristics of the structural gene for the Drosophila eye colour mutant sepia, encoding PDA synthase, a member of the Omega class glutathione S-transferases. *Biochemical Journal* 398, 451–460. doi:10.1042/BJ20060424

Kim, K., Kim, S.-H., Kim, J., Kim, H., *et al.*, 2012. Glutathione s-transferase omega 1 activity is sufficient to suppress neurodegeneration in a Drosophila model of Parkinson disease. *Journal of Biological Chemistry* 287, 6628–6641. doi:10.1074/jbc.M111.291179

Kim, Y., Lee, S., Kim, Y.H., Kim, Y.H., 2018b. Comparative analyses of susceptibility to chemicals associated with fermentation between Drosophila melanogaster and Drosophila suzukii. *Entomological Research* 48, 514–521. doi:10.1111/1748-5967.12326

Kimura, M., 2012. The Neutral Theory of Molecular Evolution, 1st ed. Cambridge University Press. doi:10.1017/CBO9780511623486

King-Jones, K., Horner, M.A., Lam, G., Thummel, C.S., 2006. The DHR96 nuclear receptor regulates xenobiotic responses in Drosophila. *Cell Metab* 4, 12–12. doi:10.1016/j.cmet.2006.06.006

Kiriishi, S., Rountree, D.B., Sakurai, S., Gilbert, L.I., 1990. Prothoracic gland synthesis of 3-dehydroecdysone and its hemolymph 3β-reductase mediated conversion to ecdysone in representative insects. *Experientia* 46, 716–721. doi:10.1007/BF01939944

Kitts, P.A., Church, D.M., Thibaud-Nissen, F., Choi, J., *et al.*, 2016. Assembly: a resource for assembled genomes at NCBI. *Nucl Acids Res* 44, D73–D80. doi:10.1093/nar/gkv1226

Kliebenstein, D.J., 2008. A Role for Gene Duplication and Natural Variation of Gene Expression in the Evolution of Metabolism. *PLoS ONE* 3, e1838–e1838. doi:10.1371/journal.pone.0001838

Knapp, E., Sun, J., 2017. Steroid signaling in mature follicles is important for Drosophila ovulation. *PNAS* 201614383. doi:10.1073/pnas.1614383114

Knapp, E.M., Li, W., Singh, V., Sun, J., 2020. Nuclear receptor Ftz-f1 promotes follicle maturation and ovulation partly via bHLH/PAS transcription factor Sim. *eLife* 9, 2410. doi:10.7554/eLife.54568

Koenig, C., Bretschneider, A., Heckel, D.G., Grosse-Wilde, E., *et al.*, 2015. The plastic response of Manduca sexta to host and non-host plants. *Insect biochemistry and molecular Biology* 63, 72–85. doi:10.1016/j.ibmb.2015.06.001

Kohli, M.K., Ware, J., Bechly, G., 2016. How to date a dragonfly: Fossil calibrations for odonates. *Palaeontologia Electronica* 19, 1FC. doi:10.26879/576

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., *et al.*, 2010. Small Peptides Switch the Transcriptional Activity of Shavenbaby During Drosophila Embryogenesis. *Science* 329, 336–339. doi:10.1126/science.1188158

Koonin, E.V., 2005. Orthologs, Paralogs, and Evolutionary Genomics. *Annu. Rev. Genet.* 39, 309–338. doi:10.1146/annurev.genet.39.073003.114725

Korytko, P.J., Scott, J.G., 1998. CYP6D1 protects thoracic ganglia of houseflies from the neurotoxic insecticide cypermethrin. *Archives of Insect Biochemistry and Physiology* 37, 57–63. doi:10.1002/(SICI)1520-6327(1998)37:1<57::AID-ARCH7>3.0.CO;2-S

Kovganko, N.V., 1999. Ecdysteroids and related compounds in fungi. *Chem Nat Compd* 35, 597–611. doi:10.1007/BF02236282

Kowanetz, K., Crosetto, N., Haglund, K., Schmidt, M., *et al.*, 2004. Suppressors of T-cell receptor signaling Sts-1 and Sts-2 bind to Cbl and inhibit endocytosis of receptor tyrosine kinases. *J. Biol. Chem.* 279, 32786–32795. doi:10.1074/jbc.M403759200

Kozlova, T., Thummel, C.S., 2003. Essential Roles for Ecdysone Signaling During Drosophila Mid-Embryonic Development. *Science* 301, 1911–1914. doi:10.1126/science.1087419

König, S., Romoth, L., Stanke, M., 2018. Comparative Genome Annotation, in: *Comparative Genomics*, Methods and Protocols. Humana Press, New York, NY, New York, NY, pp. 189–212. doi:10.1007/978-1-4939-7463-4_6

König, S., Romoth, L.W., Gerischer, L., Stanke, M., 2016. Simultaneous gene finding in multiple genomes. *Bioinformatics* 32, 3388–3395. doi:10.1093/bioinformatics/btw494

Kramer, K.J., Hopkins, T.L., 1987. Tyrosine metabolism for insect cuticle tanning. *Archives of Insect Biochemistry and Physiology* 6, 279–301. doi:10.1002/arch.940060406

Krempl, C., Heidel-Fischer, H.M., Jiménez-Alemán, G.H., Reichelt, M., *et al.*, 2016a. Gossypol toxicity and detoxification in Helicoverpa armigera and Heliothis virescens. *Insect biochemistry and molecular Biology*. doi:10.1016/j.ibmb.2016.09.003

Krempl, C., Sporer, T., Reichelt, M., Ahn, S.-J., *et al.*, 2016b. Potential detoxification of gossypol by UDP-glycosyltransferases in the two Heliothine moth species Helicoverpa armigera and Heliothis virescens. *Insect biochemistry and molecular Biology* 71, 49–57. doi:10.1016/j.ibmb.2016.02.005

Kretschmar, J.A., Baumann, T.W., 1999. Caffeine in Citrus flowers. *Phytochemistry* 52, 19–23. doi:10.1016/S0031-9422(99)00119-3

Krosch, M.N., Schutze, M.K., Armstrong, K.F., Graham, G.C., *et al.*, 2012. A molecular phylogeny for the Tribe Dacini (Diptera: Tephritidae): Systematic and biogeographic implications. *Molecular Phylogenetics and Evolution* 64, 513–523. doi:10.1016/j.ympev.2012.05.006

Kryuchkova-Mostacci, N., Robinson-Rechavi, M., 2016. Tissue-Specificity of Gene Expression Diverges Slowly between Orthologs, and Rapidly between Paralogs. *PLoS Comp Biol* 12, e1005274. doi:10.1371/journal.pcbi.1005274

Kubo, I., Komatsu, S., Asaka, Y., de Boer, G., 1987. Isolation and identification of apolar metabolites of ingested 20-hydroxyecdysone in frass of Heliothis virescens larvae. *Journal of chemical ecology* 13, 785–794. doi:10.1007/BF01020160

Kubo, I., Zhang, M., De Boer, G., Uchima, K., 1994. Location of ecdysteroid 22-O-acyltransferase in the larvae of Heliothis virescens. *Entomologia Experimentalis et Applicata* 70, 263–272. doi:10.1111/j.1570-7458.1994.tb00755.x

Kuhlmann, W., Fromme, H.G., Heege, E.M., Ostertag, W., 1968. The mutagenic action of caffeine in higher organisms. *Cancer Res* 28, 2375–2389.

Kürsteiner, R., 1961. Ueber die fluoreszierenden Stoffe (Pterine) in den Meconien der Wildrasse und der Mutanten white und rosy2 von Drosophila melanogaster. *Journal of Insect Physiology* 7, 5–31. doi:10.1016/0022-1910(61)90041-5

Labbé, R., Caveney, S., Donly, C., 2010. Genetic analysis of the xenobiotic resistance-associated ABC gene subfamilies of the Lepidoptera. *Insect Mol Biol* 20, 243–256. doi:10.1111/j.1365-2583.2010.01064.x

Lafont, R., Dauphin-Villemant, C., Warren, J.T., Rees, H., 2012. Ecdysteroid Chemistry and Biochemistry, in: *Insect Endocrinology*. Academic Press, pp. 106–176. doi:10.1016/B978-0-12-384749-2.10004-4

Lagueux, M., Hoffmann, J.A., Goltzené, F., Kappler, C., *et al.*, 1984. Ecdysteroids in Ovaries and Embryos of Locusta migratoria, in: *Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones*, Proceedings in Life Sciences. Springer, Berlin, Heidelberg, Berlin, Heidelberg, pp. 168–180. doi:10.1007/978-3-642-69922-1_18

Lam, G., Hall, B.L., Bender, M., Thummel, C.S., 1999. DHR3 Is Required for the Prepupal–Pupal Transition and Differentiation of Adult Structures during Drosophila Metamorphosis. *Developmental biology* 212, 204–216. doi:10.1006/dbio.1999.9343

Lancaster, L.T., 2020. Host use diversification during range shifts shapes global variation in Lepidopteran dietary breadth. *Nature Ecology & Evolution 2017 1:4* 112, 1–7. doi:10.1038/s41559-020-1199-1

Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi:10.1093/bioinformatics/btu531

Lavrynenko, O., Rodenfels, J., Carvalho, M., Dye, N.A., *et al.*, 2015. The ecdysteroidome of Drosophila: influence of diet and development. *Development* 142, 3758–3768. doi:10.1242/dev.124982

Leader, D.P., Krause, S.A., Pandit, A., Davies, S.A., *et al.*, 2018. FlyAtlas 2: a new version of the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucl Acids Res* 46, D809–D815. doi:10.1093/nar/gkx976

Leal, W.S., 2013. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual review of entomology* 58, 373–391. doi:10.1146/annurev-ento-120811-153635

Legal, L., David, J.R., Jallon, J.M., 1992. Toxicity and attraction effects produced by Morinda citrifolia fruits on the Drosophila melanogaster complex of species. *Chemoecology* 3, 125–129. doi:10.1007/BF01370140

Lespinet, O., Labedan, B., 2005. Orphan Enzymes? *Science* 307, 42–42. doi:10.1126/science.307.5706.42a

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* 12, 1048–1059. doi:10.1101/gr.174302

Li, D., Zhang, J., 2014. Diet Shapes the Evolution of the Vertebrate Bitter Taste Receptor Gene Repertoire. *Mol. Biol. Evol.* 31, 303–309. doi:10.1093/molbev/mst219

Li, M., Au, L.Y.C., Douglah, D., Chong, A., *et al.*, 2017a. Generation of heritable germline mutations in the jewel wasp Nasonia vitripennis using CRISPR/Cas9. *Sci. Rep.* 7, 1–7. doi:10.1038/s41598-017-00990-3

Li, S., Yu, X., Feng, Q., 2019. Fat Body Biology in the Last Decade. *Annu Rev Entomol* 64, 315–333. doi:10.1146/annurev-ento-011118-112007

Li, X., Bin Zhu, Gao, X., Liang, P., 2017b. Over-expression of UDP–glycosyltransferase gene UGT2B17 is involved in chlorantraniliprole resistance in Plutella xylostella (L.). *Pest. Manag. Sci.* 73, 1402–1409. doi:10.1002/ps.4469

Li, X., Schuler, M.A., Berenbaum, M.R., 2007. Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics. *Annual review of entomology* 52, 231–253. doi:10.1146/annurev.ento.51.110104.151104

Li, X., Zhuo, R., Tiong, S., Di Cara, F., *et al.*, 2013. The Smc5/Smc6/MAGE Complex Confers Resistance to Caffeine and Genotoxic Stress in Drosophila melanogaster. *PLoS ONE* 8, e59866. doi:10.1371/journal.pone.0059866

Li, Z., Ge, X., Ling, L., Zeng, B., *et al.*, 2014. CYP18A1 regulates tissue-specific steroid hormone inactivation in Bombyx mori. *Insect biochemistry and molecular Biology* 54, 33–41. doi:10.1016/j.ibmb.2014.08.007

Li, Z., You, L., Zeng, B., Ling, L., *et al.*, 2015. Ectopic expression of ecdysone oxidase impairs tissue degeneration in Bombyx mori. *Proceedings of the Royal Society B: Biological Sciences* 282, 20150513. doi:10.1098/rspb.2015.0513

Li-Kroeger, D., Kanca, O., Lee, P.-T., Cowan, S., *et al.*, 2018. An expanded toolkit for gene tagging based on MiMIC and scarless CRISPR tagging in Drosophila. *eLife* 7, e1002472. doi:10.7554/eLife.38709

Liu, C., Wang, J.-L., Zheng, Y., Xiong, E.-J., *et al.*, 2014. Wolbachia-induced paternal defect in Drosophila is likely by interaction with the juvenile hormone pathway. *Insect biochemistry and molecular Biology* 49, 49–58. doi:10.1016/j.ibmb.2014.03.014

Liu, Z., Chen, Y., Wang, D., Wang, S., *et al.*, 2010. Distinct Presynaptic and Postsynaptic Dismantling Processes of Drosophila Neuromuscular Junctions during Metamorphosis. *Journal of Neuroscience* 30, 11624–11634. doi:10.1523/JNEUROSCI.0410-10.2010

Lloyd, D.L., Toegel, M., Fulga, T.A., Wilkie, A.O.M., 2018. The Drosophila homologue of MEGF8 is essential for early development. *Sci. Rep.* 8, 1–10. doi:10.1038/s41598-018-27076-y

Logan-Garbisch, T., Bortolazzo, A., Luu, P., Ford, A., *et al.*, 2015. Developmental Ethanol Exposure Leads to Dysregulation of Lipid Metabolism and Oxidative Stress in Drosophila. *G3 (Bethesda)* 5, 49–59. doi:10.1534/g3.114.015040

Loppin, B., Dubruille, R., Horard, B., 2015. The intimate genetics of Drosophila fertilization. *Open Biology* 5, 150076. doi:10.1098/rsob.150076

Loveall, B.J., Deitcher, D.L., 2010. The essential role of bursicon during Drosophila development. *BMC Developmental Biology* 10, 92. doi:10.1186/1471-213X-10-92

Low, W.Y., Feil, S.C., Ng, H.L., Gorman, M.A., *et al.*, 2010. Recognition and Detoxification of the Insecticide DDT by Drosophila melanogaster Glutathione S-Transferase D1. *Journal of molecular biology* 399, 358–366. doi:10.1016/j.jmb.2010.04.020

Low, W.Y., Ng, H.L., Morton, C.J., Parker, M.W., *et al.*, 2007. Molecular Evolution of Glutathione S-Transferases in the Genus Drosophila. *Genetics* 177, 1363–1375. doi:10.1534/genetics.107.075838

Lozano, R., Thompson, M.J., Svoboda, J.A., Lusby, W.R., 1989. Profiles of free and conjugated ecdysteroids and ecdysteroid acids during pupal-adult development of Manduca sexta. *Archives of Insect Biochemistry and Physiology* 12, 63–74. doi:10.1002/arch.940120106

Lund, A.H., Duch, M., Pedersen, F.S., 1996. Increased cloning efficiency by temperature-cycle ligation. *Nucl Acids Res* 24, 800–801. doi:10.1093/nar/24.4.800

Lundin, D., Poole, A.M., Sjöberg, B.-M., Högbom, M., 2012. Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *Journal of Biological Chemistry* 287, 20565–20575. doi:10.1074/jbc.M112.367458

Lynch, J.A., 2015. The Expanding Genetic Toolbox of the Wasp Nasonia vitripennis and Its Relatives. *Genetics* 199, 897–904. doi:10.1534/genetics.112.147512

Lynch, M., 2012. The Evolution of Multimeric Protein Assemblages. *Mol. Biol. Evol.* 29, 1353–1366. doi:10.1093/molbev/msr300

Ma, W., Zhang, Z., Peng, C., Wang, X., *et al.*, 2012. Exploring the Midgut Transcriptome and Brush Border Membrane Vesicle Proteome of the Rice Stem Borer, Chilo suppressalis (Walker). *PLoS ONE* 7, e38151. doi:10.1371/journal.pone.0038151

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., *et al.*, 2012. The Drosophila melanogaster Genetic Reference Panel. *Nature* 482, 173–178. doi:10.1038/nature10811

Mackenzie, P., Bock, K.W., Burchell, B., Guillemette, C., *et al.*, 2005. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenetics and Genomics* 15, 677–685. doi:10.1097/01.fpc.0000173483.13689.56

Maddison, W.P., FitzJohn, R.G., 2015. The Unsolved Challenge to Phylogenetic Correlation Tests for Categorical Characters. *Systematic Biol.* 64, 127–136. doi:10.1093/sysbio/syu070

Maddison, W.P., Maddison, D.R., 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61 [WWW Document]. *http://www.mesquiteproject.org*. URL (accessed 4.29.20).

Maddrell, S.H.P., Gardiner, B.O.C., 1976. Excretion of alkaloids by malpighian tubules of insects. *Journal of Experimental Biology* 64, 267–281.

Magwire, M.M., Bayer, F., Webster, C.L., Cao, C., *et al.*, 2011. Successive Increases in the Resistance of Drosophila to Viral Infection through a Transposon Insertion Followed by a Duplication. *PLOS Genetics* 7, e1002337. doi:10.1371/journal.pgen.1002337

Magwire, M.M., Fabian, D.K., Schweyen, H., Cao, C., *et al.*, 2012. Genome-Wide Association Studies Reveal a Simple Genetic Basis of Resistance to Naturally Coevolving Viruses in Drosophila melanogaster. *PLOS Genetics* 8, e1003057. doi:10.1371/journal.pgen.1003057

Mahapatra, C.T., Bond, J., Rand, D.M., Rand, M.D., 2010. Identification of Methylmercury Tolerance Gene Candidates in Drosophila. *Toxicol Sci* 116, 225–238. doi:10.1093/toxsci/kfq097

Makka, T., Seino, A., Tomita, S., Fujiwara, H., *et al.*, 2002. A possible role of 20-hydroxyecdysone in embryonic development of the silkworm Bombyx mori. *Archives of Insect Biochemistry and Physiology* 51, 111–120. doi:10.1002/arch.10055

Makki, R., Cinnamon, E., Gould, A.P., 2013. The development and functions of oenocytes. *Annu Rev Entomol* 59, 405–425. doi:10.1146/annurev-ento-011613-162056

Malhotra, J.D., Kaufman, R.J., 2007. Endoplasmic reticulum stress and oxidative stress: a vicious cycle or a double-edged sword? *Antioxid. Redox Signal.* 9, 2277–2293. doi:10.1089/ars.2007.1782

Malki, A.M., Gentry, J., Evans, S.C., 2006. Differential effect of selected methylxanthine derivatives on radiosensitization of lung carcinoma cells. *Experimental oncology* 28, 16–24.

Mallott, M., Hamm, S., Troczka, B.J., Randall, E., *et al.*, 2019. A flavin-dependent monooxgenase confers resistance to chlorantraniliprole in the diamondback moth, Plutella xylostella. *Insect biochemistry and molecular Biology* 115, 103247. doi:10.1016/j.ibmb.2019.103247

Mamiya, Y., Sonobe, H., Yoshida, K., Hara, N., *et al.*, 1995. Occurrence of 3-Epi-22-Deoxy-20-Hydroxyecdysone and Its Phosphoric Ester in Diapause Eggs of the Silkworm, Bombyx-Mori. *Experientia* 51, 363–367.

Manseau, L., Baradaran, A., Brower, D., Budhu, A., *et al.*, 1997. GAL4 enhancer traps expressed in the embryo, larval brain, imaginal discs, and ovary of drosophila. *Developmental Dynamics* 209, 310–322. doi:10.1002/(SICI)1097-0177(199707)209:3<310::AID-AJA6>3.0.CO;2-L

Mao, W., Schuler, M.A., Berenbaum, M.R., 2017. Disruption of quercetin metabolism by fungicide affects energy production in honey bees (Apis mellifera). *PNAS* 114, 2538–2543. doi:10.1073/pnas.1614864114

Marais, Des, D.L., Rausher, M.D., 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454, 762–765. doi:10.1038/nature07092

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., *et al.*, 2015. CDD: NCBI's conserved domain database. *Nucl Acids Res* 43, D222–6. doi:10.1093/nar/gku1221

Markow, T.A., O'Grady, P.M., 2007. Drosophila Biology in the Genomic Age. *Genetics* 177, 1269–1276. doi:10.1534/genetics.107.074112

Maróy, P., Kaufmann, G., Dübendorfer, A., 1988. Embryonic ecdysteroids of Drosophila melanogaster. *Journal of Insect Physiology* 34, 633–637. doi:10.1016/0022-1910(88)90071-6

Marriage, T.N., King, E.G., Long, A.D., Macdonald, S.J., 2014. Fine-Mapping Nicotine Resistance Loci in Drosophila Using a Multiparent Advanced Generation Inter-Cross Population. *Genetics* 198, 45–57. doi:10.1534/genetics.114.162107

Martin-Creuzburg, D., Westerlund, S.A., Hoffmann, K.H., 2007. Ecdysteroid levels in Daphnia magna during a molt cycle: Determination by radioimmunoassay (RIA) and liquid chromatography-mass spectrometry (LC-MS). *Gen Comp Endocrinol* 151, 6–6. doi:10.1016/j.ygcen.2006.11.015

Matsushima, D., Kasahara, R., Matsuno, K., Aoki, F., *et al.*, 2019. Involvement of Ecdysone Signaling in the Expression of the doublesex Gene during Embryonic Development in the Silkworm, Bombyx mori. *Sex Dev* 13, 151–163. doi:10.1159/000502361

Matzkin, L.M., 2012. Population transcriptomics of cactus host shifts in Drosophila mojavensis. *Mol Ecol* 21, 2428–2439. doi:10.1111/j.1365-294X.2012.05549.x

Maxwell, R.A., Welch, W.H., Schooley, D.A., 2002a. Juvenile hormone diol kinase. I. Purification, characterization, and substrate specificity of juvenile hormone-selective diol kinase from Manduca sexta. *J. Biol. Chem.* 277, 21874–21881. doi:10.1074/jbc.M201510200

Maxwell, R.A., Welch, W.H., Horodyski, F.M., Schegg, K.M., *et al.*, 2002b. Juvenile hormone diol kinase. II. Sequencing, cloning, and molecular modeling of juvenile hormone-selective diol kinase from Manduca sexta. *J. Biol. Chem*. 277, 21882–21890. doi:10.1074/jbc.M201509200

Mayer, F., Mayer, N., Chinn, L., Pinsonneault, R.L., *et al.*, 2009. Evolutionary Conservation of Vertebrate Blood–Brain Barrier Chemoprotective Mechanisms in Drosophila. *Journal of Neuroscience* 29, 3538–3550. doi:10.1523/JNEUROSCI.5564-08.2009

Mazina, M.Y., Nikolenko, J.V., Fursova, N.A., Nedil'ko, P.N., *et al.*, 2015. Early-late genes of the ecdysone cascade as models for transcriptional studies. *Cell Cycle* 14, 3593–3601. doi:10.1080/15384101.2015.1100772

McBride, C.S., Arguello, J.R., O'Meara, B.C., 2007. Five Drosophila Genomes Reveal Nonneutral Evolution and the Signature of Host Specialization in the Chemoreceptor Superfamily. *Genetics* 177, 1395–1416. doi:10.1534/genetics.107.078683

McDonnell, E., Strasser, K., Tsang, A., 2018. Manual Gene Curation and Functional Annotation, in: *Fungal Genomics*, Methods in Molecular Biology. Humana Press, New York, NY, New York, NY, pp. 185–208. doi:10.1007/978-1-4939-7804-5_16

McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H., *et al.*, 2004. Shared Transcriptional Signature in Caenorhabditis elegans Dauer Larvae and Long-lived daf-2 Mutants Implicates Detoxification System in Longevity Assurance. *Journal of Biological Chemistry* 279, 44533–44543. doi:10.1074/jbc.M406207200

McKenna, D.D., Scully, E.D., Pauchet, Y., Hoover, K., *et al.*, 2016. Genome of the Asian longhorned beetle (Anoplophora glabripennis), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. *Genome Biol* 17, 227. doi:10.1186/s13059-016-1088-8

Mebs, D., Reuss, E., Schneider, M., 2005. Studies on the cardenolide sequestration in African milkweed butterflies (Danaidae). *Toxicon* 45, 581–584. doi:10.1016/j.toxicon.2004.12.017

Melk, J.P., Govind, S., 1999. Developmental analysis of Ganaspis xanthopoda, a larval parasitoid of Drosophila melanogaster. *Journal of Experimental Biology* 202, 1885–1896.

Mellgren, E.M., Johnson, S.L., 2006. pyewacket, a new zebrafish fin pigment pattern mutant. *Pigment Cell Research* 19, 232–238. doi:10.1111/j.1600-0749.2006.00311.x

Mesquita, R.D., Vionette-Amaral, R.J., Lowenberger, C., Rivera-Pomar, R., *et al.*, 2015. Genome of Rhodnius prolixus, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *PNAS* 112, 14936–14941. doi:10.1073/pnas.1506226112

Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucl Acids Res* 46, W329–W337. doi:10.1093/nar/gky384

Miller, D.E., Cook, K.R., Hemenway, E.A., Fang, V., *et al.*, 2018. The Molecular and Genetic Characterization of Second Chromosome Balancers in Drosophila melanogaster. *G3 (Bethesda)* 8, 1161–1171. doi:10.1534/g3.118.200021

Miller, D.E., Cook, K.R., Kazemi, N.Y., Smith, C.B., *et al.*, 2016. Rare recombination events generate sequence diversity among balancer chromosomes in Drosophila melanogaster. *Proc Natl Acad Sci U S A* 113, E1352–E1361. doi:10.1073/pnas.1601232113

Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE), 2010* 1–8. doi:10.1109/GCE.2010.5676129

Minakhina, S., Steward, R., 2006. Melanotic Mutants in Drosophila: Pathways and Phenotypes. *Genetics* 174, 253–263. doi:10.1534/genetics.106.061978

Mirth, C., 2005. Ecdysteroid control of metamorphosis in the differentiating adult leg structures of Drosophila melanogaster. *Developmental biology* 278, 163–174. doi:10.1016/j.ydbio.2004.10.026

Mirth, C.K., Truman, J.W., Riddiford, L.M., 2009. The Ecdysone receptor controls the post-critical weight switch to nutrition-independent differentiation in Drosophila wing imaginal discs. *Development* 136, 2345–2353. doi:10.1242/dev.032672

Misof, B., Liu, S., Meusemann, K., Peters, R.S., *et al.*, 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767. doi:10.1126/science.1257570

Misra, J.R., Horner, M.A., Lam, G., Thummel, C.S., 2011. Transcriptional regulation of xenobiotic detoxification in Drosophila. *Genes & Development* 25, 1796–1806. doi:10.1101/gad.17280911

Misra, J.R., Lam, G., Thummel, C.S., 2013. Constitutive activation of the Nrf2/Keap1 pathway in insecticide-resistant strains of Drosophila. *Insect biochemistry and molecular Biology*. doi:10.1016/j.ibmb.2013.09.005

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., *et al.*, 2014. The InterPro protein families database: the classification resource after 15 years. *Nucl Acids Res* 43, D213–D221. doi:10.1093/nar/gku1243

Mitchell, H.K., Chen, P.S., Hadorn, E., 1960. Tyrosine phosphate on paper chromatograms of Drosophila melanogaster. *Experientia* 16, 410–411. doi:10.1007/BF02178836

Mitchell, J.A., McCray, A.T., Bodenreider, O., 2003. From phenotype to genotype: issues in navigating the available information resources. *Methods of information in medicine* 42, 557–563. doi:10.1267/meth03050557

Mitchell, S.C., 2015. Xenobiotic conjugation with phosphate – a metabolic rarity. *Xenobiotica* 46, 743–756. doi:10.3109/00498254.2015.1109161

Mithöfer, A., Boland, W., 2012. Plant Defense Against Herbivores: Chemical Aspects. *Annual Review of Plant Biology* 63, 431–450. doi:10.1146/annurev-arplant-042110-103854

Modde, J.-F., Lafont, R., Hoffmann, J.A., 1984. Ecdysone metabolism in Locusta migratoria larvae and adults. *International Journal of Invertebrate Reproduction and Development* 7, 161–183. doi:10.1080/01688170.1984.10510087

Mody, K., Unsicker, S.B., Linsenmair, K.E., 2007. Fitness Related Diet-Mixing by Intraspecific Host-Plant-Switching of Specialist Insect Herbivores. *Ecology* 88, 1012–1020. doi:10.1890/06-1338

Moeller, M.E., Danielsen, E.T., Herder, R., O'Connor, M.B., *et al.*, 2013. Dynamic feedback circuits function as a switch for shaping a maturation-inducing steroid pulse in Drosophila. *Development* 140, 4730–4739. doi:10.1242/dev.099739

Mohit, P., Makhijani, K., Madhavi, M.B., Bharathi, V., *et al.*, 2006. Modulation of AP and DV signaling pathways by the homeotic gene Ultrabithorax during haltere development in Drosophila. *Developmental biology* 291, 356–367. doi:10.1016/j.ydbio.2005.12.022

Montgomery, S.L., Vorojeikina, D., Huang, W., Mackay, T.F.C., *et al.*, 2014. Genome-Wide Association Analysis of Tolerance to Methylmercury Toxicity in Drosophila Implicates Myogenic and Neuromuscular Developmental Pathways. *PLoS ONE* 9, e110375. doi:10.1371/journal.pone.0110375

Morales-Cruz, A., Amrine, K.C.H., Blanco-Ulate, B., Lawrence, D.P., *et al.*, 2015. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics* 16, 469. doi:10.1186/s12864-015-1624-z

Moreau, C.S., Bell, C.D., Vila, R., Archibald, S.B., *et al.*, 2006. Phylogeny of the Ants: Diversification in the Age of Angiosperms. *Science* 312, 101–104. doi:10.1126/science.1124891

Moribayashi, A., Kurahashi, H., Ohtaki, T., 1985. Comparative studies on ecdysone metabolism between mature larvae and pharate pupae in the fleshfly, Sarcophaga peregrina. *Archives of Insect Biochemistry and Physiology* 2, 237–250. doi:10.1002/arch.940020303

Moribayashi, A., Ohtaki, T., 1980. Inactivation and reactivation of 20-hydroxyecdysone during pupal-adult development of the flesh fly, Sarcophaga peregrina. *JJMSB* 33, 189–201. doi:10.7883/yoken1952.33.189

Morozova, T.V., Anholt, R.R.H., Mackay, T.F.C., 2006. Transcriptional response to alcohol exposure in Drosophila melanogaster. *Genome Biol* 7, R95. doi:10.1186/gb-2006-7-10-r95

Morozova, T.V., Huang, W., Pray, V.A., Whitham, T., *et al.*, 2015. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult drosophila. *BMC Genomics* 16, 539. doi:10.1186/s12864-015-2064-5

Mudgal, R., Sandhya, S., Chandra, N., 2015. De-DUFing the DUFs: Deciphering distant evolutionary relationships of Domains of Unknown Function using sensitive homology detection methods. *Biol Direct* 10, 405. doi:10.1186/s13062-015-0069-2

Mukherjee, D., Mukherjee, A., Ghosh, T.C., 2016. Evolutionary Rate Heterogeneity of Primary and Secondary Metabolic Pathway Genes in Arabidopsis thaliana. *Genome Biology and Evolution* 8, 17–28. doi:10.1093/gbe/evv217

Murali, T., Pacifico, S., Yu, J., Guest, S., *et al.*, 2011. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucl Acids Res* 39, D736–43. doi:10.1093/nar/gkq1092

Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology* 247, 536–540. doi:10.1016/S0022-2836(05)80134-2

Mustard, J.A., 2014. The buzz on caffeine in invertebrates: effects on behavior and molecular mechanisms. *Cellular and molecular life sciences : CMLS* 71, 1375–1382. doi:10.1007/s00018-013-1497-8

Nagata, M., Tsuchida, K., Shimizu, K., Narumi, Y., 1987. Physiological aspects of nm-g mutant: An ecdysteroid-deficient mutant of the silkworm, Bombyx mori. *Journal of Insect Physiology* 33, 723–727. doi:10.1016/0022-1910(87)90057-6

Nagy, A., Hegyi, H., Farkas, K., Tordai, H., *et al.*, 2008. Identification and correction of abnormal, incomplete and mispredicted proteins in public databases. *BMC bioinformatics* 9, 353. doi:10.1186/1471-2105-9-353

Najarro, M.A., Hackett, J.L., Macdonald, S.J., 2017. Loci Contributing to Boric Acid Toxicity in Two Reference Populations of Drosophila melanogaster. *G3 (Bethesda)* 7, 1631–1641. doi:10.1534/g3.117.041418

Najarro, M.A., Hackett, J.L., Smith, B.R., Highfill, C.A., *et al.*, 2015. Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. *PLOS Genetics* 11, e1005663. doi:10.1371/journal.pgen.1005663

Namiki, T., Niwa, R., Sakudoh, T., Shirai, K.-I., *et al.*, 2005. Cytochrome P450 CYP307A1/Spook: A regulator for ecdysone synthesis in insects. *Biochem Biophys Res Commun* 337, 367–374. doi:10.1016/j.bbrc.2005.09.043

Nascimento, F.F., Reis, dos, M., Yang, Z., 2017. A biologist's guide to Bayesian phylogenetic analysis. *Nature Ecology & Evolution 2017 1:4* 1, 1446–1454. doi:10.1038/s41559-017-0280-x

Nathanson, J.A., 1984. Caffeine and related methylxanthines: possible naturally occurring pesticides. *Science* 226, 184–187. doi:10.1126/science.6207592

Neafsey, D.E., Waterhouse, R.M., Abai, M.R., Aganezov, S.S., *et al.*, 2015. Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. *Science* 347, 1258522. doi:10.1126/science.1258522

Nehlig, A., 1999. Are we dependent upon coffee and caffeine? A review on human and animal data. *Neuroscience & Biobehavioral Reviews* 23, 563–576. doi:10.1016/S0149-7634(98)00050-5

Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39, 121–152. doi:10.1146/annurev.genet.39.073003.112240

Nelson, D.R., 2009. The Cytochrome P450 Homepage. *Hum Genomics* 4, 1. doi:10.1186/1479-7364-4-1-59

Nelson, D.R., 2006. Cytochrome P450 Nomenclature, 2004, in: *Cytochrome P450 Protocols*. Humana Press, Totowa, NJ, New Jersey, pp. 1–10. doi:10.1385/1-59259-998-2:1

Neuman, S.D., Bashirullah, A., 2018. Reconsidering the Passive Diffusion Model of Steroid Hormone Cellular Entry. *Developmental Cell* 47, 261–262. doi:10.1016/j.devcel.2018.10.022

Neumann, R., Peter, H.H., 1987. Insecticidal organophosphates: Nature made them first. *Experientia* 43, 1235–1237. doi:10.1007/BF01945541

Newcomb, R.D., Gleeson, D.M., Yong, C.G., Russell, R.J., *et al.*, 2005. Multiple Mutations and Gene Duplications Conferring Organophosphorus Insecticide Resistance Have Been Selected at the Rop-1 Locus of the Sheep Blowfly, Lucilia cuprina. *J Mol Evol* 60, 207–220. doi:10.1007/s00239-004-0104-x

Ngah, W.Z., Smith, J.N., 1983. Acidic conjugate of phenols in insects; glucoside phosphate and glucoside sulphate derivatives. *Xenobiotica* 13, 383–389. doi:10.3109/00498258309052276

Nguyen, L.-T., Schmidt, H.A., Haeseler, von, A., Minh, B.Q., 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi:10.1093/molbev/msu300

Ni, J.-Q., Zhou, R., Czech, B., Liu, L.-P., *et al.*, 2011. A genome-scale shRNA resource for transgenic RNAi in Drosophila. *Nat Meth* 8, 405–407. doi:10.1038/nmeth.1592

Nielsen, M.T., Klejnstrup, M.L., Rohlfs, M., Anyaogu, D.C., *et al.*, 2013. Aspergillus nidulans Synthesize Insect Juvenile Hormones upon Expression of a Heterologous Regulatory Protein and in Response to Grazing by Drosophila melanogaster Larvae. *PLoS ONE* 8, e73369. doi:10.1371/journal.pone.0073369.t003

Niepielko, M.G., Marmion, R.A., Kim, K., Luor, D., *et al.*, 2014. Chorion Patterning: A Window into Gene Regulation and Drosophila Species' Relatedness. *Mol. Biol. Evol.* 31, 154–164. doi:10.1093/molbev/mst186

Nigsch, J., Graf, U., Würgler, F.E., 1977. Caffeine toxicity in Drosophila strains having different MMS sensitivities. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 43, 57–63. doi:10.1016/0027-5107(77)90131-2

Niwa, R., Matsuda, T., Yoshiyama, T., Namiki, T., *et al.*, 2004. CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of Bombyx and Drosophila. *J. Biol. Chem.* 279, 35942–35949. doi:10.1074/jbc.M404514200

Niwa, R., Namiki, T., Ito, K., Shimada-Niwa, Y., *et al.*, 2010. Non-molting glossy/shroud encodes a short-chain dehydrogenase/reductase that functions in the "Black Box" of the ecdysteroid biosynthesis pathway. *Development* 137, 1991–1999. doi:10.1242/dev.045641

Niwa, R., Niwa, Y.S., 2014. Enzymes for ecdysteroid biosynthesis: their biological functions in insects and beyond. *Bioscience, biotechnology, and biochemistry* 78, 1283–1292. doi:10.1080/09168451.2014.942250

Niwa, R., Sakudoh, T., Matsuya, T., Namiki, T., *et al.*, 2011. Expressions of the cytochrome P450 monooxygenase gene *Cyp4g1* and its homolog in the prothoracic glands of the fruit fly *Drosophila melanogaster* (Diptera: Drosophilidae) and the silkworm *Bombyx mori* (Lepidoptera: Bombycidae). *Applied Entomology and Zoology* 46(4), 533–543. doi: 10.1007/s13355-011-0074-6

Nobler, J.D., Camp, M.J., Crowell, M.M., Shipley, L.A., *et al.*, 2018. Preferences of Specialist and Generalist Mammalian Herbivores for Mixtures Versus Individual Plant Secondary Metabolites. *Journal of chemical ecology* 45, 74–85. doi:10.1007/s10886-018-1030-5

Novoselov, A., Becker, T., Pauls, G., Reuß, von, S.H., *et al.*, 2015. Spodoptera littoralis detoxifies neurotoxic 3-nitropropanoic acid by conjugation with amino acids. *Insect biochemistry and molecular Biology* 63, 97–103. doi:10.1016/j.ibmb.2015.05.013

Nüsslein-Volhard, C., Wieschaus, E., Kluding, H., 1984. Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. *Wilhelm Roux' Archiv* 193, 267–282. doi:10.1007/BF00848156

Nygaard, S., Hu, H., Li, C., Schiøtt, M., *et al.*, 2015. Reciprocal genomic evolution in the ant-fungus agricultural symbiosis. *Nature Communications* 7, 12233–12233. doi:10.1038/ncomms12233

O'Donnell, M.J., 2009. Too much of a good thing: how insects cope with excess ions or toxins in the diet. *Journal of Experimental Biology* 212, 363–372. doi:10.1242/jeb.023739

Oakeshott, J.G., Claudianos, C., Campbell, P.M., Newcomb, R.D., *et al.*, 2005. Biochemical Genetics and Genomics of Insect Esterases. *Comprehensive Molecular Insect Science*, Comprehensive Molecular Insect Science 309–381. doi:10.1016/B0-44-451924-6/00073-9

Ode, P.J., 2006. Plant chemistry and natural enemy fitness: effects on herbivore and natural enemy interactions. *Annu Rev Entomol* 51, 163–185. doi:10.1146/annurev.ento.51.110104.151110

Ohta, T., 2001. Gene Families: Multigene Families and Superfamilies. John Wiley & Sons, Ltd, Chichester, UK. doi:10.1002/9780470015902.a0005126.pub2

Okamoto, N., Viswanatha, R., Bittar, R., Li, Z., *et al.*, 2018. A Membrane Transporter Is Required for Steroid Hormone Uptake in Drosophila. *Developmental Cell*. doi:10.1016/j.devcel.2018.09.012

Okamoto, N., Yamanaka, N., 2020. Steroid Hormone Entry into the Brain Requires a Membrane Transporter in Drosophila. *Current Biology* 30, 359–366.e3. doi:10.1016/j.cub.2019.11.085

Olsen, L.R., Gabel-Jensen, C., Nielsen, P.A., Hansen, S.H., *et al.*, 2014. Identification of a Functional Homolog of the Mammalian CYP3A4 in Locusts. *Drug Metabolism and Disposition* 42, 1153–1162. doi:10.1124/dmd.114.057430

Olsen, L.R., Gabel-Jensen, C., Wubshet, S.G., Kongstad, K.T., *et al.*, 2015. Characterization of midazolam metabolism in locusts: the role of a CYP3A4-like enzyme in the formation of 1'-OH and 4-OH midazolam. *Xenobiotica* 46, 99–107. doi:10.3109/00498254.2015.1051604

Omiecinski, C.J., Heuvel, J.P.V., Perdew, G.H., Peters, J.M., 2011. Xenobiotic Metabolism, Disposition, and Regulation by Receptors: From Biochemical Phenomenon to Predictors of Major Toxicities. *Toxicol Sci* 120, S49–S75. doi:10.1093/toxsci/kfq338

Ono, H., 2014. Ecdysone differentially regulates metamorphic timing relative to 20-hydroxyecdysone by antagonizing juvenile hormone in Drosophila melanogaster. *Developmental biology* 391, 32–42. doi:10.1016/j.ydbio.2014.04.004

Ono, H., Morita, S., Asakura, I., Nishida, R., 2012. Conversion of 3-oxo steroids into ecdysteroids triggers molting and expression of 20E-inducible genes in Drosophila melanogaster. *Biochem Biophys Res Commun* 421, 561–566. doi:10.1016/j.bbrc.2012.04.045

Ono, H., Rewitz, K.F., Shinoda, T., Itoyama, K., *et al.*, 2006. Spook and Spookier code for stage-specific components of the ecdysone biosynthetic pathway in Diptera. *Developmental biology* 298, 555–570. doi:10.1016/j.ydbio.2006.07.023

Oostra, V., Mateus, A.R.A., van der Burg, K.R.L., Piessens, T., *et al.*, 2015. Ecdysteroid Hormones Link the Juvenile Environment to Alternative Adult Life Histories in a Seasonal Insect. *American Naturalist* 184, E79–E92. doi:10.1086/677260

Ormerod, K.G., LePine, O.K., Abbineni, P.S., Bridgeman, J.M., *et al.*, 2017. Drosophila development, physiology, behavior, and lifespan are influenced by altered dietary composition. *Fly* 11, 153–170. doi:10.1080/19336934.2017.1304331

Orr, T.J., Kitanovic, S., Schramm, K.M., Skopec, M.M., *et al.*, 2020. Strategies in herbivory by mammals revisited: The role of liver metabolism in a juniper specialist (Neotoma stephensi) and a generalist (Neotoma albigula). *Mol Ecol* 96, 207. doi:10.1111/mec.15431

Ou, Q., Magico, A., King-Jones, K., 2011. Nuclear Receptor DHR4 Controls the Timing of Steroid Hormone Pulses During Drosophila Development. *PLoS Biol* 9, e1001160. doi:10.1371/journal.pbio.1001160

Ou, Q., Zeng, J., Yamanaka, N., Brakken-Thal, C., *et al.*, 2016. The Insect Prothoracic Gland as a Model for Steroid Hormone Biosynthesis and Regulation. *Cell Reports* 16, 247–262. doi:10.1016/j.celrep.2016.05.053

O'Donnell, S., Baudier, K., Marenda, D.R., 2016. Non-Nutritive Polyol Sweeteners Differ in Insecticidal Activity When Ingested by Adult Drosophila melanogaster(Diptera: Drosophilidae). *J Insect Sci* 16, 47. doi:10.1093/jisesa/iew031

O'Reilly, D.R., Miller, L.K., 1989. A Baculovirus Blocks Insect Molting by Producing Ecdysteroid UDP-Glucosyl Transferase. *Science* 245, 1110–1112. doi:10.1126/science.2505387

Pais, F.S.-M., Ruy, P. de, Oliveira, G., Coimbra, R., 2014. Assessing the efficiency of multiple sequence alignment programs. *Algorithms Mol Biol* 9, 4. doi:10.1186/1748-7188-9-4

Pais, I.S., Valente, R.S., Sporniak, M., Teixeira, L., 2018. Drosophila melanogaster establishes a species-specific mutualistic interaction with stable gut-colonizing bacteria. *PLoS Biol* 16, e2005710. doi:10.1371/journal.pbio.2005710

Pan, Y., Tian, F., Wei, X., Wu, Y., *et al.*, 2018. Thiamethoxam Resistance in Aphis gossypii Glover Relies on Multiple UDP-Glucuronosyltransferases. *Front. Physiol.* 9, 207. doi:10.3389/fphys.2018.00322

Papantonis, A., Swevers, L., Iatrou, K., 2015. Chorion Genes: A Landscape of Their Evolution, Structure, and Regulation. *Annu Rev Entomol* 60, 177–194. doi:10.1146/annurev-ento-010814-020810

Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi:10.1093/bioinformatics/bty633

Park, J.H., Schroeder, A.J., Helfrich-Forster, C., Jackson, F.R., *et al.*, 2003. Targeted ablation of CCAP neuropeptide-containing neurons of Drosophila causes specific defects in execution and circadian timing of ecdysis behavior. *Development* 130, 2645–2656. doi:10.1242/dev.00503

Parvy, J.-P., Napal, L., Rubin, T., Poidevin, M., *et al.*, 2012. Drosophila melanogaster Acetyl-CoA-Carboxylase Sustains a Fatty Acid–Dependent Remote Signal to Waterproof the Respiratory System. *PLOS Genetics* 8, e1002925. doi:10.1371/journal.pgen.1002925

Parvy, J.-P., Wang, P., Garrido, D., Maria, A., *et al.*, 2014. Forward and feedback regulation of cyclic steroid production in Drosophila melanogaster. *Development* 141, 3955–3965. doi:10.1242/dev.102020

Pasay, C., Arlian, L., Morgan, M., Gunning, R., *et al.*, 2009. The Effect of Insecticide Synergists on the Response of Scabies Mites to Pyrethroid Acaricides. *PLoS Negl Trop Dis* 3, e354. doi:10.1371/journal.pntd.0000354

Patel, D.S., Garza-Garcia, A., Nanji, M., McElwee, J.J., *et al.*, 2008. Clustering of Genetically Defined Allele Classes in the Caenorhabditis elegans DAF-2 Insulin/IGF-1 Receptor. *Genetics* 178, 931–946. doi:10.1534/genetics.107.070813

Paterson, R.R.M., Simmonds, M.S.J., Blaney, W.M., 1987. Mycopesticidal effects of characterized extracts of Penicillium isolates purified secondary metabolites (including mycotoxins) on Drosophila melanogaster and Spodoptora littoralis. *J Invertebr Pathol* 50, 124–133. doi:10.1016/0022-2011(87)90112-1

Pauchet, Y., Muck, A., Svatos, A., Heckel, D.G., 2009a. Chromatographic and electrophoretic resolution of proteins and protein complexes from the larval midgut microvilli of Manduca sexta. *Insect biochemistry and molecular Biology* 39, 467–474. doi:10.1016/j.ibmb.2009.05.001

Pauchet, Y., Wilkinson, P., van Munster, M., Augustin, S., *et al.*, 2009b. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle Chrysomela tremulae reveals new gene families in Coleoptera. *Insect biochemistry and molecular Biology* 39, 403–413. doi:10.1016/j.ibmb.2009.04.001

Pauchet, Y., Wilkinson, P., Vogel, H., Nelson, D.R., *et al.*, 2010. Pyrosequencing the Manduca sexta larval midgut transcriptome: messages for digestion, detoxification and defence. *Insect Mol Biol* 19, 61–75. doi:10.1111/j.1365-2583.2009.00936.x

Paul, F., Meng, Y., Roux, B., 2020. Identification of Druggable Kinase Target Conformations Using Markov Model Metastable States Analysis of apo-Abl. *Journal of Chemical Theory and Computation* 16, 1896–1912. doi:10.1021/acs.jctc.9b01158

Peisach, D., Gee, P., Kent, C., Xu, Z., 2003. The crystal structure of choline kinase reveals a eukaryotic protein kinase fold. *Structure* 11, 703–713. doi:10.1016/s0969-2126(03)00094-7

Penny, D., Hendy, M., 2004. Phylogenetics: Parsimony and Distance Methods, *Handbook of Statistical Genetics*. American Cancer Society, Chichester. doi:10.1002/0470022620.bbc14

Pentzold, S., Zagrobelny, M., Rook, F., Bak, S., 2014. How insects overcome two-component plant chemical defence: plant β-glucosidases as the main target for herbivore adaptation. *Biological Reviews* 89, 531–551. doi:10.1111/brv.12066

Peri, F., Roth, S., 2000. Combined activities of Gurken and decapentaplegic specify dorsal chorion structures of the Drosophila egg. *Development* 127, 841–850.

Perkins, M.C., Woods, H.A., Harrison, J.F., Elser, J.J., 2004. Dietary phosphorus affects the growth of larval Manduca sexta. *Archives of Insect Biochemistry and Physiology* 55, 153–168. doi:10.1002/arch.10133

Perry, C., Scanlan, J., Robin, C., 2019. Mining insect genomes for functionally affiliated genes. *Current Opinion in Insect Science* 31, 114–122. doi:10.1016/j.cois.2018.12.006

Peters, J.L., Cnudde, F., Gerats, T., 2003. Forward genetics and map-based cloning approaches. *Trends in Plant Science* 8, 484–491. doi:10.1016/j.tplants.2003.09.002

Peters, N.C., Thayer, N.H., Kerr, S.A., Tompa, M., *et al.*, 2013. Following the "tracks": Tramtrack69 regulates epithelial tube expansion in the Drosophila ovary through Paxillin, Dynamin, and the homeobox protein Mirror. *Developmental biology* 378, 154–169. doi:10.1016/j.ydbio.2013.03.017

Peters, R.S., Krogmann, L., Mayer, C., Donath, A., *et al.*, 2017. Evolutionary History of the Hymenoptera. *Current Biology* 27, 1013–1018. doi:10.1016/j.cub.2017.01.027

Peters, R.S., Niehuis, O., Gunkel, S., Bläser, M., *et al.*, 2018. Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea) reveals a history of rapid radiations, convergence, and evolutionary success. *Molecular Phylogenetics and Evolution* 120, 286–296. doi:10.1016/j.ympev.2017.12.005

Petersen, R.A., Zangerl, A.R., Berenbaum, M.R., Schuler, M.A., 2001. Expression of CYP6B1 and CYP6B3 cytochrome P450 monooxygenases and furanocoumarin metabolism in different tissues of Papilio polyxenes (Lepidoptera: Papilionidae). *Insect biochemistry and molecular Biology* 31, 679–690. doi:10.1016/s0965-1748(00)00174-0

Peterson, J.J., Beecher, G.R., Bhagwat, S.A., Dwyer, J.T., *et al.*, 2006. Flavanones in grapefruit, lemons, and limes: A compilation and review of the data from the analytical literature. *Journal of Food Composition and Analysis* 19, S74–S80. doi:10.1016/j.jfca.2005.12.009

Petrov, D.A., Hartl, D.L., 1998. High rate of DNA loss in the Drosophila melanogaster and Drosophila virilis species groups. *Mol. Biol. Evol.* 15, 293–302. doi:10.1093/oxfordjournals.molbev.a025926

Petrov, D.A., Hartl, D.L., 1997. Trash DNA is what gets thrown away: high rate of DNA loss in Drosophila. *Gene* 205, 279–289. doi:10.1016/S0378-1119(97)00516-7

Petrov, D.A., Lozovskaya, E.R., Hartl, D.L., 1996. High intrinsic rate of DNA loss in Drosophila. *Nature* 384, 346–349. doi:10.1038/384346a0

Petruccelli, E., Lark, A., Mrkvicka, J.A., Kitamoto, T., 2020. Significance of DopEcR, a G-protein coupled dopamine/ecdysteroid receptor, in physiological and behavioral response to stressors. *J. Neurogenet.* 265, 1–14. doi:10.1080/01677063.2019.1710144

Petruccelli, E., Li, Q., Rao, Y., Kitamoto, T., 2016. The Unique Dopamine/Ecdysteroid Receptor Modulates Ethanol-Induced Sedation in Drosophila. *Journal of Neuroscience* 36, 4647–4657. doi:10.1523/JNEUROSCI.3774-15.2016

Petryk, A., Warren, J.T., Marqués, G., Jarcho, M.P., *et al.*, 2003. Shade is the Drosophila P450 enzyme that mediates the hydroxylation of ecdysone to the steroid insect molting hormone 20-hydroxyecdysone. *Proc Natl Acad Sci U S A* 100, 13773–13778. doi:10.1073/pnas.2336088100

Pfeiffer, F., Oesterhelt, D., 2015. A manual curation strategy to improve genome annotation: application to a set of haloarchael genomes. *Life (Basel)* 5, 1427–1444. doi:10.3390/life5021427

Phalnikar, K., Kunte, K., Agashe, D., 2019. Disrupting butterfly caterpillar microbiomes does not impact their survival and development. *Proceedings of the Royal Society B* 286, 20192438. doi:10.1098/rspb.2019.2438

Phillips, M.D., Thomas, G.H., 2006. Brush border spectrin is required for early endosome recycling in Drosophila. *J Cell Sci* 119, 1361–1370. doi:10.1242/jcs.02839

Piper, M.D.W., Skorupa, D., Partridge, L., 2005. Diet, metabolism and lifespan in Drosophila. *Experimental Gerontology* 40, 857–862. doi:10.1016/j.exger.2005.06.013

Pirmohamed, M., Kitteringham, N.R., Park, B.K., 1994. The Role of Active Metabolites in Drug Toxicity. *Drug-Safety* 11, 114–144. doi:10.2165/00002018-199411020-00006

Pis, J., Girault, J.P., Grau, V., Lafont, R., 1995. Analysis of ecdysteroid conjugates: Chromatographic characterization of sulfates, phosphates and glucosides. *EJE* 92, 41–41.

Podar, M., Eads, J.R., Richardson, T.H., 2005. Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. *BMC Evolutionary Biology* 5, 42. doi:10.1186/1471-2148-5-42

Port, F., Bullock, S.L., 2016. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat Meth* 13, 852–854. doi:10.1038/nmeth.3972

Port, F., Strein, C., Stricker, M., Rauscher, B., *et al.*, 2020. A large-scale resource for tissue-specific CRISPR mutagenesis in Drosophila. *eLife* 9, 509. doi:10.7554/eLife.53865

Prince, V.E., Pickett, F.B., 2002. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* 3, 827–837. doi:10.1038/nrg928

Puinean, A.M., Foster, S.P., Oliphant, L., Denholm, I., *et al.*, 2010. Amplification of a Cytochrome P450 Gene Is Associated with Resistance to Neonicotinoid Insecticides in the Aphid Myzus persicae. *PLOS Genetics* 6, e1000999. doi:10.1371/journal.pgen.1000999

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., *et al.*, 2012. The Pfam protein families database. *Nucl Acids Res* 40, D290–D301. doi:10.1093/nar/gkr1065

Puntener, W., 1981. Manual for field trials in plant protection. CIBA-GEIGY, Basle, Switzerland.

Qiu, S., Adema, C.M., Lane, T., 2005. A computational study of off-target effects of RNA interference. *Nucl Acids Res* 33, 1834–1847. doi:10.1093/nar/gki324

Qiu, Y., Tittiger, C., Wicker-Thomas, C., Le Goff, G., *et al.*, 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A* 109, 14858–14863. doi:10.1073/pnas.1208650109

R Core Team, 2019. R: A Language and Environment for Statistical Computing.

Ramirez, M.S., Tolmasky, M.E., 2010. Aminoglycoside modifying enzymes. *Drug Resistance Updates* 13, 151–171. doi:10.1016/j.drup.2010.08.003

Ramsey, S.D., Ochoa, R., Bauchan, G., Gulbronson, C., *et al.*, 2019. Varroa destructor feeds primarily on honey bee fat body tissue and not hemolymph. *Proc Natl Acad Sci U S A* 116, 1792–1801. doi:10.1073/pnas.1818371116

Rand, M.D., Lowe, J.A., Mahapatra, C.T., 2012. Drosophila CYP6g1 and its human homolog CYP3A4 confer tolerance to methylmercury during development. *Toxicology* 300, 75–82. doi:10.1016/j.tox.2012.06.001

Rand, M.D., Vorojeikina, D., Peppriell, A., Gunderson, J., *et al.*, 2019. Drosophotoxicology: Elucidating Kinetic and Dynamic Pathways of Methylmercury Toxicity in a Drosophila Model. *Front. Genet.* 10, 1. doi:10.3389/fgene.2019.00666

Rane, R.V., Ghodke, A.B., Hoffmann, A.A., Edwards, O.R., *et al.*, 2019. Detoxifying enzyme complements and host use phenotypes in 160 insect species. *Current Opinion in Insect Science*. doi:10.1016/j.cois.2018.12.008

Rane, R.V., Oakeshott, J.G., Nguyen, T., Hoffmann, A.A., *et al.*, 2017. Orthonome – a new pipeline for predicting high quality orthologue gene sets applicable to complete and draft genomes. *BMC Genomics* 18, 360. doi:10.1186/s12864-017-4079-6

Rane, R.V., Walsh, T.K., Pearce, S.L., Jermiin, L.S., *et al.*, 2016. Are feeding preferences and insecticide resistance associated with the size of detoxifying enzyme families in insect herbivores? *Current Opinion in Insect Science* 13, 70–76. doi:10.1016/j.cois.2015.12.001

Redfern, C.P.F., 1983. Ecdysteroid synthesis by the ring gland of Drosophila melanogaster during late-larval, prepupal and pupal development. *Journal of Insect Physiology* 29, 65–71. doi:10.1016/0022-1910(83)90107-5

Rees, H.H., Isaac, R.E., 1984. Biosynthesis of Ovarian Ecdysteroid Phosphates and Their Metabolic Fate During Embryogenesis in Schistocerca gregaria, in: *Biosynthesis, Metabolism and Mode of Action of Invertebrate Hormones*, Proceedings in Life Sciences. Springer, Berlin, Heidelberg, Berlin, Heidelberg, pp. 181–195. doi:10.1007/978-3-642-69922-1_19

Regan, J.C., Brandão, A.S., Leitão, A.B., Dias, A.R.M., *et al.*, 2013. Steroid hormone signaling is essential to regulate innate immune cells and fight bacterial infection in Drosophila. *PLoS Pathog* 9, e1003720–e1003720. doi:10.1371/journal.ppat.1003720

Reis, dos, M., Thawornwattana, Y., Angelis, K., Telford, M.J., *et al.*, 2015. Uncertainty in the Timing of Origin of Animals and the Limits of Precision in Molecular Timescales. *Current Biology* 25, 2939–2950. doi:10.1016/j.cub.2015.09.066

Remnant, E.J., Good, R.T., Schmidt, J.M., Lumb, C., *et al.*, 2013. Gene duplication in the major insecticide target site, Rdl, in Drosophila melanogaster. *Proc Natl Acad Sci U S A* 110, 14705–14710. doi:10.1073/pnas.1311341110

Rewitz, K.F., O'Connor, M.B., Gilbert, L.I., 2007. Molecular evolution of the insect Halloween family of cytochrome P450s: phylogeny, gene organization and functional conservation. *Insect biochemistry and molecular Biology* 37, 741–753. doi:10.1016/j.ibmb.2007.02.012

Rewitz, K.F., Rybczynski, R., Warren, J.T., Gilbert, L.I., 2006. The Halloween genes code for cytochrome P450 enzymes mediating synthesis of the insect moulting hormone. *Biochem. Soc. Trans.* 34, 1256–1260. doi:10.1042/BST0341256

Rewitz, K.F., Yamanaka, N., O'Connor, M.B., 2010. Steroid hormone inactivation is required during the juvenile-adult transition in Drosophila. *Developmental Cell* 19, 895–902. doi:10.1016/j.devcel.2010.10.021

Rharrabe, K., Alla, S., Maria, A., Sayah, F., *et al.*, 2007. Diversity of detoxification pathways of ingested ecdysteroids among phytophagous insects. *Archives of Insect Biochemistry and Physiology* 65, 65–73. doi:10.1002/arch.20191

Rharrabe, K., Sayan, F., Lafont, R., 2009. Dietary effects of four phytoecdysteroids on growth and development of the Indian meal moth, Plodia interpunctella. *J Insect Sci* 10, 13–13. doi:10.1673/031.010.1301

Richards, S., Murali, S.C., 2015. Best Practices in Insect Genome Sequencing: What Works and What Doesn't. *Current Opinion in Insect Science* 7, 1–7. doi:10.1016/j.cois.2015.02.013

Richter, S., 2002. The Tetraconata concept: hexapod-crustacean relationships and the phylogeny of Crustacea. *Organisms Diversity & Evolution* 2, 217–237. doi:10.1078/1439-6092-00048

Riddiford, L.M., Truman, J.W., Mirth, C.K., Shen, Y.-C., 2010. A role for juvenile hormone in the prepupal development of Drosophila melanogaster. *Development* 137, 1117–1126. doi:10.1242/dev.037218

Ritz, C., Baty, F., Streibig, J.C., Gerhard, D., 2015. Dose-Response Analysis Using R. *PLoS ONE* 10, e0146021. doi:10.1371/journal.pone.0146021

Rivest, S., Forrest, J.R.K., 2020. Defence compounds in pollen: why do they occur and how do they affect the ecology and evolution of bees? *New Phytologist* 225, 1053–1064. doi:10.1111/nph.16230

Robin, C., Bardsley, L.M.J., Coppin, C., Oakeshott, J.G., 2009. Birth and Death of Genes and Functions in the β-Esterase Cluster of Drosophila. *J Mol Evol* 69, 10–21. doi:10.1007/s00239-009-9236-3

Robin, C., Battlay, P., Fournier-Level, A., 2019. What can genetic association panels tell us about evolutionary processes in insects? *Current Opinion in Insect Science* 31, 99–105. doi:10.1016/j.cois.2018.12.004

Robin, C., Daborn, P.J., Hoffmann, A.A., 2007. Fighting fly genes. *Trends in genetics: TIG* 23, 51–54. doi:10.1016/j.tig.2006.12.005

Robin, C., Russell, R.J., Medveczky, K.M., Oakeshott, J.G., 1996. Duplication and divergence of the genes of the alpha-esterase cluster of Drosophila melanogaster. *J Mol Evol* 43, 241–252. doi:10.1007/bf02338832

Robinson, G.S., Kitching, I.J., Beccaloni, G.W., Hernández, L.A.M., 2010. HOSTS - A Database of the World's Lepidopteran Hostplants.

Robinson, P.D., Morgan, E.D., Wilson, I.D., Lafont, R., 1987. The metabolism of ingested and injected [3H]ecdysone by final instar larvae of Heliothis armigera. *Physiological Entomology* 12, 321–330. doi:10.1111/j.1365-3032.1987.tb00757.x

Rogers, R.L., Hartl, D.L., 2011. Chimeric genes as a source of rapid evolution in Drosophila melanogaster. *Mol. Biol. Evol.* 29, 517–529. doi:10.1093/molbev/msr184

Rohlfs, M., 2013. Fungal secondary metabolite dynamics in fungus-grazer interactions: novel insights and unanswered questions. *Front. Microbiol.* 5, 788–788. doi:10.3389/fmicb.2014.00788

Rohlfs, M., Obmann, B., Petersen, R., 2005. Competition with filamentous fungi and its implication for a gregarious lifestyle in insects living on ephemeral resources. *Ecological Entomology* 30, 556–563. doi:10.1111/j.0307-6946.2005.00722.x

Roman, H., Ruzinski, M.M., 1990. Mechanisms of gene conversion in Saccharomyces cerevisiae. *Genetics* 124, 7–25.

Rost, B., 1999. Twilight zone of protein sequence alignments. *Protein Eng.* 12, 85–94. doi:10.1093/protein/12.2.85

Rountree, D.B., Nijhout, H.F., 1995. Hormonal control of a seasonal polyphenism in Precis coenia (Lepidoptera: Nymphalidae). *Journal of Insect Physiology* 41, 987–992. doi:10.1016/0022-1910(95)00046-W

Ruaud, A.-F., Lam, G., Thummel, C.S., 2009. The Drosophila nuclear receptors DHR3 and betaFTZ-F1 control overlapping developmental responses in late embryos. *Development* 137, 123–131. doi:10.1242/dev.042036

Rus, F., Flatt, T., Tong, M., Aggarwal, K., *et al.*, 2013. Ecdysone triggered PGRP-LC expression controls Drosophila innate immunity. *EMBO J* 32, 1626–1638. doi:10.1038/emboj.2013.100

Ryder, E., Ashburner, M., Bautista-Llacer, R., Drummond, J., *et al.*, 2007. The DrosDel Deletion Collection: A Drosophila Genomewide Chromosomal Deficiency Resource. *Genetics* 177, 615–629. doi:10.1534/genetics.107.076216

Saha, S., 2018. Long Range Sequencing and Validation of Insect Genome Assemblies, in: *Insect Genomics*, Methods and Protocols. Springer New York, New York, NY, pp. 33–44. doi:10.1007/978-1-4939-8775-7_4

Saito, J., Kimura, R., Kaieda, Y., Nishida, R., *et al.*, 2016. Characterization of candidate intermediates in the Black Box of the ecdysone biosynthetic pathway in Drosophila melanogaster: evaluation of molting activities on ecdysteroid-defective larvae. *Journal of Insect Physiology*. doi:10.1016/j.jinsphys.2016.09.012

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Sakurai, S., Warren, J.T., Gilbert, L.I., 1989. Mediation of ecdysone synthesis in Manduca sexta by a hemolymph enzyme. *Archives of Insect Biochemistry and Physiology* 10, 179–197. doi:10.1002/arch.940100303

Sakurai, S., Williams, C.M., 1989. Short-loop negative and positive feedback on ecdysone secretion by prothoracic gland in the tobacco hornworm, Manduca sexta. *Gen Comp Endocrinol* 75, 204–216. doi:10.1016/0016-6480(89)90072-5

Sanchez-Arroyo, H., Koehler, P.G., Valles, S.M., 2001. Effects of the synergists piperonyl butoxide and S,S,S-tributyl phosphorotrithioate on propoxur pharmacokinetics in Blattella germanica (Blattodea: Blattellidae). *J Econ Entomol* 94, 1209–1216. doi:10.1603/0022-0493-94.5.1209

Saner, C., Weibel, B., Würgler, F.E., Sengstag, C., 1996. Metabolism of promutagens catalyzed by Drosophila melanogaster CYP6A2 enzyme in Saccharomyces cerevisiae. *Environ. Mol. Mutagen.* 27, 46–58. doi:10.1002/(SICI)1098-2280(1996)27:1<46::AID-EM7>3.0.CO;2-C

Sann, M., Niehuis, O., Peters, R.S., Mayer, C., *et al.*, 2018. Phylogenomic analysis of Apoidea sheds new light on the sister group of bees. *BMC Evolutionary Biology* 18, 15118. doi:10.1186/s12862-018-1155-8

Santana, E., de los Reyes, T., Casas-Tintó, S., 2020. Small heat shock proteins determine synapse number and neuronal activity during development. *PLoS ONE* 15, e0233231. doi:10.1371/journal.pone.0233231

Santos-Pinto, dos, J.R.A., Perez-Riverol, A., Lasa, A.M., Palma, M.S., 2018. Diversity of peptidic and proteinaceous toxins from social Hymenoptera venoms. *Toxicon* 148, 172–196. doi:10.1016/j.toxicon.2018.04.029

Sánchez-Higueras, C., Sotillos, S., Hombría, J.C.-G., 2014. Common Origin of Insect Trachea and Endocrine Organs from a Segmentally Repeated Precursor. *Current Biology* 24, 76–81. doi:10.1016/j.cub.2013.11.010

Scanlan, J.L., Gledhill-Smith, R.S., Battlay, P., Robin, C., 2020. Genomic and transcriptomic analyses in Drosophila suggest that the ecdysteroid kinase-like (EcKL) gene family encodes the "detoxification-by-phosphorylation" enzymes of insects. *Insect biochemistry and molecular Biology* 123, 103429. doi:10.1016/j.ibmb.2020.103429

Scates, S.S., O'Neal, S.T., Anderson, T.D., 2019. Bacteria-mediated modification of insecticide toxicity in the yellow fever mosquito, Aedes aegypti. *Pestic Biochem Physiol* 161, 77–85. doi:10.1016/j.pestbp.2019.07.016

Schaeffer, S.W., Bhutkar, A., McAllister, B.F., Matsuda, M., *et al.*, 2008. Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* 179, 1601–1655. doi:10.1534/genetics.107.086074

Schirmeier, S., Klämbt, C., 2015. The Drosophila blood-brain barrier as interface between neurons and hemolymph. *Mech Dev* 138, 50–55. doi:10.1016/j.mod.2015.06.002

Schmidt, J.M., Battlay, P., Gledhill-Smith, R.S., Good, R.T., *et al.*, 2017. Insights into DDT Resistance from the Drosophila melanogaster Genetic Reference Panel. *Genetics* genetics.300310.2017. doi:10.1534/genetics.117.300310

Schmidt, J.M., Good, R.T., Appleton, B., Sherrard, J., *et al.*, 2010. Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. *PLOS Genetics* 6, e1000998. doi:10.1371/journal.pgen.1000998

Schmidt, J.M., Robin, C., 2011. An Adaptive Allelic Series Featuring Complex Gene Rearrangements. *PLOS Genetics* 7, e1002347. doi:10.1371/journal.pgen.1002347.g001

Schuler, M.A., 2010. P450s in plant-insect interactions. *BBA Proteins and Proteomics* 1814, 10–10. doi:10.1016/j.bbapap.2010.09.012

Schumann, I., Kenny, N., Hui, J., Hering, L., *et al.*, 2018. Halloween genes in panarthropods and the evolution of the early moulting pathway in Ecdysozoa. *Open Science* 5, 180888. doi:10.1098/rsos.180888

Schwartz, M.B., Kelly, T.J., Woods, C.W., Imberski, R.B., 1989. Ecdysteroid fluctuations in adult Drosophila melanogaster caused by elimination of pupal reserves and synthesis by early vitellogenic ovarian follicles. *Insect Biochemistry* 19, 243–249. doi:10.1016/0020-1790(89)90068-1

Schwarz, W.H., Brunecky, R., Broeker, J., Liebl, W., *et al.*, 2018. Handling gene and protein names in the age of bioinformatics: the special challenge of secreted multimodular bacterial enzymes such as the cbhA / cbh9A gene of Clostridium thermocellum. *World J Microbiol Biotechnol* 34, 1–10. doi:10.1007/s11274-018-2424-9

Schwedes, C., Tulsiani, S., Carney, G.E., 2011. Ecdysone receptor expression and activity in adult Drosophila melanogaster. *Journal of Insect Physiology* 57, 899–907. doi:10.1016/j.jinsphys.2011.03.027

Schwedes, C.C., Carney, G.E., 2012. Ecdysone signaling in adult Drosophila melanogaster. *Journal of Insect Physiology* 58, 293–302. doi:10.1016/j.jinsphys.2012.01.013

Schweizer, U., Braun, D., Forrest, D., 2019. The Ins and Outs of Steroid Hormone Transport Across the Plasma Membrane: Insight From an Insect. *Endocrinology* 160, 339–340. doi:10.1210/en.2018-01034

Schwentner, M., Combosch, D.J., Pakes Nelson, J., Giribet, G., 2017. A Phylogenomic Solution to the Origin of Insects by Resolving Crustacean-Hexapod Relationships. *Current Biology* 27, 1818–1824.e5. doi:10.1016/j.cub.2017.05.040

Schwentner, M., Richter, S., Rogers, D.C., Giribet, G., 2018. Tetraconatan phylogeny with special focus on Malacostraca and Branchiopoda: highlighting the strength of taxon-specific matrices in phylogenomics. *Proceedings of the Royal Society B: Biological Sciences* 285, 20181524. doi:10.1098/rspb.2018.1524

Scienski, K., Fay, J.C., Conant, G.C., 2015. Patterns of Gene Conversion in Duplicated Yeast Histones Suggest Strong Selection on a Coadapted Macromolecular Complex. *Genome Biology and Evolution* 7, 3249–3258. doi:10.1093/gbe/evv216

Scott, J.G., Buchon, N., 2019. Drosophila melanogaster as a powerful tool for studying insect toxicology. *Pestic Biochem Physiol* 161, 95–103. doi:10.1016/j.pestbp.2019.09.006

Scott, J.G., Warren, W.C., Beukeboom, L.W., Bopp, D., *et al.*, 2014. Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. *Genome Biology* 15, 466. doi:10.1186/s13059-014-0466-3

Seinen, E., Burgerhof, J.G.M., Jansen, R.C., Sibon, O.C.M., 2011. RNAi-induced off-target effects in Drosophila melanogaster: frequencies and solutions. *Briefings in Functional Genomics* 10, 206–214. doi:10.1093/bfgp/elr017

Seppey, M., Ioannidis, P., Emerson, B.C., Pitteloud, C., *et al.*, 2019. Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. *Genome Biol* 20, 555. doi:10.1186/s13059-019-1704-5

Serbus, L.R., Ferreccio, A., Zhukova, M., McMorris, C.L., *et al.*, 2011. A feedback loop between Wolbachia and the Drosophila gurken mRNP complex influences Wolbachia titer. *Journal of cell science* 124, 4299–4308. doi:10.1242/jcs.092510

Seringhaus, M.R., Cayting, P.D., Gerstein, M.B., 2008. Uncovering trends in gene naming. *Genome Biology* 9, 401. doi:10.1186/gb-2008-9-1-401

Seybold, S.J., Tittiger, C., 2003. Biochemistry and molecular biology of de novo isoprenoid pheromone production in the Scolytidae. *Annu Rev Entomol* 48, 425–453. doi:10.1146/annurev.ento.48.091801.112645

Sezutsu, H., Le Goff, G., Feyereisen, R., 2013. Origins of P450 diversity. *Phil. Trans. R. Soc. B* 368, 20120428–20120428. doi:10.1098/rstb.2012.0428

Shabab, M., Khan, S.A., Vogel, H., Heckel, D.G., *et al.*, 2014. OPDA isomerase GST16 is involved in phytohormone detoxification and insect development. *FEBS Journal* 281, 2769–2783. doi:10.1111/febs.12819

Shah, S., Yarrow, C., Dunning, R., Ben Cheek, *et al.*, 2012. Insecticide detoxification indicator strains as tools for enhancing chemical discovery screens. *Pest. Manag. Sci.* 68, 38–48. doi:10.1002/ps.2218

Sharma, V., Hecker, N., Roscito, J.G., Foerster, L., *et al.*, 2018. A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nature Communications* 9, 1–9. doi:10.1038/s41467-018-03667-1

Shingleton, A.W., Masandika, J.R., Thorsen, L.S., Zhu, Y., *et al.*, 2017. The sex-specific effects of diet quality versus quantity on morphology in Drosophila melanogaster. *Open Science* 4, 170375. doi:10.1098/rsos.170375

Shows, T.B., McAlpine, P.J., 1978. The catalog of human genes and chromosome assignments. A report on human genetic nomenclature and genes that have been mapped in man. *Cytogenet. Cell Genet.* 22, 132–145. doi:10.1159/000130925

Sieber, M.H., Thummel, C.S., 2012. Coordination of Triacylglycerol and Cholesterol Homeostasis by DHR96 and the Drosophila LipA Homolog magro. *Cell Metab* 15, 122–127. doi:10.1016/j.cmet.2011.11.011

Siegfried, B.D., Scharf, M.E., 2001. Mechanisms of Organophosphate Resistance in Insects, in: *Biochemical Sites of Insecticide Action and Resistance*. Springer, Berlin, Heidelberg, Berlin, Heidelberg, pp. 269–291. doi:10.1007/978-3-642-59549-3_13

Sievers, F., Higgins, D.G., 2019. QuanTest2: benchmarking multiple sequence alignments using secondary structure prediction. *Bioinformatics* 1079, 245. doi:10.1093/bioinformatics/btz552

Simon, J.-C., d'Alencon, E., Guy, E., Jacquin-Joly, E., *et al.*, 2015. Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics* 14, 413–423. doi:10.1093/bfgp/elv015

Singer, M.S., Lichter-Marck, I.H., Farkas, T.E., Aaron, E., *et al.*, 2014. Herbivore diet breadth mediates the cascading effects of carnivores in food webs. *Proc Natl Acad Sci U S A* 111, 9521–9526. doi:10.1073/pnas.1401949111

Sioud, M., Leirdal, M., 2007. Druggable Signaling Proteins, in: *Target Discovery and Validation Reviews and Protocols*. Humana Press, New Jersey, pp. 1–24. doi:10.1385/1-59745-208-4:1

Siviter, H., Koricheva, J., Brown, M.J.F., Leadbeater, E., 2018. Quantifying the impact of pesticides on learning and memory in bees. *Journal of Applied Ecology* 55, 2812–2821. doi:10.1111/1365-2664.13193

Smith, C.R., Morandin, C., Noureddine, M., Pant, S., 2018. Conserved roles of Osiris genes in insect development, polymorphism and protection. *Journal of Evolutionary Biology* 31, 516–529. doi:10.1111/jeb.13238

Smith, J.N., Turbert, H.B., 1964. Comparative detoxication. 11. Conjugations of 1-naphthol and some other phenols in houseflies and locusts. *Biochemical Journal* 92, 127–131. doi:10.1042/bj0920127

Smyth, K.A., Boyce, T.M., Russell, R.J., Oakeshott, J.G., 2000. MCE activities and malathion resistances in field populations of the australian sheep blowfly (Lucilia cuprina). *Heredity* 84 (Pt 1), 63–72. doi:10.1046/j.1365-2540.2000.00641.x

Snoeck, S., Pavlidi, N., Pipini, D., Vontas, J., *et al.*, 2019. Substrate specificity and promiscuity of horizontally transferred UDP-glycosyltransferases in the generalist herbivore Tetranychus urticae. *Insect biochemistry and molecular Biology*. doi:10.1016/j.ibmb.2019.04.010

Sommé-Martin, G., Colardeau, J., Beydon, P., Blais, C., *et al.*, 1990. P1 gene expression in Drosophila larval fat body: induction by various ecdysteroids. *Archives of Insect Biochemistry and Physiology* 15, 43–56. doi:10.1002/arch.940150105

Sommé-Martin, G., Colardeau, J., Lafont, R., 1988a. Conversion of ecdysone and 20-hydroxyecdysone into 3-dehydroecdysteroids is a major pathway in third instar Drosophila melanogaster larvae. *Insect Biochemistry* 18, 729–734. doi:10.1016/0020-1790(88)90082-0

Sommé-Martin, G., Colardeau, J., Lafont, R., 1988b. Metabolism and biosynthesis of ecdysteroids in the Drosophila development mutant ecd1. *Insect Biochemistry* 18, 735–742. doi:10.1016/0020-1790(88)90083-2

Sonobe, H., Ito, Y., 2009. Phosphoconjugation and dephosphorylation reactions of steroid hormone in insects. *Molecular and Cellular Endocrinology* 307, 25–35. doi:10.1016/j.mce.2009.03.017

Sonobe, H., Ohira, T., Ieki, K., Maeda, S., *et al.*, 2006. Purification, kinetic characterization, and molecular cloning of a novel enzyme, ecdysteroid 22-kinase. *J. Biol. Chem.* 281, 29513–29524. doi:10.1074/jbc.M604035200

Sonobe, H., Yamada, R., 2004. Ecdysteroids during early embryonic development in silkworm Bombyx mori: Metabolism and functions. *Zoological Science* 21, 503–516. doi:10.2108/zsj.21.503

Soucy, S.M., Huang, J., Gogarten, J.P., 2015. Horizontal gene transfer: building the web of life. *Nat Rev Genet* 16, 472–482. doi:10.1038/nrg3962

Spieth, H.T., 1979. The Virilis Group of Drosophila and the Beaver Castor. *American Naturalist* 114, 312–316. doi:10.1086/283479

Spofford, J.B., 1969. Heterosis and the Evolution of Duplications. *American Naturalist* 103, 407–432. doi:10.1086/282611

Stamboulian, M., Guerrero, R.F., Hahn, M.W., Radivojac, P., 2020. The ortholog conjecture revisited: the value of orthologs and paralogs in function prediction. *Bioinformatics* 36, i219–i226. doi:10.1093/bioinformatics/btaa468

Stanley, D., 2006. Prostaglandins and other eicosanoids in insects: Biological significance. *Annu Rev Entomol* 51, 25–44. doi:10.1146/annurev.ento.51.110104.151021

Stanley, D., Kim, Y., 2018. Prostaglandins and Other Eicosanoids in Insects: Biosynthesis and Biological Actions. *Front. Physiol.* 9, 1927. doi:10.3389/fphys.2018.01927

Staudinger, H., Ruzicka, L., 1924. Insektentötende Stoffe I. Über Isolierung und Konstitution des wirksamen Teiles des dalmatinischen Insektenpulvers. *Helvetica Chimica Acta* 7, 177–201. doi:10.1002/hlca.19240070124

Stoiber, M., Celniker, S., Cherbas, L., Ben Brown, *et al.*, 2016. Diverse Hormone Response Networks in 41 Independent Drosophila Cell Lines. *G3 (Bethesda)* 6, 683–694. doi:10.1534/g3.115.023366

Stone, E.A., 2011. Why the phylogenetic regression appears robust to tree misspecification. *Systematic Biol.* 60, 245–260. doi:10.1093/sysbio/syq098

Storelli, G., Nam, H.-J., Simcox, J., Villanueva, C.J., *et al.*, 2019. Drosophila HNF4 Directs a Switch in Lipid Metabolism that Supports the Transition to Adulthood. *Developmental Cell* 48, 200–214.e6. doi:10.1016/j.devcel.2018.11.030

Strauss, A.S., Peters, S., Boland, W., Burse, A., 2013. ABC transporter functions as a pacemaker for sequestration of plant glucosides in leaf beetles. *eLife* 2, 151. doi:10.7554/eLife.01096

Stroehlein, A.J., Young, N.D., Gasser, R.B., 2018. Improved strategy for the curation and classification of kinases, with broad applicability to other eukaryotic protein groups. *Sci. Rep.* 8, 1–10. doi:10.1038/s41598-018-25020-8

Studer, R.A., Robinson-Rechavi, M., 2009. How confident can we be that orthologs are similar, but paralogs differ? *Trends in genetics: TIG* 25, 210–216. doi:10.1016/j.tig.2009.03.004

Subramoniam, T., 2000. Crustacean ecdysteroids in reproduction and embryogenesis. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology* 125, 135–156. doi:10.1016/S0742-8413(99)00098-5

Sun, B.-F., Li, T., Xiao, J.-H., Jia, L.-Y., *et al.*, 2015. Horizontal functional gene transfer from bacteria to fishes. *Sci. Rep.* 5, 18676. doi:10.1038/srep18676

Sun, L., Li, H.-M., Seufferheld, M.J., Walters, K.R., Jr, *et al.*, 2011. Systems-Scale Analysis Reveals Pathways Involved in Cellular Response to Methamphetamine. *PLoS ONE* 6, e18215. doi:10.1371/journal.pone.0018215

Sun, R., Jiang, X., Reichelt, M., Gershenzon, J., *et al.*, 2019. Tritrophic metabolism of plant chemical defenses and its effects on herbivore and predator performance. *eLife* 8, 1. doi:10.7554/eLife.51029

Sun, W., Shen, Y.-H., Qi, D.-W., Xiang, Z.-H., *et al.*, 2012. Molecular cloning and characterization of Ecdysone oxidase and 3-dehydroecdysone-3α-reductase involved in the ecdysone inactivation pathway of silkworm, Bombyx mori. *Int. J. Biol. Sci.* 8, 125–138. doi:10.7150/ijbs.8.125

Sun, W., Shen, Y.-H., Zhou, L.-X., Zhang, Z., 2016. Ecdysone Titer Determined by 3DE-3β-Reductase Enhances the Immune Response in the Silkworm. *The Journal of Immunology* 196, 1646–1654. doi:10.4049/jimmunol.1500158

Sun, W.W., Margam, V.M.V., Sun, L.L., Buczkowski, G.G., *et al.*, 2006. Genome-wide analysis of phenobarbital-inducible genes in Drosophila melanogaster. *Insect Mol Biol* 15, 455–464. doi:10.1111/j.1365-2583.2006.00662.x

Suzanne, M., Perrimon, N., Noselli, S., 2001. The Drosophila JNK Pathway Controls the Morphogenesis of the Egg Dorsal Appendages and Micropyle. *Developmental biology* 237, 282–294. doi:10.1006/dbio.2001.0384

Suzuki, S., Yamasaki, K., Fujita, T., Mamiya, Y., *et al.*, 1996. Ovarian and hemolymph ecdysteroids in the terrestrial isopod Armadillidium vulgare (Malacostracan Crustacea). *Gen Comp Endocrinol* 104, 129–138. doi:10.1006/gcen.1996.0155

Swevers, L., 2019. An update on ecdysone signaling during insect oogenesis. *Current Opinion in Insect Science* 31, 8–13. doi:10.1016/j.cois.2018.07.003

Sykiotis, G.P., Bohmann, D., 2008. Keap1/Nrf2 Signaling Regulates Oxidative Stress Tolerance and Lifespan in Drosophila. *Developmental Cell* 14, 76–85. doi:10.1016/j.devcel.2007.12.002

Sztal, T., Chung, H., Berger, S., Currie, P.D., *et al.*, 2012. A Cytochrome P450 Conserved in Insects Is Involved in Cuticle Formation. *PLoS ONE* 7, e36544. doi:10.1371/journal.pone.0036544

Sztal, T., Chung, H., Gramzow, L., Daborn, P.J., *et al.*, 2007. Two independent duplications forming the Cyp307a genes in Drosophila. *Insect biochemistry and molecular Biology* 37, 1044–1053. doi:10.1016/j.ibmb.2007.05.017

Szuts, D., Freeman, M., Bienz, M., 1997. Antagonism between EGFR and Wingless signalling in the larval cuticle of Drosophila. *Development* 124, 3209–3219.

Škerlová, J., Lindström, H., Gonis, E., Sjödin, B., *et al.*, 2020. Structure and steroid isomerase activity of Drosophila glutathione transferase E14 essential for ecdysteroid biosynthesis. *FEBS letters* 401, 333. doi:10.1002/1873-3468.13718

Takahashi, S.Y., Okamoto, K., Sonobe, H., Kamba, M., *et al.*, 1992. In vitro synthesis of ecdysteroid conjugates by tissue extracts of the silkworm, Bombyx mori. *Zoological Science* 9, p169–174.

Takashima, S., Younossi-Hartenstein, A., Ortiz, P.A., Hartenstein, V., 2011. A novel tissue in an established model system: the Drosophila pupal midgut. *Dev. Genes Evol.* 221, 69–81. doi:10.1007/s00427-011-0360-x

Takeuchi, H., Chen, J.-H., O'Reilly, D.R., Turner, P.C., *et al.*, 2001. Regulation of Ecdysteroid Signaling: Cloning and Characterization of Ecdysone Oxidase. *J. Biol. Chem.* 276, 26819–26828. doi:10.1074/jbc.M104291200

Takeuchi, H., Rigden, D.J., Ebrahimi, B., Turner, P.C., *et al.*, 2005. Regulation of ecdysteroid signalling during Drosophila development: identification, characterization and modelling of ecdysone oxidase, an enzyme involved in control of ligand concentration. *Biochemical Journal* 389, 637–645. doi:10.1042/BJ20050498

Talamillo, A., Sánchez, J., Cantera, R., Pérez, C., *et al.*, 2008. Smt3 is required for Drosophila melanogaster metamorphosis. *Development* 135, 1659–1668. doi:10.1242/dev.020685

Tamura, K., Subramanian, S., Kumar, S., 2004. Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. *Mol. Biol. Evol.* 21, 36–44. doi:10.1093/molbev/msg236

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., *et al.*, 2015. Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Systematic Biol.* 64, 778–791. doi:10.1093/sysbio/syv033

Tan, H.M., Low, W.Y., 2018. Rapid birth-death evolution and positive selection in detoxification-type glutathione S-transferases in mammals. *PLoS ONE* 13, e0209336. doi:10.1371/journal.pone.0209336

Tan, K.L., Vlisidou, I., Wood, W., 2014. Ecdysone Mediates the Development of Immunity in the Drosophila Embryo. *Curr. Biol.* doi:10.1016/j.cub.2014.03.062

Tanaka, Y., 1998. Induction of larval ecdysis by ecdysone in the non-molting glossy (nm-g) mutant larvae of Bombyx mori. *The Journal of Sericultural Science of Japan* 67, 109–115. doi:10.11416/kontyushigen1930.67.109

Tang, H., Kambris, Z., Lemaitre, B., Hashimoto, C., 2008. A Serpin that Regulates Immune Melanization in the Respiratory System of Drosophila. *Developmental Cell* 15, 617–626. doi:10.1016/j.devcel.2008.08.017

Taubert, S., Hansen, M., Van Gilst, M.R., Cooper, S.B., *et al.*, 2008. The Mediator Subunit MDT-15 Confers Metabolic Adaptation to Ingested Material. *PLOS Genetics* 4, e1000021. doi:10.1371/journal.pgen.1000021

Terashima, J., Takaki, K., Sakurai, S., Bownes, M., 2005. Nutritional status affects 20-hydroxyecdysone concentration and progression of oogenesis in Drosophila melanogaster. *J Endocrinol* 187, 69–79. doi:10.1677/joe.1.06220

Terhzaz, S., Cabrero, P., Brinzer, R.A., Halberg, K.A., *et al.*, 2015. A novel role of Drosophila cytochrome P450-4e3 in permethrin insecticide tolerance. *Insect biochemistry and molecular Biology* 67, 38–46. doi:10.1016/j.ibmb.2015.06.002

Terriere, L.C., 1984. Induction of Detoxication Enzymes in Insects. *Annu Rev Entomol* 29, 71–88. doi:10.1146/annurev.en.29.010184.000443

Thomas, A.M., Hui, C., South, A., McVey, M., 2013. Common Variants of Drosophila melanogaster Cyp6d2 Cause Camptothecin Sensitivity and Synergize With Loss of Brca2. *G3 (Bethesda)* 3, 91–99. doi:10.1534/g3.112.003996

Thomas, G.W.C., Dohmen, E., Hughes, D.S.T., Murali, S.C., *et al.*, 2020. Gene content evolution in the arthropods. *Genome Biol* 21, 20160133. doi:10.1186/s13059-019-1925-7

Thomas, J.H., 2007. Rapid Birth–Death Evolution Specific to Xenobiotic Cytochrome P450 Genes in Vertebrates. *PLOS Genetics* 3, e67. doi:10.1371/journal.pgen.0030067

Thomas, J.H., 2006. Analysis of Homologous Gene Clusters in Caenorhabditis elegans Reveals Striking Regional Cluster Domains. *Genetics* 172, 127–143. doi:10.1534/genetics.104.040030

Thompson, A., Zakon, H.H., Kirkpatrick, M., 2016. Compensatory Drift and the Evolutionary Dynamics of Dosage-Sensitive Duplicate Genes. *Genetics* 202, 765–774. doi:10.1534/genetics.115.178137

Thompson, M.J., Feldlaufer, M.F., Lozano, R., Rees, H.H., *et al.*, 1987a. Metabolism of 26-[C-14]Hydroxyecdysone 26-Phosphate in the Tobacco Hornworm, Manduca Sexta L., to a

New Ecdysteroid Conjugate: 26-[C-14]Hydroxyecdysone 22-Glucoside. *Archives of Insect Biochemistry and Physiology* 4, 1–15. doi:10.1002/arch.940040102

Thompson, M.J., Kaplanis, J.N., Robbins, W.E., Yamamoto, R.T., 1967. 20, 26-Dihydroxyecdysone, a new steroid with moulting hormone activity from the tobacco hornworm, Manduca sexta (Johannson). *Chemical Communications (London)* 650–653. doi:10.1039/C19670000650

Thompson, M.J., Svoboda, J.A., Lozano, R., Wilzer, K.R., 1988. Profile of free and conjugated ecdysteroids and ecdysteroid acids during embryonic development of Manduca sexta (L.) following maternal incorporation of [14C]cholesterol. *Archives of Insect Biochemistry and Physiology* 7, 157–172. doi:10.1002/arch.940070303

Thompson, M.J., Svoboda, J.A., Rees, H.H., Wilzer, K.R., 1987b. Isolation and identification of 26-hydroxyecdysone 2-phosphate: An ecdysteroid conjugate of eggs and ovaries of the tobacco hornworm,Manduca sexta. *Archives of Insect Biochemistry and Physiology* 4, 183–190. doi:10.1002/arch.940040304

Thompson, M.J., Weirich, G.F., Rees, H.H., Svoboda, J.A., *et al.*, 1985. New ecdysteroid conjugate: Isolation and identification of 26-hydroxyecdysone 26-phosphate from eggs of the tobacco hornworm, Manduca sexta (L.). *Archives of Insect Biochemistry and Physiology* 2, 227–236. doi:10.1002/arch.940020302

Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., *et al.*, 2019. FlyBase 2.0: the next generation. *Nucl Acids Res* 47, D759–D765. doi:10.1093/nar/gky1003

Tian, F., Wang, Z., Li, C., Liu, J., *et al.*, 2019. UDP-Glycosyltransferases are involved in imidacloprid resistance in the Asian citrus psyllid, Diaphorina citri (Hemiptera: Lividae). *Pestic Biochem Physiol* 154, 23–31. doi:10.1016/j.pestbp.2018.12.010

Tian, X., Zhao, S., Guo, Z., Hu, B., *et al.*, 2018. Molecular characterization, expression pattern and metabolic activity of flavin-dependent monooxygenases in Spodoptera exigua. *Insect Mol Biol* 27, 533–544. doi:10.1111/imb.12392

Tobe, S.S., 1978. Reproductive physiology of Glossina. *Annu Rev Entomol* 23, 283–307. doi:10.1146/annurev.en.23.010178.001435

Tootle, T.L., 2013. Genetic insights into the in vivo functions of prostaglandin signaling. *Int J Biochem Cell Biol* 45, 1629–1632. doi:10.1016/j.biocel.2013.05.008

Tootle, T.L., Spradling, A.C., 2008. Drosophila Pxt: a cyclooxygenase-like facilitator of follicle maturation. *Development* 135, 839–847. doi:10.1242/dev.017590

Tootle, T.L., Williams, D., Hubb, A., Frederick, R., *et al.*, 2011. Drosophila Eggshell Production: Identification of New Genes and Coordination by Pxt. *PLoS ONE* 6, –e19943. doi:10.1371/journal.pone.0019943

Torrie, L.S., Radford, J.C., Southall, T.D., Kean, L., *et al.*, 2004. Resolution of the insect ouabain paradox. *Proc. Natl Acad. Sci. USA* 101, 13689–13693. doi:10.1073/pnas.0403087101

Treangen, T.J., Rocha, E.P.C., 2011. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. *PLOS Genetics* 7, e1001284. doi:10.1371/journal.pgen.1001284.t002

Trienens, M., Keller, N.P., Rohlfs, M., 2010. Fruit, flies and filamentous fungi - experimental analysis of animal-microbe competition using Drosophila melanogaster and Aspergillus mould as a model system. *Oikos* 119, 1765–1775. doi:10.1111/j.1600-0706.2010.18088.x

Trienens, M., Kraaijeveld, K., Wertheim, B., 2017. Defensive repertoire of Drosophila larvae in response to toxic fungi. *Mol Ecol*. doi:10.1111/mec.14254

Truman, J.W., 2019. The Evolution of Insect Metamorphosis. *Current Biology* 29, R1252–R1268. doi:10.1016/j.cub.2019.10.009

Tsabar, M., Eapen, V.V., Mason, J.M., Memisoglu, G., *et al.*, 2015. Caffeine impairs resection during DNA break repair by reducing the levels of nucleases Sae2 and Dna2. *Nucl Acids Res* 43, 6889–6901. doi:10.1093/nar/gkv520

Tsoupras, G., Luu, B., Hoffmann, J.A., 1982. Isolation and identification of three ecdysteroid conjugates with a C-20 hydroxy group in eggs of locusta migratoria. *Steroids* 40, 551–560. doi:10.1016/0039-128X(82)90075-7

Tsvetkov, N., Samson-Robert, O., Sood, K., Patel, H.S., *et al.*, 2017. Chronic exposure to neonicotinoids reduces honey bee health near corn crops. *Science* 356, 1395–1397. doi:10.1126/science.aam7470

Tsygankov, A.Y., 2019. TULA-family proteins: Jacks of many trades and then some. *J. Cell. Physiol.* 234, 274–288. doi:10.1002/jcp.26890

Ujihara, K., 2019. The history of extensive structural modifications of pyrethroids. *Journal of pesticide science* 44, 215–224. doi:10.1584/jpestics.D19-102

Uyeda, J.C., Zenil-Ferguson, R., Pennell, M.W., Matzke, N., 2018. Rethinking phylogenetic comparative methods. *Systematic Biol.* 106, 13410. doi:10.1093/sysbio/syy031

Uyehara, C.M., McKay, D.J., 2019. Direct and widespread role for the nuclear receptor EcR in mediating the response to ecdysone in Drosophila. *Proc Natl Acad Sci U S A* 116, 9893–9902. doi:10.1073/pnas.1900343116

van den Bosch, T.J.M., Welte, C.U., 2017. Detoxifying symbionts in agriculturally important pest insects. *Microbial Biotechnology* 10, 531–540. doi:10.1111/1751-7915.12483

van Loo, B., Bayer, C.D., Fischer, G., Jonas, S., *et al.*, 2018. Balancing Specificity and Promiscuity in Enzyme Evolution: Multidimensional Activity Transitions in the Alkaline Phosphatase Superfamily. *J. Am. Chem. Soc.* 141, 370–387. doi:10.1021/jacs.8b10290

Van Oss, S.B., Carvunis, A.-R., 2019. De novo gene birth. *PLOS Genetics* 15, e1008160. doi:10.1371/journal.pgen.1008160

Vanderplanck, M., Decleves, S., Roger, N., Decroo, C., *et al.*, 2016. Is non-host pollen suitable for generalist bumblebees? *Insect Sci.* 6, 562. doi:10.1111/1744-7917.12410

Venken, K.J.T., Bellen, H.J., 2014. Chemical mutagens, transposons, and transgenes to interrogate gene function in Drosophila melanogaster. *Methods* 68, 15–28. doi:10.1016/j.ymeth.2014.02.025

Venken, K.J.T., Schulze, K.L., Haelterman, N.A., Pan, H., *et al.*, 2011. MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat Meth* 8, 737–743. doi:10.1038/nmeth.1662

Verma, P., Tapadia, M.G., 2015. Early gene Broad complex plays a key role in regulating the immune response triggered by ecdysone in the Malpighian tubules of Drosophila melanogaster. *Mol. Immunol.* 66, 325–339. doi:10.1016/j.molimm.2015.03.249

Vigne, P., Frelin, C., 2010. Hypoxia modifies the feeding preferences of Drosophila. Consequences for diet dependent hypoxic survival. *BMC Physiol* 10, 8. doi:10.1186/1472-6793-10-8

Visanuvimol, L., Bertram, S.M., 2011. How dietary phosphorus availability during development influences condition and life history traits of the cricket, Acheta domesticus. *Journal of Insect Science* 11, 1–17. doi:10.1673/031.011.6301

Vissers, J.H.A., Manning, S.A., Kulkarni, A., Harvey, K.F., 2016. A Drosophila RNAi library modulates Hippo pathway-dependent tissue growth. *Nature Communications* 7, 10368–6. doi:10.1038/ncomms10368

Vorojeikina, D., Broberg, K., Love, T.M., Davidson, P.W., *et al.*, 2017. Glutathione S-Transferase Activity Moderates Methylmercury Toxicity During Development in Drosophila. *Toxicol Sci* 157, 211–221. doi:10.1093/toxsci/kfx033

Wang, H., Shi, Y., Wang, L., Liu, S., *et al.*, 2018. CYP6AE gene cluster knockout in Helicoverpa armigera reveals role in detoxification of phytochemicals and insecticides. *Nature Communications* 9, 4820. doi:10.1038/s41467-018-07226-6

Wang, L., Dankert, H., Pietro Perona, Anderson, D.J., 2008. A common genetic target for environmental and heritable influences on aggressiveness in Drosophila. *Proc Natl Acad Sci U S A* 105, 5657–5663. doi:10.1073/pnas.0801327105

Wang, S., Liu, S., Liu, H., Wang, J., *et al.*, 2010. 20-hydroxyecdysone Reduces Insect Food Consumption Resulting in Fat Body Lipolysis During Molting and Pupation. *J Mol Cell Biol* 2, 128–138. doi:10.1093/jmcb/mjq006

Wang, S., Yang, C., Tu, H., Zhou, J., *et al.*, 2017. Characterization and Metabolic Diversity of Flavonoids in Citrus Species. *Sci. Rep.* 7, 10549. doi:10.1038/s41598-017-10970-2

Wang, S.F., Ayer, S., Segraves, W.A., Williams, D.R., *et al.*, 2000. Molecular Determinants of Differential Ligand Sensitivities of Insect Ecdysteroid Receptors. *Molecular and cellular biology* 20, 3870–3879. doi:10.1128/MCB.20.11.3870-3879.2000

Wang, X., Bo, J., Bridges, T., Dugan, K.D., *et al.*, 2006. Analysis of Cell Migration Using Whole-Genome Expression Profiling of Migratory Cells in the Drosophila Ovary. *Developmental Cell* 10, 483–495. doi:10.1016/j.devcel.2006.02.003

Wang, X.G., Ma, S.Y., Chang, J.S., Shi, R., *et al.*, 2019. Programmable activation of Bombyx gene expression using CRISPR/dCas9 fusion systems. *Insect Sci.* 26, 983–990. doi:10.1111/1744-7917.12634

Warren, J.T., Gilbert, L.I., 1986. Ecdysone metabolism and distribution during the pupal-adult development of Manduca sexta. *Insect Biochemistry* 16, 65–82. doi:10.1016/0020-1790(86)90080-6

Warren, J.T., Petryk, A., Marqués, G., Jarcho, M., *et al.*, 2002. Molecular and biochemical characterization of two P450 enzymes in the ecdysteroidogenic pathway of Drosophila melanogaster. *Proc Natl Acad Sci U S A* 99, 11043–11048. doi:10.1073/pnas.162375799

Warren, J.T., Petryk, A., Marqués, G., Parvy, J.-P., *et al.*, 2004. Phantom encodes the 25-hydroxylase of Drosophila melanogaster and Bombyx mori: a P450 enzyme critical in ecdysone biosynthesis. *Insect biochemistry and molecular Biology* 34, 991–1010. doi:10.1016/j.ibmb.2004.06.009

Warren, J.T., Steiner, B., Dorn, A., Pak, M., *et al.*, 1986. Metabolism of Ecdysteroids During the Embryogenesis of Manduca Sexta. *Journal of Liquid Chromatography* 9, 1759–1782. doi:10.1080/01483918608076716

Warren, J.T., Yerushalmi, Y., Shimell, M.J., O'Connor, M.B., *et al.*, 2006. Discrete pulses of molting hormone, 20-hydroxyecdysone, during late larval development of Drosophila melanogaster: Correlations with changes in gene activity. *Developmental Dynamics* 235, 315–326. doi:10.1002/dvdy.20626

Wasbrough, E.R., Dorus, S., Hester, S., Howard-Murkin, J., *et al.*, 2009. The Drosophila melanogaster sperm proteome-II (DmSP-II). *J Proteomics* 73, 2171–2185. doi:10.1016/j.jprot.2010.09.002

Waterhouse, R.M., Zdobnov, E.M., Kriventseva, E.V., 2011. Correlating traits of gene retention, sequence divergence, duplicability and essentiality in vertebrates, arthropods, and fungi. *Genome Biology and Evolution* 3, 75–86. doi:10.1093/gbe/evq083

Webb, T.J., Powls, R., Rees, H.H., 1996. Characterization, fractionation and kinetic properties of the enzymes of ecdysteroid 3-epimerization and phosphorylation isolated from the midgut cytosol of the cotton leafworm, Spodoptera littoralis. *Insect biochemistry and molecular Biology* 26, 809–816. doi:10.1016/S0965-1748(96)00030-6

Webb, T.J., Powls, R., Rees, H.H., 1995. Enzymes of ecdysteroid transformation and inactivation in the midgut of the cotton leafworm, Spodoptera littoralis: properties and developmental profiles. *Biochemical Journal* 312 (Pt 2), 561–568. doi:10.1042/bj3120561

Weirich, G.F., Thompson, M.J., Svoboda, J.A., 1991. Enzymes of ecdysteroid 3-epimerization in midgut cytosol of Manduca sexta: pH optima cosubstrate kinetics, and sodium chloride effect. *Insect Biochemistry* 21, 65–71. doi:10.1016/0020-1790(91)90065-M

Weirich, G.F., Thompson, M.J., Svoboda, J.A., 1986. In vitro ecdysteroid conjugation by enzymes of Manduca sexta midgut cytosol. *Archives of Insect Biochemistry and Physiology* 3, 109–126. doi:10.1002/arch.940030202

Weiszmann, R., Hammonds, A.S., Celniker, S.E., 2009. Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. *Nat Protoc* 4, 605–618. doi:10.1038/nprot.2009.55

Weng, J.-K., Philippe, R.N., Noel, J.P., 2012. The Rise of Chemodiversity in Plants. *Science* 336, 1667–1670. doi:10.1126/science.1217411

Werren, J.H., Baldo, L., Clark, M.E., 2008. Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol* 6, 741–751. doi:10.1038/nrmicro1969

Wheeler, M.W., Park, R.M., Bailer, A.J., 2006. Comparing median lethal concentration values using confidence interval overlap or ratio tests. *Environ Toxicol Chem* 25, 1441–1444. doi:10.1897/05-320R.1

Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., *et al.*, 2011. Episodic radiations in the fly tree of life. *PNAS* 108, 5690–5695. doi:10.1073/pnas.1012675108

Wieschaus, E., Nüsslein-Volhard, C., 2016. The Heidelberg Screen for Pattern Mutants of Drosophila: A Personal Account. *Annu. Rev. Cell Dev. Biol.* 32, 1–46. doi:10.1146/annurev-cellbio-113015-023138

Wieschaus, E., Nüsslein-Volhard, C., Jürgens, G., 1984. Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster. *Wilhelm Roux' Archiv* 193, 296–307. doi:10.1007/BF00848158

Wilbrandt, J., Misof, B., Panfilio, K.A., Niehuis, O., 2019. Repertoire-wide gene structure analyses: a case study comparing automatically predicted and manually annotated gene models. *BMC Genomics* 20, 1–12. doi:10.1186/s12864-019-6064-8

Wilding, C.S., 2018. Regulating resistance: CncC:Maf, antioxidant response elements and the overexpression of detoxification genes in insecticide resistance. *Current Opinion in Insect Science* 27, 89–96. doi:10.1016/j.cois.2018.04.006

Wilding, M., Nachtschatt, M., Speight, R., Scott, C., 2017. An Improved and General Streamlined Phylogenetic Protocol Applied to the Fatty Acid Desaturase Family. *Molecular Phylogenetics and Evolution*. doi:10.1016/j.ympev.2017.07.012

Wilkenfeld, S.R., Lin, C., Frigo, D.E., 2018. Communication between genomic and non-genomic signaling events coordinate steroid hormone actions. *Steroids* 133, 2–7. doi:10.1016/j.steroids.2017.11.005

Wilkinson, C.F., 1986. Xenobiotic Conjugation in Insects, in: *Xenobiotic Conjugation Chemistry*, ACS Symposium Series. American Chemical Society, Washington, DC, pp. 48–61. doi:10.1021/bk-1986-0299.ch003

Will, I., Das, B., Trinh, T., Brachmann, A., *et al.*, 2020. Genetic Underpinnings of Host Manipulation by Ophiocordyceps as Revealed by Comparative Transcriptomics. *G3 (Bethesda)* g3.401290.2020. doi:10.1534/g3.120.401290

Williams, D.R., Chen, J.H., Fisher, M.J., Rees, H.H., 1997. Induction of Enzymes Involved in Molting Hormone (Ecdysteroid) Inactivation by Ecdysteroids and an Agonist, 1,2-Dibenzoyl-1-tert-butylhydrazine (RH-5849). *J. Biol. Chem.* 272, 8427–8432. doi:10.1074/jbc.272.13.8427

Williams, D.R., Fisher, M.J., Smagghe, G., Rees, H.H., 2002. Species specificity of changes in ecdysteroid metabolism in response to ecdysteroid agonists. *Pestic Biochem Physiol* 72, 91–99. doi:10.1006/pest.2001.2588

Williams, R.T., 1959. Detoxication mechanisms: the metabolism and detoxication of drugs, toxic substances and other organic compounds, 2nd ed. Chapman & Hall.

Williams, R.T., 1951. The metabolism of drugs and toxic substances. *Annual Review of Biochemistry* 20, 441–464. doi:10.1146/annurev.bi.20.070151.002301

Willingham, A.T., Keil, T., 2004. A tissue specific cytochrome P450 required for the structure and function of Drosophila sensory organs. *Mech Dev* 121, 1289–1297. doi:10.1016/j.mod.2004.04.017

Willoughby, L., Batterham, P., Daborn, P.J., 2007. Piperonyl butoxide induces the expression of cytochrome P450 and glutathione S-transferase genes in Drosophila melanogaster. *Pest. Manag. Sci.* 63, 803–808. doi:10.1002/ps.1391

Willoughby, L., Chung, H., Lumb, C., Robin, C., *et al.*, 2006. A comparison of Drosophila melanogaster detoxification gene induction responses for six insecticides, caffeine and phenobarbital. *Insect biochemistry and molecular Biology* 36, 934–942. doi:10.1016/j.ibmb.2006.09.004

Winberg, J.O., McKinley-McKee, J.S., 1998. Drosophila melanogaster alcohol dehydrogenase: mechanism of aldehyde oxidation and dismutation. *Biochemical Journal* 329 (Pt 3), 561–570. doi:10.1042/bj3290561

Wingen, A., Carrera, P., Ekaterini Psathaki, O., Voelzmann, A., *et al.*, 2017. Debris buster is a Drosophila scavenger receptor essential for airway physiology. *Developmental biology* 430, 52–68. doi:10.1016/j.ydbio.2017.08.018

Wojtasek, H., Leal, W.S., 1999. Degradation of an alkaloid pheromone from the pale-brown chafer, Phyllopertha diversa (Coleoptera: Scarabaeidae), by an insect olfactory cytochrome P450. *FEBS letters* 458, 333–336. doi:10.1016/S0014-5793(99)01178-3

Wong, K.M., Suchard, M.A., Huelsenbeck, J.P., 2008. Alignment uncertainty and genomic analysis. *Science* 319, 473–476. doi:10.1126/science.1151532

Wölfle, S., Trienens, M., Rohlfs, M., 2009. Experimental evolution of resistance against a competing fungus in Drosophila melanogaster. *Oecologia* 161, 781–790. doi:10.1007/s00442-009-1414-x

Wright, R.C., Nemhauser, J., 2019. Plant Synthetic Biology: Quantifying the "Known Unknowns" and Discovering the 'Unknown Unknowns'. *Plant Physiology* 179, 885–893. doi:10.1104/pp.18.01222

Wu, C., Chakrabarty, S., Jin, M., Liu, K., *et al.*, 2019a. Insect ATP-Binding Cassette (ABC) Transporters: Roles in Xenobiotic Detoxification and Bt Insecticidal Activity. *International Journal of Molecular Sciences 2019, Vol. 20, Page 1064* 20, 2829. doi:10.3390/ijms20112829

Wu, K., Zhang, J., Zhang, Q., Zhu, S., *et al.*, 2015. Plant phenolics are detoxified by prophenoloxidase in the insect gut. *Sci. Rep.* 5, 16823. doi:10.1038/srep16823

Wu, L., Jia, Q., Zhang, X., Zhang, X., *et al.*, 2019b. CYP303A1 has a conserved function in adult eclosion in Locusta migratoria and Drosophila melanogaster. *Insect biochemistry and molecular Biology* 113, 103210. doi:10.1016/j.ibmb.2019.103210

Xiao, J.-H., Yue, Z., Jia, L.-Y., Yang, X.-H., *et al.*, 2013. Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biology* 14, R141. doi:10.1093/molbev/msr121

Xu, H., Lee, S.-J., Suzuki, E., Dugan, K.D., *et al.*, 2004. A lysosomal tetraspanin associated with retinal degeneration identified via a genome-wide screen. *EMBO J* 23, 811–822. doi:10.1038/sj.emboj.7600112

Xu, K., Zheng, X., Sehgal, A., 2008. Regulation of feeding and metabolism by neuronal and peripheral clocks in Drosophila. *Cell Metab* 8, 289–300. doi:10.1016/j.cmet.2008.09.006

Yamada, M., Murata, T., Hirose, S., Lavorgna, G., *et al.*, 2000. Temporally restricted expression of transcription factor betaFTZ-F1: significance for embryogenesis, molting and metamorphosis in Drosophila melanogaster. *Development* 127, 5083–5092.

Yamada, R., Sonobe, H., 2003. Purification, Kinetic Characterization, and Molecular Cloning of a Novel Enzyme Ecdysteroid-phosphate Phosphatase. *Journal of Biological Chemistry* 278, 26365–26373. doi:10.1074/jbc.M304158200

Yamada, R., Yamahama, Y., Sonobe, H., 2005. Release of ecdysteroid-phosphates from egg yolk granules and their dephosphorylation during early embryonic development in silkworm, Bombyx mori. *Zoological Science* 22, 187–198. doi:10.2108/zsj.22.187

Yamanaka, N., Marqués, G., O'Connor, M.B., 2015. Vesicle-Mediated Steroid Hormone Secretion in Drosophila melanogaster. *Cell* 163, 907–919. doi:10.1016/j.cell.2015.10.022

Yamanaka, N., Rewitz, K.F., O'Connor, M.B., 2013. Ecdysone Control of Developmental Transitions: Lessons from Drosophila Research. *Annual review of entomology* 58, 497–516. doi:10.1146/annurev-ento-120811-153608

Yang, J., McCart, C., Woods, D.J., Terhzaz, S., *et al.*, 2007. A Drosophila systems approach to xenobiotic metabolism. *Physiol Genomics* 30, 223–231. doi:10.1152/physiolgenomics.00018.2007

Yang, R.S.H., Wilkinson, C.F., 1973. Sulphotransferases and phosphotransferases in insects. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* 46, 717–726. doi:10.1016/0305-0491(73)90116-8

Yang, Y., Yoo, C.G., Guo, H.-B., Rottmann, W., *et al.*, 2017a. Overexpression of a Domain of Unknown Function 266-containing protein results in high cellulose content, reduced recalcitrance, and enhanced plant growth in the bioenergy crop Populus. *Biotechnol Biofuels* 10, 1. doi:10.1186/s13068-017-0760-x

Yang, Y., Yoo, C.G., Winkeler, K.A., Collins, C.M., *et al.*, 2017b. Overexpression of a Domain of Unknown Function 231-containing protein increases O-xylan acetylation and cellulose biosynthesis in Populus. *Biotechnol Biofuels* 10, 12. doi:10.1186/s13068-017-0998-3

Yang, Z., 1996. Phylogenetic analysis using parsimony and likelihood methods. *J Mol Evol* 42, 294–307. doi:10.1007/BF02198856

Yassin, A., Debat, V., Bastide, H., Gidaszewski, N., *et al.*, 2016. Recurrent specialization on a toxic fruit in an island Drosophila population. *Proc Natl Acad Sci U S A* 113, 4771–4776. doi:10.1073/pnas.1522559113

Yoshiyama, T., Namiki, T., Mita, K., Kataoka, H., *et al.*, 2006. Neverland is an evolutionally conserved Rieske-domain protein that is essential for ecdysone synthesis and insect growth. *Development* 133, 2565–2574. doi:10.1242/dev.02428

Young, N.J., Webster, S.G., Jones, D.A., Rees, H.H., 1991. Profile of embryonic ecdysteroids in the decapod crustacean, Macrobrachium rosenbergii. *Invertebrate Reproduction & Development* 20, 201–212. doi:10.1080/07924259.1991.9672200

Young, N.J., Webster, S.G., Rees, H.H., 1993. Ecdysteroid profiles and vitellogenesis in Penaeus monodon (Crustacea: Decapoda). *Invertebrate Reproduction & Development* 24, 107–117. doi:10.1080/07924259.1993.9672340

Younus, F., Chertemps, T., Pearce, S.L., Pandey, G., *et al.*, 2014. Identification of candidate odorant degrading gene/enzyme systems in the antennal transcriptome of Drosophila melanogaster. *Insect biochemistry and molecular Biology* 53, 30–43. doi:10.1016/j.ibmb.2014.07.003

Yu, Q.-Y., Fang, S.M., Zhang, Z., Jiggins, C.D., 2016. The transcriptome response of Heliconius melpomene larvae to a novel host plant. *Mol Ecol*. doi:10.1111/mec.13826

Yu, S.J., 2008. Detoxification Mechanisms in Insects, in: *Encyclopedia of Entomology*. Springer, Dordrecht, Dordrecht, pp. 1187–1201. doi:10.1007/978-1-4020-6359-6_891

Yueh, C., Rettenmaier, J., Xia, B., Hall, D.R., *et al.*, 2019. Kinase Atlas: Druggability Analysis of Potential Allosteric Sites in Kinases. *J. Med. Chem.* 62, 6512–6524. doi:10.1021/acs.jmedchem.9b00089

Zahiri, R., Holloway, J.D., Kitching, I.J., Lafontaine, J.D., *et al.*, 2011. Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). *Systematic Entomology* 37, 102–124. doi:10.1111/j.1365-3113.2011.00607.x

Zelensky, A.N., Sanchez, H., Ristic, D., Vidic, I., *et al.*, 2013. Caffeine suppresses homologous recombination through interference with RAD51-mediated joint molecule formation. *Nucl Acids Res* 41, 6475–6489. doi:10.1093/nar/gkt375

Zelhof, A.C., Yao, T.P., Evans, R.M., McKeown, M., 1995. Identification and characterization of a Drosophila nuclear receptor with the ability to inhibit the ecdysone response. *Proc Natl Acad Sci U S A* 92, 10477–10481. doi:10.1073/pnas.92.23.10477

Zeng, B., Zhan, S., Wang, Y., Huang, Y., *et al.*, 2016. Expansion of CRISPR targeting sites in Bombyx mori. *Insect biochemistry and molecular Biology* 72, 31–40. doi:10.1016/j.ibmb.2016.03.006

Zeng, Sen, R., Wen, Z., Niu, G., Berenbaum, M.R., 2013. Aflatoxin B1: Toxicity, bioactivation and detoxification in the polyphagous caterpillar Trichoplusia ni. *Insect Sci.* 20, 318–328. doi:10.1111/1744-7917.12007

Zhang, G., Luo, Y., Li, G., Wang, L., *et al.*, 2014a. DHRSX, a novel non-classical secretory protein associated with starvation induced autophagy. *Int J Med Sci* 11, 962–970. doi:10.7150/ijms.9529

Zhang, H.H., Luo, M.J., Zhang, Q.W., Cai, P.M., *et al.*, 2019. Molecular characterization of prophenoloxidase-1 (PPO1) and the inhibitory effect of kojic acid on phenoloxidase (PO) activity and on the development of Zeugodacus tau (Walker) (Diptera: Tephritidae). *Bull Entomol Res* 109, 236–247. doi:10.1017/S0007485318000470

Zhang, S.-Q., Che, L.-H., Li, Y., Liang, D., *et al.*, 2018. Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nature Communications* 9, 205. doi:10.1038/s41467-017-02644-4

Zhang, X., Carter, M.S., Vetting, M.W., San Francisco, B., *et al.*, 2016. Assignment of function to a domain of unknown function: DUF1537 is a new kinase family in catabolic pathways for acid sugars. *PNAS* 113, E4161–9. doi:10.1073/pnas.1605546113

Zhang, X., Goodsell, J., Norgren, R.B., Jr, 2012. Limitations of the rhesus macaque draft genome assembly and annotation. *BMC Genomics* 13, 222. doi:10.1186/1471-2164-13-206

Zhang, Z., Elser, J.J., Cease, A.J., Zhang, X., *et al.*, 2014b. Grasshoppers Regulate N:P Stoichiometric Homeostasis by Changing Phosphorus Contents in Their Frass. *PLoS ONE* 9, e103697. doi:10.1371/journal.pone.0103697

Zhao, J., Xu, L., Sun, Y., Song, P., *et al.*, 2019. UDP-Glycosyltransferase Genes in the Striped Rice Stem Borer, Chilo suppressalis (Walker), and Their Contribution to Chlorantraniliprole Resistance. *International Journal of Molecular Sciences 2019, Vol. 20, Page 1064* 20, 1064. doi:10.3390/ijms20051064

Zhao, T., Gu, T., Rice, H.C., McAdams, K.L., *et al.*, 2008. A Drosophila Gain-of-Function Screen for Candidate Genes Involved in Steroid-Dependent Neuroendocrine Cell Remodeling. *Genetics* 178, 883–901. doi:10.1534/genetics.107.082487

Zheng, Y., Wang, J.-L., Liu, C., Wang, C.-P., *et al.*, 2011. Differentially expressed profiles in the larval testes of Wolbachia infected and uninfected Drosophila. *BMC Genomics* 12, 595–12. doi:10.1186/1471-2164-12-595

Zhou, X., Zhou, B., Truman, J.W., Riddiford, L.M., 2004. Overexpression of broad: a new insight into its role in the Drosophila prothoracic gland cells. *Journal of Experimental Biology* 207, 1151–1161. doi:10.1242/jeb.00855

Zhou, Y., Fu, W.-B., Si, F.-L., Yan, Z.-T., *et al.*, 2019. UDP-glycosyltransferase genes and their association and mutations associated with pyrethroid resistance in Anopheles sinensis (Diptera: Culicidae). *Malar J* 18, 907. doi:10.1186/s12936-019-2705-2

Zhu, F., Parthasarathy, R., Bai, H., Woithe, K., *et al.*, 2010. A brain-specific cytochrome P450 responsible for the majority of deltamethrin resistance in the QTC279 strain of Tribolium castaneum. *Proc Natl Acad Sci U S A* 107, 8557–8562. doi:10.1073/pnas.1000059107

Zhuo, R., 2014. Caffeine sensitivity in Drosophila melanogaster. University of Alberta, Alberta. doi:10.7939/R3N33X

Ziemann, M., Eren, Y., El-Osta, A., 2016. Gene name errors are widespread in the scientific literature. *Genome Biol* 17, 80. doi:10.1186/s13059-016-1044-7

Zimmer, C.T., Garrood, W.T., Singh, K.S., Randall, E., *et al.*, 2018. Neofunctionalization of Duplicated P450 Genes Drives the Evolution of Insecticide Resistance in the Brown Planthopper. *Current Biology* 28, 1–7. doi:10.1016/j.cub.2017.11.060

Zirin, J., Cheng, D., Dhanyasi, N., Cho, J., *et al.*, 2013. Ecdysone signaling at metamorphosis triggers apoptosis of Drosophila abdominal muscles. *Developmental biology* 383, 275–284. doi:10.1016/j.ydbio.2013.08.029

Zirin, J., Hu, Y., Liu, L., Yang-Zhou, D., *et al.*, 2020. Large-Scale Transgenic Drosophila Resource Collections for Loss- and Gain-of-Function Studies. *Genetics* 214, 755–767. doi:10.1534/genetics.119.302964

Appendices

Appendix 1: Ubiquitous misexpression of EcKLs in *Drosophila melanogaster*

A1.1. Introduction

In order to discover candidate EcKLs that might catabolise developmentally essential endogenous substrates (ala. Chapter 5), I conducted a small screen using the very strong, ubiquitous *tub-GAL4* driver and a collection of 14 pre-existing UAS-ORF lines from the Zurich ORFeome Project (FlyORF; Bischof *et al.* 2013) and the Drosophila Protein Interaction Map transgenic fly resource (DPiM; Guruharsha *et al.* 2014) for 13 EcKLs, to see if misexpression of any of these genes disrupted normal development.

A1.2. Materials and Methods

UAS-ORF responder lines for 13 EcKLs were obtained from FlyORF at the University of Zurich (Bischof *et al.* 2013) or the DPiM transgenic fly resource at the Bangalore Fly Resource Center (Guruharsha *et al.* 2014); FlyORF lines contain 3xHA-tagged ORFs, while DPiM lines contain FLAG-HA-tagged ORFs. FlyORF EcKL lines used were F002982 (*CG31300*), F002821 (*CG10560*), F002832 (*CHKov2*; '*CHKov2*^{1}') and F002521 (*CG9259*). DPiM lines used were 817 (*CG10562*), 854 (*CHKov2*; '*CHKov2*^{2}'), 1332 (*JhI-26*), 2262 (*CG14314*), 2439 (*CG5644*), 2866 (*CG10514*), 3380 (*CG6830*), 3774 (*CG31102*), 3915 (*CG31087*) and 3916 (*CG31288*). *w*^{1118} (VDRC stock 60000) was used as a wild-type control in the absence of true matched genetic backgrounds. *tub-GAL4*/TM3, *actGFP*, *Ser*^{1} and *phm-GAL4*/CyO, *actGFP* (Guittard *et al.* 2011) were a kind gift of Philip Batterham (The University of Melbourne).

UAS-ORF responder (and *w*^{1118}) males were crossed to *tub*-GAL4/TM3, *actGFP*, *Ser*^{1} females, which were allowed to lay on lab media food, and the offspring were left to develop at 25 ºC for 14 days. Adult offspring were collected after eclosion and scored for the presence or absence of the TM3, *actGFP*, *Ser*^{1} balancer chromosome (and the TM6B, *Antp*^{Hu}, *Tb*^{1} chromosome in the case of DPiM line 3774). UAS-*CG5644* males were also crossed to *phm-GAL4*/CyO, *actGFP* females; adult offspring were scored for the presence or absence of the CyO, *actGFP* chromosome. The 'binom.test' function in

R was used to test deviations from expected Mendelian ratios (1:1 for all crosses except those involving DPiM line 3774, which was 1:3)—significant deviations against misexpression genotypes were considered evidence for developmental lethality due to misexpression of the EcKL ORF in question.

A1.3. Results

Misexpression of *CG6830* (Dro24-2, DS = 3) and *CG5644* (Dro42-0, DS = 0) caused developmental arrest before the adult stage, but was only semi-lethal in both cases (Fig. A1.1A). Crosses involving the UAS-*CG31102* construct resulted in significantly more misexpression individuals than expected, which might be a result of lower developmental viability of the TM3, *actGFP*, Ser^{1}/TM6B, $Antp^{Hu}$, Tb^{1} genotype. The UAS-*CG5644* line was further crossed to the *phm-GAL4* driver to test if misexpression in the ecdysteroidogenic prothoracic gland also arrested development. This was also semi-lethal, with 66 misexpression adults to 115 balancer adults (proportion = 0.365, 95% CI: 0.295–0.439, p = 0.0003)

While misexpression of *CG31288* did not appear to affect developmental progression, I did notice a visible phenotype for *tub>CG31288* adult flies—their meconium was a dark grey, almost black colour, compared to the yellow-green of non-misexpressing flies (Fig. A1.1B–D). This colour change was apparent both within the posterior abdomen of recently eclosed flies (Fig. A1.1B) and when the meconium had been excreted and left on empty puparia (Fig. A1.1C–D).

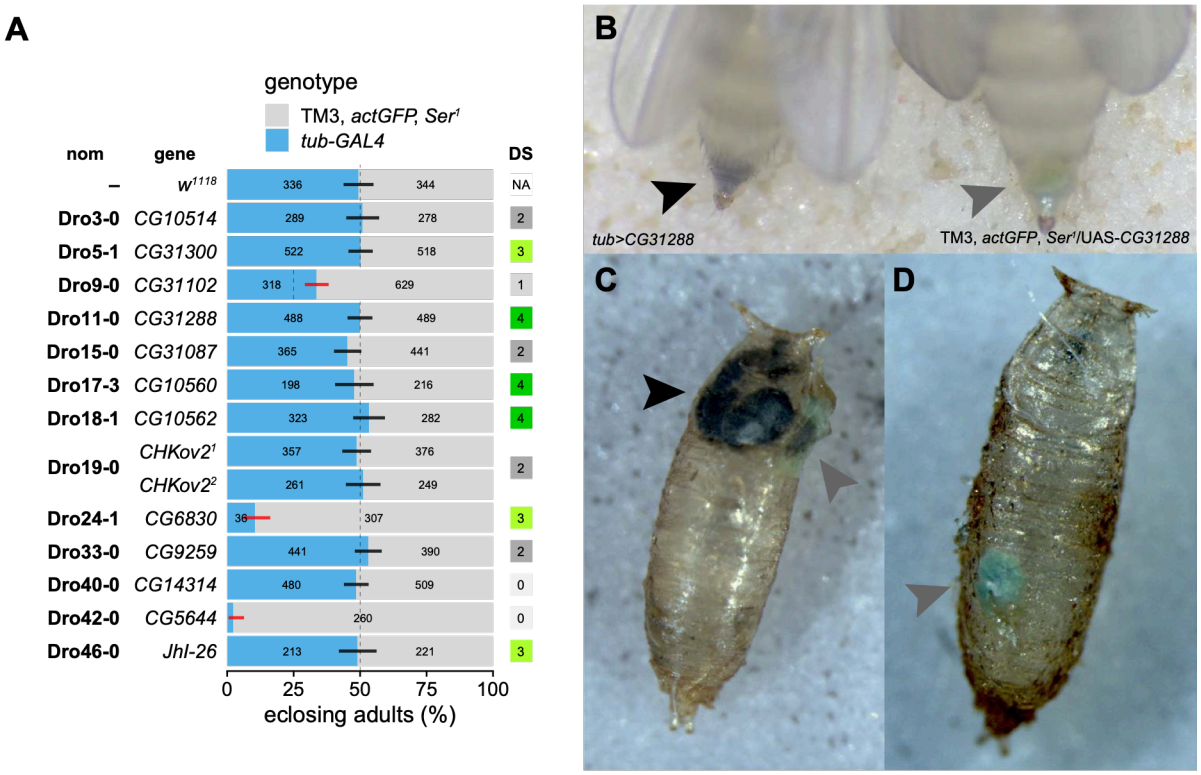


Figure A1.1. (A) Egg-to-adult viability of the misexpression of 13 EcKLs (and one control genotype, w^{1118}), estimated from the adult genotypic ratios of offspring from crosses between *tub-GAL4*/TM3, *actGFP*, Ser^{1} females and UAS-responder males at 25 ºC. The dashed line indicates the expected 1:1 genotypic ratio (or the 1:3 genotypic ratio for *CG31102*) if both genotypes per cross are equally developmentally viable; error bars are 99.7% confidence intervals (95% CI adjusted for 15 tests) for the proportion of *GAL4*-containing heterozygotes; black and red bars indicate non-significant or significant deviations, respectively, from expected genotypic ratios after correction for multiple tests. Numbers on the bars are the number of adults of each genotype (for numbers greater than six). nom, *Drosophila* EcKL nomenclature (Chapter 2); DS, detoxification score; NA, not applicable for w^{1118}. (B) Meconium of freshly eclosed female flies, visible within the posterior tip of the abdomen before excretion. Black arrowhead, dark meconium of a *tub>CG31288* fly; grey arrowhead, wild-type (green) meconium of a TM3, *actGFP*, Ser^{1}/UAS-*CG31288* fly. (C) Meconium deposited on the empty puparium of a fly from the cross between *tub-GAL4*/TM3, *actGFP*, Ser^{1} and UAS-*CG31288*. Black arrowhead, dark meconium of a *tub>CG31288* fly; grey arrowhead, wild-type (green) meconium of a TM3, *actGFP*, Ser^{1}/UAS-*CG31288* fly. (D) Yellow-green meconium deposited on the empty puparium of a wild-type w^{1118} fly (grey arrowhead).

A1.4. Discussion

Developmental arrest from ubiquitous misexpression of *CG6830* (Dro24-1/Dip4/subfamily H) or *CG5644* (Dro42-0/Dip14/subfamily D) is consistent with a hypothesis that the encoded enzymes of these genes can use developmentally essential endogenous molecules as substrates. However, whether this reflects the native function of either gene in development, or simply promiscuous enzymatic activity, is an open question.

In the case of *CG6830*, whose detoxification score (DS; Chapter 2) is 3, this may be a case of promiscuous activity of a xenobiotic-metabolising enzyme acting on an essential molecule, as has been hypothesised for *CG31104* (Dro5-2) and *CG13658* (Dro5-5; see Chapter 4). Interestingly, *CG6830* is a dual-domain EcKL—the unique enzymatic and structural properties of these members of the gene family have yet to be studied.

In the case of *CG5644*, whose DS is 0, this may indeed reflect its native function in development, and partially penetrant developmental arrest upon misexpression in the PG cells of the ring gland may be consistent with an ecdysteroid substrate. *CG5644* appears to be expressed only in the nervous system (Leader *et al.* 2018), with no expression before the end of metamorphosis (Graveley *et al.* 2011). Curiously, *CG5644* is also expressed in migratory cells in the adult ovary (Wang *et al.* 2006), despite negligible expression in the whole ovary (Leader *et al.* 2018), raising the possibility that CG5644 could be the ecdysteroid 22-kinase responsible for the E 22-phosphate present in the ovary (Grau *et al.* 1995; Pis *et al.* 1995). *CG5644* is the sole member of the D subfamily of EcKLs in *D. melanogaster*, a clade that—while not highly retained across all insect taxa—tends to be present in single-copy (Chapter 3), consistent with a dosage-sensitive function (Waterhouse *et al.* 2011).

For other genes, a lack of developmental arrest is notable. Ubiquitous misexpression of *JhI-26* (Dro46-0) did not cause developmental lethality, consistent with previous research (Liu *et al.* 2014). Misexpression of *CG14314* (Dro40-0/Dip13/subfamily C) also did not cause arrest, despite its high retention at single-copy across insects, suggesting it might not encode an enzyme that catabolises an essential endogenous molecule.

Also of note, a meconium colour-change phenotype was observed for adults misexpressing *CG31288* (DS = 4, Dro11-0/Dip4/subfamily H), with dark grey/black meconium instead of the wild-type yellow-green colour. As *CG31288* is a strong detoxification candidate gene, this may simply reflect promiscuity of its encoded enzyme; however, understanding what compound(s) are being metabolised in the meconium by CG31288 may offer clues as to its possible xenobiotic substrates. Meconium is formed from the remnants of the larval and pupal midguts during metamorphosis (Hakim *et al.* 2010; Takashima *et al.* 2011) and contains a number of different compounds, from pigments to ecdysteroids (Kürsteiner 1961; Schwartz *et al.* 1989). To the best of my knowledge, it is not known what causes the yellow-green colour of *D. melanogaster* meconium, but hydroxylated pterins (Kürsteiner 1961) are good candidates for CG31288 substrates.

Appendix 2: *CG31098* may be required for choriogenesis in *Drosophila melanogaster*

A2.1. Introduction

In *Drosophila melanogaster* (Diptera: Drosophilidae), 26 of the 51 EcKL genes in this species are found within a large cluster on chromosome 3R (Fig. 4.1). In 2012, my supervisor Charlie Robin conducted a cross between the w^{1118};; *Df(3R)Exel6202*, *P{*w^{+mC}*=XP-U}Exel6202*/TM6B, $Antp^{Hu}$, Tb^{1} (BL7681) and w^{1118};; *Df(3R)BSC852*/TM6C, Sb^{1}, cu^{1} (BL27923) genotypes, which contain chromosomal deficiencies that overlap within this large cluster, deleting or disrupting seven EcKLs when in *trans*: *CG31098* (Dro8-0), *CG31102* (Dro9-0), *CG31097* (Dro10-0), *CG31288* (Dro11-0), *CG13659* (Dro5-7), *CG31370* (Dro5-8) and *CG31436* (Dro5-10). *Df(3R)Exel6202* spans 222.7 kb, starting 20 bp upstream of the transcriptional start site of *CG31098*, while *Df(3R)BSC852* spans 37.4 kb, completely deleting the *CG31098* locus and ending 15 bp downstream of the transcriptional start site of *CG31436* (Fig. A2.1A). Charlie found that *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygote females had bloated abdomens and failed to lay eggs; examination of their ovaries showed that oocytes had defects in their dorsal appendages (DAs).

In 2013–14, during my MSc (Genetics) research project, I replicated this result, finding that the *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygote genotype was semi-viable to the adult stage, with 34–35% viability compared to each deficiency-over-balancer genotype (X^2 = 122.52, $p < 2.2\times10^{-16}$; Fig. A2.1B). As found before by Charlie, *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygote females had bloated abdomens compared to the parental genotypes (Fig. A2.1C) and failed to lay eggs even after mating with many males. Ovaries dissected from mated *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygote females were malformed compared to the parental genotypes, with short or missing DAs and a lack of normal patterned chorion (eggshell; Fig. A2.1D). Male *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygotes did not appear to be infertile when mated to w^{1118} females.

Of the seven genes disrupted in *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygotes,

only one—*CG31098* (Dro8-0)—has appreciable expression in the adult ovary (Leader *et al.* 2018), making this a leading candidate gene for this phenotype.

A2.2. Materials and Methods

Fly lines were obtained from the Bloomington Drosophila Stock Center (BDSC): w^{1118} (BL5905), w^{1118};; *Df(3R)BSC848*/TM6C, Sb^1, cu^1 (BL29027), w^{1118}; *Df(3R)FDD-0247625*/TM6C, Sb^1, cu^1 (BL27403), w^{1118};; *Df(3R)Exel6202*, $P\{w^{+mC}$=*XP-U}Exel6202*/TM6B, $Antp^{Hu}$, Tb^1 (BL7681) and w^{1118};; *Df(3R)BSC852*/TM6C, Sb^1, cu^1 (BL27923).

Crosses to generate transheterozygotes were carried out on yeast-cornmeal-molasses media (http://bdsc.indiana.edu/information/recipes/molassesfood.html) at 25 ºC. The 'chisq.test' function in R was used to perform a Chi-squared test of multinomial goodness of fit (for MSc data). The 'MultinomCI' function in the *DescTools* package in R was used to calculate confidence intervals for multinomial proportions (for MSc data). Virgin transheterozyote (or w^{1118} control) females were mated to w^{1118} males for 24 hours and allowed to lay on agar juice plates (Appendix 3.2) supplemented with yeast paste for 12 hours. Female flies were dissected in Ringer's solution (3 mM $CaCl_2$, 182 mM KCl, 46 mM NaCl, 10 mM Tris base, pH adjusted to 7.2 with HCl).

A2.3. Results

To test the hypothesis that disruption of *CG31098* produces these reproductive defects, I put two additional deficiencies, *Df(3R)FDD-0247625* and *Df(3R)BSC848*, in *trans* with *Df(3R)Exel6202*. *Df(3R)FDD-0247625* and *Df(3R)BSC848* both span 59.2 kb, ending 19 bp upstream of the transcriptional start site of *CG31098*, overlapping *Df(3R)Exel6202* by 1 bp—similar to *Df(3R)BSC852*, they act as complete loss-of-function alleles for *CG31098*, but leave intact the other six EcKL genes disrupted in *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygotes (Fig. AS2.1A). *Df(3R)Exel6202*/*Df(3R)BSC848* transheterozygotes and *Df(3R)Exel6202*/*Df(3R)FDD-0247625* transheterozygotes were viable (quantitative data was not recorded for these crosses), and showed very similar infertility defects to *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygotes, with abnormal oocyte morphology and DA defects (Fig. A2.1E).

Regular chorionic patterning (see Niepielko *et al*. 2014) appeared particularly disrupted, with normal patterning visible in mature eggs of wild-type females but not those of transheterozygous females (Fig. A2.1F).

Curiously, when trying to recapitulate the cross between BL7681 and BL27923 during my PhD, I could not produce viable *Df(3R)Exel6202*/*Df(3R)BSC852* adults (quantitative data was not recorded for these crosses); however, I could still compare these phenotypes to pre-recorded phenotypic data from 2013–2014, revealing some differences. *Df(3R)Exel6202*/*Df(3R)BSC848* and *Df(3R)Exel6202*/*Df(3R)FDD-0247625* females—while never laying fertilised or viable eggs after mating—did lay a small number of (abnormal) eggs on agar juice plates, unlike *Df(3R)Exel6202*/*Df(3R)BSC852* females observed during my MSc project, which never laid any eggs. Eggs of *Df(3R)Exel6202*/*Df(3R)BSC848* and *Df(3R)Exel6202*/*Df(3R)FDD-0247625* females also appeared slightly less abnormal than those of *Df(3R)Exel6202*/*Df(3R)BSC852* females, with the former usually having distinct, relatively long DAs, and the latter typically having very stunted or absent DAs.

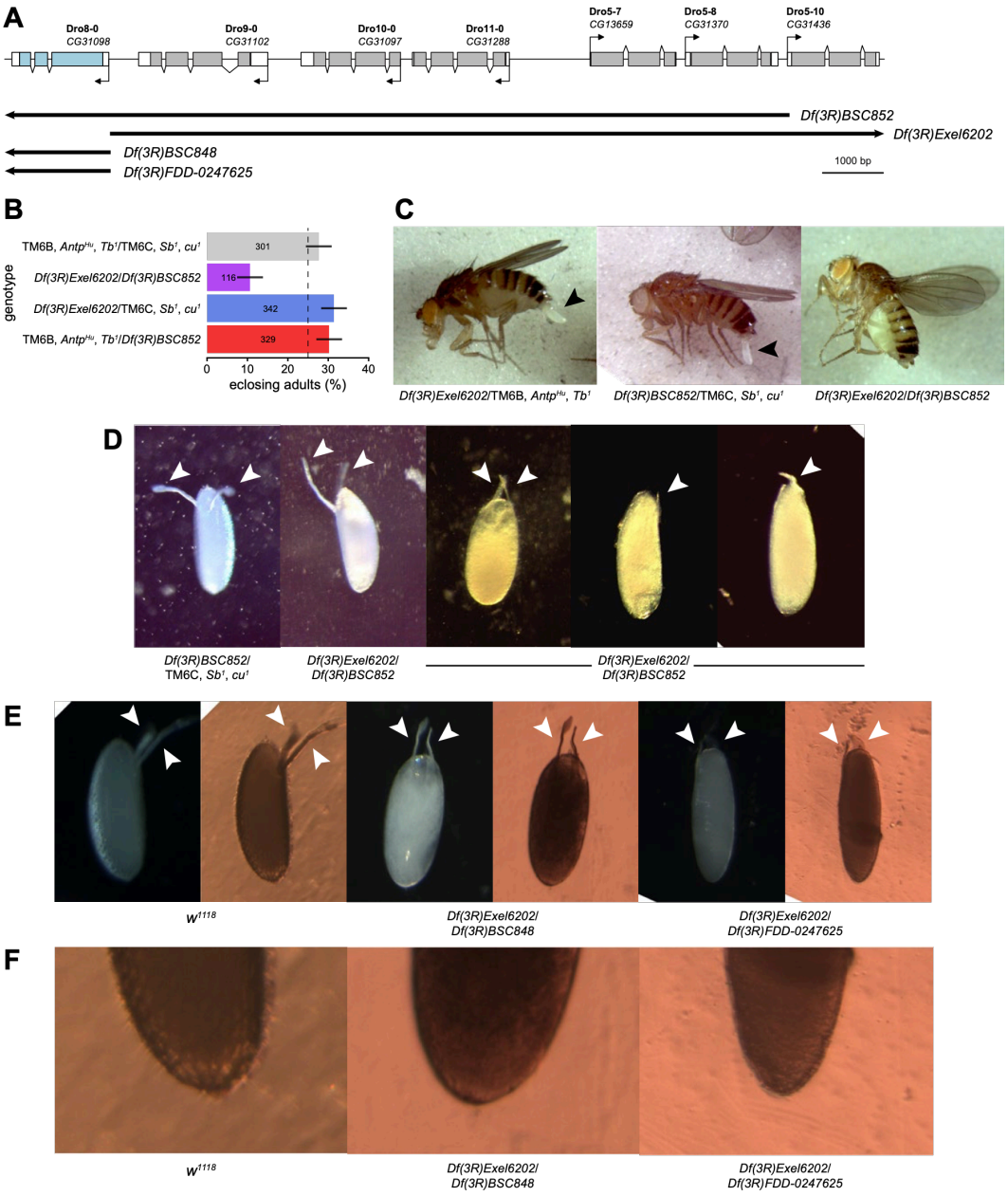


Figure A2.1. (A) Genes disrupted by the *Df(3R)BSC852*, *Df(3R)Exel6202*, *Df(3R)BSC848* and *Df(3R)FDD-0247625* deficiencies (black lines; arrowhead indicates where deficiency continues out of frame). *CG31098* (Dro8-0) has blue exons, all other genes have grey exons; UTRs are white. (B) Adult genotypic ratios from a cross between w^{1118};; *Df(3R)Exel6202*, *P{w*$^{+mC}$*=XP-U}Exel6202*/TM6B, *Antp*^{Hu}, *Tb*^{1} females and w^{1118};; *Df(3R)BSC852*/TM6C, *Sb*^{1}, *cu*^{1} males, conducted during my MSc project. (C) Female flies of the parental w^{1118};; *Df(3R)Exel6202*, *P{w*$^{+mC}$*=XP-U}Exel6202*/TM6B, *Antp*^{Hu}, *Tb*^{1} and w^{1118};; *Df(3R)BSC852*/TM6C, *Sb*^{1}, *cu*^{1} genotypes, and *Df(3R)Exel6202*/*Df(3R)BSC852* transheterozygotes. Note the extruding mature eggs of the parental genotypes (black arrowheads) and the bloated abdomen of the transheterozygote. Photos taken during my MSc project. (D) Oocytes dissected from w^{1118};; *Df(3R)Exel6202*, *P{w*$^{+mC}$*=XP-U}Exel6202*/TM6B, *Antp*^{Hu}, *Tb*^{1}, w^{1118};; *Df(3R)BSC852*/TM6C, *Sb*^{1}, *cu*^{1} and *Df(3R)Exel6202*/*Df(3R)BSC852* females. White arrowheads indicate the dorsal appendages (DAs). Different colours are due to different lighting conditions. Photos taken during my MSc project. (E) Dark-field (images 1, 3 and 5) and backlit (images 2,4 and 5) images of oocytes dissected from w^{1118}, *Df(3R)Exel6202*/*Df(3R)BSC848* and *Df(3R)Exel6202*/*Df(3R)FDD-0247625* females. White arrowheads indicate the dorsal appendages (DAs). (F) Zoomed-in parts of the backlit images of (E). Note the regular chorionic patterning of the egg from the wild-type (w^{1118}) female and the lack of such pattern on oocytes from transheterozyote females.

A2.4. Discussion

A2.4.1. *CG31098* is likely responsible for the reproductive defects seen in deficiency transheterozygotes

These additional experiments strongly suggest that disruption of *CG31098* (Dro8-0) produces the reproductive defects originally associated with the transheterozygous *Df(3R)Exel6202/Df(3R)BSC852* genotype, as it is likely the only gene disrupted in all four deficiencies. Whether the *Df(3R)Exel6202* deficiency is a hypo- or amorphic allele of *CG31098* remains to be determined—while it removes most of the promoter of the gene, *CG31098* may still be expressed at low levels. As such, a true loss-of-function genotype for *CG31098* might produce even more severe phenotypes than those seen here.

I noticed differences in the phenotypes between *Df(3R)Exel6202/Df(3R)BSC848* and *Df(3R)Exel6202/Df(3R)FDD-0247625* females and *Df(3R)Exel6202/Df(3R)BSC852* females, with the former genotypes seemingly having slightly less severe defects, with longer DAs and the ability to lay some eggs—whether this is due to genotypic or experimental (such as different lab media or photoperiod) differences is unclear. A reason to suspect that experimental differences may be behind the phenotypic differences is that *Df(3R)Exel6202/Df(3R)BSC852* transheterozygotes were seemingly inviable during this project, while previously they eclosed with a viability around one-third of their siblings. Barring explanations such as stock mislabelling or other human errors, some environmental variables are likely to have changed between 2013–2014 and 2016, given that a different laboratory space was used for each project, as well as different fly media: the experiments during my MSc project were conducted on yeast-cornmeal media (BDSC 'cornmeal media': https://bdsc.indiana.edu/information/recipes/bloomfood.html) and experiments during this project were conducted on yeast-cornmeal-molasses media (https://bdsc.indiana.edu/information/recipes/molassesfood.html). Nutrition can alter many aspects of *D. melanogaster* physiology and development (Ormerod *et al.* 2017; Piper *et al.* 2005; Shingleton *et al.* 2017), raising the possibility that *Df(3R)Exel6202/Df(3R)BSC852* transheterozygotes are partially viable on the first media, but inviable on the second media.

Regardless of this inconsistency, the reduction in viability of the

Df(3R)Exel6202/Df(3R)BSC852 genotype—either partial in my MSc project or complete during this project—suggests that *CG31098* might be involved in development as well as reproduction. *CG31098* is expressed in essentially all tissues (Leader *et al.* 2018) and throughout the *D. melanogaster* life cycle (Graveley *et al.* 2011) However, given that the relative viabilities of the *Df(3R)Exel6202/Df(3R)BSC848* and *Df(3R)Exel6202/Df(3R)FDD-0247625* genotypes compared to their balancer-containing offspring were not determined, I cannot reject the hypothesis that either of the six other genes disrupted in *Df(3R)Exel6202/Df(3R)BSC852* transheterozygotes are required for development. While loss-of-function alleles of *CG13659* (Dro5-7), *CG31370* (Dro5-8) and *CG31436* (Dro5-10) can individually complement *Df(3R)BSC852* (Chapter 4), the other three genes—*CG31102* (Dro9-0), *CG31097* (Dro10-0) and *CG31288* (Dro11-0)—might be required for development. Indeed, RNAi knockdown of *CG31102* was found to arrest development before the adult stage (Chapter 2), although it is unclear if this is due to on- or off-target knockdown.

CG31098 is a member of the Dro8 clade in *Drosophila* and the Dip5 clade in Diptera, which is perfectly retained in all 35 dipteran genomes annotated in Chapter 3, at very low copy-number—only *Culex quinquefasciatus* (Nematocera: Culicidae) has more than one copy of the gene, with three paralogs. Such strong conservation suggests an important function in the biology of Diptera, consistent with its likely role in reproduction and chorion formation demonstrated here, and the molecular function of *CG31098* deserves much further study.

A2.4.2. Possible molecular functions of *CG31098*

In this work, viable offspring were never produced by transheterozygous deficiency females; this was likely due to both an arrest in oocyte maturation and defects in choriogenesis, as the micropyle—the chorionic structure that allows sperm penetration (Loppin *et al.* 2015)—was likely not functional in these oocytes, and the oocytes themselves may not have been mature enough to undergo fertilisation regardless.

The regulation and evolution of choriogenesis is well understood in many insect model species, including *D. melanogaster* (Papantonis *et al.* 2015), and disruption of many pathways and individual genes can produce similar defects to those seen here, such as DNA damage repair (Alexander *et al.* 2016), prostaglandin synthesis (Tootle

& Spradling 2008; Tootle *et al.* 2011), *gurken* and *decapentaplegic* (Peri & Roth 2000; Serbus *et al.* 2011), *tramtrack* (Ge *et al.* 2015; Peters *et al.* 2013) and the JNK pathway (Suzanne *et al.* 2001). However, an intriguing hypothesis that I would like to propose is that CG31098 functionally interacts with Ftz-f1.

Ftz-f1 is a nuclear receptor that acts as a competency factor in various developmental and reproductive processes in insects, including metamorphosis (Broadus *et al.* 1999) and reproduction (Knapp *et al.* 2020). Notably, *ftz-f1* expression is regulated in a complex manner by ecdysteroids, in that the ecdysteroid titre must sequentially rise to a peak and then substantially decline before it is induced (Akagi *et al.* 2016; Yamada *et al.* 2000)—this suggests that ecdysteroid inactivation may be directly linked to *ftz-f1* expression and function. Recently, *ftz-f1* was found to play a role in the maturation of oocytes, including the formation of the chorion and the dorsal appendages (Knapp *et al.* 2020). Overexpression or knockdown of the ecdysteroid-inactivating enzyme Cyp18a1 does not appear to affect *ftz-f1* expression in the ovary (Knapp *et al.* 2020), suggesting other factors may control its regulation by ecdysteroids. That disruption of *CG31098* leads to chorionic and possibly ovulation defects raises the possibility that this EcKL regulates the expression of *ftz-f1*, possibly by locally repressing ecdysteroid signalling in the ovary. Ecdysteroid 22-phosphate conjugates are known to be present in the *D. melanogaster* ovary (Grau *et al.* 1995; Grau & Lafont 1994; Pis *et al.* 1995), raising the possibility that CG31098 might be an ecdysteroid 22-kinase.

Future experiments on *CG31098* should test this hypothesis by exploring the expression of *ftz-f1* in *CG31098*-deficient ovaries and seeing if transgenic expression of *ftz-f1* can rescue the disruption of *CG31098*, as has been done for *Cyp18a1* mutants during metamorphosis (Rewitz *et al.* 2010). Independent *CG31098* null mutants should also be generated to validate the phenotypic effects of the transheterozygous *CG31098* disruption genotypes generated here, and UAS-*CG31098* constructs should be constructed to test if they can rescue *CG31098* mutant phenotypes.

Appendix 3: Supplementary Methods

A3.1. Single fly DNA extraction protocol

Single flies were placed in a 200 μL PCR tube and crushed for 10 sec with a 100 μL pipette tip filled with 50 μL of squishing buffer (10 mM Tris.HCl (pH = 8.2), 1 mM EDTA, 25 mM NaCL, 200 μg/mL Proteinase K), then the buffer liquid was completely voided into the tube. Tubes were incubated at 37 ºC for 30 min then 85 ºC for 5 min. DNA preps were stored at -20 ºC. Adapted from http://francois.schweis-guth.free.fr/protocols/Single_fly_DNA_prep.pdf.

A3.2. Agar juice plates

For 1 L media (100 plates): 12.5 g sucrose and two lots of 10 g agar were each separately dissolved in 250 mL dH_2O (750 mL total) in the microwave, and added to 250 mL of boiled apple juice (final concentrations: 2% w/v agar, 3.125% w/v sucrose, 25% v/v apple juice). 10 mL of media was aliquoted into 60 mm plates and stored at 4 ºC until use.

A3.3. Yeast-sucrose fly media

For 100 mL media (20 vials): 5 g inactive yeast (Macro Foods), 5 g sucrose and 1 g agar were added to 100 mL of dH_2O and boiled in the microwave until agar had fully dissolved (significant foaming). Liquid was cooled to 60 ºC, after which 1.74 mL 10% Tegosept in 95% EtOH and 0.94 mL of propionic acid and orthophosphoric acid mix was added (final concentrations: 5% w/v inactive yeast, 5% w/v sucrose, 1% w/v agar, 0.38% v/v propionic acid, 0.039% v/v orthophosphoric acid, 0.174% w/v Tegosept, 1.65% v/v EtOH). 5 mL of media was aliquoted into each vial and stored at 4 ºC until use.