Adaptive evolution of Annual ryegrass (*Lolium rigidum*, L.) to herbicides in agroecosystems

Jefferson Paril
ORCID: 0000-0002-5693-4123

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

December 2020

Faculty of Science
School of BioSciences
The University of Melbourne

Abstract

Significance: The rapid evolution of herbicide resistance in weeds provides a tractable system to study adaptive evolution. Investigations into the dynamics of herbicide resistance development in agroecosystems will advance our understanding of evolutionary biology, as well as guide the formulation of recommendations to control herbicide-resistant weeds.

Materials and Methods: Accurate and efficient strategies to characterise and monitor weed populations are required. Populations can be monitored directly by gathering phenotype and genotype data; as well as indirectly by predicting phenotypes using genomic prediction. This thesis investigated the adaptive evolution of weed populations with respect to herbicides. We used *Lolium rigidum* (rigid/annual ryegrass) as the model weed species because of its ability to evolve resistance to many herbicides with different modes of action. The herbicides used were clethodim (fatty acid biosynthesis inhibitor), glyphosate (aromatic amino acid biosynthesis inhibitor), sulfometuron (branched-chain amino acid biosynthesis inhibitor), terbuthylazine (photosystem II disruptor), and trifluralin (tubulin disruptor).

Results: The results of this research can be summarised as:

1. Characterising a landscape by sampling more populations using low-resolution low-cost pool sequencing (Pool-seq) will yield better QTL detection and genomic prediction accuracies than sampling less populations with high-resolution high-cost individual sequencing.
2. *L. rigidum* populations were collected across South-East Australia, and a colorimetric phenotyping method (instaGraminoid) was developed to cost-effectively quantify herbicide resistance using photographs.
3. Pool-sequencing coupled with genome complexity reduction (Pool-ddRADseq) with MseI and NsiI restriction enzymes, followed by filtering at 99% read and mapping accuracies, 227X minimum depth, and 0.001 minor allele frequency were predicted to result in accurate allele frequency estimates while retaining 600,000 SNPs.
4. Genome-wide association experiments using instaGraminoid-derived phenotype data and Pool-seq-derived genotype data show evidence of highly polygenic herbicide resistance traits and the genetic basis of cross-resistance.
5. The putative resistance alleles are more likely to have originated from standing genetic variation than from *de novo* mutations.
6. Genomic prediction using population-level data can be useful in herbicide resistance monitoring and management by identifying potentially resistant populations at the early stages of resistance evolution.

Conclusion and recommendations: Herbicide resistance is a polygenic trait with many loci conferring cross-resistance. It is rapidly evolving from the rich reservoir of standing genetic variation. Herbicide application remains to be a cost-effective approach to control weeds, but it will inevitably become ineffective with the evolution of cross-resistance. I propose a genomically-informed rotation of chemical and non-chemical control strategies. This involves using genomic prediction to cost-effectively and rapidly monitor herbicide resistances. Depending on the predicted levels of resistance, different herbicide mixtures or non-chemical weed control methods can be recommended.

Declaration

I declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution. This thesis consists only of my original work towards the degree of Doctor of Philosophy except where due reference has been made in the text and as indicated in the preface. This thesis is less than the 100,000 word-limit for a Doctoral Degree thesis excluding tables, bibliographies and appendices.

Jefferson Paril
December 2020

Preface

This thesis includes one original paper published in a peer reviewed journal and one submitted for review. The ideas, development and writing up of all the papers in this thesis were the principal responsibility of myself, the student, working within the School of BioSciences under the supervision of Alexandre Fournier-Level. These papers are listed below:

Thesis Chapter	Publication Title	Status	Journal	Co-authors
2	Optimising sampling design and sequencing strategy for the genomic analysis of quantitative traits in natural populations	Revisions submitted in 2021	Molecular Ecology Resources	Alexandre Fournier-Level, and David J. Balding
3	instaGraminoid, a novel colorimetric method to assess herbicide resistance, identifies patterns of cross-resistance in Annual ryegrass	Published in 2019	Plant Phenomics	Alexandre Fournier-Level

The data generated from this research project are stored in a secure University of Melbourne server.

This work is partially funded by the Research Grant Support Scheme from the Faculty of Science, and the Computation Biology Research Initiative of the University of Melbourne.

Acknowledgements

My sincerest thanks to my supervisor, Alex Fournier-Level, for his mentorship, patience, and insights during my PhD.

I also express my gratitude to my co-supervisor, David Balding, for his guidance and helpful discussions.

I appreciate the help and support from Roger Cousens and Kay Hodgins of my PhD committee. I also thank Steve Elefteriadis for the technical support in the greenhouse, Steve Wilcox for his technical support in DNA library preparation, as well as the farmers and agronomists of South-East Australia for allowing us to take samples from their paddocks.

Thanks to my lab mates, as well as to the Golz, Haydon, Andrianopoulos, and Murray labs, especially Ben, Emma, Harshini, Irma, Andhika, Andrew, Tara, Tania, Yu, Mike T., Qiwei, Maheshika, Xiang and Ted.

Finally, special thanks to my wonderful partner, Rhona, and my family for the love and support.

……..

In accordance to the tradition for dissertations concerning evolutionary biology, I present an apt quote from the field's founder:

"*In Staffordshire, on the estate of a relation where I had ample means of investigation, there was a large and extremely barren heath, which had never been touched by the hand of man; but several hundred acres of exactly the same nature had been enclosed twenty-five years previously and planted with Scotch fir. The change in the native vegetation of the planted part of the heath was most remarkable, more than is generally seen in passing from quite different soil to another: not only the proportional numbers of the heath plants were wholly changed, but twelve species of plants (not counting grasses and carices) flourished in the plantations, which could not be found on the heath.*"
- Darwin 1859, p69

In the face of anthropogenic environmental changes, may it be the establishment of a farm or the application of herbicides, nature will adapt and variability is the key to success.

……..

Table of Contents

Abstract....2
Declaration....3
Preface....4
Acknowledgements....5
Table of Contents....6
List of Tables....9
List of Figures....10
List of Tables in the Appendices....14
List of Figures in the Appendices....15
1. Introduction....20
1.1. Aims and scope....23
1.2. Herbicide resistance evolution in weeds....24
1.2.1. Weed biology....24
1.2.2. Weed control....25
1.2.3. Herbicides....26
1.2.4. Herbicide resistance....28
1.2.5. *Lolium rigidum*: the world's most herbicide resistant weed species....31
1.3. Quantitative herbicide resistance....33
1.3.1. Herbicide resistance as a quantitative trait....33
1.3.2. High-throughput phenomics....34
1.3.3. Quantifying herbicide resistance from RGB photographs....36
1.4. Genomic characterisation of a non-model species....39
1.4.1. Sequencing-based genomic characterisation....39
1.4.2. Reduced-representation and pool sequencing....41
1.4.3. Application in *Lolium rigidum*....42
1.5. Genome-wide association and genomic prediction of herbicide resistance in weeds....44
1.5.1. History and theory of GWAS and GP....44
1.5.2. Phenotype and genotype data acquisition for GWAS and GP....46
1.5.3. Population structure....48
1.5.4. GWAS models....49
1.5.5. Genomic prediction models....55
1.6. Evolutionary paths towards herbicide resistance in *L. rigidum* populations across SE Australia....60
1.6.1. Effective population size....60
1.6.2. Mutation rate and fitness cost....60
1.6.3. Genomic signatures of selection....61
1.6.4. Genetic differentiation and gene flow....62
1.6.5. Inferring likely models of herbicide resistance evolution....62
2. Optimising sampling design and sequencing strategy for the genomic analysis of quantitative traits in natural populations....65
2.1. Abstract....65
2.2. Introduction....65
2.3. Materials and methods....67
2.3.1. Workflow overview....67
2.3.2. Landscape simulations....67
2.3.3. Genome-wide association and trait prediction based on polygenic scores....69
2.3.4. Sampling strategy optimisation....70
2.3.5. Implementation....71
2.4. Results....71
2.4.1. GPAS model representatives and the relationships of GPAS performance with landscape and sampling parameters....71
2.4.2. How many populations to sample and when to use Indi-seq or Pool-seq?....72
2.4.3. Which populations to select under different landscape scenarios?....74
2.5. Discussion....77
2.5.1. How many populations to sample?....77
2.5.2. When to use Indi-seq or Pool-seq?....77

2.5.3. Which populations to select under different landscape scenarios?....78
2.6. Conclusion....79
3. Quantitative measures of herbicide resistance in weeds....82
3.1. Introduction....82
3.2. instaGraminoid, a novel colorimetric method to assess herbicide resistance, identifies patterns of cross-resistance in annual ryegrass....82
3.2.1. Abstract....82
3.2.2. Introduction....82
3.2.3. Materials and Methods....84
3.2.4. Results....88
3.2.5. Discussion....92
3.2.6. Conclusion....95
3.3. Materials and Methods....96
3.3.1. Weed populations and herbicides....96
3.3.2. Validation of instaGraminoid on field populations collected as live plants....97
3.3.3. South-East Australian populations characterisation....97
3.3.4. Intra-population characterisation of 8 SE Australian populations....97
3.4. Results....98
3.4.1. Validation of instaGraminoid on field populations collected as live plants....98
3.4.2. South-East Australian populations characterisation....98
3.4.3. Intra-population characterisation of 8 SE Australian populations....100
3.5. Discussion....101
3.5.1. Quantitative herbicide resistance measurement using instaGraminoid....101
3.5.2. Intra-population herbicide resistance....102
3.5.3. Herbicide resistance in SE Australia....102
3.6. Conclusion....102
4. Accurate Allele Frequencies from Pool-ddRADseq....104
4.1. Introduction....104
4.2. Materials and Methods....105
4.2.1. Genetic materials and pooling....105
4.2.2. DNA extraction and sequencing....105
4.2.3. Reference genome and sequence alignment....106
4.2.4. Unbiased reduced-representation of the Lolium genome through ddRADseq....106
4.2.5. Accurate allele frequency estimation from Pool-ddRADseq....107
4.3. Results....109
4.3.1. Reference genome and sequence alignment....109
4.3.2. Unbiased reduced representation of the Lolium genome through ddRADseq....109
4.3.3. Accurate allele frequency estimated from Pool-ddRADseq....111
4.4. Discussion....116
4.4.1. Genome complexity reduction for high and unbiased genome coverage....116
4.4.2. SNP filtering to retain a genome-saturating number of SNPs with accurate allele frequency estimates....117
4.5. Conclusion....118
5. Evolutionary quantitative genetics of herbicide resistance in weeds....120
5.1. Introduction....120
5.2. Materials and Methods....121
5.2.1. Plant material....121
5.2.2. Phenotyping and genotyping....122
5.2.3. GWAS....123
5.2.4. Gene ontology analysis....123
5.2.5. Composite likelihood of convergent evolutionary models....124
5.2.6. GP accuracy....125
5.3. Results....126
5.3.1. GWAS and gene ontology enrichment analysis....126
5.3.2. Composite likelihood of convergent evolutionary models....132
5.3.3. GP accuracy....134
5.4. Discussion....135
5.4.1. Polygenicity of herbicide resistance traits and the genetic basis of cross-resistance....135
5.4.2. Herbicide resistance evolution in *Lolium rigidum* across South-East Australia....136

5.4.3. Genomic prediction using Pool-seq data for rapid and cost-effective herbicide resistance assessment....137
5.5. Conclusion....138
6. General discussion and the evolution of herbicide resistance in SE Australian *Lolium rigidum* populations....141
6.1. A model of herbicide resistance evolution in *Lolium rigidum* across SE Australia....141
6.2. Polygenic herbicide resistance....143
6.3. Rapid monitoring of herbicide resistance with genomic prediction and Pool-seq....144
6.4. Recommendations to improve GWAS power and GP accuracy....144
6.4.1. Improvements in phenotyping....145
6.4.2. Improvements in genotyping....145
6.4.3. Improvements in sampling and genotype-to-phenotype mapping....145
6.5. Recommendation to control herbicide resistance in weeds....146
6.6. Conclusion....147
7. Bibliography....149
8. Appendices....187
Chapter 2 Appendix....187
Chapter 4 Appendix....194
Chapter 5 Appendix....199

List of Tables

Table 1.1. Herbicides classifications based on mode of action.....27
Table 1.2. List of some of the known genetic bases of herbicide resistance.....30
Table 1.3. A contingency table for genome wide association study using pooled genotype data describing the number of alleles under different categories.....54
Table 1.4. The types of selective sweeps expected under the different origins of favourable alleles (adapted from Olson-Manning et al. 2012 and Hermisson and Pennings 2005).....61
Table 2.1. List of variables used in landscape simulations.....67
Table 3.1. List of herbicides used to characterise herbicide resistance in two *Lolium rigidum* populations.....84
Table 3.2. Plant health metrics derived from photographs of above-ground tissues of *Lolium rigidum* plants.....86
Table 3.3. Proportion of resistant plants to the four herbicides. Chi-square test (df=1) was applied to test for independence between the two populations.....88
Table 3.4. Proportion of plants correctly classified using the instaGraminoid colorimetric models trained on the populations in the row and validated in populations in the column (all these correlations are different based on Tukey's honest significant difference grouping at α=0.05).....90
Table 3.5. Fraction of entries correctly classified using the instaGraminoid colorimetric models trained on herbicide treatments in the row and validated in herbicide treatments in the column (Letters correspond to Tukey's honest significant difference grouping at α=0.05).....90
Table 3.6. List of the 67 *Lolium rigidum* populations sampled from South-East Australia in 2016 and 2018.....96
Table 3.7. List of herbicides used to characterise herbicide resistance in *Lolium rigidum* populations.....97
Table 5.1. List populations used in the genomic analysis of herbicide resistance of *Lolium rigidum* across South-East Australia, and the corresponding herbicide resistance traits measured (details of the populations are found in Chapter 3 Table 3.6).....122
Table 5.2. List of putatively non-neutral scaffolds for each herbicide resistance trait and the sets of populations used to estimate the likelihoods of different coalescent models of convergent evolution.....125
Table 5.3. Summary of GWAS results and gene ontology enrichment analysis.....128
Table 5.4. List of candidate cross-resistance genes with descriptions obtained from The UniProt Consortium (2019) (continued in the next two pages).....130

List of Figures

Figure 1.1. The number of unique cases of herbicide resistance identified globally from 1945 to 2019 (adapted from Heap, 2020).....29
Figure 1.2. *Lolium rigidum* specimen collected from Murrumbidgee, New South Wales, Australia in 2018 (Scale bar 10 cm).....32
Figure 1.3. Distribution of phenotypic values resembles continuous quantitative variation as the number of loci controlling the trait increases (illustration idea came from Roff (1997); but the assumptions here are biallelic loci with equal additive effects and generated by randomly sampling 10,000 individuals).....33
Figure 1.4. Modelling individual plant survival as a function of continuous quantitative health traits.....34
Figure 1.5. Imaging techniques and their respective operating electromagnetic wavelength range. The sine wave of varying wavelength depicts the electromagnetic spectrum.....35
Figure 1.6. Simulated logistic model describing the relationship between the binary mortality data and the quantitative herbicide resistance values computed from the various image-derived metrics and their corresponding estimated effects.....38
Figure 1.7. Illumina sequencing-by-synthesis via bridge amplification on a lawn of oligonucleotides using fluorescent reversible terminator dNTPs. Panel A: attachment of the adapter ligated DNA into the surface of the flow cell. Panel B: bridge amplification. Panel C: clusters of clonal DNA. Panel D: fluorescence emission in a cluster following laser excitation after DNA polymerase activity using primers and four reversible terminator dNTPs. Panel E: simultaneous fluorescence emissions from multiple clusters. Panel F: fluorescence image across the whole flow cell is captured for every cycle.....40
Figure 1.8. Sequencing costs per megabase from 2001 to 2020 (data from the US National Human Genome Research Institute; Wetterstrand 2020).....41
Figure 1.9. Summary statistics of the *Lolium perenne* genome assembly (adapted from Byrne, 2015).....43
Figure 1.10. An illustration of power analysis. H_0 and H_a are the null and alternative hypotheses, respectively, defined by specific probability density functions, with the latter having a non-centrality parameter proportional to the effect size, $\beta \neq 0$ and sample size. The quantile value, q is the minimum effect size at the false positive rate threshold, α.....47
Figure 1.11. Illustration of the relationships between individuals i, j, and k.....49
Figure 1.12. Expected genomic signatures of selection around a beneficial allele found at the central position when this allele originated from *de novo* mutation (red dashed line), or standing genetic variation (green dash-dotted line), or multiple beneficial alleles from different genetic backgrounds (blue solid line) (adapted from Barrett and Schluter 2008).....61
Figure 1.13. Models of convergent herbicide resistance evolution. The red line represents the allele frequency of the resistance allele (q_i), where q_a is the frequency in the ancestral population and q_{pop1}, q_{pop2}, and q_{pop3} are the frequencies in the 3 resistant populations. The red triangle represents the onset of selection from an herbicide. The red x mark represents a mutation event. The blue dashed line represents migration.....63
Figure 2.1. Panel A: Causal allele diffusion gradients across the simulated landscape. Panel B: Systematic sampling strategy across the simulated landscape (red circles represent sampled populations, while the other colours represent variability in the different populations). Note: For one population sampling, all populations across the landscape were sampled independently.....68
Figure 2.2. Relationships between the proportion of the landscape sampled and GPAS performance. Dots represent means and whiskers indicate ±1 standard deviation from the mean.....72
Figure 2.3. Pool-GPAS can outperform Indi-GPAS under the same sequencing capacity. Panel A: Difference in true positive rate (TPR) between Indi-GPAS and Pool-GPAS

models as the ratio between the number of populations characterised through Indi-seq and Pool-seq increases. Panel B: Difference in false positive rate (FPR) between Indi-GPAS and Pool-GPAS models as the ratio between the number of populations characterised through Indi-seq and Pool-seq increases. Vertical red lines: The range of ratios between the two vertical red lines correspond to cases where Pool-GPAS can outperform Indi-GPAS in our study, i.e. Pool-GPAS has higher TPR (points below the dashed line in panel A) and lower FPR than Indi-GPAS (points above the dashed line in panel B). Moving along the x-axis involves modifying the sequencing depths for Indi-seq or Pool-seq and not the number of individuals or pools sampled per population, since 384 individuals per population and 5 pools per population were kept constant in the simulations.....73
Figure 2.4. Differences in QTL detection and polygenic score prediction accuracies between Indi-GPAS and Pool-GPAS per population. AUC = Area under the ROC curve. RMSE = root mean square error of the polygenic score prediction. The term "*ns*" beside the bars indicates non-significant differences based on HSD at p<0.05. 74
Figure 2.5. Violin plots of the area under the receiver operating curve (AUC; measure of QTL detection accuracy), root mean square error (RMSE; measure of polygenic score prediction accuracy), mean number of polymorphic causal loci per population, and causal allele frequencies across the landscape divided into 10 rows under uniform, unidirectional, and bidirectional causal allele diffusion. Each row is perpendicular to the causal allele diffusion gradient. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05).....75
Figure 2.6. Effects of sampling across the landscape on QTL detection accuracy under varying number of simulated QTL, selection intensities, migration rates, and GPAS models for unidirectional (top graphs) and bidirectional (bottom graphs) causal allele diffusion gradients. QTL detection accuracy was measured using the area under the curve which describes the relationship between power and false positive rate, where high values mean high QTL detection accuracies.....76
Figure 2.7. Effects of sampling across the landscape on prediction accuracy under varying number of simulated QTL, selection intensities, migration rates, and GPAS models for unidirectional (top graphs) and bidirectional (bottom graphs) causal allele diffusion gradients. Polygenic score prediction accuracy was measured using root mean square error where low values mean high accuracies.....76
Figure 3.1. Workflow using instaGraminoid to extract the quantitative herbicide resistance estimates from plant photographs.....85
Figure 3.2. Architecture of the instaGraminoid package.....87
Figure 3.3. Image processing results for four sample plant photographs. (w_0 refers to the expected white background pixel value, and KP refers to ×1000 pixels).....89
Figure 3.4. Histogram of the metric effects used for measuring the quantitative resistance of individual plants. (Top 2 metrics are labelled: $\Delta GREEN_FRACTION_{14\text{-}07}$ - change in green fraction from day 14 to day 7, and $\Delta GREEN_FRACTION_{14\text{-}00}$ - change in green fraction from day 14 to day 0; and uninformative features with standard deviations < 0.02 were omitted).....89
Figure 3.5. Capacity of the colorimetric model to determine resistance using logistic regression. (The histograms show the distribution of the quantitative resistance estimates in the susceptible [bottom] and resistant [top] plants. The red line is the logistic regression function relating the quantitative herbicide resistance to the observed survival data).....90
Figure 3.6. Patterns of herbicide resistance based on (A) binary survival data and (B) quantitative resistance estimates showing the significant correlations.....91
Figure 3.7. Simulated change in herbicide resistance as affected by herbicide rotation and resistance correlations. (Herbicide rotation: sulfometuron, terbuthylazine, glyphosate and trifluralin).....92
Figure 3.8. Logistic regression model for sulfometuron resistance of live plants collected from Lexton, Rokewood, and Skipton. (The histograms show the distribution of the

quantitative sulfometuron resistance estimates in the susceptible [bottom] and resistant [top] plants. The red line is the logistic function relating the quantitative herbicide resistance to the observed survival data).....98
Figure 3.9. Herbicide resistance variation in *Lolium rigidum* populations across South-East Australia collected in November 2018. Herbicide resistances were measured as whole population survival rates. The points are the phenotyped populations. The histogram on the bottom left of every map shows the distribution of herbicide resistance based on the survival rate 28 days after herbicide application.....99
Figure 3.10. Correlations between herbicide resistance traits. The term “ns” refers to non-significant correlation at p<0.05.....100
Figure 3.11. Intra-population quantitative herbicide resistance variations in 8 South-East Australian *Lolium rigidum* populations. Letters above each population-herbicide class refer to Tukey’s honest significant grouping at α=0.05 (CLETH, GLYPH, SULFO and TERBU refers to clethodim, glyphosate, sulfometuron and terbuthylazine resistances, respectively).....101
Figure 4.1. Breadth and depth of sequencing across the Lolium perenne genome based on simulated ddRADseq using different enzyme pairs (only a small subset is shown) and at varying fragment size ranges (letters refer to Tukey’s HSD grouping at α=0.05).....110
Figure 4.2. Breadth and depth of MinIon sequencing coverage using MseI-NsiI, PstI-NsiI and MseI-PstI restriction enzyme pairs (letters refer to Tukey’s HSD grouping at α=0.05).....110
Figure 4.3. Breadth and depth of sequencing coverage as the number of reads increases.....111
Figure 4.4. Allele frequency accuracy and precision simulations.....112
Figure 4.5. Effect of different sequencing depth and minimum allele frequency filtering parameters on allele frequency estimation accuracy and the number of SNPs retained.....113
Figure 4.6. Optimising the compromise between allele frequency accuracy and number of SNPs as functions of minimum depth alone (E[Freq. Dev.]: expected deviation from actual allele frequency, SNP count: number of SNPs).....114
Figure 4.7. Optimising the compromise between allele frequency accuracy and number of SNPs as functions of minimum depth and minimum allele frequency (DEPTH: minimum sequence depth, MAF: minimum allele frequency, E[Freq. Dev.]: expected deviation from actual allele frequency, SNP count: number of SNPs). 115
Figure 5.1. Distributions of the absolute values of the allele effects of loci above the Bonferroni threshold from 8 intra-population GWAS experiments.....127
Figure 5.2. Overlap between candidate gene lists and scaffolds containing at least one herbicide resistance-associated locus. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.....129
Figure 5.3. Relative composite likelihood profile across scaffolds per herbicide resistance trait where the strongest genetic effects were mapped. Model 1 (black): all selected populations have independent mutations of the beneficial allele. Model 2 (blue): all selected populations share the beneficial allele via migration. Model 3 (red): the beneficial allele was standing in the ancestor of all selected populations. Model 4 (green): the two resistant populations inside the parenthesis share beneficial allele via migration, while the other population has an independent mutation at the same locus. Model 5 (yellow): the beneficial allele was standing in the ancestor of the two resistant populations inside the parenthesis, while the other population has an independent mutation at the same locus. Gray vertical line marks the position of the resistance QTL.....133
Figure 5.4. Genomic prediction accuracies per herbicide resistance trait using leave-one-out cross-validation. The models tested are RIDGE for ridge regression, LASSO for least absolute shrinkage and selection operator regression, and MIXED_FST is linear mixed model regression using F_{ST} as the random covariate and restricted maximum likelihood estimation of variance components. R2 is the adjusted coefficient of determination, and RMSE is the root mean square deviation of the prediction from the observation.....134
Figure 5.5. Accuracy of predicting the mean quantitative sulfometuron resistance of Lexton,

Rokewood, and Skipton populations using the GWAlpha-derived allelic effects from Inverleigh and Urana populations. *RMSE* is the root mean square error between predicted and observed values and *n* is the number of populations: one for Lexton and two each for Rokewood and Skipton.....135
Figure 6.1. Model of the adaptive evolution of weeds to herbicides in agroecosystems.....142
Figure 6.2. Genomically-informed rotation of chemical and non-chemical control strategies.. .146

List of Tables in the Appendices

Table S2.1. List of parameters used in Quantinemo2 simulations.....187
Table S2.2. Mean comparison using Tukey's honest significant differences, and the linear (full-factorial) model-derived adjusted coefficients of the GPAS models.....188
Table S4.1. Parameters used in bwa mem for the alignment of Illumina DNA sequencing reads. 194
Table S5.1. List of neutral scaffold used in the estimation of likelihoods of various convergent evolutionary models.....199
Table S5.2. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for clethodim resistance using scaffold_3626|ref0003919 and the following populations: Susceptibles: ACC01,ACC05,ACC06 and Resistants: [ACC27,ACC34] & ACC36.....222
Table S5.3. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for clethodim resistance using scaffold_3626|ref0003919 and the following populations: Susceptibles: ACC06,ACC38,ACC50 and Resistants: [ACC08,ACC09],ACC11.....222
Table S5.4. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for glyphosate resistance using scaffold_198|ref0030880 and the following populations: Susceptibles: ACC01,ACC02,ACC15 and Resistants: ACC18,[ACC47,ACC50].....223
Table S5.5. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for glyphosate resistance using scaffold_198|ref0030880 and the following populations: Susceptibles: ACC38,ACC39,ACC55 and Resistants: [ACC04,ACC09],ACC24.....223
Table S5.6. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for sulfometuron resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC01,ACC19,ACC21 and Resistants: ACC14,[ACC51,ACC53].....224
Table S5.7. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for sulfometuron resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC03,ACC20,ACC57 and Resistants: [ACC11,ACC12],ACC53.....224
Table S5.8. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for terbuthylazine resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC07,ACC21,ACC26 and Resistants: [ACC04,ACC06],ACC55.....225
Table S5.9. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for terbuthylazine resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC07,ACC21,ACC26 and Resistants: ACC09,[ACC60,ACC62].....225

List of Figures in the Appendices

Figure S2.1. Representative plots showing the linear relationships between the actual phenotypic values and the polygenic scores computed from single-SNP estimates using Indi-GPAS and Pool-GPAS models.....189
Figure S2.2. Violin plots of the true positive rate (TPR) for the landscape parameters, GPAS models and the number of populations sampled. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. levels with the same letter are not significantly different at p<0.05).....190
Figure S2.3. Violin plots of the root mean square error (RMSE) for the landscape parameters, GPAS models and the number of populations sampled. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. levels with the same letter are not significantly different at p<0.05).....191
Figure S2.4. Violin plots of GPAS performances across the landscape under unidirectional causal allele diffusion gradient for 10, 50 and 100 simulated QTL. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05).....192
Figure S2.5. Violin plots of GPAS performances across the landscape under unidirectional causal allele diffusion gradient for each level of the simulated selection intensities. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05). 192
Figure S2.6. Violin plots of GPAS performances across the landscape under unidirectional causal allele diffusion gradient for each level of the simulated migration rates. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05).....193
Figure S2.7. Violin plots of GPAS performances across the landscape under bidirectional causal allele diffusion gradient for 10, 50 and 100 simulated QTL. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05).....193
Figure S2.8. Violin plots of GPAS performances across the landscape under bidirectional causal allele diffusion gradient for each level of the simulated selection intensities. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05).....194
Figure S2.9. Violin plots of GPAS performances across the landscape under bidirectional causal allele diffusion gradient for each level of the simulated migration rates. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at p<0.05).....194
Figure S4.1. Distribution of the proportions of Illumina reads mapped to the *Lolium perenne* draft reference genome.....195
Figure S4.2. Simulated depth of RADseq across the *Lolium perenne* genome using pair-wise combinations of 8 restriction enzymes (letters refer to Tukey's HSD grouping at α=0.05).....196

Figure S4.3. Breadth of sequencing coverage across the genome.....196
Figure S4.4. Depth of sequencing coverage across the genome.....197
Figure S4.5. Theoretical minimum number of SNPs required to saturate the *Lolium rigidum* genome assuming a genome size of 2Gb as a function of linkage block size (purple area under the curve and dashed lines indicate the range of SNP counts required based on empirical *Lolium perenne* linkage block sizes ranging from 0.5-3kb.....197
Figure S4.6. Number of SNPs with at least 10X, 100X and 1000X depths as functions of the number of sequencing reads.....198
Figure S4.7. Responses of allele frequency accuracy and number of SNPs to varying stringency of the SNP filtering parameters: minimum sequencing depth, minimum base quality, and minimum allele frequency (MAF).....198
Figure S4.8. Distribution of base calling qualities of the Illumina sequencing data.....199
Figure S4.9. Model comparisons between linear and polynomial models of 1st, 2nd, 3rd and 4th degrees using F-test (RSS: residual sum of squares), Akaike information criterion (AIC) and Bayesian information criterion (BIC).....199
Figure S5.1. Manhattan plots for clethodim resistance using the merged *Lolium ridum* dataset and the linear mixed model with FST as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....201
Figure S5.2. Manhattan plots for clethodim resistance using ACC13 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....202
Figure S5.3. Manhattan plots for clethodim resistance using ACC49 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....203
Figure S5.4. Manhattan plots for glyphosate resistance using the merged *Lolium ridum* dataset and the linear mixed model with FST as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....204
Figure S5.5. Manhattan plots for glyphosate resistance using ACC31 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....205
Figure S5.6. Manhattan plots for glyphosate resistance using ACC62 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....206
Figure S5.7. Manhattan plots for glyphosate resistance using Inverleigh *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....207

Figure S5.8. Manhattan plots for glyphosate resistance using Urana *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....208
Figure S5.9. Manhattan plots for sulfometuron resistance using the merged *Lolium ridum* dataset and the linear mixed model with FST as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....209
Figure S5.10. Manhattan plots for sulfometuron resistance using ACC11 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....210
Figure S5.11. Manhattan plots for sulfometuron resistance using ACC59 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....211
Figure S5.12. Manhattan plots for sulfometuron resistance using Inverleigh *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....212
Figure S5.13. Manhattan plots for sulfometuron resistance using Urana *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....213
Figure S5.14. Manhattan plots for terbuthylazine resistance using the merged *Lolium ridum* dataset and the linear mixed model with FST as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....214
Figure S5.15. Manhattan plots for terbuthylazine resistance using ACC09 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....215
Figure S5.16. Manhattan plots for terbuthylazine resistance using ACC54 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....216
Figure S5.17. Manhattan plots for terbuthylazine resistance using Inverleigh *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni

threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....217
Figure S5.18. Manhattan plots for terbuthylazine resistance using Urana *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -log10(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).....218
Figure S5.19. Venn diagram of candidate clethodim resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.....219
Figure S5.20. Venn diagram of candidate glyphosate resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.....220
Figure S5.21. Venn diagram of candidate sulfometuron resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.....221
Figure S5.22. Venn diagram of candidate terbuthylazine resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.....222

Chapter 1

Introduction and Literature Review

1. Introduction

The remarkable propensity of life to survive is arguably best appreciated in terms of the evolution of xenobiotic resistance. Xenobiotics are any foreign substance capable of entering an organism (Mason et al. 1965), but most contemporary definitions restrict this to toxic substances (Schwass and Finley 1985; Juchau and Chen 1998; Croom 2012; Evans 2013), these include antibiotics and pesticides. Xenobiotic resistance is one of the most tractable and important systems to study quantitative trait evolution. Xenobiotic resistance is increasingly found to be polygenic and measured quantitatively (Tabashnik and Cushing 1989; Humphry et al. 2002; Busi et al 2013; Frey et al. 2015; Najarro et al. 2015; Cui et al. 2015; Lukačišinová and Bollenbach 2017). The intense selection pressures imposed by xenobiotics coupled with the large population sizes and short-generation times of the pests and pathogens they target, drive evolution at an accelerated rate (Neve and Powles 2005; Hermsen et al. 2012; Fardisi et al. 2019).

The overuse and misuse of antibiotics in clinical and agricultural settings have led to the evolution of superbugs (Lindsay and Holden 2004; Häusler 2006; Ashworth et al. 2014). These are strains of bacteria that cannot be controlled by most or even all of the currently used antibiotics (Davies and Davies 2010; Adegoke et al. 2016). The same trend is happening with insecticide resistance (Lemon 1994; Battlay et al. 2016; South and Hastings 2018; Dusfour et al. 2019) and herbicide resistance (Beckie and Tardif 2012; Beckie et al. 2019; Heap 2020). Although this seemingly perpetual race between humanity developing new xenobiotics and pathogens and pests evolving resistance to them, is a sustainable business model for the entities developing these xenobiotics, the limited availability of new antibiotics to fight diseases and new pesticides to control pests can be catastrophic. Therefore, there is a need to understand the genetic basis of xenobiotic resistance and the evolutionary processes that lead to it, so that strategies can be developed to delay resistance evolution.

Among all the pesticides, herbicides are the most intensively and extensively used globally (Lichtfouse 2015; FAO 2020; Roser 2020). Weeds can cause as much as 50% crop yield losses (Yaduraju and Rao 2013; Peter 2016), and in Australia, weeds cause $708 million loss in revenue per year (Llewellyn et al. 2016). The intense selection pressure weeds face in agroecosystems, their large effective population sizes and high genetic variability, enable the rapid and repeated evolution of herbicide resistance (Manalil 2014; Baucom 2016; Baucom 2019). Elucidating the genetic bases of and the evolutionary mechanisms leading to herbicide resistance will aid the development of improved strategies to control and mitigate crop yield loss.

Before we can start discussing the approaches available to dissect the genetic basis and the evolutionary mechanisms involved in the development of herbicide resistance, it is valuable to discuss the history of evolution and quantitative genetics. An in-depth analysis of this history until the 1970s was provided by Provine in his book “The origins of theoretical population genetics” (Provine 2001). A continuation of this history was provided by Charlesworth and Charlesworth in their review entitled “Population genetics from 1966 to 2016” (Charlesworth and Charlesworth 2017). Various other historical accounts and reviews are mentioned below.

Emergence of the evolutionary theories: from Darwin to Fisher and Wright

Among the evolutionary theories in the 19th century, the Lamarckian evolution of heritable somatic variations in response to the environment, was the most attractive even decades after Darwin published his book “On the origin of species by natural selection or the preservation of favoured races in the struggle for life” (Darwin 1859). However, after years of genetics research, the Lamarckian view as the general phenomenon to explain evolution was eventually superseded by Darwin’s theory of evolution through natural selection (Wright 1931). Darwin recognised the stochastic nature of heritable variation on which selection acts on in a continuous and gradual manner.

After the rediscovery of Mendel’s work in 1900, the conflict between Biometricians and Mendelians arose. Biometricians have been firm advocates of continuous evolution. They

supported Darwin's theory of gradual continuous evolutionary change (Provine 2001). On the other hand, Mendelians were advocates of discontinuous or discrete evolution. In 1918, Fisher merged these two schools of thought with his paper "The Correlation between Relatives on the Supposition of Mendelian Inheritance" (Fisher 1918). He rigorously demonstrated that Mendelism and Darwinian selection on small continuous variation were complementary. This was followed by the corroboration of the work by Haldane in his 1924 paper: "A Mathematical Theory of Natural and Artificial Selection-I" (Haldane 1924), and culminating with Wrights' 1931 paper: "Evolution in Mendelian populations" (Wright 1931). These works have laid the Mendelian basis of quantitative variation and the mutation-selection model of evolution. By the middle of the 20th century, the theoretical body of knowledge on population genetics stemming from Fisher, Wright and Haldane continued to expand building on empirical data using quantitative traits and fitness measures, together with the discovery of the chromosomal and biochemical basis of genetic variation (Charlesworth and Charlesworth 2017).

The 1950's saw the debate between the "classical" and the "balance" view of variability, where Muller and Dobzhansky, respectively were the most well-known proponents of (Charlesworth and Charlesworth 2017). The classical view postulates that functional wild type alleles dominate populations and deleterious mutant alleles are at low frequencies. On the other hand, the balance view posits that many genes have multiple functional alternative alleles which are maintained at intermediate frequencies by balancing selection. Before nucleotide sequencing, electrophoretic variants, i.e. proteins with varying mobility on agarose gel under an electric field, were shown to be abundant and the balance view was thought to have prevailed. However, Lewontin and Hubby (1966) pointed out that this does not necessarily imply balancing selection, it can also indicate neutral or nearly neutral variants.

Selectionists and neutralists

Wright in the 1930s recognised the importance of random genetic drift in shaping genetic variation within populations (Wright 1931; Wright 1938), but it was overshadowed by the "selectionist" view of the predominance of loci under selection. The 1960s to the 1980s saw the resurgence of the "neutralist" view, which posits that most of the molecular variation is neutral or nearly neutral such that the difference between and within species are under the control of drift and mutation rather than selection (Kimura 1968; Kern and Hahn 2018; Jensen et al. 2019; Kimura 2020). The utility of the neutralist view was demonstrated by Kimura in 1968 with his use of the "molecular clock" where the rate of sequence substitution between species is equivalent to the neutral mutation rate (Kimura 1968).

The neutral theory of evolution was rigorously formulated by Kimura and Ohta in the early 1970s (Ohta and Kimura 1971; Kimura and Ohta 1974). The first methods of interrogating nucleotide sequence variation developed in the 1970s were instrumental in revealing the abundance of neutral variants and the rarity of nonsynonymous variation in *Drosophila melanogaster* (Langley et al. 1982; Aquadro et al. 1986). This was confirmed in 1983 by the ground-breaking work of Kreitman using the Maxam–Gilbert technique (differential purine-pyrimidine cleavage) to sequence 11 copies of the alcohol dehydrogenase locus of *Drosophila melanogaster* (Kreitman 1983). Powerful tests to study the molecular evolution of natural populations were developed, including tests of departure from neutrality, stemming from Ewen's concept of the scaled mutation rate, $\theta = 4 N_e \mu$, where N_e is the effective population size (i.e. the size of an idealised panmictic population with the same inbreeding rate as the observed population (Crow and Kimura 1970)) and μ is the mutation rate. Watterson, Nei and Tajima developed methods to estimate this parameter and applied them on empirical sequence data (Watterson 1975; 1978; Nei and Tajima 1981; Tajima 1983; Tajima 1989). The neutral theory opened the door for a better appreciation of the role of drift in natural populations of finite size (Jensen et al. 2019), which is particularly relevant in ecology where endangered species are expected to have reduced effective population sizes. From the backdrop of neutral theory came the coalescent theory (Kingman 1982). It simplified the analysis of neutral sequence variation by enabling backward-in-time simulations instead of forward-in-time simulations, with the subsequent incorporation of the effects of linkage disequilibrium (LD), selection, migration, and demographic factors (Wakeley 2009).

Neutral theory has not been without opposition. Gillespie argued that the effects of neutral

evolution can also be explained by models of selection in spatially and temporally variable environments (Gillespie 1994). Kern and Hahn in 2018 have concluded that a large body of work using genomic sequence data rejects the neutral theory since genomes appear to be largely shaped directly and indirectly by selection. However, Jensen et al in 2019 heavily criticised the narrow definition of neutral theory and the weak and flawed empirical evidence presented by Kern and Hahn (2018). They argue that current empirical data are evidence of the straightforward extension of Kimura's neutral theory, demonstrating its continued importance rather than rejecting it.

Dissecting the genetic basis of quantitative traits

Parallel to developments in population genetics is the rapid advancement of quantitative genetics. Galton, Weldon and Pearson were the pioneers of quantitative genetics prior to Fisher's synthesis (Provine 2001; Benirschke 2004). Fisher's 1918 paper "The Correlation between Relatives on the Supposition of Mendelian Inheritance" is not only the landmark paper which established evolution in the light of Mendelian genetics but it is also the seminal paper on modern quantitative genetics.

Before genotyping technologies could be used to characterise genomic variants, the classical approaches to quantitative genetics analysis consisted of estimating the number of loci controlling the trait (Castle 1921; Wright 1984), and performing variance components analysis to estimate heritability. Total phenotypic variance can be partitioned into its various genetic, environmental, and residual components by exploiting the relationships between parents, offsprings, and siblings, usually generated from artificial crosses in plants and animals (Roff 1997; Hallauer 2007).

Linkage disequlibrium (LD) mapping, first performed by Sturtevant in 1913 on the 6 sex-linked qualitative traits of *Drosophila melanogaster* (Sturtevant 1913), was the precursor to quantitative trait loci (QTL) mapping using genome-wide markers. The advent of next generation of sequencing (NGS) technologies in the 21st century saw a huge leap in genotyping throughput. This paved the way for the elucidation of the genetic basis of quantitative traits using QTL mapping and genome-wide association studies (GWAS). QTL mapping exploits the variation derived from synthetic crosses, i.e. biparental populations, and multi-parental population, to detect major QTL controlling quantitative traits (Zeng 2005). GWAS, on the other hand, uses the natural variation found in wild and domesticated populations, and even advanced generational synthetic populations. Tanaka's group in 2002 reported the first genome-wide association study, where they identified a candidate locus associated with susceptibility to myocardial infarction in humans (Ozaki et al. 2002; Ikegawa 2012).

Despite having resolved the problem of missing heritability by recognising the small effects of numerous loci below stringent significance thresholds (Manolio et al. 2009; Jian Yang et al. 2010; Shi et al. 2016), our understanding of the genetic basis of quantitative traits remains largely incomplete (Visscher et al. 2012; Callaway 2017; Boyle et al. 2017). In response to this, an "omnigenic" view of complex traits was put forward by Boyle et al in 2017. This view posits that the gene regulatory networks controlling a trait must be sufficiently interconnected such that most of the heritability can be explained by mechanisms outside the core pathways. This is supported by the observations that association signals for most complex traits are spread across the genome and unbiased towards known trait associated-genes.

Alongside GWAS, methods for predicting quantitative traits without the need to perform significance tests, i.e. genomic prediction (GP) or genomic selection, were conceived. At the beginning of the 21st century, Whittaker et al (2000), and Meuwissen et al (2001) in response to the inefficient use of dense genomic data in QTL mapping experiments, proposed an approach which circumvents candidate loci detection to directly predict genotypic values using genome-wide marker data. In contrast to QTL mapping and GWAS, GP aims to predict quantitative traits without the need to further identify and characterise the genes and the molecular mechanisms involved in the expression of the phenotype.

1.1. Aims and scope

This research project is an attempt at improving our understanding of the adaptive evolution of weeds in agroecosystems by elucidating the genetic and evolutionary bases of herbicide resistance. We have chosen *Lolium rigidum* (rigid/annual ryegrass) populations across South-East (SE) Australia as the species and landscape to study. *L. rigidum* is the world's most successful herbicide-resistant weed species having evolved resistance to the highest number of herbicides with different modes of action, and it is particularly problematic in SE Australian cropping systems. Arable land in this region is very prone to soil erosion particularly wind erosion (Terry 1945). Because of this, low to no tillage practices are preferred, which incentivises the use of herbicides even more as a convenient and cost-effective weed control method. This has led to multiple herbicide-resistant *L. rigidum* populations and has contributed to ranking Australia as the country with the second highest number of reported cases of herbicide resistance after the United States of America (Heap 2020).

Adaptive evolution can be studied in real-time through experimental evolution coupled with sequencing (evolve-and-resequence experiments) (Garland and Rose 2010; Schlötterer et al. 2015). This approach to study evolution is unarguably the gold standard in terms of generating repeatable and verifiable scientific hypotheses for basic research. However, it is not amenable to species or traits that are not adaptable to controlled laboratory conditions such as large organisms with long generation time, endangered species or humans. An alternative approach is to infer the most likely models of adaptive evolution using genomic information from present-day populations which holds the signature of their evolutionary histories (Boake et al. 2002; Kelley et al. 2006). The limitations of this approach are the absence of perfect replications, and the very limited, if any, time resolution. Despite these limitations, analysing genomic signatures of evolutionary and demographic histories has been fruitful (Kelley et al. 2006; Duforet-Frebourg et al. 2016; Nielsen et al. 2018; Cammen et al. 2018; Pankin et al. 2018; Greenbury et al. 2020).

GWAS and genomic scans for signatures of selection are complementary approaches to detect putative QTL controlling fitness-related quantitative traits (You et al. 2018; Igoshin et al. 2019; Yasumizu et al. 2020). Usually it is difficult to untangle demographic and selective effects in genomic data (Li et al. 2012). Performing independent experiments across populations exhibiting variation for the trait of interest to detect the underlying QTL is needed. These putative QTL are then used to infer likely evolutionary models that are expected to differ from those explaining the variation of the genomic background, i.e. not implicated in the trait. Since control or ancestral population data are rarely available, the neutral expectations are generated through numeric simulations. Deviations from these neutral or null expectations relative to the empirical data provides the information to estimate the likelihood of the different models of evolution.

What are the genetic bases of and the evolutionary mechanisms leading to herbicide resistance in agroecosystems? This is the question we answered with this research project. This can be broken down into three research questions:

1. How do we construct efficient sampling, phenotyping, genotyping, association analysis, and genomic prediction frameworks?
2. What are the genetic bases of herbicide resistance traits?
3. How does herbicide resistance evolve in agroecosystems?

1.2. Herbicide resistance evolution in weeds

Among all the pests and diseases of crop plants, weeds can generate the highest potential loss in crop productivity (Oerke 2006). This is because weeds are on the same trophic level as the crop plants they outcompete and have been evolving with since the dawn of agriculture approximately 12,000 years ago. In addition to the direct impact of weeds on crops due to competition, weeds can also harbour other crop pests and diseases (Capinera 2005; Linde et al. 2016). Weeds can cause as much as 34% to 50% crop yield reduction (Oerke 2006; Yaduraju and Rao 2013; Peter 2016). However, this is often under-appreciated, because unlike the conspicuous physical damage caused by insect pests and diseases, the visual manifestation of weed damage is subtle. Additionally, there are many ways to control weeds effectively, ranging from mechanical and chemical to biological strategies. However, this is starting to change due to the over-reliance on chemical control over the past several decades resulting in the evolution of herbicide-resistant weeds.

1.2.1. Weed biology

A weed is generally defined as any plant growing in an unwanted area (Monaco et al. 2002). The term “weed” does not have any botanical meaning. It is a social construct where classifying a plant as weed is dependent upon the situation and the people involved (Monaco et al. 2002; Jussaume et al. 2019). However, it is advantageous to restrict this definition in agricultural context, i.e. a weed species is any plant species growing on agricultural land that is not a crop. Weed scientists have devised weed classification systems which are useful in the context of weed management. These include classifications based on lifespan and leaf morphology. Based on lifespan, weeds can be classified as annual, biennial or perennial (Monaco et al. 2002). Annuals senesce and die after flowering. Similarly, biennials die after flowering but with an intermediate period of quiescence. Perennials are capable of flowering multiple times and continue to grow vegetatively for years. Based on leaf morphology, weeds can be classified as monocots or dicots. Monocot weeds, e.g. grasses and sedges have smaller and narrower leaves than dicots, also referred to as broadleaf weeds (Monaco, et al. 2002).

Weeds compete for the same resources crops need, i.e. light, water, carbon dioxide and soil nutrients. Weeds possess characteristics that enable them to outcompete crops. Baker in 1965 listed 14 characteristics that the “ideal” weed possesses (Baker 1965). These characteristics can be summarised as seed longevity, dormancy and robust germination characteristics (corresponds to items 1 to 3 in Baker’s list); quick and efficient vegetative growth and early flowering (corresponds to the 4th in the list); robust sexual or asexual reproduction (corresponds to items 5 to 13); and the ability to monopolise resources by rosette formation and allelopathy (corresponds to Baker’s last item on the list).

Their robust seed characteristics enable weeds to persist in the field after every cropping season. Seed dormancy is the inability of a viable seed to germinate under favourable conditions (Finch-Savage and Leubner-Metzger 2006). Weed seeds can be dormant for a period of time, resulting in staggered germination and a continual replenishment of the weed seedbank in the soil. Weed seeds can remain viable for years, which is attributable to dormancy, hardseededness, and the presence of secondary metabolites which defer decomposition by microorganisms and herbivory by insects (Roberts 1964; Egley and Chandler 1983; Hendry et al. 1994).

Once germinated, weeds grow quickly, such that the early stages of crop and weed development are the most important period in determining the magnitude of crop yield loss (Bridges 1992; Martin et al. 2001; Monaco et al. 2002). Weeds can flower earlier than crops to avoid their seeds being harvested together with crops, thereby replenishing the soil seedbank (Zhao et al. 2018). Early flowering and maturation can also allow weeds to escape unfavourable conditions that may come later in the cropping season.

The fecundity and flexible reproductive strategies of weeds allow them to spread quickly and persist in agricultural fields despite aggressive control measures. Each individual weed plant can produce hundreds to hundreds of thousands of viable seeds (Stevens 1932; Monaco et al

2002). Some weeds can reproduce by both sexual and asexual means. Sexual reproduction via cross-pollination is favoured when adapting to changing or stressful environments, while, asexual reproduction and self-pollination are favoured under stable favourable conditions (Yang and Kim 2016).

Weeds can outcompete crops by effectively monopolising access to sunlight and nutrients or at the very least by reducing the amount of light and nutrients the crops uptake. Weeds do this by occupying the same area as crops, thereby sharing the same rhizosphere and atmosphere, by out-growing the crops to shade them from direct sunlight (Liu et al. 2009), and by generating an unfavourable soil environment for crop seed germination and growth through allelopathy (Monaco et al. 2002).

In addition to Baker's list of "ideal" weed characteristics is the ability of weeds to develop resistance to biotic and abiotic stresses. This characteristic not only enables weeds to survive and reproduce under unfavourable conditions, but also gives them the ability to harbour other crop pests and diseases (Capinera 2005; Zimdahl 2007).

1.2.2. Weed control

Weeds can be controlled with mechanical, physical, chemical, biological, and cultural strategies. Integrated weed management (IWM) brings these strategies together for a holistic approach to weed control which optimises crop yield while minimising the negative impacts on human health and the environment (Swanton and Murphy 1996).

Mechanical weed control involves the physical removal or displacement of weeds from the field. This is the oldest method of weed control which includes hand-weeding, hoeing, and ploughing. It also includes the newest methods, i.e. automated robotic weed control systems to selectively remove weeds (Slaughter et al. 2008). This can be very effective in removing established weeds and exhausting the weed seedbank with recurrent ploughing prior to planting to induce repeated weed seed germination. However, it is the most labour-intensive method and its application can be limited in areas prone to soil erosion.

Physical weed control involves physical barriers including non-chemical and non-biological strategies to kill weeds and their seeds or to render the environment unsuitable for germination and growth. This includes mulching the soil with plastic sheets (Bennett et al. 1966) or with dead plant material (Crutchfield et al. 1986), soil sterilisation (Horowitz et al. 1983), high temperature treatment with flame (Carter et al. 1960; Parish 1990) or steam and hot water (Banks and Sandral 2007), and high frequency electromagnetic radiation (Menges and Wayland 1974). Flame and radiation treatments are not popular, while hot water or steam treatments are gaining traction in Australia as an alternative to chemical control. This approach can be very effective in high-value horticultural cropping systems like orchards; however its application to large-scale agronomy can be very costly.

Chemical weed control involves the use of synthetic chemical compounds or herbicides to kill weeds by disrupting integral biological processes. Herbicides can be classified in a number of ways including by chemical family, mode of action, specificity (i.e. broad-spectrum or selective), and by timing of application (i.e. pre-emergence or post-emergence). Herbicides can inhibit the function of key enzymes in amino acid synthesis (e.g. sulfonylureas and glycines) and fatty acid biosynthesis (e.g. cyclohexanedione and thiocarbamates). Others cause disruption in photosynthesis (e.g. triazines and diphenylethers), cell division (e.g. dinitroanilines and tetrazolinone) and auxin response (e.g. benzoic acids and pthalamates). While others have unclear and probably multiple target sites (e.g. organic arsenicals). Some herbicides selectively target specific weeds leaving certain crops unharmed (e.g. 2,4-D to kill broadleaf weeds in cereal crops); while others are non-selective or broad-spectrum which indiscriminately can kill both weeds and crops (e.g. glyphosate). Herbicides can be applied pre-emergence, i.e. before the weeds germinate (e.g. trifluralin) and post-emergence, i.e. after germination or once the weeds have been established (e.g. terbuthylazine). Herbicide application is the most convenient, flexible, and cost-effective method of weed control. However, the efficacy of herbicides has been eroding due to the evolution of herbicide-resistant weeds (Heap 2020), and the over-reliance on herbicides have contributed significantly to environmental degradation and

water pollution (Nitschke and Schüssler 1998).

Biological weed control involves the use of insects, livestocks, aquatic animals, and plant pathogens to selectively eliminate weeds and leave the crops unaffected. This also includes crop allelopathy against weeds and the breeding of more competitive crops. Herbivory by insects, grazing/browsing livestock, and fishes are routinely used to control weeds (Willis and Memmott 2005; Schaffner et al. 2020; Silvestri et al. 2020). The idea of using pathogens to control weeds have been around for a long time (Wilson 1969; Harding and Raizada 2015), but has not gained much traction primarily due to the inability to mass-produce the inocula for large-scale use (Charudattan and Dinoor 2000). Crops breeders are developing new cultivars with improved competitiveness and allelopathic effects against weeds (Bertholdsson 2005; Belz 2007; Worthington and Reberg-Horton 2013). These cultivars are designed to outcompete weeds by having more vigorous growth and allelopathy which suppresses weed growth.

Finally, cultural strategies of weed control are combinations of the methods mentioned above but excluding chemical control (Monaco et al. 2002). These include crop rotation, flooding, increased crop density, shallow sowing depth, optimal fertiliser application, exclusion of weed seeds from crop seed stocks, and the use of competitive or better adapted cultivars (Gunsolus 1990; Upadhyaya and Blackshaw 2007).

1.2.3. Herbicides

The first herbicides include sodium chloride, copper sulfate, iron sulfate, sulfuric acid, sodium arsenite, carbon bisulfide, petroleum, sodium chlorate, borates, and dinitrophenol salts (Timmons 1970; Zimdahl 2007; Lichtfouse 2015). Large scale adoption of herbicides started with the discovery of 2,4-D (2,4-dichlorophenoxyacetic acid) during the 1942-1944 period, and the use of herbicides has increased ever since (Timmons 1970). A significant contributor to this wide adoption was the development of herbicide-resistant crops which greatly simplify weed management, e.g. triazine-resistant canola and most notably glyphosate-resistant or Roundup ready™ maize, cotton and soybean (Reddy and Nandula 2012).

Herbicide discovery and commercialisation is a long and resource-intensive process. New herbicides can be developed through random screening for herbicidal activity and through targeted screening (Kudsk and Streibig 2003). Weed scientists and biochemists have been developing herbicides with the recent additional consideration for health and environmental safety (Edwards and Hannah 2014). In 2020, nine new herbicides have recently been developed, one of which is a new mode of action (Condon 2020).

One of the most widely adopted herbicide classification systems is the classification based on mode of action or target site; however there are variations in the grouping and naming schemes (Table 1.1). The Herbicide Resistance Action Committee (HRAC) grouping is the most widely used worldwide. The Weed Science Society of America (WSSA) grouping is used only in the United States and Canada. The Australian classification system (AUS) is used only in Australia. Classification based on timing of application, i.e. pre-emergence or post-emergence is also widely adopted because of its practical utility.

Table 1.1. Herbicides classifications based on mode of action.

HRAC Group*	WSSA Group**	AUS Group†	Mode of Action	Target Pathway	Examples
A	17, 25, 26	A	Acetyl CoA carboxylase (ACCase) inhibition	Fatty acid biosynthesis	Cyclohexanedione and phenylpyrazole (e.g. Clethodim)
B	2	B	Acetolactate synthase (ALS) of Acetohydroxy acid synthase (AHAS) inhibition	Branched amino acid biosynthesis	Imidazolinones and sulfonylureas (e.g. Sulfometuron)
C1, C2, C3	5, 6, 7	C	Photosystem II (PSII) inhibition	Photosynthetic electron transport chain	Phenyl Carbamates and triazine (e.g. Terbuthylazine)
D	22	L	Photosystem I (PSI) inhibition	Photosynthetic electron transport chain	Bipyridyliums
E	14	G	Protoporphyrinogen oxygenase (PPO) inhibition	Chlorophyll and heme biosynthesis	Diphenylethers and triazolinones
F1, F2, F3	11, 12, 13, 27	F, H, Q	Phytoene desaturase, 4-hydroxyphenylpyruvate dioxygenase or unknown protein inhibition	Carotenoid biosynthesis	Amitrole, pyrazoles and triketones
G	9	M	5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) inhibition	Aromatic amino acid biosynthesis	Glycines (i.e. Glyphosate)
H	10	N	Glutamine synthase inhibition	Glutamine biosynthesis	Phosphinic acids
I	18	R	Dihydropteroate synthase inhibition	Folic acid and purine biosynthesis	Carbamates
K1, K2, K3	3, 15, 23	D, E, K	Tubulin depolymerisation enhancement or very long chain fatty acid synthase (VLCFAS) inhibition	Cell division	Benzamides, tetrazolinone and dinitroanilines (e.g. Trifluralin)
L	20, 21	I, O, Z	Cellulose synthase inhibition	Cell wall biosynthesis	Benzamides and nitriles
M	24	Z	Oxidative phosphorylation uncoupling	ATP	Dinitrophenols
N	8, 16	J	Fatty acid elongase inhibition	Fatty acid biosynthesis	Benzofuranes, phosphorodithioates and thiocarbamates
O	4	I	Transport inhibitor response protein 1 stimulation	Auxin response	Benzoic acids and quinoline carboxylic acids (e.g. 2,4-D)
P	19	P	Auxin transport inhibition	Auxin response	Pthalamates and semicarbazones
Z	17, 25, 26	Z	Unknown or poorly understood mechanisms	Unknown or poorly understood mechanisms	Organic arsenicals and arylaminopropionic acids

* HRAC: Herbicide Resistance Action Committee; ** WSSA: Weed Science Society of America; † AUS: Australian herbicide classification system

Herbicides are applied in the field in a variety of ways which aim to optimise efficacy, cost-efficiency, and minimise the negative impacts on the environment and non-target organisms. Chemical companies in collaboration with academic institutions empirically determine the optimal dosage of herbicides for commercial use (Wolf et al. 2000). These recommended doses vary for different crops, target weeds, and location. Herbicide application dose is usually expressed in terms of active ingredient (a.i.) per hectare, e.g. 720 g a.i./ha of glyphosate.

Commercially available herbicides come in a variety of forms including powder, granules, and concentrates which can be water-soluble, emulsifiable, or may require adjuvants to improve homogeneity, adherence to plant tissues and plant absorption. Some herbicides can be mixed with other herbicides and pesticides for more cost-efficient field application. These can be applied in the field using sprayers or by soil incorporation. Sprayers can be as simple as hand-held or knapsack sprayers or as sophisticated as tractors with boom sprayers and aircraft sprayers. Soil incorporation is critical for herbicides which easily volatilise and photodegrade. Personal protective equipment is required to protect the personnel involved.

Herbicides can be pre-emergent, post-emergent, or both. Pre-emergence herbicides (e.g. trifluralin) are applied prior to weed seed germination to inhibit seedling growth. Post-emergence herbicides (e.g. clethodim, glyphosate, & terbuthylazine) are applied to the field with established or growing weeds to be absorbed by the roots and/or leaves. Some herbicides can be used as both pre- and post-emergence herbicides (e.g. sulfometuron). Herbicide manufacturers can specify the optimal time of the day to apply herbicides which can be early morning, midday, late afternoon or other times in between. The rule of thumb is to apply herbicides early in the morning or late afternoon to minimise degradation and volatilisation which can reduce efficacy. However, a 2013 study on common post-emergence herbicides concluded that midday rather than early morning or late evening is the optimal time of application (Stopps et al. 2013). The high temperature and high light stresses during midday probably compounded on the adverse effects of herbicides following successful herbicide absorption.

Herbicides have contributed greatly to the ability of agricultural production to feed the ever-increasing world population. Together with mechanisation, herbicides helped lift many people out of the drudgery of manual labour in the field (Monaco et al. 2002). Additionally, herbicides have simplified and reduced the cost of weed management which is especially advantageous in low to no tillage cropping systems where soil erosion is a big problem, e.g. dry arid plains like in Australia, and mountainous or sloping areas (Weston 1990; D'Emden et al. 2006; Ito et al. 2007).

Despite these positive outcomes, over-reliance on herbicides in the past several decades have resulted in adverse biological and environmental effects. Herbicide run-off contributes to water pollution (Nitschke and Schüssler 1998). Commercially available herbicides are considered to be non-toxic to humans but can be toxic to other non-target organisms (Prado et al. 2009; Hasenbein et al. 2017). Additionally, herbicide application in agroecosystems imposes extreme selective pressure on weed populations resulting in the evolution of herbicide resistance (Jasieniuk et al. 1996; Délye 2013; Délye et al. 2013; Baucom 2016; Baucom 2019).

1.2.4. Herbicide resistance

Herbicide resistance has exacerbated weed-induced crop yield losses (Orson 1999; Llewellyn et al. 2016). Herbicide-resistant weeds are recalcitrant to herbicide treatment at the recommended or historical field application rates. They have evolved in response to the repeated and sustained selection pressures from herbicides (Jasieniuk et al. 1996; Powles and Yu 2010). The first report of herbicide resistance in a weed species was recorded in 1957, i.e. 2,4-D resistance in spreading dayflower (*Commelina diffusa*) (Hilton 1957). However, the first confirmed report was recorded in 1968, i.e. simazine resistance in common groundsel (*Senecio vulgaris*) (Ryan 1970; Monaco et al. 2002). As of the year 2020, there has been a total of 515 unique cases of weed herbicide resistance globally (Figure 1.1) (Heap 2020). This list spans 263 weed species and resistance 23 out of the 26 known herbicide modes of action (Heap 2020).

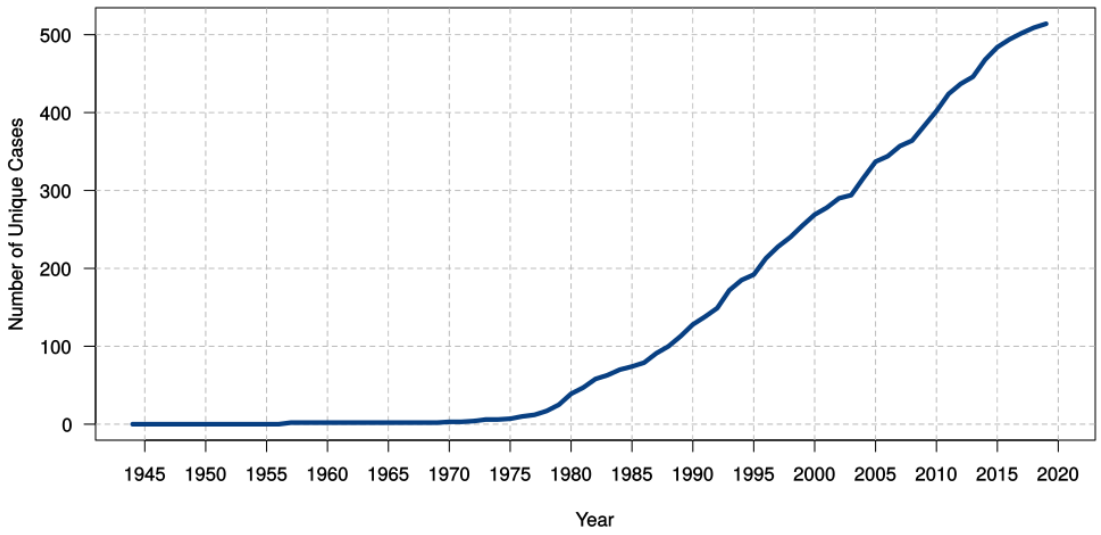


Figure 1.1. The number of unique cases of herbicide resistance identified globally from 1945 to 2019 (adapted from Heap, 2020).

The molecular bases of herbicide resistance in weeds need to be elucidated in order to develop strategies to control this problem. Herbicide resistance mechanisms can be broadly categorised as target site resistance (TSR) and non-target site resistance (NTSR). TSR involves target site alterations, and over-expression, while NTSR includes all the other mechanisms (Délye et al. 2013).

TSR conferred by mutations in the target protein inhibits herbicide molecules from binding while maintaining protein functions. On the other hand, TSR due to the overexpression or over-production of target proteins compensates for the detrimental effects of the herbicide. Most TSR are monogenic and dominant to semi-dominant (Délye et al. 2013), since the presence of at least one TSR allele or herbicide-incompatible target site can at least partially restore the inhibited biochemical pathways.

NTSR involves reduction in herbicide penetration and translocation, enhanced herbicide metabolism, reactive oxygen species neutralisation, and all other mechanisms that are not under TSR (Délye 2013). NTSR are therefore likely to be complex and polygenic. This is the dominant type of resistance to HRAC Groups A and G, and the only type of resistance identified for HRAC Groups D, N, O and Z (Beckie and Tardif 2012; Délye 2013). A cytochrome P450 gene, CYP81A10v7 was found to endow *Lolium rigidum* resistance to multiple herbicides with different modes of action (Han et al. 2020). Together with cytochrome P450 gene family, glutathione S-transferase and glycosyltransferase gene families induce the breakdown of herbicide molecules (Yuan et al. 2007). These NTSR mechanisms can affect a broader range of herbicides with different modes of action and therefore can promote cross-resistance.

Some of the identified genetic bases of herbicide resistance in weeds are listed in Table 1.2. These include the bases of TSR and NTSR for ACCase-, ALS-, PSII-, PPO-, carotenoid biosynthesis-, EPSPS-, and cell division-inhibiting herbicides. Most TSR are conferred by single-base mutations (mostly substitution) that reduce the ability of the herbicide molecules to bind to the altered target sites. The exact mutations that confer NTSR are usually harder to identify than those conferring TSR.

Table 1.2. List of some of the known genetic bases of herbicide resistance.

Gene	HRAC Group	Weed species	Description
Acetyl-CoA carboxylase (ACCase)	A	*Avena fatua, Alopecurus myosuroides, Lolium rigidum, Lolium multiflorum, and Phalaris paradoxa*	Target site resistance (TSR) conferred by amino acid substitutions: Ile-541-Leu, & Ile-1769-Leu (Christoffers et al. 2002); Ile-1781-Leu, Asp-2078-Gly, & Cys-2088-Arg (Yu et al. 2013); Trp-2027-Cys, Ile-2041-Asn, Asp-2078-Gly, & Gly-2096-Ala (Délye et al. 2008); Non-synonymous substitutions at Leu-1781, Gly-2078, Cys-2027, & Asn-2041 (Petit et al. 2010); Ile-1781-Leu, Trp-2027-Cys, Ile-2041-Asn, Asp-2078-Gly, & Cys-2088-Arg (Yu et al. 2007); Ile-1781-Leu, & Gln-1756-Glu (Zhang and Powles 2006); Ile-418-Leu (White et al. 2005); Asp-2078-Gly, & Ile-2041-Asn (Hochberg et al. 2009); and Asp-2078-Gly (Kaundun 2010).
Acetolactate synthase (ALS) or acetohydroxy acid synthase (AHAS)	B	*L. rigidum, Descurainia sophia, Amaranthus tuberculatus, and Setaria viridis*	TSR: Pro-197-Ala, Pro-197-Arg, Pro-197-Gln, Pro-197-Leu, Pro-197-Ser, & Trp-574-Leu (Yu et al. 2008); Pro-197-Thr (Yang et al. 2016); Ser-653-Asp, Ser-653-Thr, & Trp-574-Leu (Patzoldt and Tranel 2007); and Ser-653-Asn, Ser-653-Thr, Ser-653-Ile, Gly-654-Asp (Laplante et al. 2009).
Photosystem II protein D1-encoding gene (psbA)	C	*Amaranthus hybridus, Portulaca oleracea, Poa annua, Kochia scoparia, Senecio vulgaris, Chenopodium album, and Capsella bursa-pastoris*	TSR: Ser-228-Gly (Hirschberg and McIntosh 1983); Ser-264-Thr (Masabni et al. 1999); Val-219-Ile (Mengistu et al. 2000); Ile-219-Val, & Gly-264-Ser (Mengistu et al. 2005); Asn-266-Thr (Park and Mallory-Smith 2006); Ala-251-Val (Mechant et al. 2008); and Phe-255-Ile (Perez-Jones et al. 2009).
Protoporphyrinogen oxidase (PPX2L)	E	*Amaranthus tuberculatus*	TSR: Gly-210 deletion (Patzoldt et al. 2006).
5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)	G	*Eleusine indica, Amaranthus palmeri, Amaranthus hybridus, and Euphorbia heterophylla*	TSR: Pro-106-Ser (Baerson 2002); EPSP gene amplification (Gaines et al. 2010); Thr-102-Ile, Ala-103-Val & Pro-106-Ser (Garcia et al. 2019); and double mutations: Thr102Ile + Pro106Thr (Mendes et al. 2020)
α-tubulin	K	*Eleusine indica, and Setaria viridis*	TSR: Thr-239-Ile (Anthony et al. 1998); Thr-239-Ile, & Leu-136-Phe (Délye et al. 2004); and Thr-239-Ile, & Met-268-Thr (Yamamoto 1998).
Cytochrome P450 mono-oxygenases (P450s)	A, B, C	*Alopecurus myosuroides, and L. rigidum*	Non-target site resistance (NTSR) by enhanced detoxification/metabolism of herbicides (De Prado et al. 1997; Hall et al. 1995; Gaines et al. 2014; Duhoux et al. 2017).
Glutathione S-transferases (GST)	C	*Abutilon theophrasti, and L. rigidum*	NTSR: increased GST activities (Anderson and Gronwald 1991; Gaines et al. 2014; Duhoux et al. 2017).
P450s, ABC transporter, and glycosyl-transferase	B	*Descurainia sophia, and L. rigidum*	NTSR: upregulation of 4 P450s, 3 ABC transporters, & 1 glycosyltransferase genes (Yang et al. 2016; Gaines et al. 2014; Duhoux et al. 2017).
Unknown	F, G	*A. tuberculatus, L. rigidum, and Conyza bonariensis*	NTSR: enhanced metabolic breakdown of mesotrione (Kaundun et al. 2017); increased expression of EPSPS gene (Baerson et al. 2002); and reduced glyphosate translocation, & elevated constitutive EPSPS transcript level (Dinelli et al. 2008).

Most of the studies cited in Table 1.2 focused on specific genes, with effects assumed *a priori*, i.e.TSR genes and well-known detoxifying genes such as cytochrome P450 and GST. The exceptions are the transcriptomic analyses in *Descurainia sophia* (Yang et al. 2016), and *Lolium rigidum* (Duhoux et al. 2017; Gaines et al. 2014) which identified putative NTSR genes involved in resistance to ALS- and ACCase-inhibitors. The genetic bases of TSR listed in Table 1.2 can explain a significant portion of the herbicide resistance variation in some populations. However, these have been insufficient to explain the vast majority of the observed variation in herbicide resistance in many weed species and populations (Beckie and Tardif 2012; Délye et al. 2013; Délye 2013; Shaner and Beckie 2014). For example, target-site EPSPS Pro-106 mutations were insufficient to induce glyphosate resistance in *Echinochloa colona* (Han et al. 2016), and many experiments have failed to recapture known TSR mutations or have shown evidence for the abundance of NTSR genes (Yu et al. 2009; Délye et al. 2011; Yu et al. 2013; Scarabel et al. 2015; Rey-Caballero et al. 2017; Bai et al. 2019). Most of these previous studies on uncovering the genetic basis of herbicide resistance are under-powered because these involved only two groups, the susceptible and the resistant without consideration for the intermediate phenotypes.

The narrow focus on TSR genes is insufficient to control herbicide resistance evolution in weeds. This probably resulted in the misguided early recommendation to reduce herbicide dose (Gorddard et al. 1995). This misconception that low herbicide application rates or doses that result in low weed mortality can delay herbicide resistance evolution has been proven wrong by numerous low-dose selection experiments (Yu et al. 2013; Busi et al. 2013; Lagator et al. 2013; Busi et al. 2016). These suboptimal doses result in NTSR (Yu et al. 2013). Resistance to a single herbicide is likely a complex combination of TSR and NTSR mechanisms and requires mutations in many different loci. This results in polygenic resistance traits as in the case of diclofop-methyl resistance in *Lolium rigidum* (Gaines et al. 2014); and tribenuron-methyl resistance in flixweed (Yang et al. 2016).

The evolution of both TSR and NTSR mechanisms need to be elucidated in order to generate science-based, theoretically- and empirically-informed herbicide resistance control strategies. Genetic studies without prior assumptions of which genetic elements confer resistance are rare. Therefore there is a need for a genome-wide approach to detect resistance-associated genomic variation to improve our understanding of herbicide resistance beyond TSR and well-characterised NTSR genes. At the population and landscape levels, there is a need to study how herbicide resistance evolves.

1.2.5. *Lolium rigidum*: the world's most herbicide resistant weed species

Lolium rigidum, commonly known as annual ryegrass or rigid ryegrass or Wimmera grass is an agriculturally-relevant weed species to study the evolutionary quantitative genetics of herbicide resistance. *L. rigidum* is considered as the world's most herbicide-resistant weed (Heap 2014). It has developed resistance to the highest number of herbicides with different modes of action (i.e. 13 modes of action as of 2020) (Heap 2020). It is the first weed species reported to have evolved resistance to glyphosate (Powles et al. 1998). The combination of its fecundity and cross-pollinated nature resulting in large genetically diverse populations gives this weed species a high adaptive potential.

Lolium rigidum (Gaudin, 1811) belongs to the grass family Poaceae. Growth habits range from prostrate to erect, with erect being the dominant type (Figure 1.2). It can grow up to 1 meter in height. Leaf blades are 5-25 cm long, 3-5 mm wide, green and glabrous (CABI 2017). Auricles are small and narrow, while ligules are short and white to translucent. Leaf sheaths are glabrous which can have purple colouration at the base. Roots are extensive and fibrous. Inflorescence is a spike (length ≤ 30 cm) with 10-12 florets (10-25 mm long) in each spikelet (CABI 2017). Each floret is generally awnless with equally-sized palea and lemma, and three yellow anthers (Kloot 1983). It is a diploid with chromosome number 2n=2x=14 (Terrell 1966; Monaghan 1980). It has an estimated genome size of 2Gb similar to that of the closely-related forage crop *Lolium perenne* (Byrne et al. 2015). It is cross-compatible with other members of the Lolium genus, i.e. *L. multiflorum* and *L. perenne* (Kloot 1983). This genus is a complex of cross-compatible species which can produce fertile hybrids and the distinction between species is often blurred

(Naylor 1960; Terrell 1966; Kloot 1983).

It is a highly-competitive, self-incompatible, wind-pollinated, annual, C3 weed species (Monaghan 1980; McCraw et al. 1983; CABI 2017). It can produce up to 45,000 seeds m^{-2} in infested wheat fields (Gill 1996). These seeds can have varying levels of dormancy ensuring the sustained germination in the field and the replenishment of the soil seedbank (Goggin et al. 2012). A density of 300 *L. rigidum* plants m^{-2} can cause significant reduction in rapeseed and cereal crop yields from below 10% to more than 50% (Lemerle et al. 1995). Additionally, its seeds can be infected by *Clavibacter toxicus* which causes livestock poisoning (Riley and McKay 1991; Ophel et al. 1993).

It is native to the Mediterranean region, i.e. Europe and northern Africa. It has spread across the temperate crop-growing regions around the world. In the 19th century, it was introduced to Australia as a forage crop (Kloot 1983). After years of artificial and natural selection, it has adapted to local conditions and has become the major weed in the wheat-growing regions of Australia (Reeves 1976; Medd et al. 1985; Powles and Matthews 1992).



Figure 1.2. *Lolium rigidum* specimen collected from Murrumbidgee, New South Wales, Australia in 2018 (Scale bar 10 cm).

1.3. Quantitative herbicide resistance

How do we cost-effectively measure herbicide resistance of weeds in a manner which reflect the polygenic nature of these traits? The identification and measurement of herbicide resistance is the first step towards improving our understanding of the genetic and evolutionary bases of herbicide resistance in weeds. We have established in the previous section the importance of broadening the search space for the genetic basis of herbicide resistance to accommodate the likely polygenic nature of these traits. Weeds may die or survive after herbicide application. Measuring the survival rate of a population after a period of time following the application of herbicide at some relevant dose is the most basic way of quantifying these traits. However, the cascade of biochemical reactions within the plant prior to death or survival are complex rather than binary. This is the challenge that needs to be overcome, i.e. to measure the complex and polygenic nature of herbicide resistance traits for each individual plant.

1.3.1. Herbicide resistance as a quantitative trait

Despite the fact that many herbicides were developed to target a specific protein to disrupt a specific biochemical pathway, the cascade of biochemical reactions leading to plant death involves more than that one target protein and pathway. It follows that the assumption that TSR is the only major mechanism of herbicide resistance is naive. Biochemical processes are sufficiently interconnected such that functional mutations in genes outside the target pathway can have significant effects on the eventual death or survival of plants. All biotic and abiotic stresses including herbicides, induce stress responses which involve the metabolism of carbohydrates, proteins and lipids, as well as the detoxification of reactive oxygen species (Liu et al. 2007; Garg and Manchanda 2009; Rojas et al. 2014). This means that the numerous genes and pathways involved in resistance are expected to produce a quantitative phenotype. This is illustrated in Figure 1.3, as the number of biallelic loci with equal additive effects increases, the range of phenotypic values becomes continuous rather than discrete. Therefore binary phenotype, i.e. death or survival after herbicide treatment, is insufficient to express the complex nature of herbicide resistance. We need to measure herbicide resistance as a quantitative trait.

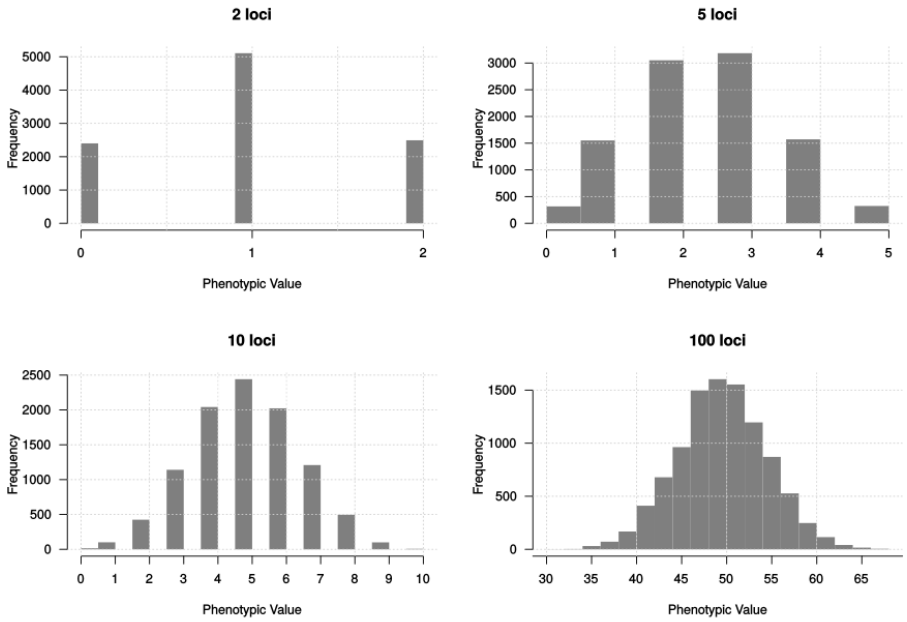


Figure 1.3. Distribution of phenotypic values resembles continuous quantitative variation as the number of loci controlling the trait increases (illustration idea came from Roff (1997); but the assumptions here are biallelic loci with equal additive effects and generated by randomly sampling 10,000 individuals).

The quantitative measurement of herbicide resistance needs to be an unbiased representation of the plant's ability to survive the abiotic stress brought about by herbicide treatment. The concentrations of the enzymes and metabolites involved can be measured; however, *a priori* knowledge of the pathways involved is assumed which can bias the measurement and may not be available. The concentration of the herbicide can also be measured but this will only measure the rate of its breakdown and absorption, while other mechanisms are unaccounted for, e.g. target-site incompatibility, over-expression of target enzyme, and reduced translocation. Various plant health-related quantitative traits can be measured ranging from destructive biomass assessment to sophisticated remote-sensing.

1.3.2. High-throughput phenomics

Exploiting high-throughput phenotyping technologies will allow us to determine suitable traits which reflect the level of herbicide resistance of individual plants. Logistic modelling can be used to assess the suitability of different quantitative plant health-related traits (Figure 1.4). These traits can have variable ranges and the point of maximum slope can deviate from the median depending on the underlying relationship. Multi-dimensional phenotype data can be used to build this logistic model. These data can be measured at multiple scales from cell, tissue, organ, to whole plant levels, as well as at a single to multiple time points.

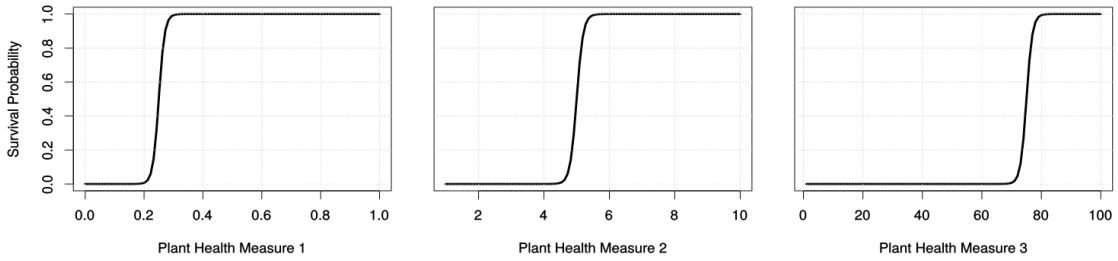


Figure 1.4. Modelling individual plant survival as a function of continuous quantitative health traits.

Various high-throughput plant phenotyping platforms generate data at different spatial and temporal resolutions to cater to a wide variety of applications. Cell and tissue phenotyping produce high-resolution data but usually at lower throughput (i.e. more measurements per plant but less plants) and involve destructive sampling (Zhao et al. 2019). On the other hand, organ and whole plant level phenotyping generate lower resolution data but usually at higher throughput (i.e. less measurements per plant but more plants) and involve both destructive and non-destructive sampling (Zhao et al. 2019; Yang et al. 2020). Traits can be measured at a specific time point of interest (e.g. 2-leaf stage or flowering stage), and at multiple time points which may span multiple stages of development. Measurements can be taken directly or manually from the plants and indirectly or in an automated manner with imaging techniques.

The classical trait used to measure plant health is the growth rate (Smythies 1834; Liggett et al. 1893). This can be in terms of changes in plant height, leaf dimensions, canopy density, and biomass per unit time, among many others. Direct or manual measurement of these growth rates is the most straightforward strategy, e.g. collecting whole plant samples followed by oven-drying and weighing, as well as using measuring tapes, rulers or callipers to measure plant height and leaf dimensions. However, these are resource- and labour-intensive. Additionally, the destructive nature of biomass measurements is undesirable when resources are limiting and replicating entries are difficult or impractical. To address these issues, automated to semi-automated phenotyping strategies through plant imaging are being developed. For example, it has been demonstrated that plant biomass can be accurately predicted from plant photographs (Tackenberg 2007).

Imaging techniques allow high-throughput non-destructive phenotyping at the organ and whole plant levels; while cell and tissue trait measurements are destructive at the sample preparation step (Zhao et al. 2019). Morphological and physiological changes can reveal the state of a plant's health. Healthy plants exhibit a sigmoidal growth pattern which reaches some genotype-specific maximum trait value under a specific environment. For most plants, healthy leaves

appear dark green indicating high chlorophyll content. Actively dividing and growing cells of healthy plants are turgid with water (Taiz and Zeiger 2002). On the other hand, unhealthy, diseased or stressed plants deviate from these states. These plants exhibit growth retardation which may eventually lead to death. The plant may senesce which visually manifests as chlorosis and necrosis of leaves and other photosynthetic tissues as chlorophyll gets broken down (Hörtensteiner 2006; Christ and Hörtensteiner 2014; Wang and Blumwald 2014). Heat, light, and osmotic stresses result in wilting (Taiz and Zeiger 2002). Lesions, and spots may appear as signs and symptoms of insect damage or pathogen infection. Various stresses can induce early flowering in some plant species (Takeno 2016; Kazan and Lyons 2016). All forms of plant stresses induce the over-production of reactive oxygen species which inflict oxidative damage on cell and organelle membranes as well as nucleic acids (Turrens 2003; Bailey-Serres and Mittler 2006; Huang et al. 2019).

These health-related morphological and physiological responses to stress can be accurately and cost-effectively captured with plant imaging instruments. These instruments measure light reflectance, absorption, and emittance at different wavelengths. The types of information they generate can be broadly categorised as colorimetric and spatio-tomographic. Colorimetric information are predictors of the concentrations or densities of various plant-associated substances, e.g. water, chlorophyll, and starch. Spatio-tomographic information describes the external and internal morphologies of plant cells, tissues, organs and whole plants. These instruments can use a few specific wavelengths or a wide range of the electromagnetic spectrum (EMS). Figure 1.5 shows the different imaging techniques and their operating wavelength range.

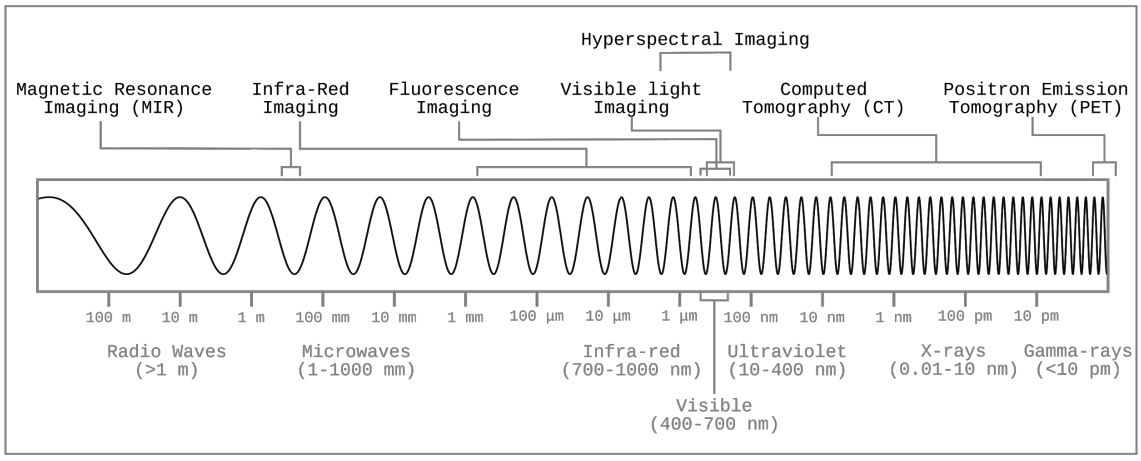


Figure 1.5. Imaging techniques and their respective operating electromagnetic wavelength range. The sine wave of varying wavelength depicts the electromagnetic spectrum.

Instruments which primarily capture colorimetric data include the visible light camera (red-green-blue or RGB camera), soil plant analysis development (SPAD) chlorophyll meter, hyperspectral imaging instruments, and chlorophyll fluorescence sensors. RGB cameras measure the visible light reflecting off of plant tissues and generate quantitative representations of what the human eyes generally perceive. SPAD chlorophyll meter, on the other hand, is a low cost hand-held instrument which measures the transmittance of two wavelengths of light at the red (650 nm) and infrared (940 nm) range through leaves (Yuan et al. 2016). The difference in the transmittance of these wavelengths is proportional to the chlorophyll content. Measuring a wider range of the EMS through hyperspectral imaging (400-2500 nm) can yield more information on plant health (Mishra et al. 2017; Ge et al. 2019). At the visible range (400-700 nm), photosynthetic pigments are the dominant sink. At near-infrared wavelengths (700-1100 nm), light absorption by plant dry matter dominates; while at short-wave infrared wavelengths (1100-2500 nm), water inside plant tissues absorb the light (Mishra et al. 2017). Hyperspectral imaging has been shown to yield reliable measures of plant health and other physiological traits (Bock et al. 2010; Mishra et al. 2017; García-Sánchez et al. 2017; Ge et al. 2019). Chlorophyll fluorescence imaging, e.g. pulse amplitude modulation imaging (Imaging-PAM) induces chlorophyll fluorescence by emitting 460 nm blue light, and then measures the red fluorescence at 620 nm (Wang P et al. 2017). It has been demonstrated that Imaging-PAM can reliably and

rapidly distinguish herbicide resistant and susceptible plants under greenhouse and field conditions (Wang P et al. 2017). This imaging technique requires a period of dark-adaptation prior to chlorophyll excitation and fluorescence measurements, which can be a challenge to implement for high-throughput applications (Schneider et al. 2019).

Instruments which primarily capture spatio-tomographic data include light detection and ranging (Lidar), magnetic resonance imaging (MRI), X-ray computed tomography (CT), and positron emission tomography (PET) scanners. Lidar uses a laser at some specific wavelength to measure distances based on the time it takes for the light to get reflected back and the variation in the wavelength of the reflected light (Wandinger 2005). Lidar-based phenotyping platforms have been shown to accurately measure leaf area, leaf orientation, and canopy volume (Wang H et al. 2017; Thapa et al. 2018; Panjvani et al. 2019; Jin et al. 2020). MRI uses strong magnetic fields and pulses of different radio frequencies to induce discriminating radio wave emittance to reveal internal structures within an opaque material (Berger 2002). MRI scanners can be used for non-destructive root trait measurements without the need to uproot the plant (Pflugfelder et al. 2017). CT uses multiple X-ray images of a material to build a 3D model and elucidate its internal structures (Kalender 2011). PET senses the internal structures of a material by measuring the attenuation of photons produced by the annihilation of positrons emitted by a radioactive material with very short half-lives, e.g. carbon-11, nitrogen-13, oxygen-15, and fluorine-18 (Bailey et al. 2005). Garbout el at (2012) and Lafond et al (2015) demonstrated that CT and PET scanners can be used to accurately measure root and shoot morphologies.

These imaging instruments vary considerably in terms of procurement and operating costs, as well as portability. Colorimetric instruments are simpler and therefore more portable than spatio-tomographic instruments. Spatio-tomographic instruments are sophisticated and usually bulky pieces of equipment and incur not only high procurement costs but also high operating and maintenance costs. Despite the advantages and the state-of-the-art technologies offered by spatio-tomographic instruments, as well as hyperspectral and fluorescence imaging instruments, the investments to acquire and maintain them can be financially prohibitive. On the other hand, SPAD meters are cheap but the light spectrum they capture is very narrow and small leaves especially narrow grass leaves are hard to reliably measure. The accessibility and simplicity of taking RGB photographs of plants, as well as the density of information that can be extracted from these photographs make RBG cameras one of the most attractive imaging instruments to measure plant health and more specifically the herbicide resistance of individual plants.

1.3.3. Quantifying herbicide resistance from RGB photographs

Bioassays to measure resistance should be simple, utilising basic and low-cost equipment, while maximising reliability (R4P Network 2016). Among the imaging instruments described, RGB cameras are the most cost-effective. RGB cameras have become ubiquitous and powerful yet low-cost. Precise measurements can be made from photographs taken using standardised lighting and the same backdrop across plants and herbicide treatments. Temporal information can be extracted by taking pictures at multiple time-points before and after herbicide application.

The resulting images need to be processed and analysed to yield accurate information. Colour constancy within and among photographs need to be guaranteed. The plants need to be accurately distinguished from the background, and the image artefacts removed. There are a number of free and open-source computer vision libraries for image processing and analysis, one of the most popular is OpenCV (OpenCV 2017). It can be used to yield multidimensional data of colour-based metrics across spatial and temporal dimensions.

Variation in illumination within and among photographs is rectified by performing colour correction. Various colour correction algorithms exist, where some perform better than others under different spatial and spectral image characteristics (Gijsenij and Gevers 2011). Reference white-coloured objects photographed together with the plants can be used to colour-correct the images by simple RGB scaling (Sharma and Bala 2017):

$$X_{corrected} = \frac{255\,X}{x_{white}},$$

where $X_{corrected}$ and X are the corrected and uncorrected matrices of pixel values in one of the channels in the RGB colour space (i.e. red, green or blue), respectively, while x_{white} is the mean pixel value of the white reference on the same channel. Other scaling methods exist which use different colour spaces, one of which is based on the LMS (long, medium, short) colour space which represents the sensing ability of the long, medium and short cone cells of the human eyes (Sharma and Bala 2017).

Discriminating between the plant, and other objects in the photograph involves image segmentation. This involves partitioning the image into homogeneous regions punctuated by areas of discontinuity separating each region (Cheng et al. 2001). This can be performed on monochromatic and coloured representation of images using histogram thresholding, feature clustering, fuzzy techniques and neural networks (Pal and Pal 1993; Cheng et al. 2001). The most straightforward and least computationally expensive is monochromatic histogram thresholding. This uses a monochromatic representation of the image and one to multiple threshold pixel values to partition the image (Pal and Pal 1993; Chao et al. 2009). These thresholds can be set manually based on a few image segmentation tests or heuristically using clustering algorithms (Nakib et al. 2007; Chao et al. 2009; Tan and Isa 2011).

Various features or metrics can be extracted from the colour-corrected, background- and artefact-subtracted plant images. These include statistical features, global transformation and series expansion features, as well as geometrical and topological features (Kumar and Bhatia 2014). Many of these are hard to interpret or may not have any biological interpretations. Using colour-based metrics, notably those that potentially measure the degree and progression of chlorosis and necrosis are preferable.

To generate a single quantitative measure of herbicide resistance, these image-derived multidimensional data need to be mapped to the observed mortality data (Figure 1.6). This can be achieved with logistic regression modelling:

$$y_{binary} = \left(1 + e^{-X\beta}\right)^{-1},$$

where y_{binary} is the binary plant mortality data, e is Euler's number or the base of natural logarithm, X is the matrix of image-derived metrics including the intercept, and β is the corresponding vector of effects. The quantitative herbicide resistance trait estimate can be defined as:

$$y_{quantitative} = X\,\hat{\beta},$$

which expands to:

$$y_i = \hat{b}_0 + x_{i1}\hat{b}_1 + x_{i2}\hat{b}_2 + \ldots + x_{im}\hat{b}_m,$$

where $\hat{\beta}$ is the vector of estimated effects, y_i is the quantitative herbicide resistance phenotype of the ith plant, $\hat{b}_0$ is the estimated intercept or the risk threshold value beyond which the plant is classified as resistant, $\hat{b}_j$ is the estimated effect of the jth metric, and x_{ij} is the value of the jth metric of the ith plant. This logistic model can be solved using elastic-net regularisation, which can perform simultaneous effect estimation and variable selection (Friedman et al. 2010).

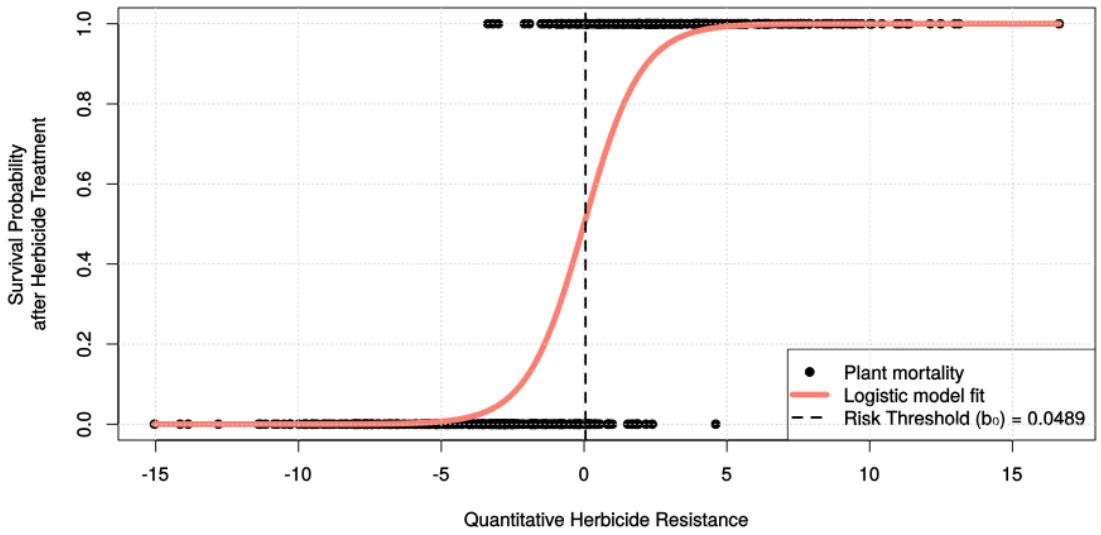


Figure 1.6. Simulated logistic model describing the relationship between the binary mortality data and the quantitative herbicide resistance values computed from the various image-derived metrics and their corresponding estimated effects.

1.4. Genomic characterisation of a non-model species

How do we accurately and cost-effectively characterise the large genome of a non-model species for quantitative and population genetics studies? Specifically, how do we efficiently extract accurate genomic information from the large and complex *Lolium rigidum* genome? Genomic characterisation is the next logical step after phenotypic characterisation on the road to uncovering the genetic and evolutionary bases of herbicide resistance in weeds.

Various molecular markers have been developed over the past three decades to assess diversity, to study population and landscape dynamics, to build linkage maps, and ultimately to link them with observable phenotypes (Schlötterer 2004). Accurate genome-wide molecular marker information is fundamental in building genotype-to-phenotype maps (e.g. QTL mapping, GWAS, and genomic prediction models), and models of evolution (e.g. selective sweep scans and coalescent models). Extracting this information from large ($\geq 1\times10^9$ bases) and complex (chromosomal rearrangements and abundant repetitive elements) genomes is challenging (Lynch and Conery 2003; Peterson et al. 2012; Andrews et al. 2016). Next-generation sequencing technologies offer cost-effective strategies to do this. Their high-throughput allows us to characterise genomes of multiple individuals in parallel. Accuracy of the resulting data is dependent upon the depth or coverage of sequencing, i.e. the number of times a nucleotide base is read, such that higher depth means more replication to ascertain the validity of the sequence (Anand et al. 2016; Song et al. 2016; Kishikawa et al. 2019; Tilk et al. 2019). The breadth of coverage or the fraction of the genome characterised is dependent upon how much of the genome was represented in the sample. Whole-genome sequencing is ideal but financially prohibitive for species with large genomes and resource-limited projects (Futschik and Schlötterer 2010).

1.4.1. Sequencing-based genomic characterisation

Pioneering works on nucleic acid sequencing occurred during the 1970s (Griffin and Griffin 1993; Sanger 1998; Heather and Chain 2016). The first genome sequencing was performed by Sanger and Coulson in 1975 on the bacteriophage φX174 (or PhiX, the positive control genome as it is now commonly known) using the “plus and minus” technique (Sanger and Coulson 1975). The first generation sequencing methods were published in 1977. Sanger, Nicklen and Coulson developed the “chain termination method” of DNA sequencing (Sanger et al 1977); while Maxam and Gilbert developed the “chemical method” (Maxam and Gilbert 1977). The chemical method was the first widely adopted DNA sequencing technique but quickly lost popularity as the advantages of the chain termination method or Sanger sequencing became more apparent (Heather and Chain 2016). Sanger sequencing has gone through several changes which improved reliability and safety, including the replacement of radiolabelling with fluorescence-based detection (Smith et al. 1985; Ansorge et al. 1986; Ansorge et al. 1987), and the use capillary-based electrophoresis (Swerdlow and Gesteland 1990; Luckey et al. 1990). Sanger sequencing has proven itself over the years as a robust and accurate method so much so that it has become the gold standard in nucleotide sequencing.

The next generation of sequencing technologies include 454 Roche pyrosequencing, Solexa/Illumina cluster sequencing, sequencing by oligonucleotide ligation and detection (ABI SOLiD) system, Ion torrent semiconductor-based sequencing, PacBio single molecule real time (SMRT) sequencing and Oxford nanopore sequencing (Heather and Chain 2016; Mardis 2017). These platforms enable massively parallel sequencing, real-time sequence output, and single molecule sequencing. Among these, Illumina’s sequencing technologies dominate at ~75% market share as of 2020 (Truong 2020).

Illumina sequencing platforms provide massively parallel sequencing-by-synthesis via bridge amplification on a lawn of oligonucleotides using fluorescent reversible terminator deoxyribonucleotides (Figure 1.7). Target DNA molecules are ligated with barcode sequences and sequences complementary to the oligonucleotides bound to the surface of the flow cell (Fedurco et al. 2006; Heather and Chain 2016). The barcode sequences allow for multiplexing, i.e. samples from multiple individuals are pooled, then sequenced in parallel, and afterwards the resulting sequences are separated. The ligated DNA molecules attach to the lawn of

oligonucleotides and are sequenced with fluorescent reversible terminator dNTPs (each of the four nucleotide bases fluoresce with a specific colour) (Bentley et al. 2008). DNA sequence information is extracted from the series of flow cell images. Each DNA sequence read is currently limited to 300 bp and 600 bp for single-end and paired-end reads, respectively (Illumina Inc. 2020). As of 2020, Illumina sequencing platforms have run times ranging from 7 to 84 hours, generating 4 million to 20 billion reads or 0.2 to 3000 Gigabases of output (Kumar et al. 2019).

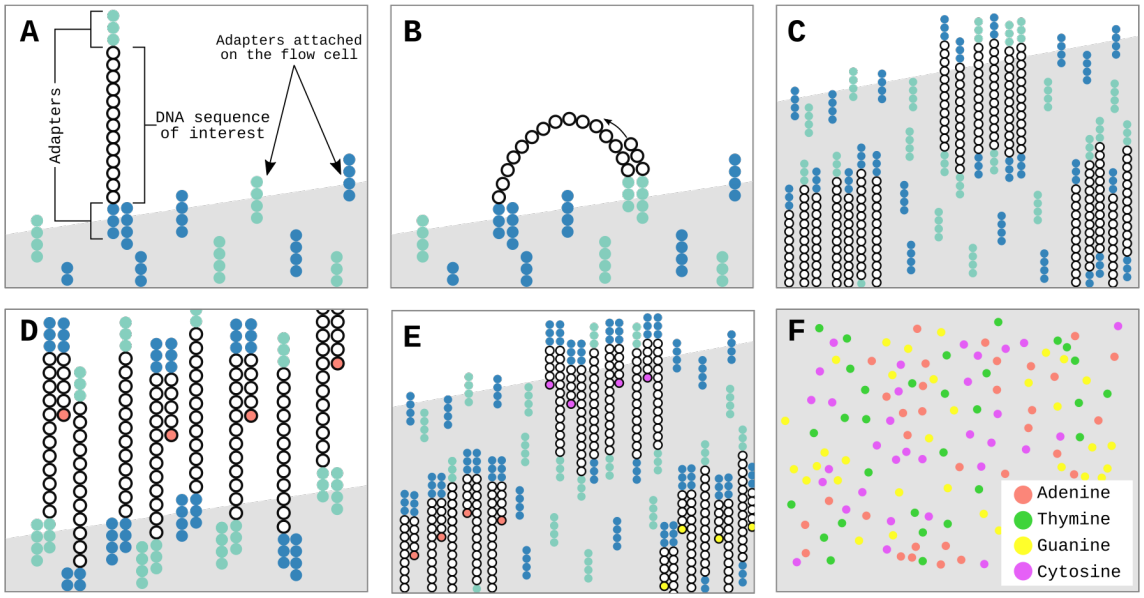


Figure 1.7. Illumina sequencing-by-synthesis via bridge amplification on a lawn of oligonucleotides using fluorescent reversible terminator dNTPs. Panel A: attachment of the adapter ligated DNA into the surface of the flow cell. Panel B: bridge amplification. Panel C: clusters of clonal DNA. Panel D: fluorescence emission in a cluster following laser excitation after DNA polymerase activity using primers and four reversible terminator dNTPs. Panel E: simultaneous fluorescence emissions from multiple clusters. Panel F: fluorescence image across the whole flow cell is captured for every cycle.

Innovations in sequencing technology have boosted throughput and dramatically reduced cost (Figure 1.8). These innovations can be summarised as the elimination of bacterial cloning steps, and the processing of millions of cyclical sequencing reactions in parallel (Park and Kim 2016). These have enabled the assembly of reference genomes of many species, whole-genome, transcriptome and exome sequencing, as well as population and landscape genomics (Park and Kim 2016). A reference genome or genome assembly is a reconstruction of the linear sequence of nucleotides of an organism's genome from fragments of sequenced DNA. This can be constructed using *de novo* (i.e. from scratch) and comparative (i.e. aligning sequences to the reference genome of a related species) approaches (Pop 2009). Whole-genome, transcriptome and exome sequencing refer to the characterisation of the entire nuclear DNA, transcribed ribonucleic acid (RNA), and messenger RNA (mRNA), respectively.

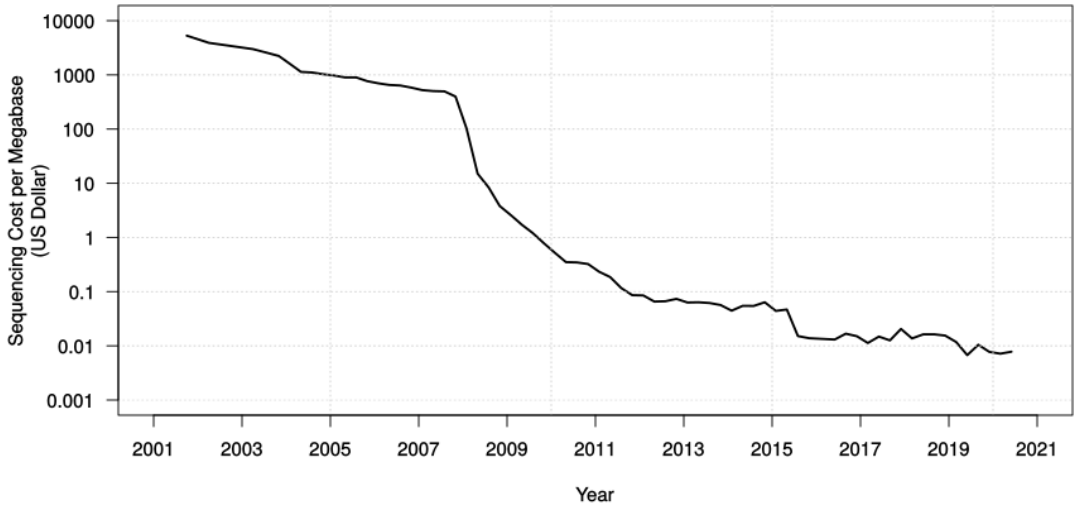


Figure 1.8. Sequencing costs per megabase from 2001 to 2020 (data from the US National Human Genome Research Institute; Wetterstrand 2020).

1.4.2. Reduced-representation and pool sequencing

The advent of next generation sequencing technologies has enabled cost-effective acquisition of accurate genome-wide molecular marker data for quantitative and population genetics studies (Ekblom and Galindo 2011; Li et al. 2017; Wordsworth et al. 2018). The entire genome can be sequenced (i.e. whole-genome sequencing) or only a fraction of it but with the aim of having an unbiased genome-wide representation (i.e. reduced-representation sequencing). Reducing the breadth of genome coverage with reduced-representation sequencing increases the depth of coverage per locus thereby increasing the accuracy of the sequence information, but at the cost of an incomplete representation of the genome (Elshire et al. 2011; Poland and Rife 2012). In characterising populations, individuals can be sequenced independently (i.e. individual sequencing; Indi-seq) or groups of individuals can be pooled and sequenced as a single sample (i.e. pool sequencing; Pool-seq). Reducing the resolution of the genomic data with Pool-seq will increase the number of individuals that can be represented which improves allele frequency estimates but at the cost of losing haplotype information (Futschik and Schlötterer 2010; Fariello et al. 2017). The choice of the appropriate genomic breadth and data resolution depends on genome size and accessibility, the target samples size, number of populations, sequencing depth required to achieve accurate genotype data, and the resulting costs.

Whole-genome sequencing generates a complete or nearly complete representation of the genome. There is little chance of missing nucleotide variation with this approach. Its feasibility for genomic characterisation for various applications including genome-wide association, genomic prediction, evolutionary and population genetics studies, predominantly depends on the genome size (Scheben et al. 2017; Li et al. 2017). In species with small genomes such as *Arabidopsis thaliana* (135Mb), whole-genome sequencing can be used to characterise multiple individuals at high depth, but with considerable investment from a global consortium (1001 Genomes Consortium et al. 2016). Acquiring whole-genome information from large and complex genomes for multiple individuals is challenging and costly. Sequencing cost is the limiting factor as the genome size, sample size and sequencing depth required increase.

Reduced-representation sequencing offers an alternative to whole-genome sequencing for large complex genomes and large sample sizes. Specific subsets of the genome can be targeted through exome capture, transcriptome sequencing (RNAseq), and restriction site-associated DNA sequencing (RADseq). Exome capture reduces genome complexity by sequencing only the protein-coding regions (Albert et al. 2007; Warr et al. 2015). Although exome capture-derived genomic information is highly reproducible, the initial investment in developing probes can be financially straining and the variation in non-coding regions is not represented (Hirsch et

al. 2014). Additionally, exome libraries constructed from different probe capture platforms vary considerably in coding sequence coverage (Barbitoff et al. 2020). Similarly, RNAseq represents only the transcribed genomic regions and under-represents regulatory sequences (Zhao et al. 2019). In addition, RNAseq data is tissue-specific and highly sensitive to environmental conditions (Hirsch et al. 2014; Lopez-Maestre et al. 2016). Restriction site-associated DNA sequencing (RADseq) represents genomic regions flanked by restriction sites (Baird et al. 2008). One or more restriction enzymes can be used to digest the genomic DNA, followed by size selection, and high-throughput multiplex sequencing (Baird et al. 2008; Peterson et al. 2012). Among these genome complexity reduction methods, RADseq is the most attractive due to the lower investment required and lower bias in genome representation compared to exome capture and RNAseq (Andrews et al. 2016).

Individual sequencing (Indi-seq) can generate high-resolution population-level genomic data. This is a common strategy for quantitative and population genetics studies in natural populations (Holliday et al. 2010; Willing et al. 2010; Cappa et al. 2013; Pallares et al. 2014; Johnston and Schoen 2017; Perrier et al. 2018; Hansson et al. 2018). Depth of sequence coverage per individual needs to be sufficiently high to allow accurate variant-calling which limits the number of individuals that can be included. As a consequence, most of the raw sequence data provide redundant information. Sacrificing sample size over variant-calling accuracy can result in under-powered analyses. Accurate population-level genomic information entail high sequencing depth per individual and large sample sizes which necessitate high costs. The high level of detail captured with Indi-seq is often not cost-effective for the genomic characterisation of large groups of individuals.

Pool sequencing (Pool-seq) is a cost-effective alternative to Indi-seq (Futschik and Schlötterer 2010; Schlötterer et al. 2014; Anand et al. 2016). It captures low resolution information, i.e. allele frequency information from groups of individuals instead of allelic identities per individual. It sacrifices sequence information redundancy per individual in favour of large sample sizes. As a consequence, haplotype information is absent in Pool-seq data, which makes the estimation of linkage disequilibrium (LD) and recombination rates challenging. The discrete sequencing reads can be used to estimate LD information and recombination rates, hence the range is limited by the read length of the sequencing platform used (Feder et al. 2012). Pool-seq can generate more accurate allele frequency estimates than Indi-seq because of the larger sample sizes that can be sequenced for the same cost (Futschik and Schlötterer 2010). To maximise the accuracy of variant calling and allele frequency estimation, Pool-seq guidelines and recommendations have been proposed (Schlötterer et al. 2014; Fracassetti et al. 2015; Anand et al. 2016). Among these guidelines, large pool size and high sequencing coverage appear to be the most important factors. Pool sizes of at least 25 (Gautier et al. 2013; Fracassetti et al. 2015) or 40 individuals (Schlötterer et al. 2014); and depths greater than 50X (Zhu et al. 2012) to 65X (Gautier et al. 2013) have been recommended. Pool-seq has been used successfully in ecology (Boitard et al. 2012; Cheng et al. 2012; Bastide et al. 2013; Nielsen et al. 2018), crop breeding (Beissinger et al. 2014; Bélanger et al. 2016) and quantitative genetics studies (Macgregor et al. 2006; Knight et al. 2009; Jinliang Yang et al. 2015; Fournier-Level et al. 2017; Micheletti and Narum 2018; Guo et al. 2018). The cost-effectiveness of Pool-seq will have the highest impact on poorly funded research involving non-model species (Schlötterer et al. 2014).

1.4.3. Application in *Lolium rigidum*

The combination of reduced-representation sequencing and Pool-seq offers a cost-effective approach to the genomic characterisation of *Lolium rigidum* populations. *L. rigidum* is a non-model yet agriculturally-important plant species with a large genome (~2 Gb) and large effective population sizes due its high fecundity and self-incompatibility. Its large genome can be rendered tractable with RADseq-based genome complexity reduction; while its large populations can be represented effectively with Pool-seq. The resulting sequencing information needs to be aligned or mapped to a reference genome. This is to put each read in the right context or to append positional information so that valid statistics can be estimated between reads and pools.

L. rigidum does not currently have a published reference genome. However, the genome of a closely related species, *Lolium perenne* a forage crop, has been assembled with some preliminary gene annotations (Byrne et al. 2015). The summary statistics of this genome

assembly is presented in Figure 1.9. Given that these two species are cross-compatible and capable of producing fertile offsprings (Terrell 1966; Kloot 1983), this genome assembly can be used to align the *L. rigidum* sequencing reads.

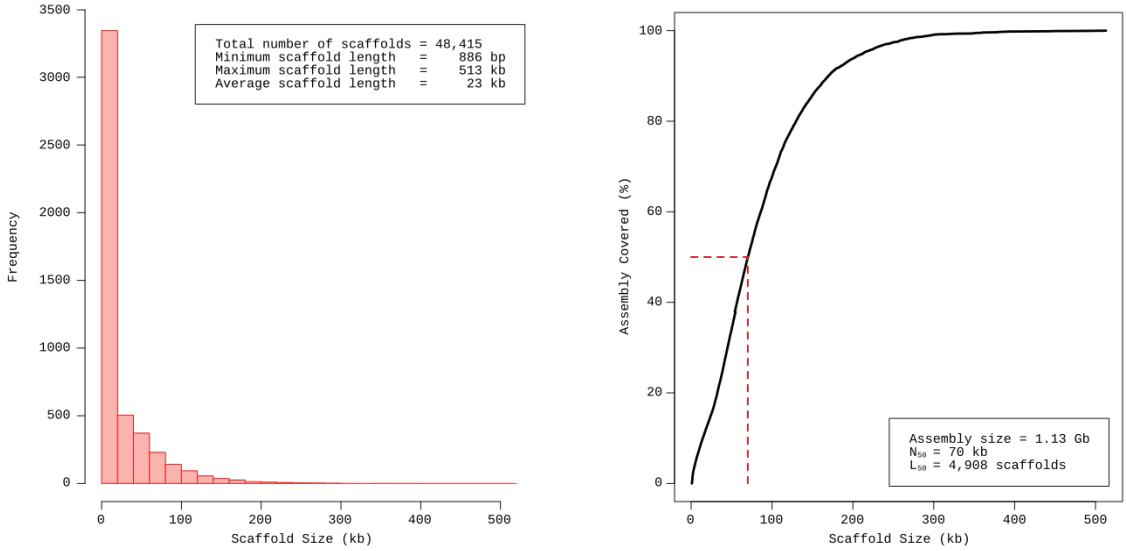


Figure 1.9. Summary statistics of the *Lolium perenne* genome assembly (adapted from Byrne, 2015).

1.5. Genome-wide association and genomic prediction of herbicide resistance in weeds

How do we bring together phenotype and genotype information to gain insights into the genetic bases of herbicide resistance in weeds? Are herbicide resistance traits in weeds truly polygenic or do major genes which confer target-site resistance dominate? What are the genetic bases of herbicide cross-resistance? Weeds have come up with adaptive solutions to the recurrent application of herbicides through the evolution of herbicide resistance. To understand the mechanisms through which these adaptive solutions have emerged, the genetic bases of these adaptations need to be identified first. Given the random process through which new mutations arise and the massive search space to perform this random walk as well as the interconnected nature of biochemical processes, these adaptive solutions are likely to be numerous and highly variable.

Our understanding of the genetic bases of herbicide resistances in weeds is largely incomplete. Most of the previous studies are under-powered due to small sample sizes which involved only two groups, the susceptible and the resistant without consideration for the intermediate phenotypes. The explanatory power of previously discovered genetic variants is insufficient (White et al. 2005; Délye et al. 2009; Délye 2013; Kumar et al. 2015). These variants were found through experiments with the assumption that herbicide resistance is a qualitative or binary trait controlled by one to a few loci, which includes the known target protein and a few well-characterised detoxification genes. To improve the explanatory power of genotype-to-phenotype maps, a genome-wide approach coupled with a quantitative treatment of resistance traits is needed.

Bringing together genomic and quantitative phenotype data to build genotype-to-phenotype maps of herbicide resistance involves using the *de facto* methods of genome-wide association study (GWAS) and genomic prediction (GP) (Gondro et al. 2013). GWAS is a population-level experiment, designed to search across the genome for markers that are significantly associated with phenotypes of interest. GP is the prediction of phenotypic values using genome-wide marker data and a genotype-to-phenotype model. The objective of GWAS is to identify putative quantitative trait loci (QTL), while GP is to predict phenotypes. Despite different objectives, the underlying models for both methods are the same. Predictions can be made with GWAS-derived allelic effects and QTL can be detected using GP-derived allelic effects.

1.5.1. History and theory of GWAS and GP

In the early 20th century, Fisher laid the foundation of quantitative genetics (Fisher 1918), Morgan, Sturtevant, and Muller deduced that mechanisms of linkage and recombination (Sturtevant 1913; Morgan et al. 1915; Muller 1916); and in the 1950s, Henderson developed the mixed model equations (Henderson et al. 1959). Their work constitutes the theoretical basis of GWAS and GP. However it was only at the beginning of the 21st century that GWAS and GP became feasible owing to the advent of cost-effective genotyping technologies. The first GWAS was reported in 2002 by Tanaka's group (Ozaki et al. 2002; Ikegawa 2012), and GP was conceived independently by Whittaker et al in 2000 and Meuwissen et al in 2001.

Fisher introduced the concept of variance and pioneered its use to deconstruct the phenotype into its various components. In his seminal paper (Fisher 1918), Fisher showed that phenotypic variation can be expressed as the sum of the additive genetic variance and non-additive genetic variance or dominance deviations as a consequence of heterozygosity. This has been extended to include environmental effects as well as the interaction between the genotype and environment (GxE) and interactions between loci (epistasis) (Roff 1997; Walsh 2006). This variance decomposition of the phenotype, y can be expressed as:

$$y = G+E+(G\times E) = A+D+I+E+(G\times E),$$

where G is the genotype effect, E is the environmental effect, $(G\times E)$ is the interaction between genotype and environment, A is the additive effect, D is the dominance effect, and I is the epistatic effect. This has enabled the estimation of heritability beyond phenotype means between relatives. Specifically, heritability (h^2) can be defined as:

$$h^2_{broad-sense} = \frac{\sigma^2_G}{\sigma^2_P}, \text{ and } h^2_{narrow-sense} = \frac{\sigma^2_A}{\sigma^2_P},$$

where σ^2_G , σ^2_P, and σ^2_A are the genotype, phenotype and additive genetic variances, respectively. The confirmation of the chromosomal theory of inheritance (Morgan 1910; Sturtevant 1913), and DNA as its molecular basis (Avery et al. 1944; Khanna 2010), as well as the discovery of the structure of DNA (Watson and Crick 1953) have enabled the elucidation of the molecular bases of σ^2_G and σ^2_A.

Linkage analysis is the estimation of the relative distances of genetic markers from one another to build a linkage map using pedigree information (Dexter 1914; March 1999; Ott et al. 2015). Genetic markers can be phenotype- or DNA-based. The linkage or association between a pair of genetic markers decreases as a result of eukaryotic recombination during meiosis via chromosomal crossover events (Morgan et al. 1915). It follows that more crossover events between two loci implies greater physical distance between them. Initially, phenotype markers were used to build linkage maps (Taylor 1970; Friend and Hopwood 1971). Sturtevant used 6 sex-linked qualitative traits in *Drosophila melanogaster* to generate the first linkage map (Sturtevant 1913). Restriction enzymes and polymerase chain reaction among other innovations in molecular biology, have enabled the construction of DNA marker-based linkage maps (Botstein et al. 1980; Rollini et al. 1985; Lander and Botstein 1986). These maps have been useful for breeders aiming to reduce linkage drag (i.e. favourable loci linked to unfavourable ones) by providing estimates of the minimum population size to generate enough recombination events to result in favourable linkage breaks (Brown et al. 1989; Tanksley et al. 1989).

Linkage maps have enabled quantitative trait loci (QTL) mapping, i.e. the identification of genetic markers linked or associated with quantitative traits using pedigree information (Sax 1923; Lander and Botstein 1986; Lander and Botstein 1989). The resulting QTL map resolutions have been low, despite the increase in the number of genetic markers from thousands to hundreds of thousands from the late 20th to the early 21st century (Hästbacka et al. 1992; Talbot et al. 1999; Meuwissen and Goddard 2000; Fan and Xiong 2002). This is due to the large linkage blocks present in genealogically shallow populations, e.g. synthetic populations and pedigree data dating back to only a few generations, used in QTL mapping (Meuwissen and Goddard 2000). Linkage blocks are large in these populations as a consequence of small population sizes and limited recombination events. Increasing the number of generations prior to mapping will reduce the linkage block size and improve mapping resolution.

Linkage disequilibrium mapping (Hästbacka et al. 1992; Meuwissen and Goddard 2000) or it's contemporary iteration the genome-wide association study (GWAS) (Kruglyak 1999; Ozaki et al. 2002; Hirschhorn and Daly 2005; Ikegawa 2012), exploits the large number of historical recombinations that have occurred in natural populations to generate high resolution QTL maps. While QTL mapping assumes relatedness in its family-based experimental design, GWAS assumes unrelated individuals in its population-based design (Lippert 2003). Since a truly panmictic population does not exist in nature because some individuals are more likely to mate than others as a consequence of spatial and temporal distributions, and selective forces. Hence, with natural population comes the problem of population structure which results in spurious associations (Sul et al. 2018). These associations are caused by relatedness rather than linkage with the true causal variants (Pritchard and Rosenberg 1999). Clustering individuals based on relatedness followed by intra-cluster GWAS can alleviate this problem but reduces power.

A statistical approach using linear mixed modelling developed by Henderson from 1959 to 1975 can effectively control for population structure while retaining high QTL detection power. Random parameters which are associated with a probability distribution are "mixed" with fixed parameters without an associated probability distribution into a single linear model. The linear mixed model can be expressed as:

$$y = X\beta + Zu + \varepsilon,$$

where y is the vector of trait values, X is the incidence matrix of fixed effects β, Z is the incidence matrix of the random effects u which accounts for population structure, ε is the vector of residual effects, $u \sim N(0, G)$, $\varepsilon \sim N(0, R)$, and $Cov(u, \varepsilon) = 0$. Maximising the joint density

function of y and u, with respect to β and u yields Henderson's mixed model equations (Henderson et al. 1959; Henderson 1975):

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix} .$$

The variance components as well as the fixed and random effects are estimated using maximum likelihood or Bayesian methods.

GWAS is the genome-wide interrogation of genetic variants for significant associations with phenotypes using dense molecular markers. Presently, SNPs using SNP chips or genotyping-by-sequencing are the most widely used type of molecular markers owing to their abundance, relatively uniform distribution across genomes and co-dominant nature (Vignal et al. 2002; Li et al. 2008; Yu et al. 2014). In the context of elucidating the genetic basis of quantitative traits or the detection of QTL, the associations between SNPs and phenotype are usually quantified using linear mixed models. The significance of SNP associations are calculated and some threshold value is set to qualify which SNPs are likely to be true QTL through various multiple comparison methods (Fadista et al. 2016; Kaler and Purcell 2019). These threshold values aim to minimise false positive rate by transforming the desired significance level as a function of the number of independent multiple comparisons being tested, which in the case of GWAS, involves testing tens of thousands to millions of SNP associations.

GP is the prediction of phenotypic values using genomic information and a pre-trained genotype-to-phenotype model. The same linear models are used in GWAS and GP; however GP circumvents significance testing and proceeds directly to phenotype prediction assuming at least one of the SNPs is linked to each of the causal genetic variants (Meuwissen et al. 2001). GP was in part born out of the frustration of Meuwissen, Hayes and Goddard on the suboptimal use of genomic information to predict quantitative traits as a consequence of using only the statistically significant QTL. They have pioneered the use of linear mixed models and Bayesian approach to estimate all SNP effects across the genome simultaneously. GP has proven itself instrumental in increasing the genetic gain per year in plant and animal breeding (Crossa et al. 2017; Moeinizade et al. 2019; Voss-Fels et al. 2019; Georges et al. 2019).

1.5.2. Phenotype and genotype data acquisition for GWAS and GP

The sample size, number of SNPs, the types, accuracy and structure of phenotype and genotype data need to be considered to plan successful and cost-effective GWAS and GP experiments. The power to detect QTL and the accuracy of phenotype prediction both increase with increasing sample size and SNP density (Long and Langley 1999; de Roos et al. 2009). Phenotype and genotype data can be at the individual or group (pool or population) levels, where the former provides high resolution information at the cost of high resource requirements, and the latter provides low-cost low-resolution data (Cutler and Jensen 2010; Schlötterer et al. 2014). Accurate phenotype and genomic information are achieved with high number of replications and high sequencing depth (Giovagnoli and Sebastiani 1989; Es and Es 1993; Sims et al. 2014). Population structure adds confounding effects on the data, which can lead to spurious results (Pereira et al. 2018; Sul et al. 2018).

How many samples per population do we need and which populations do we select to achieve high QTL detection power and genomic prediction accuracy? Larger sample size and denser marker information improve the genetic and phenotypic variation captured to accurately identify loci linked to causal genomic variation, and to construct more general genomic prediction models. In GWAS, the number of samples to detect a QTL with a specific effect and/or heritability can be estimated with power analysis; while in GP, the variation in the training population needs to be representative of the variation present in populations whose phenotypes need to be predicted (Asoro et al. 2011).

Power in GWAS is defined as the probability of detecting true QTL, i.e. loci linked to the causal variants, with a specific criterion. In other words, power is the probability of identifying an effect β with false positive rate (Type I error rate) less than α ($P<\alpha$) (Ball 2013). A simple generic illustration is presented in Figure 1.10.

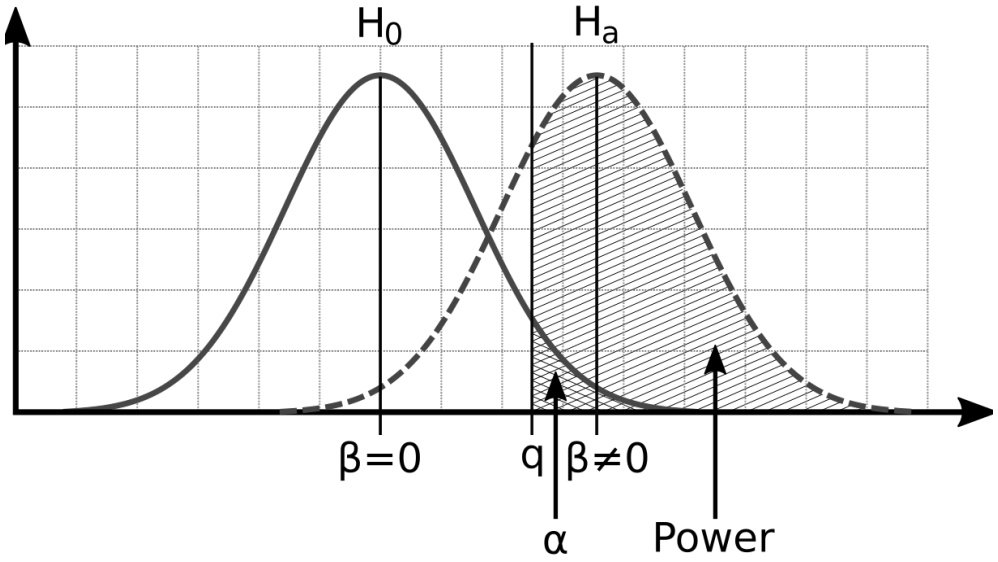


Figure 1.10. An illustration of power analysis. H_0 and H_a are the null and alternative hypotheses, respectively, defined by specific probability density functions, with the latter having a non-centrality parameter proportional to the effect size, β≠0 and sample size. The quantile value, q is the minimum effect size at the false positive rate threshold, α.

Power analysis can be analytical or simulation-based (Ball 2013; Wang and Xu 2019). In QTL mapping and non-mixed model GWAS, the standard approach for power analysis is used. Power is calculated as the area under the non-central chi-square distribution (alternative hypothesis, i.e QTL has an effect, df=1) to the right of the quantile defined by the p-value threshold on the chi-square distribution (null hypothesis, i.e. QTL has no effect, df=1). The non-centrality parameter is proportional to the QTL effect we wish to detect and the sample size. The non-centrality parameter can be computed as:

$$NCP = \frac{n r^2 h^2}{1 - r^2 h^2},$$

where n is the sample size, r^2 is the LD correlation between the SNP and the causal variant, and h^2 is heritability of the QTL, i.e. the proportion of the phenotypic variance explained by the causal variant. Also, if Hardy-Weinberg equilibrium is assumed then,

$$h^2 = 2q(1-q)\beta^2,$$

where q is the minor allele frequency and β is the standardised effect size of the allele (Visscher et al. 2017). The expected QTL detection power is computed for a number of QTL effect sizes and sample sizes. The cost of phenotype and genotype characterisation for a given sample size is weighted against the expected power and a compromise is sought.

In GWAS using linear mixed models to control for population structure, the non-centrality parameter is difficult to define. An analytical approach has been developed but Monte Carlo simulation-based power analyses are usually performed (Feng et al. 2011; Visscher et al. 2017; Shin and Lee 2015; Wang and Xu 2019). Simulation-based power analysis involves simulating multiple populations for a range of QTL effects. GWAS are then performed on these populations under various sample sizes. Power is defined as the frequency of detecting the simulated QTL at a specified α.

Genomic prediction accuracy increases as the training population size, marker density, and the relatedness between the training and validation sets increase (Jannink et al. 2010; Asoro et al. 2011). The number of SNPs need to be proportional to the effective population size (N_e) of the training population and the genome size of the species. It has been demonstrated in a simulation study that a training population size of $N_e L$, where L is the genome size in Morgans, is required to yield accurate genomic prediction accuracies (Meuwissen 2009). Additionally, it has been recommended that $8 N_e L$ SNPs are needed for an optimal prediction accuracy

(Solberg et al. 2008). Finally, accuracy is maximised when the training population represents the overall diversity of the species (de Roos et al. 2009).

Once the appropriate number of populations have been collected, the phenotyping approach needs to be robust and reliable to generate accurate phenotype information. Appropriate replications and controls should be included. Linear mixed models or hierarchical models can be used to control for confounding effects such as temporal and spatial heterogeneity. Outliers can be removed, the data transformed or heavy-tailed error distribution used to accommodate them (West 1984; Dai et al. 2016). Unbalanced datasets can be fit using linear mixed models (Schielzeth and Nakagawa 2013; Schielzeth et al. 2020) or the missing data can be imputed using various imputation methods (Nakagawa and Freckleton 2008; Penone et al. 2014).

Once the appropriate marker density has been used to characterise the samples, the resulting genotype information needs to be processed to filter-out low quality reads and/or impute missing information. Genotyping-by-sequencing approach enables cost-effective acquisition of genome-wide and potentially dense genotype data. Inherent to this approach are low quality sequencing reads and reads that map poorly to the reference genome. These data are filtered out using minimum quality score thresholds usually expressed in terms of “phred scores” (Q),

$$Q = -10\log_{10}(p),$$

where p is the probability of incorrect base call (Ewing and Green 1998). Similar to the treatment of missing phenotype data, missing genotype data can be imputed. The simplest approach is the major allele method where each missing data point is replaced with the major allele of the locus (Jung et al. 2007). More sophisticated methods such as haplotype reconstruction (Jung et al. 2007) and neural networks (Chen and Shi 2019; Kojima et al. 2020) can perform more reliable imputation.

1.5.3. Population structure

Accounting for structure in the data can improve the power of GWAS and accuracy of GP. Merging multiple populations have the potential to improve power as some populations may have polymorphic causal variants that are fixed in others. However, structure in the data can result in spurious associations which are due to relatedness rather than the underlying functional molecular basis (Yu et al. 2006). Hence, GWAS and GP models are generally constructed on a per population basis. And even with intra-population datasets, population structure may exist which needs to be accounted for. Fitting relatedness factors into linear mixed models in GWAS and GP is a well-established approach to do this (Devlin and Roeder 1999; Yu et al. 2006; Kang et al. 2010; Zhou and Stephens 2012; Sul et al. 2018).

Various relatedness factors or relatedness matrices exist. The most fundamental of which derives from pedigree or coancestry information, i.e. a square matrix is constructed where each element, r_{ij} corresponds to the degree of relationship between individuals i and j. This coancestry is the probability that at a locus a randomly drawn allele in individual i is identical-by-descent to a randomly drawn allele in individual j (Wright 1921; Malecot 1948). This means that ½ refers to parent-offspring and full-sib relationships, ¼ to half-sibs, ⅛ to first cousins, and so on. Inbreeding coefficient, F may also be used. F is the probability that at a random locus, the alleles from parents i and j are identical-by-descent (Wright 1921). To illustrate, if individual k is the progeny of individuals i and j, then the inbreeding coefficient of individual k, $F_k = r_{ij}$ (Figure 1.11).

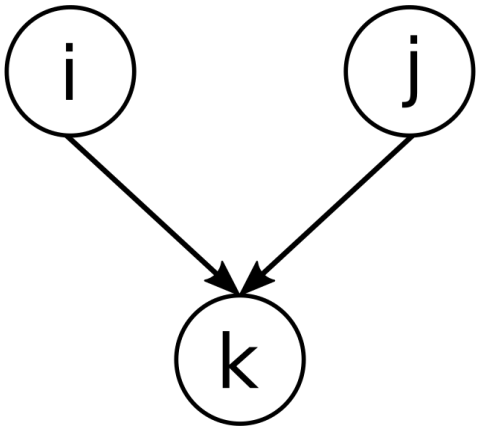


Figure 1.11. Illustration of the relationships between individuals i, j, and k.

The fixation index, F_{ST} which measures heterozygosity, can be used to define relatedness between groups of individuals within populations or between populations in a landscape (Wright 1949). An $F_{ST}=0$ describes an unstructured population undergoing random mating; while an $F_{ST}=1$ describes a completely structured population where each population is completely different from the one another. Another approach to describe population structure is through the realised relationship matrix,

$$G G^T,$$

where G is the $n \, x \, m$ matrix of genotype data with n individuals and m loci, can also be used. Different variations of this realised relationship matrix exist such as the standardised relationship matrix (Zhou and Stephens 2012) and sparse genetic relationship matrix where elements below some threshold value, e.g. 0.01 are shrunk to zero for computational efficiency (Yang et al. 2010; Zaitlen et al. 2013; Jiang et al. 2019).

1.5.4. GWAS models

Simple linear model

In GWAS, the magnitude and direction of association between a SNP and a phenotype are estimated using statistical models. Generally, these models are built and assessed one locus or one SNP at a time. Most GWAS models are parametric and linear. A simple linear regression model for GWAS can be expressed as,

$$y_i = \beta_0 + (x_{i,j} \cdot \beta_j) + \epsilon_i,$$

with

$$y_i \sim Normal\left(\mu_i = \sum_{j=1}^{m} (x_{i,j} \cdot \beta_j), \sigma^2\right) \text{ or } \epsilon_i \sim Normal(0, \sigma^2),$$

where y_i is the phenotype of the ith individual which is normally distributed with mean μ_i and variance σ^2, β_0 is the y-intercept or the bias in $\mu_.$, $x_{i,j}$ is the dosage of the jth SNP in the ith individual, β_j is the effect of the jth SNP, and ϵ_i is the random residual effect associated with the ith individual and normally distributed with mean 0 and variance σ^2. In matrix notation this is,

$$y = X\beta + \varepsilon,$$

with

$$y \sim Normal(X\beta, \sigma^2 I) \text{ or } \varepsilon \sim Normal(0, \sigma^2 I),$$

where y is the vector of phenotype values, X is an $(n+1) \times m$ matrix of SNP dosages for n individuals and m SNPs preceded by a vector of ones, β is the vector of the bias and SNP effects, and ε is the vector or random residual effects. The vector of phenotype values, y follows a multivariate normal distribution with the vector of means equal to $X\beta$ and variance-covariance matrix equal to $\sigma^2 I$ where I is an $n \times n$ identity matrix, i.e. diagonal elements are

ones and off-diagonal elements are zeros. The vector of residual effects also follow a multivariate normal distribution with a vector of zero means and the same variance covariance matrix as y.

The aim is to estimate the parameters of the model, i.e. the SNP effects, β and residual variance, σ^2. The estimates of these parameters are usually written with a hat symbol to distinguish them from the actual parameters of the model, i.e. $\hat{\beta}, \widehat{\sigma^2}$. The probability density of y or the likelihood of the parameters given y is defined as,

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i-\mu)^2}{2\sigma^2}},$$

or

$$f(y) = L(\beta, \sigma^2|y) = 2\pi^{-n/2} det(\sigma^2 I)^{-1/2} e^{(-1/2)(y-X\beta)^T(\sigma^2 I)^{-1}(y-X\beta)},$$

following the probability density function of the normal distribution. This likelihood function is usually log-transformed since it is easier to work with sums than products. Thus, the log-likelihood of the parameters (β, σ^2) can be expressed as,

$$logL(\beta, \sigma^2|y) = \frac{-n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(y-X\beta)^T(y-X\beta).$$

Linear mixed model

To control for population structure and other structures in the data, linear mixed models (LMM) (also referred to as mixed linear models (MLM) or hierarchical models) are used. The basic form of LMM in GWAS is,

$$y = X\beta + g + \varepsilon,$$

with

$$y \sim Normal(\mu = X\beta, V_\theta = \sigma_g^2 K + \sigma^2 I),$$
$$g \sim Normal(0, \sigma_g^2 K),$$

and

$$\varepsilon \sim Normal(0, \sigma^2 I),$$

where y is the vector of phenotype values, $X = [1\, G]$ is the $(n+1) \times m$ matrix of SNP dosages (G) preceded by a vector of ones, β is the vector of the bias and fixed SNP effects, g is the vector of random genetic effects, ε is the vector of random residual effects. The phenotype values follow a multivariate normal distribution with vector of means μ and variance-covariance matrix V_θ. The random genetic effects also follow a multivariate normal distribution with a vector of zero means and variance-covariance matrix equivalent to the product of the genetic variance, σ_g^2 and the $n \times n$ population structure matrix, K which is usually proportional to GG^T. The random residual effects also follow a multivariate normal distribution with a vector of zero means and variance-covariance matrix $\sigma^2 I$. The probability density of y or the likelihood of the parameters β and $\theta = \{\sigma_g^2, \sigma^2\}$ given y is,

$$f(y) = L(\beta, \theta|y) = 2\pi^{-n/2} det(V_\theta)^{-1/2} e^{(-1/2)(y-X\beta)^T(\sigma^2 I)^{-1}(y-X\beta)},$$

following the probability density function of the normal distribution. Therefore the log-likelihood is,

$$logL(\beta, \theta|y) = \frac{-n}{2}\log(2\pi) - \frac{1}{2}\log(det(V_\theta)) - \frac{1}{2}(y-X\beta)^T V_\theta^{-1}(y-X\beta).$$

Alternatively, if we define $\gamma = \sigma_g^2/\sigma^2$, and $H_\gamma = I + \gamma K$ we can simplify maximum likelihood

estimation (Hartley and Rao 1967) and define the log-likelihood function as,

$$logL(\beta,\gamma,\sigma^2|y)=\frac{-n}{2}\log(2\pi\sigma^2)-\frac{1}{2}\log(det(H_\gamma))-\frac{1}{2\sigma^2}(y-X\beta)^T H_\gamma^{-1}(y-X\beta).$$

These parameters can be estimated using the frequentist method of maximum likelihood estimation, and Bayesian methods of accounting for prior beliefs.

Maximum likelihood estimation for the simple linear model

Maximum likelihood estimation (MLE) for the simple linear regression model is as follows,

$$\hat{\beta},\widehat{\sigma^2}=argmax_{(\hat{\beta},\widehat{\sigma^2})}\,logL(\beta,\sigma^2|y),$$

which is read as: find the arguments $\hat{\beta},\widehat{\sigma^2}$ which maximise the log-likelihood function. The derivative of the log-likelihood function with respect to β is,

$$\frac{\nabla logL(\beta,\sigma^2)}{\nabla\beta} = \frac{1}{\sigma^2}X^T y - \frac{1}{\sigma^2}X^T X\beta \ ,$$

And the partial derivative with respect to σ^2 is,

$$\frac{\partial logL(\beta,\sigma^2)}{\partial\sigma^2} = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y-X\beta)^T(y-X\beta) \ .$$

These are equated to zero and solved for the parameters,

$$\hat{\beta}=\left(X^T X\right)^{-1}X^T y=X^T\left(XX^T\right)^{-1}y,$$

and

$$\widehat{\sigma^2}=\frac{1}{n}\left(y-X\hat{\beta}\right)^T\left(y-X\hat{\beta}\right).$$

These are the maximum likelihood estimates of the SNP effects and residual variance.

Maximum likelihood estimation for the linear mixed model

MLE for the linear mixed model is similar,

$$\hat{\beta},\hat{\theta}=argmax_{(\hat{\beta},\hat{\theta})}\,logL(\beta,\theta|y),$$

or

$$\hat{\beta},\hat{\gamma},\widehat{\sigma^2}=argmax_{(\hat{\beta},\hat{\gamma},\widehat{\sigma^2})}\,logL(\beta,\gamma,\sigma^2|y),$$

where $\theta=\{\sigma_g^2,\sigma^2\}$ and $\gamma=\sigma_g^2/\sigma^2$. The derivative of the log-likelihood function with respect to β is,

$$\frac{\nabla logL(\beta,\theta)}{\nabla\beta} = X^T V_\theta^{-1} y - X^T V_\theta^{-1} X\beta \ ,$$

or

$$\frac{\nabla logL(\beta,\gamma,\sigma^2)}{\nabla\beta} = -\frac{1}{\sigma^2}X^T H_\gamma^{-1} y - \frac{1}{\sigma^2}X^T H_\gamma^{-1} X\beta$$

where $V_\theta=\sigma_g^2 K+\sigma^2 I$ and $H_\gamma=I+\gamma K$. And the partial derivative of $logL(\beta,\theta\vee y)$ with respect to θ is,

$$\frac{\partial logL(\beta,\theta)}{\partial\theta_i} = -\frac{1}{2}trace\left(V_\theta^{-1}\frac{\partial V_\theta^{-1}}{\partial\theta_i}\right) + \frac{1}{2}(y-X\beta)^T V_\theta^{-1}\frac{\partial V_\theta^{-1}}{\partial\theta_i}(y-X\beta) \ ,$$

where

$$\frac{\partial V_\theta^{-1}}{\partial \theta_i} = \begin{Bmatrix} I, & if\ \theta_i = \sigma^2 \\ K, & if\ \theta_i = \sigma_g^2 \end{Bmatrix};$$

while the partial derivative of $logL(\beta, \gamma, \sigma^2 \vee y)$ with respect to σ^2 is,

$$\frac{\partial logL(\beta, \gamma, \sigma^2)}{\partial \sigma^2} = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)^T H_\gamma^{-1}(y - X\beta) \quad .$$

The derivatives of the alternative log-likelihood $logL(\beta, \gamma, \sigma^2 \vee y)$ with respect to β and σ^2 are usually used,

$$\hat{\beta} = \left(X^T H_\gamma^{-1} X\right)^{-1} X^T H_\gamma^{-1} y,$$

and

$$\widehat{\sigma^2} = \frac{1}{n}(y - X\beta)^T H_\gamma^{-1}(y - X\beta).$$

These parameter estimates cannot be solved analytically because they require γ, i.e. the ratio of genetic variance over residual variance. The maximum likelihood estimate of γ can be obtained with optimisation algorithms including gradient descent and derivative-free methods. This generally works by systematically choosing values for $\hat{\gamma}$ from some predefined search space (i.e. the set of all values $\hat{\gamma}$ can be), use this to compute H_γ, $\hat{\beta}$ and $\widehat{\sigma^2}$, then evaluate the log-likelihood function, $logL(\beta, \gamma, \sigma^2 \vee y)$ for multiple iterations until convergence, i.e. the log-likelihood function is maximised. Thus the likelihood function essentially becomes a function of γ alone.

ML estimation results in negatively biased variance components, i.e. underestimated variances because of their dependence on the distribution of the fixed effects (β),

$$E\left(\widehat{\sigma^2}\right) = \frac{n-1}{n}\sigma^2.$$

Experiments with small sample sizes will suffer from larger bias than experiments with more samples. Restricted maximum likelihood (REML) estimation generates unbiased variance estimates by estimating them from the residual effects of the model, i.e. after accounting for the fixed effects (Patterson and Thompson 1971). To illustrate with the simple linear model, let

$$y = X\beta + \varepsilon, \text{ with}$$

$$y \sim Normal(X\beta, \sigma^2 I).$$

We need to find K, such that,

$$K^T y = K^T X\beta + K^T \epsilon,$$

$$K^T X = 0, \text{ and}$$

$$K^T y \sim Normal(0, K^T \sigma^2 IK),$$

then $\widehat{\sigma^2}$ does not depend on β.

Bayesian linear regression

Bayesian inference is based on Bayes' theorem (Bayes and Price 1763),

$$P(\theta|y) = \frac{P(y|\theta)\,P(\theta)}{P(y)} \quad ,$$

where $P(\theta|y)$ is referred to as the posterior probability which in the context of GWAS is the probability of the SNP effects and variance components given the phenotype data, $P(y|\theta)$ is

the likelihood function, $P(\theta)$ is referred to as the prior probability which is the assumed distribution of the parameters being estimated, and $P(y)$ is referred to as the evidence which is the probability of the phenotypes. Since the $P(y)$ is a fixed term, i.e. the given data do not change, then the Bayes' theorem can be expressed simply as,

$$P(\theta|y) \propto P(y|\theta)P(\theta) \quad .$$

The prior distribution, $P(\theta)$ is first defined. This can be uninformative uniform distribution or informative based on the expectation that for example the SNP effects are normally distributed with most SNPs having small to no effect. Different prior distributions may be used for each of the parameters being estimated, i.e. different search spaces and their respective probability distributions for the SNP effects, population structure effects, and variance components. In cases where no analytical solution for the posterior distribution is available given the prior distribution, a conjugate prior distribution is used instead. A conjugate prior is in the same probability distribution family as the posterior distribution (Murphy 2007).

The same likelihood function is used as in MLE. The parameters are then drawn from the prior distribution and plugged into Bayes' formula to generate a posterior distribution. The prior distribution is then updated based on how well the parameters fit the given data. These series of steps are repeated multiple times until convergence or until some criteria are met, e.g. the maximum number of iterations is reached, using various Markov Chain Monte Carlo (MCMC) algorithms such as the Gibbs sampler (Geman and Geman 1984), and Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970). The first several posterior distribution results are usually excluded because these are usually outliers especially when uninformative priors are used. This is called the burn-in step.

Statistical significance tests

The statistical significance of SNP-to-phenotype associations, i.e. the estimated SNP effects, $\widehat{\beta}_j$ can be assessed using t-tests, and likelihood-ratio tests (LRT) with correction for multiple testing, among many other tests. These tests compare the null hypothesis H_0, i.e. the SNP has no effect on the phenotype, and the alternative hypothesis H_a, i.e. the SNP has an effect on the phenotype. Generally, the distribution of $\widehat{\beta}_j$ *per se* is not used as the null distribution, instead the distribution of the t-statistic or the likelihood-ratio test statistic is used. The H_0 distributions of these test statistics are used to generate p-values, i.e. the probability of rejecting the null hypothesis given the null hypothesis is true, also known as the family-wise error rate (FWER), or type I error rate, or false positive rate. A SNP is significantly associated with the phenotype if its corresponding p-value is equal to or lower than some significance level threshold usually denoted by α. The significance threshold or p-values need to be corrected since GWAS involves tens of thousands to millions of SNPs. Using the usual significance thresholds of 1% and 5% are too lenient, e.g. at $\alpha = 1\%$ for a GWAS with 1 million SNPs this equates to 10,000 spurious associations.

The t-statistic is computed as,

$$t = \frac{\widehat{\beta}_j}{\widehat{\sigma_{\beta j}}}, \text{ and } t \sim T_{df=n-2} \text{ or } t^2 \sim \chi^2_{df=1},$$

where $\widehat{\beta}_j$ is the estimated effect of the j[th] SNP, and $\widehat{\sigma_{\beta j}} = \sqrt{\widehat{\sigma^2}\left(X^T X\right)^{-1}_{jj}}$ is the estimated standard error of j[th] SNP effect. This test statistic follows Student's t distribution with n-2 degrees of freedom, alternatively its square is χ^2-distributed with one degree of freedom.

The likelihood-ratio test statistic is expressed as,

$$LRT = 2\left(\frac{max(L(\theta|H_a))}{max(L(\theta|H_0))}\right), \text{ and } LRT \sim \chi^2_{df=1},$$

where the null hypothesis, H_0 assumes the effect of the SNP in the phenotype is zero, $\beta_j = 0$; while the alternative hypothesis, H_a assumes a non-zero effect, $\beta_j \neq 0$, and this test statistic is χ^2-distributed with one degree of freedom (Hartley and Rao 1967).

The corrected genome-wide significance level threshold can be defined with Bonferroni correction, Benjamini-Hochberg false discovery rate, and permutation tests. Bonferroni correction calculates the p-value threshold as,

$$\alpha / m,$$

where α is usually 5% and m is the number of SNPs (Bonferroni 1936). This assumes independence between SNPs, however because of linkage, Bonferroni correction is often too stringent which can result in elevated false negative rate (Marees et al. 2018). The Benjamini-Hochberg false discovery rate (FDR) is a less conservative method of controlling for false positives. The FDR procedure is as follows, let $p_1 \leq p_2 \leq ... p_m$ be the ordered p-values for m SNPs, find the largest k, such that,

$$p_k \leq \frac{k}{m}\alpha,$$

and reject the null hypothesis for SNPs with p_i where $i = \{1,2,3,...,k\}$ (Benjamini and Hochberg 1995). Permutation method generates a realised distribution of the p-values given the empirical data where phenotype and genotype data are permuted thousands to millions of times, which effectively removes real associations. The percentile corresponding to α, on this realised p-value distribution defines the significance threshold (Gao et al. 2008).

Pool GWAS: genome-wide association using pool data

So far all the models and tests described above assume individual genotype data. Similar linear models can be used on pooled genotype data, however this is uncommon in pool GWAS. Contingency tests using two extreme pools are commonly practised. These include Fisher's exact test and Cochran-Mantel-Haenszel test. Contingency tests involve calculating the probability of the frequency distribution described by a contingency table given its marginal probabilities. The null hypothesis is the absence of association between two categorical variables which in terms of GWAS means there is no association between the alleles at a locus and the categorical phenotypes. To illustrate, consider the 2×2 contingency table below (Table 1.3).

Table 1.3. A contingency table for genome wide association study using pooled genotype data describing the number of alleles under different categories.

	Allele A1	Allele A2	Total allele counts per pool
Positive Pool	a	b	R_1=a+b
Negative Pool	c	d	R_2=c+d
Total counts per allele	C_1=a+c	C_2=b+d	n=a+b+c+d

Fisher's exact test (Fisher 1922), calculates a conditional probability of obtaining this matrix given the marginal sums,

$$P = \frac{R_1!R_2!C_1!C_2!}{a!b!c!d!n!}.$$

This probability is then computed for all possible values of a, b, c, and d given constant R_1, R_2, C_1, and C_2 marginal sums. Finally, the probabilities equal to or lower than that of the original data are added up to generate the p-value, i.e. the probability of obtaining this contingency table given the marginal sums purely by chance.

Cochran-Mantel-Haenszel test (Cochran 1954; Mantel and Haenszel 1959) can be used, if the

observed contingency table is replicated such that a_i, b_i, c_i, d_i are the allele counts, R_{1i}, R_{2i}, C_{1i}, C_{2i} are the marginal sums, and n_i is the total allele count for the i[th] replication. The odds of individuals in the "Positive Pool" having allele A1 is $(a/R_1)/(b/R_1)$, while the odds of individuals in the "Negative Pool" having allele A1 is $(c/R_2)/(d/R_2)$. The ratio of these odds across k replicates is,

$$R=\left(\frac{\sum_{i=1}^{k} a_i d_i}{n_i}\right)\div\left(\frac{\sum_{i=1}^{k} b_i c_i}{n_i}\right).$$

The null hypothesis is that the two odds are equal, i.e. $R=1$, while the alternative hypothesis is $R\neq 1$. The test statistic to compute the p-value is,

$$Q=\left(\sum_{i=1}^{k}\left(a_i-\frac{R_{1i}C_{1i}}{n_i}\right)\right)^2\div\left(\frac{\sum_{i=1}^{k} R_{1i}C_{1i}R_{2i}C_{2i}}{n_i^2(n_i-1)}\right),$$

where $Q\sim\chi^2_{df=1}$ (Kuritz, Landis, and Koch 1988).

More recently a parametric test for more than two extreme pools, GWAlpha (Fournier-Level et al. 2017) has been developed. GWAlpha models the distribution of allele frequencies across pools sorted by mean phenotype values as a Beta distribution. The shape parameters of this distribution, θ_1 and θ_2, where each distribution corresponds to the frequency of an allele per locus across pools, are estimated using least squares or maximum likelihood methods. The test statistic of the i[th] allele in a locus is computed as,

$$\hat{\alpha}_i=2\sqrt{p_i(1-p_i)}\left(\hat{\mu}_i-\hat{v}_i\right)/\sigma_y,$$

where p_i is the frequency of the i[th] allele, $\hat{\mu}_i$ is the mean of the Beta distribution of the i[th] allele across the k pools, i.e. $\mu_i=\hat{\theta}_1/(\hat{\theta}_1+\hat{\theta}_2)$, $\hat{v}_i$ is the mean of the Beta distribution of the additive inverse of the i[th] allele frequency, i.e. $1-p_i$, across the pools. This test statistic follows a normal distribution with parameters estimated using maximum likelihood, and from which empirical p-values are computed.

Missing heritability

The significance level thresholds in GWAS are usually very stringent to minimise false positives at the cost of elevated false negative rate. This resulted in the problem of missing heritability, where only a small fraction of the phenotypic variability can be explained by the QTL that were identified (Manolio et al. 2009; Zuk et al. 2012). This has been resolved by recognising the small effects of numerous loci below stringent significance thresholds (Manolio et al. 2009; Yang et al. 2010; Shi et al. 2016). In response to the poor explanatory power or the poor predictive ability of models built using these GWAS-detected QTL and the suboptimal use of dense molecular marker information, a phenotype prediction approach using all marker information, i.e. genomic prediction was developed.

1.5.5. Genomic prediction models

Models for genomic prediction are similar to that of GWAS, with some key differences. The effects of all SNPs included in the study are estimated simultaneously instead of one at a time. And since there are more SNP effects (p) that need to be estimated than the number of individuals or observations (n), then a simple linear model is insufficient. Ordinary least squares regression or maximum likelihood estimation for a simple linear model does not have enough degrees of freedom to estimate the full model, i.e. there are insufficient observations to estimate all the effects. Multicollinearity often exists among the SNPs, i.e. SNPs in the X matrix are

correlated, which results in singular $X^T X$ which means it cannot be inverted and β cannot be estimated. Various approaches have been developed to circumvent this $n \ll p$ problem, including ridge regression (RR), least absolute shrinkage and selection operator (LASSO) regression, elastic-net regularisation, linear mixed modelling with ML or REML estimation, Bayesian linear regression, and machine learning approaches such as random forest and neural networks.

Ridge regression

Ridge regression (RR) attempts to generate a parsimonious linear model by shrinking the estimated effects, $\hat{\beta}$ towards zero using a shrinkage or tuning parameter $\lambda \geq 0$, which can be thought of as $\lambda = \sigma^2 / \sigma_g^2$ (Whittaker et al. 2000; Searle et al. 2009; Endelman 2011). The linear model is a fully random model which can be expressed as,

$$y = X\beta + \epsilon,$$

with

$$\beta \sim Normal\left(0, \sigma_g^2 I\right), \text{ and } \epsilon \sim Normal\left(0, \sigma^2 I\right),$$

where y is the vector of phenotype values, X is the $n \times p$ matrix of SNP data which is centred, i.e. coded as -1, 0, 1 or dosage based 0, 1, 2 if y is standardised, since the vector of normally distributed SNP effects (β) are centred on zero with common variance σ_g^2, ϵ is the vector of normally distributed residual effects with mean zero and variance σ^2, and I is an identity matrix. The aim is to minimise the cost function defined as the residual sum of squares with Lagrangian penalty on the values of $\hat{\beta}$. This penalty is called L2 since the Lagrangian multiplier, λ is multiplied with the Euclidean norm or the square of SNP effects, $||\beta||_2^2$, i.e. $\lambda\left(\beta^T \beta\right)$, which is why RR is also called the L2 regularised regression,

$$\hat{\beta} = argmin_{(\hat{\beta})}\left(y - X\beta\right)^T \left(y - X\beta\right) + \lambda ||\beta||_2^2.$$

The estimates of the SNP effects or the best linear unbiased predictors (BLUP; since these are random effects), $\hat{\beta}$ are computed as,

$$\hat{\beta} = \left(X^T X - \lambda I\right)^{-1} X^T y.$$

Various optimisation algorithms can be used to find the optimal λ which minimises the cost function.

LASSO regression

Least absolute shrinkage and selection operator (LASSO) regression attempts to generate an even more parsimonious model than RR by allowing some effects to be exactly zero. The LASSO model is similar to that of RR. The cost function for LASSO uses the L1 penalty $\lambda ||\beta||_1$, i.e. the Lagrangian multiplier is multiplied to the absolute value of the SNP effects (Tibshirani 1996), instead of the L2 penalty in RR. LASSO regression or L1 regularisation can be expressed as,

$$\hat{\beta} = argmin_{(\hat{\beta})}\left(y - X\beta\right)^T \left(y - X\beta\right) + \lambda ||\beta||_1.$$

If λ=0, then none of the estimated SNP effects are zero. More effects are set to zero as λ increases, and as $\lambda \rightarrow \infty$ all the effects becomes zero. In terms of variance components ($\lambda = \sigma^2 / \sigma_g^2$), as residual variance increases or as genetic variance decreases, λ increases and less SNPs contribute to the phenotype. The $\hat{\beta}$ and λ which minimise the cost function are determined using various quadratic optimisation algorithms. LASSO regression generates a sparse $\hat{\beta}$ which can improve the interpretability of the model.

Elastic-net regression

Elastic-net regression combines the L1 and L2 penalties of RR and LASSO into a single framework (Friedman et al. 2010). Elastic-net regularisation can be expressed as,

$$\hat{\beta} = argmin_{(\hat{\beta})}(y - X\beta)^T(y - X\beta) + \lambda\left((1-\alpha)\frac{1}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1\right),$$

where $\alpha \in [0,1]$ is the compromise between RR and LASSO. An $\alpha = 0$ corresponds to a L2 penalty or RR, while $\alpha = 1$ corresponds to L1 penalty or LASSO regression. The estimates of the SNP effects $\hat{\beta}$ are solved using various quadratic optimisation algorithms for a range of decreasing λ starting from λ_{max} where all SNP effects are zero. The λ which minimises the cost function and the corresponding $\hat{\beta}$ are determined.

Linear mixed modelling with maximum likelihood estimation

Another approach to circumvent this $n \ll p$ problem in GP is to use linear mixed models and maximum likelihood estimation of the random SNP effects. This linear mixed model can be expressed as,

$$y = X\beta + Zu + \epsilon,$$
with
$$y \sim Normal(\mu = X\beta, V_\theta = \sigma_g^2 K + \sigma^2 I),$$
$$u \sim Normal(0, \sigma_g^2 K),$$
and
$$\varepsilon \sim Normal(0, \sigma^2 I),$$

where y is the vector of n phenotype values, X is the design matrix for the fixed effects β which in this case is just the phenotype mean and therefore X is simply a vector of ones, Z is the $n \times p$ matrix of SNP data, u is the vector of random SNP effects, and ϵ is the vector of random residual effects. The phenotype values follow a multivariate normal distribution with vector of means μ and variance-covariance matrix V_θ. The random SNP effects also follow a multivariate normal distribution with a vector of zero means and variance-covariance matrix equivalent to the product of the genetic variance, σ_g^2 and the $n \times n$ population structure matrix, K which is usually proportional to ZZ^T. The random residual effects also follow a multivariate normal distribution with a vector of zero means and variance-covariance matrix $\sigma^2 I$, where I is an $n \times n$ identity matrix. The log-likelihood function is defined differently than previously shown for LMM in GWAS,

$$logL(\beta, \sigma_g^2, \sigma^2 \vee y) = \frac{-n}{2}\log(2\pi\sigma_g^2) - \frac{1}{2}\log(det(H)) - \frac{1}{2\sigma_g^2}(y - X\beta)^T H^2 (y - X\beta),$$

where $\gamma = \sigma^2/\sigma_g^2$, $H = K + \gamma I$, the fixed and random effects are estimated as,

$$\begin{bmatrix}\hat{\beta}\\ \hat{u}\end{bmatrix} = \begin{bmatrix}X^T X & X^T Z\\ Z^T X & Z^T Z + \gamma K^{-1}\end{bmatrix}^{-1}\begin{bmatrix}X^T y\\ Z^T y\end{bmatrix}.$$

The variance components σ_g^2 and σ^2 which maximise this log-likelihood function are determined using optimisation algorithms, and in the process the fixed effects (also called the best linear unbiased estimators, BLUE) and random effects (also called the best linear unbiased predictors, BLUP) are estimated (Wang et al. 2018). This is the maximum likelihood estimation approach. Restricted maximum likelihood estimation can also be performed.

Bayesian linear regression

The SNP effects and variance components can be estimated using Bayesian approach under

the same linear model specifications as described above. For RR, the prior of the SNP effects, β is normally distributed with mean zero and the probability of SNPs having exactly zero is infinitesimally small. For LASSO, the prior of the SNP effects, β follows the Laplace distribution where many SNP effects are close to zero such that some are approximately zero. For LMM, the SNP effects u, can assume various prior distributions, e.g. inverted chi-square distribution as in Bayes A, mixture of 0 (with probability π) and inverted chi-square distribution (with probability $1-\pi$) as in Bayes B, mixture of 0 (with probability π) and normal distribution (with probability $1-\pi$) as in Bayes C, and similar to Bayes C but π is another parameter that needs to be estimated with a uniformly distributed prior as in Bayes Cπ (Wang et al. 2018).

The SNP effects and variance components are estimated using various MCMC algorithms. Briefly, using these priors, a set of parameters are proposed, i.e. sampled from the prior distribution. The posterior distribution, i.e. the probability of the SNP effects and variance components being estimated given the phenotype data, is calculated and used to update the prior for the next iteration. These steps are repeated multiple times until some criteria are met and the best parameter values are selected.

Machine learning methods

All the linear models described so far, accounts for the additive genetic effects of the SNPs and allocates the non-additive genetic effects, i.e. dominance and epistatic effects into the error term. Incorporating these non-additive genetic effects into a linear model is intractable, i.e. the 2-way, 3-way, and so on SNP interaction matrix will be massive and likely to be highly multicollinear, thereby exacerbating the $n \ll p$ problem. Non-parametric machine learning methods have the potential to solve this problem without the need to explicitly define the genotype-to-phenotype model while having the ability to quantify additive, dominance and epistatic genetic effects as well as SNP-covariate interactions, e.g. genotype-environment (GxE) interactions (Lubke et al. 2013; Li et al. 2018).

Exploiting the rapidly accumulating phenomics and genomics information driven by the increasing accessibility to massively high-throughput phenotyping and genotyping platforms (Reuter et al. 2015; Yang et al. 2020), machine learning methods can elucidate the genetic basis of traits beyond additive effects. Hence, these model-free non-parametric methods have the potential to improve the explanatory power of genotype-to-phenotype maps. Machine learning methods applied to genomic prediction include random forest (Ogutu et al. 2011; Li et al. 2018; Abdollahi-Arpanahi et al. 2020), and deep learning (González-Camacho et al. 2012; Montesinos-López et al. 2018; Waldmann 2018; Ma et al. 2018; Abdollahi-Arpanahi et al. 2020; Zingaretti et al. 2020).

Polygenic Scores

Alternatively, genomic predictions can be made using allele effects derived from one SNP at a time GWAS experiments. This is known as polygenic risk score analysis (Torkamani et al. 2018). This was the first form of genomic prediction in humans (International Schizophrenia Consortium 2009). The canonical procedure to predict these polygenic risk scores is to only use the disease-associated loci (Dudbridge 2013). The polygenic score is computed as,

$$s = G\beta,$$

where G is the $n \times p$ matrix of genotype information, and β is the vector of p allele effects of the significantly associated loci. More contemporary methods of estimating polygenic risk scores or polygenic scores in general, are the same as the GP methods discussed above, i.e. employing trait and genome-wide marker regression, where the allelic effects are estimated using various methods from penalised regression to Bayesian estimation (Abraham and Inouye 2015).

Prediction accuracy assessment

Prediction accuracy or predictive ability can be quantified using Pearson's product moment correlation, mean absolute error and root mean square error. Pearson's product moment correlation (r) is defined as,

$$r = \frac{Cov(y, \hat{y})}{\sigma_y \sigma_{\hat{y}}} = \frac{\sum_{i=1}^{n} (y_i - \mu_y)(\hat{y}_i - \mu_{\hat{y}})}{\sqrt{\sum_{i=1}^{n} (y_i - \mu_y)^2 \sum_{i=1}^{n} (\hat{y}_i - \mu_{\hat{y}})^2}},$$

where y and $\hat{y}$ are the vectors of observed and predicted phenotype values, respectively, y_i and $\hat{y}_i$ are the observed and predicted phenotype value, respectively of the ith individual, μ_y is the mean of y, $\mu_{\hat{y}}$ is the mean of $\hat{y}$, and n is the number of individuals or observations. Mean absolute error (MAE) is defined as,

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \| y - \hat{y} \|_1.$$

Root mean square deviation (RMSE) is defined as,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y - \hat{y})^2}.$$

Pearson's correlation coefficient is the common metric used in animal and plant breeding since ranking the breeding lines is more important than predicting close to the actual phenotypic values. Large errors have more weight in RMSE than MAE. In this sense, RMSE is stricter than MAE.

Cross-validation is needed to assess the genomic prediction accuracy or the predictive ability of a genomic prediction model. Cross-validation assesses how models generalise to independent data sets (Stone 1974). This involves dividing the data (simulated or empirical) into two mutually exclusive sets, i.e. the training set and the validation or testing set. The genomic prediction model is built or trained using the phenotype and genotype data of the training set. This model is then used to predict the phenotypes of the validation set. To quantify the uncertainty of prediction accuracies, multiple iterations of cross-validation are performed using different training and validation sets. Prediction accuracy is measured for each training-validation set pair, and the accuracies are summarised across iterations, e.g. the mean and standard deviation are calculated.

There are various ways to divide the dataset to train the model and predict the phenotypes. Some of the commonly used ones in genomic prediction are the k-fold cross-validation where k is usually 10, and leave-p-out cross-validation where p is usually 1. In 10-fold cross-validation, the dataset is randomly partitioned into 10 equally-sized mutually exclusive sets (Wong 2015). The model is trained on the 9 sets and validated on the remaining set, for all set combinations. This is repeated multiple times to generate different set compositions for every replication. In leave-one-out cross-validation, each individual is used as the validation set for the model trained on the rest of the data for each iteration (Wong 2015).

1.6. Evolutionary paths towards herbicide resistance in *L. rigidum* populations across SE Australia

The shared ancestry of all life on earth implies a common set of evolutionary constraints tied to the molecular basis of life which results in convergent solutions to many evolutionary problems. These convergent evolutionary solutions are exemplified by the adaptive evolution of weeds in agroecosystems, specifically in the evolution of herbicide resistance (Baucom 2019). Herbicide resistance evolution is a tractable system to study adaptive evolution in real-time (Baucom 2016). Herbicide resistance-conferring alleles arose from standing genetic variation, and/or *de novo* mutations, which may have been distributed across the landscape through gene flow, and pushed towards fixation by selection pressure from herbicides and random genetic drift. These evolutionary forces leave behind genomic signatures which can be detected and analysed to infer the likely evolutionary paths towards herbicide resistance. Insights that will be gained attempting to answer the following questions will improve our understanding of adaptive evolution in general, as well as aid in the development of better weed and herbicide resistance control strategies. What is the predominant source of herbicide resistance alleles, standing genetic variation or *de novo* mutations? How much is the contribution of gene flow to the observed variation, specifically are resistance alleles appearing independently in each population or do they originate from one or a few source populations?

1.6.1. Effective population size

Identifying standing genetic variation or *de novo* mutations as the source of favourable alleles requires estimates of the initial frequency of favourable alleles at the start of selection, q_0 (Olson-Manning et al. 2012). If q_0 is sufficiently higher than the expected frequency of a new mutation, $1/2N_e$, then favourable alleles are more likely to have come from standing genetic variation rather than from *de novo* mutations. The effective population size, N_e is the size of an idealised panmictic population with the same inbreeding rate as the observed population (Crow and Kimura 1970).

Estimates of N_e rely on the scaled mutation rate, $\theta = 4 N_e \mu$, where μ is mutation rate. This scaled mutation rate can be estimated in two ways: first as the sequence diversity (Tajima's $\hat{\pi}$) and second as the number of polymorphic sites (Watterson's estimator $\hat{\theta}$) (Xu and Fu 2004). These estimators can be calculated from Pool-seq data corrected for pool sizes and sequencing coverage (Kofler et al. 2011). Large sample sizes estimate present-day N_e better than small sample sizes because of the higher probability of sampling new and therefore rare variants (Tennessen and Akey 2011; Keinan and Clark 2012; Nelson et al. 2012). Small sample sizes, on the other hand, estimate historical N_e better, since older and more common variants are predominantly detected (Olson-Manning et al. 2012).

1.6.2. Mutation rate and fitness cost

The ultimate source of all genetic variation is stochastic, i.e. random mutations (Nei 1980; Barton 2010). These mutations may confer beneficial, detrimental, or neutral effects on fitness as well as various pleiotropic effects on multiple traits. The resulting mutation-selection balance is the mechanism which maintains quantitative trait variability even under constant selection pressures (Zhang and Hill 2005).

Mutation rate (μ) can be defined as the average number of new random single-base mutations per individual genome per generation (Crow and Kimura 1970; Crow 1997). It can be estimated by counting the number of mutant individuals, or the number of distinct mutation events in multi-generation experiments (Drummond et al. 2002; Fu and Huai 2003). In plants, estimates of mutation rates range from 1.0×10^{-9} to 1.0×10^{-7} (Kovalchuk et al. 2000; Wang et al. 2019).

Genetic variants which confer resistance to an herbicide can exhibit pleiotropic effects which can have beneficial, detrimental or neutral effects on fitness. There is an increased chance of losing a potentially favourable allele if its fitness cost is high (Ralph and Coop 2015). It follows

that alleles conferring resistance to multiple herbicides which also positively affects fitness in herbicide-free environments are highly favourable.

1.6.3. Genomic signatures of selection

The effects of selection can be detected as regions in the genome where variations significantly deviate from neutral expectations. Neutral polymorphisms around a beneficial allele are lower if it originated from *de novo* mutation than if it originated from standing genetic variation. Figure 1.12 depicts this hitchhiking of neutral polymorphisms linked to the beneficial allele under selection. The extent of the valley of low neutral allelic variation around the selected site decreases as the age of the beneficial allele increases, since a beneficial allele present in the ancestral population has been exposed to more recombination events than a new mutation.

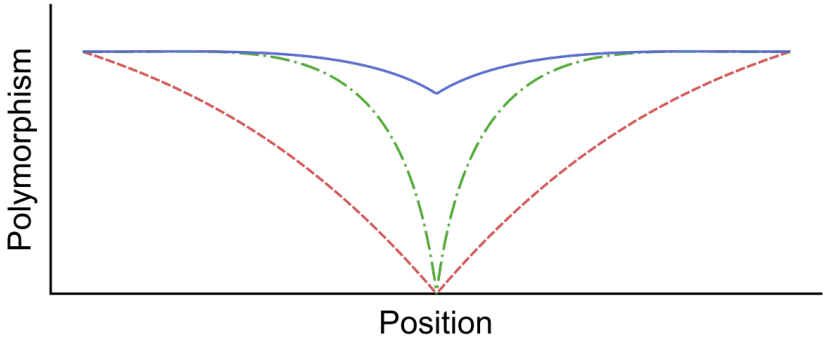


Figure 1.12. Expected genomic signatures of selection around a beneficial allele found at the central position when this allele originated from *de novo* mutation (red dashed line), or standing genetic variation (green dash-dotted line), or multiple beneficial alleles from different genetic backgrounds (blue solid line) (adapted from Barrett and Schluter 2008).

These genomic signatures of selection can be a highly pronounced reduction in allelic variation around the causal allele referred to as hard sweep (Maynard and Haigh 2007), or alternatively a soft sweep (Hermisson and Pennings 2005), where this reduction is less pronounced and harder to detect (Olson-Manning et al. 2012). These sweeps can inform us of the origin of favourable alleles as shown in Table 1.4. Detecting these signatures of selection requires estimates of recombination rate across the genome.

Table 1.4. The types of selective sweeps expected under the different origins of favourable alleles (adapted from Olson-Manning et al. 2012 and Hermisson and Pennings 2005).

	Standing genetic variation	**New recent mutation**
Single Origin	Soft sweep	Hard sweep
Multiple origins / Independent parallel mutation events	Soft sweep	Soft sweep

Recombination rate (r) can be defined as “the probability that a transmitted haplotype constitutes a new combination of alleles different from that of either parental haplotype” (Clarke and Cardon 2005). Low recombination rate around a new beneficial mutation can hinder fixation in the presence of tightly linked unfavourable mutations (genetic load or linkage drag) (Charlesworth et al. 1990; Schoen et al. 1998; Charlesworth 2012). It follows that selection favours genomic regions of high recombination (Charlesworth 2012). High recombination can also be unfavourable if it promotes breaking the linkage between favourable alleles, however this is less common than the previous case (Kirkpatrick and Barton 2006; Hoffmann and Rieseberg 2008). Recombination rate can be estimated from linkage disequilibrium (LD) information (Clarke and Cardon 2005) by maximising the likelihood function:

$$L(r)=(\hat{p}_{ab}-\hat{D}r)^{n_{ab}}(\hat{p}_{aB}+\hat{D}r)^{n_{aB}}(\hat{p}_{Ab}+\hat{D}r)^{n_{Ab}}(\hat{p}_{AB}-\hat{D}r)^{n_{AB}},$$

where $\hat{p}_{ij}$and $n_{ij}$are the frequency and number of haplotype *ij*, respectively, and $\hat{D} = \hat{p}_{ij} - \hat{p}_i \hat{p}_j$. LD information can be extracted from Pool-seq data using sequencing read information since haplotype information is unavailable (Kofler et al. 2011). It follows that long-range LD is not captured with Pool-seq using short-read sequencing.

1.6.4. Genetic differentiation and gene flow

Genetic differentiation among and within resistant and susceptible populations as affected by gene flow are critical in identifying the likely origin of herbicide resistance alleles. Highly differentiated resistant populations suggest independent mutations. On the other hand, less differentiated resistant populations suggest standing genetic variation as the origin of the resistance allele or the spread of the resistance allele through gene flow.

Genetic differentiation or the genetic distance between groups of individuals can be estimated using Nei's genetic distance (D), and the commonly used fixation index (F_{ST} sometimes referred to as G_{ST}). Nei's genetic distance is defined as $D = -log_e I$, where I is the normalized identity of genes between two populations (Nei 1972). F_{ST} is a measure of heterozygosity which is adapted to define relatedness between groups of individuals within and among populations (Wright 1949). However, equating heterozygosity to relatedness can yield misleading results (Jost 2008). These genetic differentiations between populations are highly affected by gene flow.

Gene flow influences herbicide resistance evolution across a landscape by decreasing genetic differentiation between populations, the magnitude of which depends on local selection pressures and population sizes (García-Ramos and Kirkpatrick 1997; Olson-Manning et al. 2012; Santangelo et al. 2018). High gene flow decreases differentiation between populations; however sufficiently high local selection pressures and/or small population sizes can increase differentiation between populations even in the presence of high gene flow (Jain and Bradshaw 1966). Therefore the strength of gene flow can be inferred by assessing differentiation within and among resistant and susceptible populations in the context of estimated N_e and selection pressure. In addition, maladaptive gene flow originating from populations free from the selection pressure, i.e. border and feral populations, can selectively favour gene duplication which can mask the effects of the maladaptive immigrant alleles (Olson-Manning et al. 2012; Yanchukov and Proulx 2012).

1.6.5. Inferring likely models of herbicide resistance evolution

Elucidating the most likely evolutionary paths taken by herbicide-resistant populations requires estimates of the genetic differentiation within and among resistant and susceptible populations, and the assessment of the genomic signatures of selection. Resistance alleles may have originated from standing genetic variation or *de novo* mutations and spread across the landscape through migration via seeds or pollen.

If standing genetic variation was the origin of the resistance alleles, then we expect to observe the following. Genetic differentiation is low among resistant populations; but high between resistant and susceptible populations. Fitness costs of the resistance alleles are low. In terms of genomic signature per population, soft sweeps are expected, i.e. the region of reduced genetic variation surrounding the resistance allele occurs on multiple different backgrounds. These are indicators of a common ancestral population harbouring the resistance alleles prior to selection pressure from herbicides. This origin of resistance alleles facilitates rapid adaptive evolution (Busi et al. 2013).

If resistance-conferring mutations arose independently across populations and in parallel with the selection pressure from herbicides, we expect the following. Estimates of N_e and genetic differentiation between resistant populations are expected to be high; while gene flow is low. In terms of genomic signature per population, hard sweeps are expected, since the background surrounding the resistance allele is likely to have come from a single individual. Furthermore, if gene flow is high then there is an increased chance of a single or few source populations from

which the resistance alleles originated. However, it should be noted that resistance alleles may originate from standing genetic variation in some populations and from new mutations in others in the same landscape.

Five models of herbicide resistance evolution are illustrated in Figure 1.13. These depict different scenarios for the convergent evolution of herbicide resistance across multiple populations. In the independent mutations model, the resistance allele independently appeared via *de novo* mutation in each of the populations. In the migration model, the resistance allele appeared via *de novo* mutation in one population and spread to the other populations through migration. In the standing variant model, the resistance allele was present in the ancestral population. The last two models are combinations of migration and mutation models; and standing variant and mutation models.

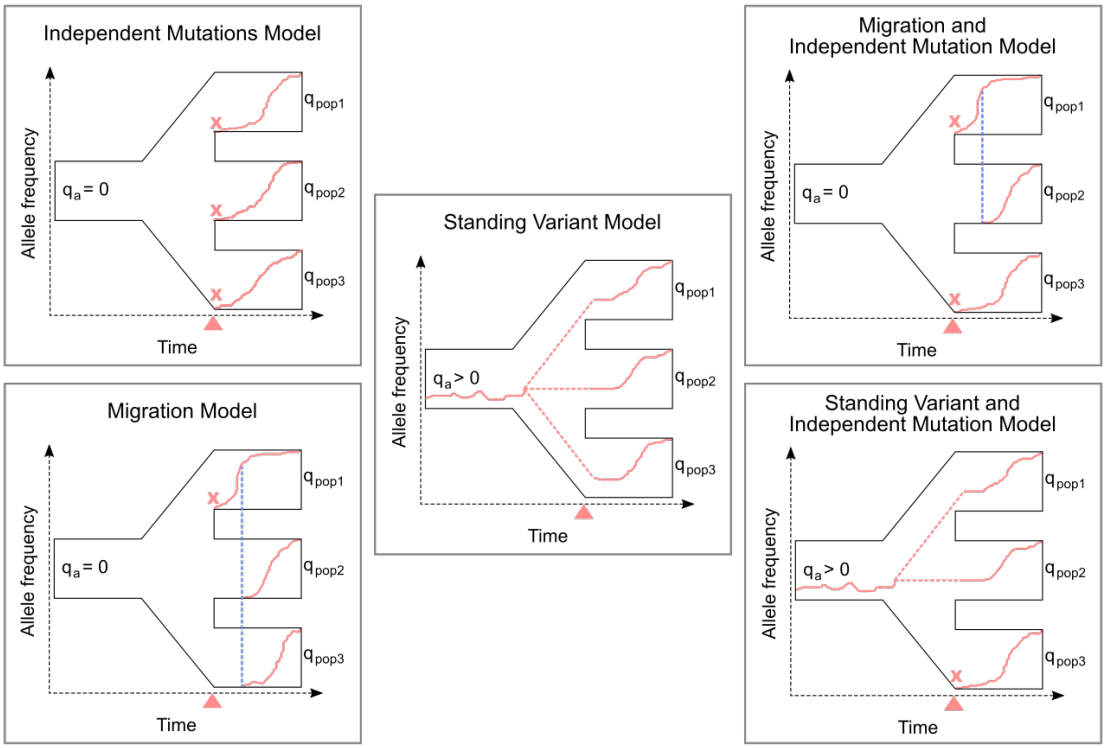


Figure 1.13. Models of convergent herbicide resistance evolution. The red line represents the allele frequency of the resistance allele (q_i), where q_a is the frequency in the ancestral population and q_{pop1}, q_{pop2}, and q_{pop3} are the frequencies in the 3 resistant populations. The red triangle represents the onset of selection from an herbicide. The red x mark represents a mutation event. The blue dashed line represents migration.

The probability of various evolutionary models can be estimated using likelihood maximisation and likelihood-free approaches. The dmc method is a likelihood maximisation approach which uses the coalescent framework to yield insights into the demographic and selective forces which have shaped the observed signatures of selection (Lee and Coop 2017). Approximate Bayesian Computation (ABC) (Pritchard et al. 1999; Beaumont et al. 2002) is a likelihood-free approach which uses summary statistics and extensive simulations coupled with an acceptance-rejection algorithm to assign probabilities to the different models of evolution. As the complexity of evolutionary models increases, explicit likelihood functions become inaccessible (Beaumont 2019). Hence, likelihood-free approaches like ABC offer the only tractable solution. However, as model complexity, sample sizes and sites involved increase, the simulation step becomes more computationally intensive.

Chapter 2

Optimising sampling design and sequencing strategy for the genomic analysis of quantitative traits in natural populations

2. Optimising sampling design and sequencing strategy for the genomic analysis of quantitative traits in natural populations

This chapter was submitted to the scientific journal Molecular Ecology Resources in October 2020. The preprint version is at https://doi.org/10.22541/au.161429335.52330927/v1. The authors are Jefferson F. Paril (PhD candidate), David J. Balding (co-supervisor), and Alexandre Fournier-Level (main supervisor). The aim of this study is to provide recommendations on the optimal sampling strategies to use under different landscape scenarios for genome-wide association and genomic prediction experiments. Specifically, we addressed the following three questions. How many populations do we need to sample to yield optimal QTL detection and genomic prediction accuracies? Under which landscape-specific circumstances should we use Indi-seq or Pool-seq? And which populations to select under different landscape scenarios?

2.1. Abstract

Mapping the genes underlying ecologically-relevant traits in natural populations is fundamental to develop a molecular understanding of species adaptation. Current sequencing technologies enable the characterisation of a species' genetic diversity across the landscape or even over its whole range. The relevant capture of the genetic diversity across the landscape is critical for a successful genetic mapping of traits and there are no clear guidelines on how to achieve an optimal sampling and which sequencing strategy to implement. Here we determine through simulation, the sampling scheme that maximises the power to map the genetic basis of a complex trait in an outbreeding species across an idealised landscape and draw genomic predictions for the trait, comparing individual and pool sequencing strategies. Our results show that QTL detection power and prediction accuracy are higher when more populations over the landscape are sampled and this is more cost-effectively done with pool sequencing than with individual sequencing. Additionally, we recommend sampling populations from areas of high genetic diversity. As progress in sequencing enables the integration of trait-based functional ecology into landscape genomics studies, these findings will guide study designs allowing direct measures of genetic effects in natural populations across the environment.

2.2. Introduction

Understanding how the molecular variation within species supports the evolution of functional traits is a central goal in ecology through the determination of so-called genotype-to-phenotype map. Genomic information for a species can then be leveraged to understand and eventually predict population fitness under a range of eco-evolutionary scenarios. Unfortunately, this genotype-to-phenotype map is only available for a handful of traits, and primarily in model organisms. With the improved accessibility of sequencing technologies, genome-wide association studies (GWAS) and genomic prediction (GP) are becoming straightforward approaches to understand and predict complex traits (Gondro et al. 2013). We thus coined the term GPAS (Genomic Prediction and Association Studies) to denote genome-wide association studies designed to both identify quantitative trait loci (QTL) and predict traits from genomic data. These studies rely on cost-effective, high-throughput sequencing and share the same well-established linear modelling framework. However, how to sample natural populations to train accurate GPAS models that are representative of the genetic diversity of a species is far from obvious. More insights are needed to develop an optimal strategy and move away from *ad hoc* field sampling.

Research in ecological genomics deals with the challenge of characterising the genetic basis of traits across multiple natural populations. This requires collecting sufficient genotype and experimental phenotype data to represent the species' diversity, which in practice is often performed with limited resources. This raises the problem of how to sample across a landscape to capture representative genetic variation while constrained by the total sequencing depth attainable for a given budget. Here, we address the question: how do we allocate a fixed sequencing capacity so that the genetic information captured over the landscape leads to an

optimal GPAS performance?

It is becoming easier to genotype genome-wide markers for large numbers of individuals, either through whole-genome sequencing or complexity reduction approaches such as restriction site-associated DNA sequencing (RADseq) (Baird et al. 2008) in the case of large and complex genomes. Increasing the density and number of markers for GPAS has the potential to increase QTL detection power and prediction accuracy (de Roos et al. 2009; Long & Langley, 1999). Despite the declining costs of sequencing, genotyping every individual of every population across a landscape is usually not feasible, and phenotyping remains resource-consuming. As a cost-effective alternative to sequencing individuals (Indi-seq), sequencing pools of individuals (Pool-seq) (Schlötterer et al. 2014) has gained popularity in ecology (Bastide et al. 2013; Cheng et al. 2012; Nielsen et al. 2018), evolution (Boitard et al. 2012; Fournier-Level et al. 2019; Fracassetti et al. 2015), and breeding (Beissinger et al. 2014; Bélanger et al. 2016) supported by developments in quantitative genetics (Fournier-Level et al. 2017; Guo et al. 2018; Knight, Saccone et al. 2009; Macgregor et al. 2006; Micheletti & Narum, 2018; Jinliang Yang et al. 2015).

Indi-seq generates high-resolution genomic data of a population; while Pool-seq yields low-resolution data in favour of cost reduction. Indi-seq yields individual allele information after variant calling, while Pool-seq generates allele frequency estimates for a group of individuals. Identifying when best to use one over the other is important. Pool-seq was shown to be at least as accurate as Indi-seq in estimating genome-wide allele frequencies (Fracassetti et al. 2015; Gautier et al. 2013; Rellstab et al. 2013; Zhu et al. 2012), but it is also prone to biases in genome representation (i.e. unequal representation of different genomic regions) when sample size and depth of coverage are low (i.e. <40 individuals per pool and <50X depth) (Cutler & Jensen, 2010; Schlötterer et al. 2014). Pool-seq also loses haplotype and linkage disequilibrium (LD) information (Fariello et al. 2017) which limits the number of quantitative and population genetics models that can be used and requires the design of novel analysis methods (Cutler & Jensen, 2010). Pool-seq is more cost-effective than Indi-seq since it requires less sequencing effort to generate the same genome-wide allele frequency data (Futschik & Schlötterer, 2010; Gautier et al. 2013), particularly for non-model organisms where individuals cannot be maintained indefinitely and used in multiple experiments. Additionally, Pool-seq can include more individuals, grouped into one or a few pools and sequenced at a high depth.

There is no research on the optimal sampling strategy across a landscape for GPAS, comparing Indi-seq and Pool-seq. Quantitative genetics studies have established that large sample sizes and diverse mapping and training populations improve the power of GWAS (Visscher et al. 2012; Visscher et al. 2017; Gurdasani et al. 2019; Wojcik et al. 2019) and the accuracy of GP (Asoro et al. 2011; Bernal-Vasquez et al. 2014; Bustos-Korts et al. 2016; Rincent et al. 2017; Akdemir & Isidro-Sánchez, 2019; Edwards et al. 2019). However, the spatial component of sampling across the landscape is lacking in these studies. In landscape genomics, in addition to the recommendation to sample as many populations as possible (Santos & Gaiotto, 2020), the spatial extent of the species' dispersal should be accounted for (Riginos et al. 2016) and stratified sampling is commonly performed to represent the different environmental clines (Hoel, 1943; Li et al. 2017; Williams & Brown, 2019; Pais et al. 2020). Landscape genomics associates genotype with the environment (Li et al. 2017); while GPAS associates genotype with phenotype (Gondro et al. 2013). It follows that in GPAS, sampling every environmental cline may not be needed as long as the genetic diversity is well-represented in the samples. Additionally, landscapes and populations with low differentiation or structure allow for less samples to represent the total diversity sufficiently. Finally, it is not known whether the cost-effectiveness of Pool-seq, being the preferred genotyping approach in landscape genomics over Indi-seq (Santos & Gaiotto, 2020), retains the same cost-effectiveness in landscape-wide GPAS.

Field researchers often adapt techniques initially developed for model organisms or crops in highly controlled environments; however, with natural populations having evolved in natural environments, devising the optimal sampling strategy becomes non-trivial. Individuals and pools can be sampled from one to a few populations or from a large number of populations. Identifying which populations warrant higher resolution (i.e. individual-resolution genotype data with Indi-seq instead of group-resolution genotype data with Pool-seq), requires some prior knowledge of

the spatial distribution of genetic variability across the landscape. To address this question, we simulated landscapes under different trait architectures and population genetics scenarios with the aim of providing recommendations on the optimal sampling strategies. Specifically, we aim to answer the following three questions. How many populations do we need to sample to yield optimal GPAS performance? Under which landscape-specific circumstances should we use Indi-seq or Pool-seq? And which populations to select under different landscape scenarios?

2.3. Materials and methods

2.3.1. Workflow overview

We first simulated landscapes inhabited by multiple populations of an outbreeding species. For every simulated landscape, migration between adjacent pairs of populations is uniform, and there is one trait of interest which is positively correlated with fitness. This trait is controlled by QTL with purely additive effects. The distribution across the landscape of the corresponding favourable alleles is affected by their population of origin, selection intensity and migration rate. We then simulated a stratified sampling strategy and performed Indi-seq and Pool-seq. Next, GPAS models were built to identify the QTL and predict the phenotypes of individuals and pools. Finally, we assessed how the number of populations sampled, the genotyping strategy (Indi-seq and Pool-seq), selection intensity, migration rate, and the distribution of the favourable alleles across the landscape in relation to the sampled populations affected QTL detection and genomic prediction accuracies.

2.3.2. Landscape simulations

The landscapes were simulated using quantiNemo2 (Neuenschwander et al. 2018), the variables are listed in Table 2.1, and the fixed parameters are listed in Appendix Table S2.1. We simulated the simple and common mating system consisting of a hermaphrodite species capable of both self- and cross-fertilisation (1/n and 1-1/n probabilities, respectively; n is the population size) with discrete non-overlapping generations. This sufficiently represents allogamous and mixed mating systems, including that of plants with considerable outcrossing, and many animals species. Variation for a quantitative phenotype over a landscape was simulated as a function of migration rate, number of QTL controlling the trait, causal allele diffusion gradient (the distribution of the favourable alleles across the landscape as it migrates from the populations of origin), and selection intensity with 3 levels for each of these variables (Table 2.1).

Table 2.1. List of variables used in landscape simulations.

Parameter	Levels
Number of populations per landscape	100
Number of hermaphroditic individuals per population	1,000
Number of loci per individual	10,000
Number of generations	200
Number of loci controlling the trait	10, 50, and 100
Migration rate per population per generation	0.0001, 0.001, and 0.01
Causal allele diffusion gradient	Uniform, unidirectional, and bidirectional
Selection intensity	0.50, 0.90, and 0.95
Number of replicates per landscape	5

Each landscape consisted of 100 populations arrayed in a uniform square lattice without barriers. Migration was modelled using a 2-dimensional stepping-stone model with bidirectional gene flow with a uniform rate into the 8 adjacent populations and absorbing boundaries.

The quantitative trait was determined by additive QTL with effects sampled from a χ^2 distribution with 1 degree of freedom to generate a cumulative heritability of 0.5. At the initial step of the simulation, all the causal alleles had a frequency (q_0) of 0.01 in the populations of origin and 0 elsewhere. Under the uniform allele diffusion gradient, all populations had q_0=0.01. Under the unidirectional gradient, one boundary row had q_0=0.01 for all 10 populations in that row and 0 elsewhere. Under the bidirectional gradient, two opposite boundary rows had q_0=0.01 in each of the populations and 0 elsewhere. For clarity, these causal allele diffusion gradients are illustrated in Figure 2.1 Panel A.

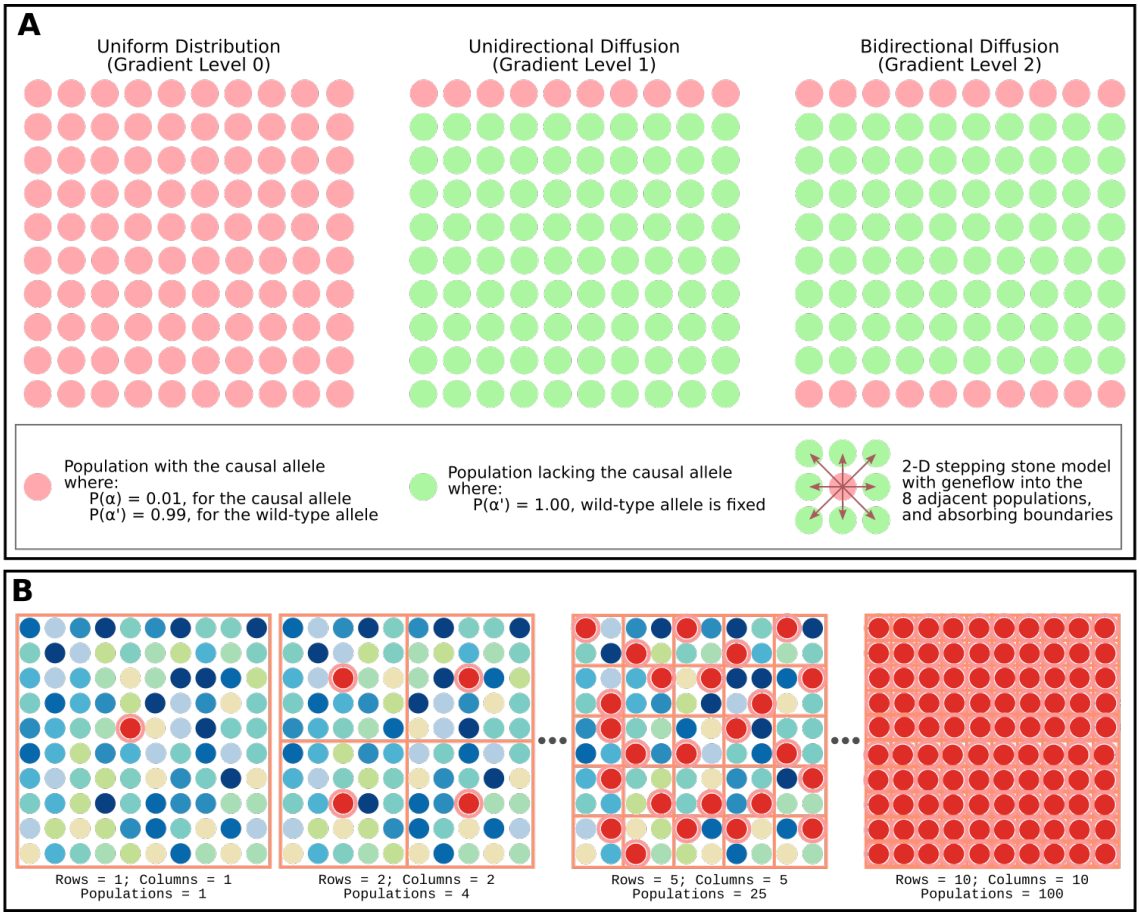


Figure 2.1. Panel A: Causal allele diffusion gradients across the simulated landscape. Panel B: Systematic sampling strategy across the simulated landscape (red circles represent sampled populations, while the other colours represent variability in the different populations). Note: For one population sampling, all populations across the landscape were sampled independently.

Selection was simulated using a generalised logistic model (Richards 1959) as

$$1/w = 1 + e^{y_c - y},$$

where w is fitness, y is the quantitative trait ($y \in R$), $y_c = y_{min} + s(y_{max}-y_{min})$, with y_{max} and y_{min} the maximum and minimum possible trait values, and s is the selection intensity. This selection intensity variable is the minimum relative phenotype value (ranging from 0 to 1) for which survival rate is at least 50%, assuming that the phenotype is positively correlated with fitness. It follows that a selection intensity of 0.50 or 0.95 means that individuals at the top 50th or 95th percentile, respectively, contribute more to the next generation than those below. We used high selection intensities, i.e. $s \in \{0.5, 0.90, 0.95\}$, to simulate pressing anthropogenic selection pressures, e.g. herbicide or insecticide pressure.

The 10,000 biallelic loci were randomly distributed across a large genome with 7 chromosomes and a total length of 2×10^9 base-pairs and 750 centimorgans. Two hundred generations were simulated to allow the causal alleles to migrate across the landscape and generate the allele frequency distributions specific to each of the three causal allele diffusion gradients. Phenotype values in the final generation were scaled in the 0 to 1 interval and used for the GPAS experiments.

2.3.3. Genome-wide association and trait prediction based on polygenic scores

GPAS was performed on all the populations. Indi-seq and Pool-seq data were assumed to have been generated without genotyping error. We used established tools for Indi-seq data, and developed a suite of tools for Pool-seq data. For Indi-seq data, 384 individuals were sampled per population, simulating four 96-well sample plates or a single 384-well plate commonly used in high-throughput molecular biology workflows. For Pool-seq data, 5 pools per population were sampled, where each pool consisted of 100 individuals. This corresponds to a high power design that was shown to be optimal to capture QTL association (Fournier-Level et al. 2017). GPAS models were trained within each population sampled and cross-validated on all other populations to assess prediction accuracies. For each population, 384 individuals and 5 pools were used to train and validate the models. This corresponds to an external cross-validation with equally-sized mutually exclusive training and validation sets.

Allele effects were estimated using 6 Indi-seq-based GPAS (Indi-GPAS) and 3 Pool-seq-based GPAS (Pool-GPAS) models. The 6 Indi-GPAS models consisted of: efficient mixed-model association expedited model (EMMAX; Kang et al. 2010), genome-wide complex trait analysis (GCTA; Jiang et al. 2019), and genome-wide efficient mixed-model analysis (GEMMA; Zhou & Stephens, 2012), in combination with 2 types of genetic relationship matrices: GCTA-derived sparse genetic relationship matrix (GRM; off diagonals <0.05 were set to zero; Zaitlen et al. 2013) and GEMMA-derived standardised relatedness matrix (STD; Zhou & Stephens, 2012) (i.e. EMMAX(GRM), EMMAX(STD), GCTA(GRM), GCTA(STD), GEMMA(GRM), and GEMMA(STD)). The 3 Pool-GPAS models consisted of the genome-wide estimation of additive effects based on trait quantile distribution from Pool-seq data (GWAlpha; Fournier-Level et al. 2017), and linear mixed models (LMM) with random pairwise genetic covariance matrix determined by F_{ST} derived using either Hivert's (Hivert et al. 2018) or Weir and Cockerham's method (Weir & Cockerham, 1984) (i.e. GWAlpha, LMM (Fst Hivert), and LMM (Fst Weir & Cockerham)). The variance components of the LMM were estimated using restricted maximum likelihood.

Phenotype predictions were derived from polygenic scores, i.e. the sum of the products of estimated allele effects and allele dosages for Indi-GPAS or allele frequencies for Pool-GPAS. For the Indi-GPAS models and GWAlpha, this involved a two-step approach. For each training set, the polygenic scores of the training set (s_{train}) were calculated as:

$$s_{\text{train}} = \mathrm{X}_{\text{train}}\,\beta, \qquad \text{eq. 2}$$

where X_{train} is the dosage or frequency of alleles in the training set, and β is the vector of estimated SNP effects. The polygenic scores and actual phenotype values have a linear relationship (Figure S1; mean adjusted R^2=0.97±0.0031 with 1,000 individuals per population for Indi-seq and mean adjusted R^2=0.99±0.0006 with 5 pools per population for Pool-seq) as expected under the additive model used to simulate the phenotypes. These polygenic scores were regressed against the actual phenotype values of the training set $\left(y_{\text{train}}\right)$,

$$y_{\text{train}} = \alpha_0 + \alpha_1 s_{\text{train}}, \qquad \text{eq. 3}$$

where α_0 is the intercept, and $\alpha_1$is the slope. The polygenic scores of the validation set, $s_{\text{valid}} = \mathrm{X}_{\text{valid}}\,\beta$, were transformed into the predicted phenotype values ($y_{\text{predicted}}$) using

$$y_{\text{predicted}} = \alpha_0 + \alpha_1 s_{\text{valid}}. \qquad \text{eq. 4}$$

For the Pool-GPAS linear mixed models, the predicted phenotypes ($y_{\text{predicted}}$) were calculated as:

$$y_{\text{predicted}} = X_{\text{valid}}\,\beta, \qquad \text{eq. 5}$$

where X_{valid} is the matrix of allele frequencies of the validation set, and β is the estimated allelic effects from the GPAS model built using the training set. The trained models were validated on all populations in the landscape.

GPAS performance was measured using three GWAS metrics, and one phenotype prediction metric. The GWAS metrics were:

1. area under the receiver operating curve (AUC) (Fawcett, 2006),
2. true positive rate (TPR) which was defined as the fraction of causal QTL with a significantly associated SNP within 1 kbp, and
3. false positive rate (FPR) which was defined as the fraction of the significantly associated SNPs with no causal QTL within 1 kbp, unless it tags a true QTL through a chain of associated SNPs each less than 1kb apart; multiple associated SNPs within 1kbp were counted as one.

The family-wise type I error rate was set at α=0.05. The metric for phenotype prediction is the root mean square error (RMSE) between actual and predicted phenotype values:

$$\text{RMSE} = \sqrt{\sum \frac{(y - \hat{y})^2}{n}}, \qquad \text{eq. 6}$$

where y is the actual phenotypes, $\hat{y}$ is the predicted phenotypes and n is the number of observations.

2.3.4. Sampling strategy optimisation

A total of 405 landscapes were simulated using all combinations of the 4 landscape parameters allowed to vary with 3 levels each and 5 replicates (Table 2.1). For each landscape, Indi-GPAS and Pool-GPAS experiments were performed for each population independently. This constitutes the intra-population dataset. The landscape was divided into equally sized rectangular regions, and the approximately central population was selected from each region to simulate a stratified sampling strategy. This is illustrated in Figure 2.1 Panel B. This constitutes the inter-population dataset. AUC and RMSE were averaged across the populations sampled. TPR and FPR were calculated using the cumulative number of true and false positive candidate loci across the populations sampled. AUC was used to measure the accuracy of QTL detection per population, while TPR and FPR were used to measure QTL detection accuracy of multiple populations.

The single best performing modelling framework was identified for each genotyping scheme (Indi-GPAS and Pool-GPAS) based on AUC and RMSE for independent populations tests using Tukey's honest significant difference (HSD mean comparison) at α=0.05.

How many populations do we need to sample to yield optimal GPAS performance?

To determine how many populations to sample to yield optimal GPAS performance, we used the inter-population dataset. We assessed the suitability of the four metrics (i.e. mean AUC, mean RMSE, TPR and FPR) to address this question by visualising their relationships with the number of populations sampled. Additionally, we compared the expected performance of Indi-GPAS and Pool-GPAS under the same sequencing capacity constraint. The Indi-GPAS experiments we simulated included 384 individuals per population, while Pool-GPAS included only 5 pools per population. Assuming a 5X sequencing depth per individual for Indi-seq (Brouard et al. 2017) and the recommended 50X depth per pool for Pool-seq (Schlötterer et al. 2014), these equate to a sequencing depth of 1,920X per base per population for Indi-seq and only 250X for Pool-seq. This means that for the sequencing capacity required to characterise one population through Indi-seq, approximately 7 populations ($\lfloor 1920/250 \rfloor$) could be characterised through Pool-seq.

Under which landscape-specific circumstances should we use Indi-seq or Pool-seq?

The second main question we addressed was which landscape-specific circumstances warrant Indi-GPAS or Pool-GPAS? Specifically, if we were to perform GPAS on one population, which

sequencing strategy (Indi-seq or Pool-seq) is better, and how does the optimal choice vary with the polygenicity of the trait, selection intensity, and gene flow? The intra-population dataset was used together with AUC and RMSE as the GPAS performance metrics.

Which populations to select under different landscape scenarios?

To determine which populations to select to best capture the genetic basis of a trait and yield accurate trait predictions, we used AUC and RMSE as the GPAS metrics and the intra-population dataset to test the effect of the three causal allele diffusion gradients. The populations were classified into 10 groups, where each group represents a row perpendicular to the causal allele diffusion gradient. The top row refers to populations 1 to 10, the second row to populations 11 to 20, and so on. The general trends and landscape parameter-specific trends in GPAS performance across the landscape were visualised using violin plots and means compared using Tukey's HSD (α=0.05). Linear mixed models fitted linear and quadratic relationships (using second degree polynomial fit) between GPAS performance and the row groups. The row group was treated as a numeric variable, and nested within each level of the parameters: number of QTL, selection intensity, migration rate, and GPAS model.

2.3.5. Implementation

The landscapes were simulated using quantiNemo2 (Neuenschwander et al. 2018). The genome and QTL information were simulated in R (R Core Team 2018). The quantiNemo outputs were parsed using R and Julia (Zappa Nardelli et al. 2018). GEMMA (Zhou and Stephens 2012), EMMAX (Kang et al. 2010), GCTA (Jiang et al. 2019), and Plink (Purcell et al. 2007) were used for Indi-GPAS. GWAlpha.jl was used for Pool-GPAS. The R package violinplotter was used to generate violin plots with HSD mean comparison grouping. The GNU shell (Free Software Foundation 2016), Spartan (Lafayette and Wiebelt 2017), Slurm (Yoo et al. 2003), and GNU parallel (Tange 2011) were used extensively. The workflow is available in the gitlab repository: https://github.com/jeffersonfparil/GPAS-landscape-simulation.git.

2.4. Results

2.4.1. GPAS model representatives and the relationships of GPAS performance with landscape and sampling parameters

GEMMA (STD) and GWAlpha showed the best GPAS performances, with >79% AUC and <5.9% RMSE (Appendix Table S2.2). Therefore, these two frameworks were selected as the representatives of Indi-GPAS and Pool-GPAS models, respectively. Overall, Indi-GPAS performed better than Pool-GPAS.

Factors increasing statistical power to identify causal loci through GPAS included a lower number of QTL controlling the trait, more intense selection, higher migration among populations, and more populations sampled (Appendix Figure S2.2). Accuracy in phenotype predictions improved as the number of QTL controlling the trait increases, as selection intensity decreases, and as migration rate increases (Appendix Figure S2.3). Accuracy was unaffected by increasing the number of populations sampled since each model was trained independently for each population. In addition, power and accuracy are high when QTL diffuses across the landscape uniformly.

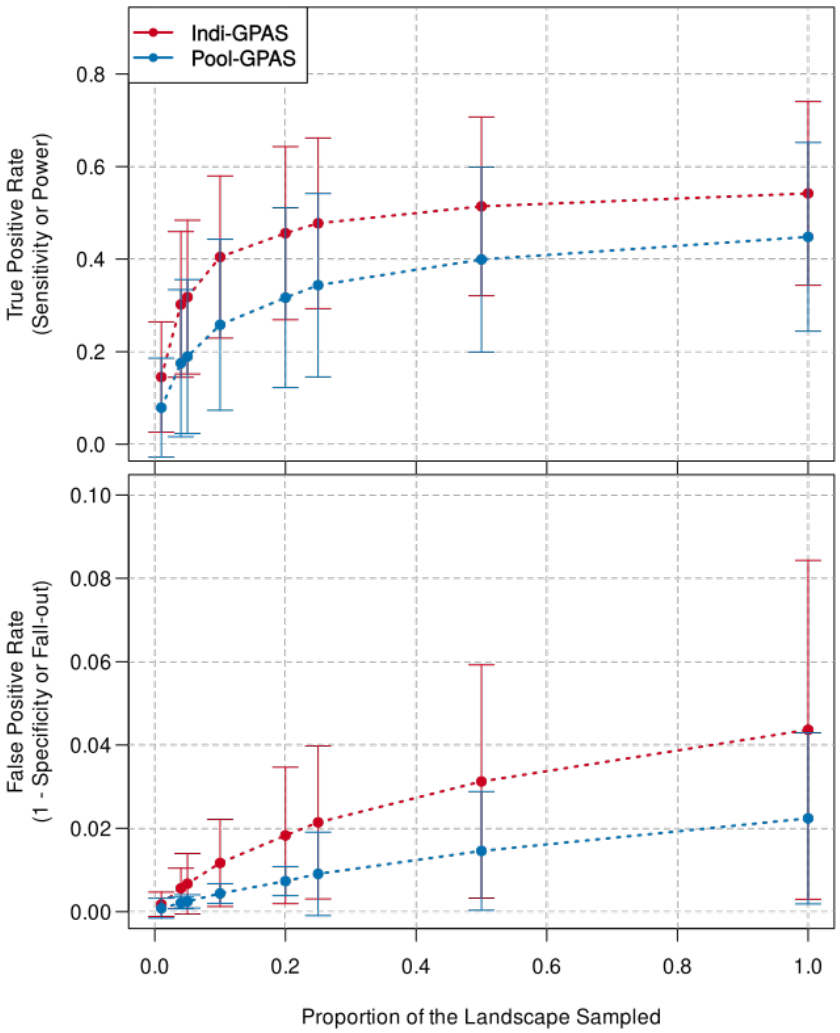


Figure 2.2. Relationships between the proportion of the landscape sampled and GPAS performance. Dots represent means and whiskers indicate ±1 standard deviation from the mean.

2.4.2. How many populations to sample and when to use Indi-seq or Pool-seq?

TPR and FPR increase logarithmically as the number of populations sampled increases (Figure 2.2), so there is no optimum based on these metrics.

Indi-GPAS achieves greater power than Pool-GPAS at the cost of a higher false positive rate (Figure 2.2). However, Pool-GPAS can outperform Indi-GPAS under the assumptions detailed in the materials and methods section, where for every population characterised with Indi-seq, approximately 7 populations can be characterised with Pool-seq. Under this 1:7 ratio, Indi-GPAS on 10 populations yield an average TPR of 0.388 and FPR of 0.0150; for the same sequencing capacity Pool-GPAS can be performed on 70 populations, yielding an average TPR of 0.418 and FPR of 0.0115. We explored a range of ratios deviating from this 1:7 ratio. This is because the 5X depth requirement for variant calling in Indi-seq and 50X depth for allele frequency estimation in Pool-seq depend on the species of interest and the resources available. Lower ratios, e.g. 1:8 to 1:10, mean even more populations can be characterised with Pool-seq for every population characterised with Indi-seq. Using our simulated data to explore various ratios, we find that there exists a range where Pool-GPAS can outperform Indi-GPAS, i.e. TPR is higher and FPR is lower for Pool-GPAS than Indi-GPAS (Figure 2.3). This shows that characterising more of the landscape at low resolution can be better than characterising a small portion of the landscape at high resolution.

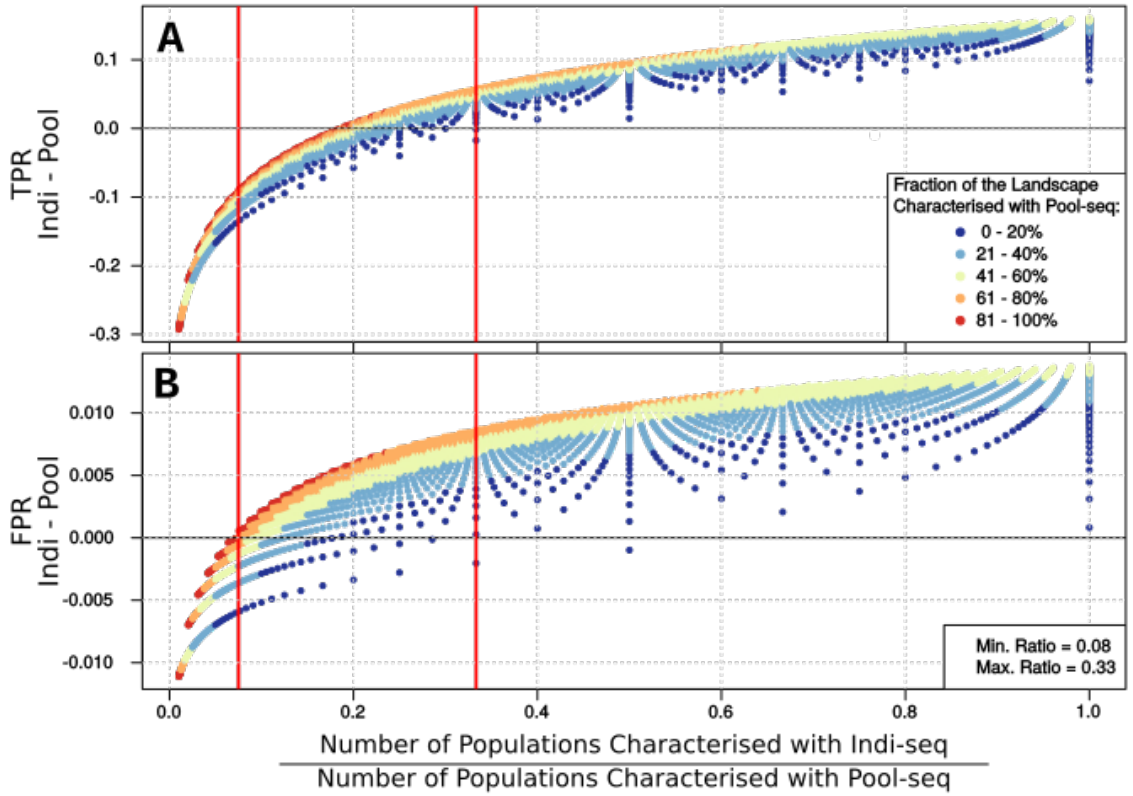


Figure 2.3. Pool-GPAS can outperform Indi-GPAS under the same sequencing capacity. **Panel A**: Difference in true positive rate (TPR) between Indi-GPAS and Pool-GPAS models as the ratio between the number of populations characterised through Indi-seq and Pool-seq increases. **Panel B**: Difference in false positive rate (FPR) between Indi-GPAS and Pool-GPAS models as the ratio between the number of populations characterised through Indi-seq and Pool-seq increases. **Vertical red lines**: The range of ratios between the two vertical red lines correspond to cases where Pool-GPAS can outperform Indi-GPAS in our study, i.e. Pool-GPAS has higher TPR (points below the dashed line in panel A) and lower FPR than Indi-GPAS (points above the dashed line in panel B). Moving along the x-axis involves modifying the sequencing depths for Indi-seq or Pool-seq and not the number of individuals or pools sampled per population, since 384 individuals per population and 5 pools per population were kept constant in the simulations.

If we were to perform GPAS on one population, Indi-GPAS is better than Pool-GPAS. However in cases where selection intensity is high (i.e. 0.90 to 0.95) or migration rate is high (i.e. 0.01) Pool-GPAS performance is not significantly different from Indi-GPAS in terms of prediction accuracy (Figure 2.4).

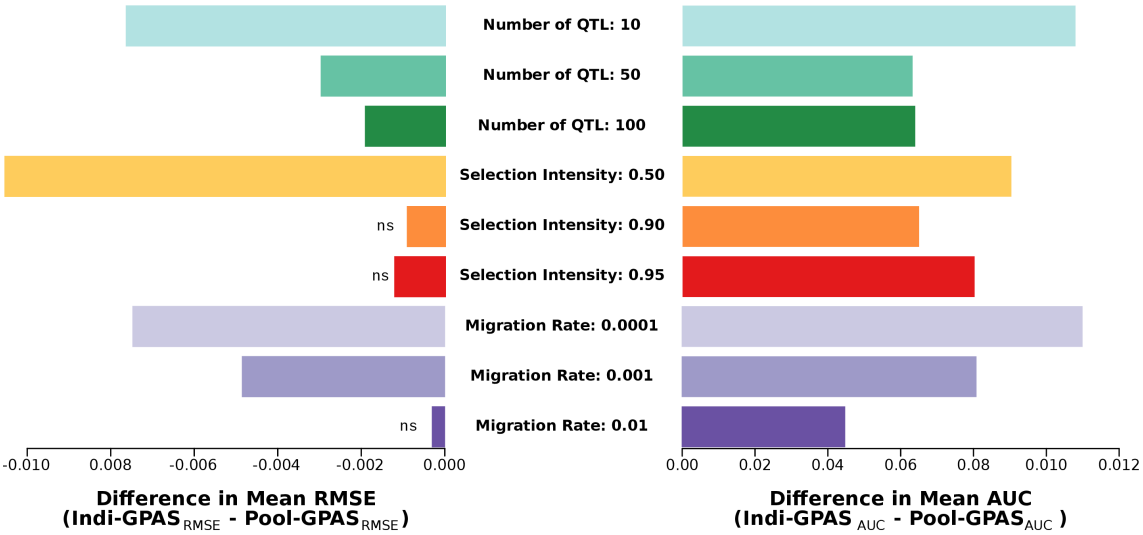


Figure 2.4. Differences in QTL detection and polygenic score prediction accuracies between Indi-GPAS and Pool-GPAS per population. AUC = Area under the ROC curve. RMSE = root mean square error of the polygenic score prediction. The term "*ns*" beside the bars indicates non-significant differences based on HSD at $p<0.05$.

2.4.3. Which populations to select under different landscape scenarios?

GPAS performance is maximised in populations with high genetic variability which at the landscape level, means sampled close to the place of origin of the causal allele (Figure 2.5). This area of high genetic variability is characterised by intermediate causal allele frequencies which translate into populations with high phenotypic variability. In the absence of a causal allele diffusion gradient (i.e. uniform causal allele distribution), no row seems to be optimal for sampling, except for some slightly better performance from populations in the middle rows. Under unidirectional gradient (i.e. causal alleles originated from the top row and diffused downwards hence a single diffusion front) and in terms of QTL detection accuracy, sampling the populations from the top row is optimal; however, in terms of prediction accuracy, the populations in the middle rows appear to be better. Under bidirectional gradient (i.e. causal alleles originated from the top and bottom rows hence two diffusion fronts) both QTL detection and prediction accuracies are optimal in the populations from the top and bottom rows. These trends across the landscape coincide with the trends in the mean number of polymorphic QTL per population and causal allele frequencies.

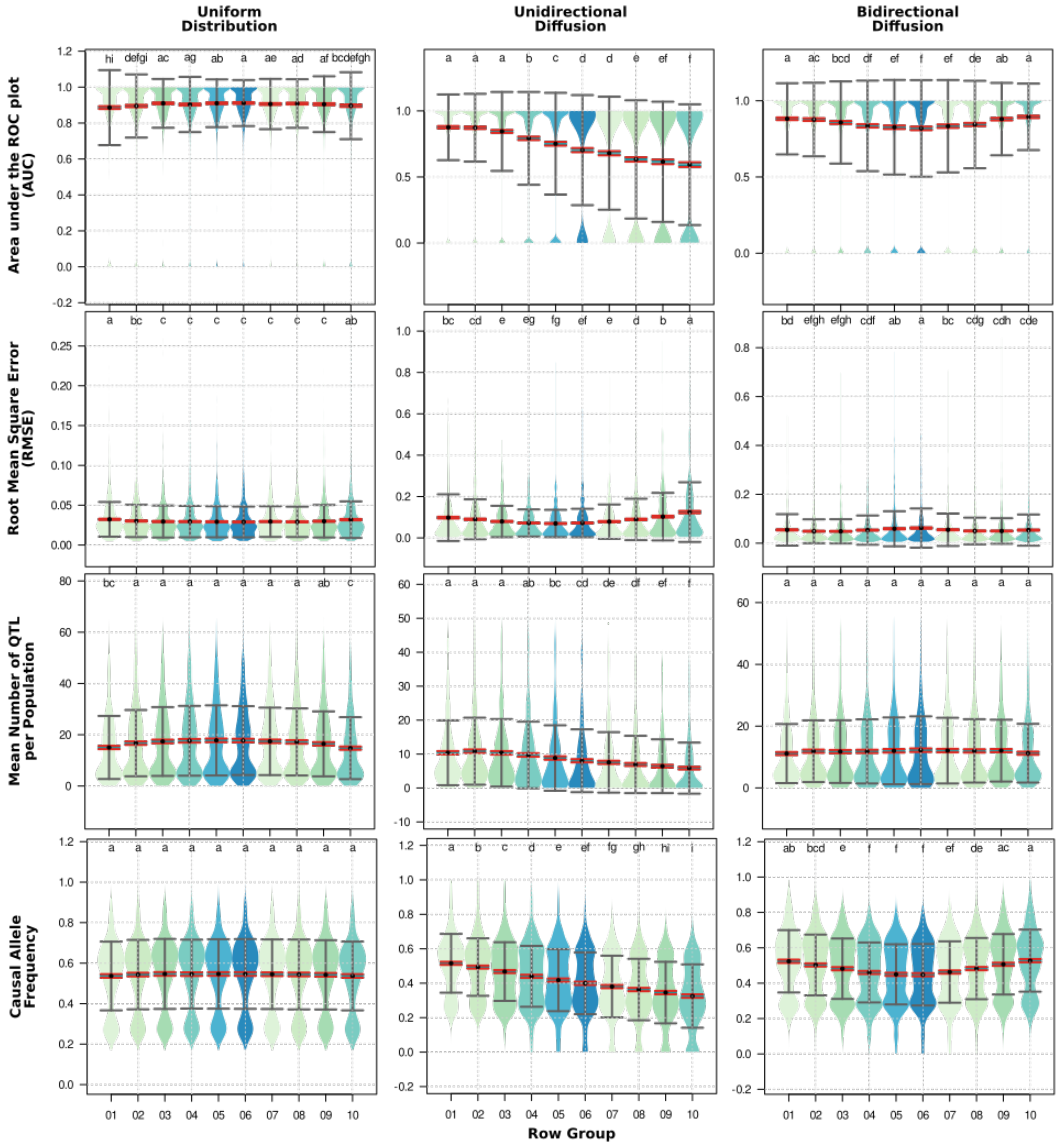


Figure 2.5. Violin plots of the area under the receiver operating curve (AUC; measure of QTL detection accuracy), root mean square error (RMSE; measure of polygenic score prediction accuracy), mean number of polymorphic causal loci per population, and causal allele frequencies across the landscape divided into 10 rows under uniform, unidirectional, and bidirectional causal allele diffusion. Each row is perpendicular to the causal allele diffusion gradient. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

In the presence of causal allele diffusion gradients, the relationships between QTL detection or prediction accuracies and the sampling location (defined as rows perpendicular to the diffusion gradient) appear to be quadratic, except for QTL detection accuracy under unidirectional causal allele diffusion, for which the relationship is linear (Figure 2.5). In terms of GWAS accuracy as measured by AUC, sampling near the diffusion fronts becomes less important (i.e. slope under unidirectional gradient and curvature under bidirectional gradient are reduced) as the number of QTL increases, as selection intensity decreases, and as migration rate increases (Figure 2.6 columns 1-3). In addition, sampling near the diffusion fronts is more important for Pool-GPAS than Indi-GPAS (Figure 2.6 column 4), in other words, power diminishes quicker for Pool-GPAS than Indi-GPAS as we move away from areas of high diversity.

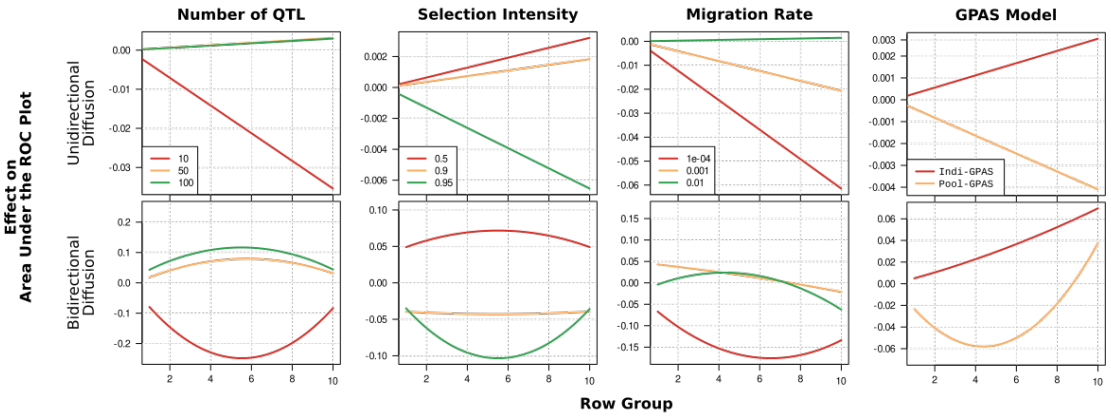


Figure 2.6. Effects of sampling across the landscape on QTL detection accuracy under varying number of simulated QTL, selection intensities, migration rates, and GPAS models for unidirectional (top graphs) and bidirectional (bottom graphs) causal allele diffusion gradients. QTL detection accuracy was measured using the area under the curve which describes the relationship between power and false positive rate, where high values mean high QTL detection accuracies.

In terms of prediction accuracy as measured by RMSE, the degree to which the middle rows (i.e. areas of high genetic and phenotypic variability) are the optimal sampling locations under unidirectional diffusion decreases (i.e. curvature becomes less severe) as the number of QTL increases, as selection intensity decreases, and as migration rate increases (Figure 2.7 top graphs). Also, sampling from the middle rows under a single diffusion front is slightly more important for Indi-GPAS than Pool-GPAS. On the other hand, the degree to which the top and bottom rows are optimal under bidirectional diffusion decreases (i.e. curvature becomes less severe) as the number of QTL, selection intensity, and migration rate increase (Figure 2.7 bottom graphs). Also, sampling from the top and bottom rows under two diffusion fronts is more important for Pool-GPAS than Indi-GPAS, in other words, similar to that of power, accuracy diminishes quicker for Pool-GPAS than Indi-GPAS as we move away from areas of high diversity.

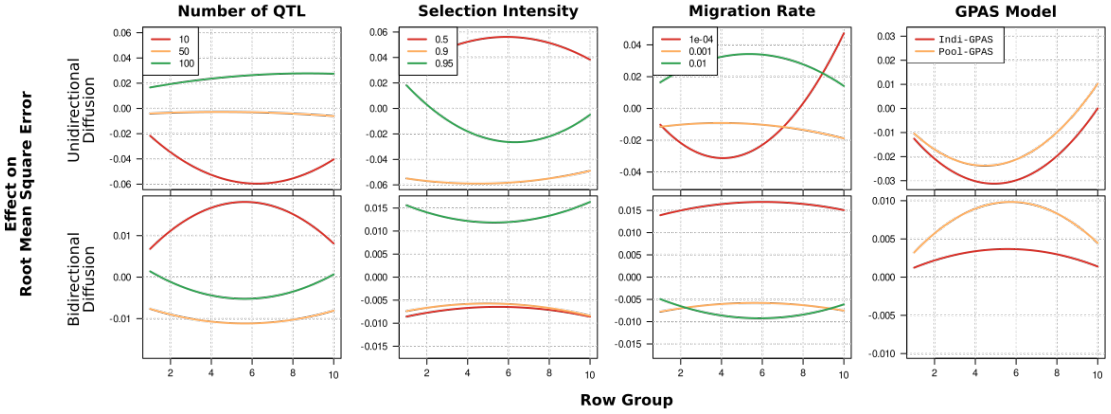


Figure 2.7. Effects of sampling across the landscape on prediction accuracy under varying number of simulated QTL, selection intensities, migration rates, and GPAS models for unidirectional (top graphs) and bidirectional (bottom graphs) causal allele diffusion gradients. Polygenic score prediction accuracy was measured using root mean square error where low values mean high accuracies.

The trends in GPAS performance across the landscape correlate with the trends in genetic variability (expressed in terms of causal allele frequency, i.e. frequencies closer to 0.5 indicates higher diversity; Appendix Figures S2.4 to S2.9). Opposite trends are observed between causal allele diffusion gradients for RMSE as selection intensity increases, i.e. in the middle of the landscape, RMSE is minimised for unidirectional causal diffusion but maximised for bidirectional causal allele diffusion (Appendix Figure S2.5 and S2.8).

2.5. Discussion

GPAS has the potential to extend the scope of genomic studies in ecology and evolution beyond environment association and niche modelling (Dormann et al. 2012; Exposito-Alonso et al. 2018; Fournier-Level et al. 2011). The predicted environmental range of individual genotypes determined through genome-environment associations (Manel et al. 2018) can be complemented by the phenotype predictions of GPAS (Cotto et al. 2017). This can be transformational for the way we monitor invasive species or assess the adaptive potential of endangered ones. We provide recommendations for optimising the sampling strategy to maximise the power to detect QTL and the accuracy of quantitative phenotype prediction in natural populations of any outbreeding species. We stress the importance of capturing sufficient representation of the genetic variability present over the landscape by sampling populations from areas of high genetic diversity. On a per population basis or if only one population were to be sampled, we recommend using Indi-seq over Pool-seq. We have not considered phenotyping costs here, but if it is high, then it would increase the attractiveness of Indi-seq to maximise information per unit cost. However, similar to a study on the estimation of population differentiation (Goudet & Büchi, 2006) and a meta analysis of several landscape genomics studies (Santos & Gaiotto, 2020), w demonstrated the value of shallow but extensive genotyping with Pool-seq in maximising the number of populations that can be analysed without compromising power. This is especially true if the aim is to predict phenotypes of some future populations for the rapid and timely monitoring of invasive and threatened species.

2.5.1. How many populations to sample?

Our results emphasise the need to sample as many populations as possible from regions of high genetic diversity. The power to detect QTL is maximised if all the populations in the landscape were included in the study. This is possible for endangered species with a small number of populations in the wild. However this is not feasible for species with a healthier number of populations. The best populations to sample are located in areas of high genetic diversity which manifests as areas with high trait variability where the causal alleles are at intermediate frequencies.

Our results show a diminishing return in terms of GPAS power when increasing the number of populations sampled. This is consistent with a study by Selmoni et al. (2020) on sampling strategy optimisation for landscape genomics which found that sampling an intermediate number of sites can perform as well as maximising the number of sites sampled. This study only considered Indi-seq and tested different sample sizes per population, and our approach is comparable because using Indi-seq equates to a high-resolution characterisation of the landscape and Pool-seq to a low-resolution one. Our analysis extends this result further because it is independent of the number of individuals sampled per population. The cost-effectiveness of Pool-seq allows for more populations to be sampled and included in the study than Indi-seq.

2.5.2. When to use Indi-seq or Pool-seq?

The number of individuals per population selected for our Indi-GPAS simulations (384 individuals) exceeds the sample size of most ecological studies (e.g. 100-200 individual samples per population in birds (Hansson et al. 2018; Perrier et al. 2018), <100 samples per population in trees (Cappa et al. 2013; Holliday et al. 2010), ~100 samples per population in mammals (Johnston et al. 2011; Pallares et al. 2014), and <20 samples per population in fish (Willing et al. 2010)). Thus for most experiments, the power of Indi-GPAS is expected to be lower than in our simulations. On the contrary, the power of Pool-GPAS is expected to remain the same since five pools per population was found to be optimal (Fournier-Level et al. 2017) and each pool can include a non-limiting number of individuals. The sequencing capacity required for Indi-seq is always higher than for Pool-seq, and more populations can be

characterised with Pool-seq than with Indi-seq under the same budgetary constraints (Schlötterer et al. 2014). Therefore, the range of the number of populations sampled where Pool-GPAS outperforms Indi-GPAS is likely to be even broader than reported here, as long as the genomic characterisation approach yields accurate genomic data.

Sufficient depth of coverage is required to correct for sequencing errors inherent to current high-throughput sequencing technologies. Indi-seq can provide high-resolution genomic information for a population, but comes at a high cost. A given genomic region needs to be sequenced at least 5 times for each individual to correct for sequencing errors and yield accurate basecalling information (Brouard et al. 2017). Additionally, many individuals are required to accurately represent a population. This is only resource-effective when the individuals are part of an association panel and the genomic information can be leveraged across several research projects (Robin et al. 2019). On the other hand, Pool-seq generates low-resolution genomic information on a population that is cost-effective while maintaining high power. To maximise the accuracy of allele frequency estimation (i.e. to correct for sequencing errors and sufficiently represent the pool), Pool-seq guidelines and recommendations have been proposed (Schlötterer et al. 2014; Fracassetti et al. 2015; Anand et al. 2016). Pool sizes of at least 25 (Gautier et al. 2013; Fracassetti et al. 2015) to 40 individuals (Schlötterer et al. 2014), and depths greater than 50X (Zhu et al. 2012) to 65X (Gautier et al. 2013) have been recommended. Hundreds of individuals can be pooled to yield accurate allele frequency data (Schlötterer et al. 2014). This means that in an outbreeding species, a few pools consisting of hundreds of individuals each can represent a population better than a few individuals.

Pool-seq is more widely used than Indi-seq in ecological and evolutionary studies because the focus is generally on populations rather than individuals, and because of its cost-effectiveness (Futschik & Schlötterer, 2010). In contrast, Cutler and Jensen (2010) concluded that Indi-seq should be preferred over Pool-seq for many applications due to the loss of haplotype and LD information. They focused on applications in human and model organisms, whereas Pool-seq has its highest impact for high-throughput data acquisition in non-model species of critical ecological and economical importance.

2.5.3. Which populations to select under different landscape scenarios?

We have shown that sampling from genetically diverse populations maximises GPAS performance. Capturing greater genetic diversity was shown to increase the power to detect causal loci (Alqudah et al. 2020; Rosenberg et al. 2010; Wojcik et al. 2019). Similarly, populations which represent the overall diversity found in the landscape or are similar to the validation populations, improve prediction accuracies (Akdemir & Isidro-Sánchez, 2019; Asoro et al. 2011; Edwards et al. 2019). Populations with high genetic diversity were found along the diffusion fronts, i.e. the areas where the causal alleles migrate from their site of origin into the neighbouring populations. The rate at which GPAS performance decreases as we sample farther away from the diffusion fronts correlates with the decrease in genetic diversity at the causal loci. In the absence of prior genomic information, the areas of high genetic diversity coincide with regions of high phenotypic diversity. Gaining prior information on the location of these areas of high genetic diversity and causal allele diffusion fronts or more broadly the landscape of adaptive genetic diversity (Eckert & Dyer, 2012) is key to an optimal sampling strategy.

When the causal allele diffusion gradient is unknown and a uniform causal allele distribution is assumed, there is a small advantage in choosing populations in the middle of the landscape. This can be explained by the absorbing boundaries used in the migration model which simulates a restricted range whereby alleles going beyond the border are lost. This reflects the phenomenon in fringe populations where migration regularly occurs beyond the suitable environmental niche of the species and the migrants fail to survive (Sexton et al. 2009). This is expected to apply to organisms with restricted range such as corals (Guan et al. 2015), 24% of coral reef fish species (Hawkins et al. 2006), and organisms in hydrothermal vents (Mullineaux et al. 2018). However, in cases where non-uniform causal allele distribution is assumed (e.g. temperature response in lodgepole pine and interior spruce (Liepe et al. 2015); and white spruce (Hornoy et al. 2015)), these populations along the borders of the species range are

important to sample because they may carry unique advantageous variants, especially in the context of climate change adaptation (Geber & Eckhart, 2005; Bridle & Vines, 2007; le Roux , 2009; Pais et al. 2020).

When considering the genetic architecture of a trait, there is less power to detect QTL but higher prediction accuracy for highly polygenic traits than for traits controlled by fewer loci. As a consequence, if we expect the polygenic trait to be controlled by relatively few loci, and we are more focused on mapping its genetic basis than on phenotype prediction, then sufficient power to detect QTL can be achieved with less populations sampled. Biotic stress resistance traits were often shown to be oligogenic, for example resistance to *Pseudomonas syringae* in *Arabidopsis thaliana* (Atwell et al. 2010), and resistance to *Cronartium ribicola* in sugar pine (Weiss et al. 2020). On the other hand, if we expect the trait to be highly polygenic, and we are more focused on phenotype prediction for monitoring purposes, then we have to sample as many populations as possible to sufficiently represent the variation across the landscape and maximise prediction accuracy. This is typically the case for abiotic stress resistance traits such as aluminium and proton tolerance in *Arabidopsis thaliana* (Nakano, et al. 2020), and coral bleaching resistance in *Acropora millepora* (Fuller et al. 2020) As the number of loci controlling the trait increases, the selection pressure acting on each locus decreases (Walsh & Lynch, 2018). This reduces the power to detect QTL since the individual contribution of each QTL decreases as more loci control the trait (Wang & Xu, 2019). This in turn reduces the proportion of polymorphic QTL within populations: if the majority of the QTL have small effects, they have a higher chance of getting lost due to drift than QTL with large effects. On the other hand, prediction accuracy increases since the rate at which genetic variance decreases due to directional selection is reduced as the number of QTL increases. Genetic variance should eventually become zero under constant stabilising or directional selection, but the rate of this reduction becomes slower with an increased number of loci controlling the trait (Crow & Kimura, 1970). Hence, GWAS and GP complement each other to achieve high QTL detection accuracy or high prediction accuracy for quantitative traits controlled by any number of loci.

When considering selection intensity, there is higher power to detect QTL but lower prediction accuracy in populations under intense selection. As a consequence, populations under high selection pressure can be selected for high-power GWAS: for example, weed and insect pest populations in agricultural areas including herbicide-resistant *Lolium rigidum* (Powles et al. 1998), and Bt-resistant *Helicoverpa armigera* (Jin et al. 2018). On the other hand, populations under low selection pressure can be selected for GP. For example, non-target organisms including *Drosophila melanogaster* populations which are resistant to Bt toxins (Babin et al. 2020) and imidacloprid (Fournier-Level et al. 2019) Increasing the selection intensity increases QTL detection power, since the effect of individual QTL becomes greater within each population (Wang & Xu, 2019) and less QTL alleles are lost due to drift. However, the predictive ability will be reduced by the Bulmer effect (Bulmer, 1971), where covariance between loci (partially explained by linkage disequilibrium (Walsh & Lynch, 2018)) is reduced after selection. Increasing selection intensity magnifies this reduction resulting in diminished additive genetic variance and less predictive models. This further solidifies the complementary nature of GWAS and GP and the utility of performing both with GPAS.

When considering migration rate, there is more power to detect QTL and greater prediction accuracy in populations experiencing high migration than in reproductively isolated ones. As a consequence, if gene flow is high, then less populations need to be sampled to sufficiently represent the variation across the landscape. This is the case for highly mobile organisms including birds (Pulido, 2007); and fishes (Brodersen et al. 2008), On the other hand, if gene flow is low which results in highly structured landscapes, then more populations need to be sampled to sufficiently represent the landscape variation. This can be the case in landscapes with considerable natural or artificial barriers to migration, and in species with reproductive structures impeding outbreeding, e.g. cleistogamous plants including *Crotalaria micans* (Etcheverry et al. 2003) and *Vigna caracalla* (Etcheverry et al. 2008). Increasing migration rate decreases differentiation between populations allowing for more causal alleles to be shared This leads to higher additive genetic variance per population, resulting in higher power and more accurate predictions (Liu et al. 2020).

2.6. Conclusion

Genome-wide association studies and genomic prediction (GPAS) are poised to complement existing methodologies in ecology in evolution. GPAS provides powerful tools to dissect the genetic basis of ecologically important quantitative traits including fitness and to rapidly monitor natural populations including invasive and threatened species. Understanding how the number of population samples and the different landscape properties affect the QTL detection power and phenotype prediction accuracies is integral to planning population collections for GPAS experiments. We recommend sampling as many populations as possible from areas of high genetic diversity. We also recommend Pool-seq whenever Indi-seq is too costly; since sampling more populations at the cost of lower resolution can be better than characterising a small number of populations at high resolution. The complementary nature of GWAS and GP allows good QTL detection power or prediction accuracy under low to high trait polygenicity and selection intensity. In the absence of prior information on the areas of high genetic diversity, we recommend against sampling populations at the border of the species' range.

Chapter 3

Quantitative measures of herbicide resistance in weeds

3. Quantitative measures of herbicide resistance in weeds

3.1. Introduction

Phenotyping strategies need to be deployed with the capacity to assess herbicide resistance of weed populations in high detail, quickly and accurately. The most popular way of assessing herbicide resistance is through assays which involve sowing seeds and applying herbicides at specific stages of development under varying doses ranging below and above recommended field application rates (Burgos 2015). These assays usually yield binary (dead/alive) survival data. However, herbicide resistance is increasingly found to be polygenic since combinations of multiple target-site resistance (TSR) mechanisms and non-target-site resistance (NTSR) mechanisms contribute to resistance traits (Powles and Yu 2010; Délye et al. 2013; Délye 2013; Baucom 2016). Measuring individual plant resistance as a continuous quantitative trait allows for more sensitive and potentially more informative assessment of resistance variation than canonical survival assays. These quantitative measurements will enable the use of powerful statistical approaches to dissect the genetic architecture of herbicide resistance.

In this chapter, we developed instaGraminoid, a novel colorimetric method to assess the quantitative herbicide resistance of individual plants using cheap ubiquitous RGB cameras (Paril and Fournier-Level 2019). We then used this method to describe the landscape of herbicide resistances across SE Australia.

3.2. instaGraminoid, a novel colorimetric method to assess herbicide resistance, identifies patterns of cross-resistance in annual ryegrass

3.2.1. Abstract

Herbicide resistance in agricultural weeds is a global problem with an increasing understanding that it is caused by multiple genes leading to quantitative resistance. These quantitative patterns of resistance are not easy to decipher with mortality assays alone, and there is a need for straightforward and unbiased protocols to accurately assess quantitative herbicide resistance. *instaGraminoid* - a computer vision and statistical analysis package was developed as an automated and scalable method for quantifying herbicide resistance. The package was tested in rigid ryegrass (*Lolium rigidum*), the most noxious and highly resistant weed in Australia and the Mediterranean region. This method provides quantitative measures of the degree of chlorosis and necrosis of individual plants which was shown to accurately reflect herbicide resistance. We were able to reliably characterise resistance to four herbicides with different sites of action (glyphosate, sulfometuron, terbuthylazine and trifluralin) in two *L. rigidum* populations from South-East Australia. Cross-validation of the method across populations and herbicide treatments showed high repeatability and transferability. Significant positive correlations in resistance of individual plants were observed across herbicides, which suggest either the accumulation of herbicide-specific resistance alleles in single genotypes (multiple stacked resistance) or the presence of general broad-effects resistance alleles (cross-resistance). We used these quantitative estimates of cross-resistance to simulate how resistance development under an herbicide rotation strategy is likely to be higher than expected. The software package, *instaGraminoid* is freely available at https://gitlab.com/jeffersonfparil/instaGraminoid.git and https://gitlab.com/jeffersonfparil/instaGraminoid_executables.git.

3.2.2. Introduction

Weeds are a major issue in modern cropping systems with yield losses due to infestation ranging from 7.5% to 10.5% in important crops i.e. wheat, rice, maize, potato, soybean and cotton (Oerke 2006; Villalobos and Fereres 2016). This typically represents a revenue loss of

$708 million per year in Australia alone (Llewellyn et al. 2016). Herbicide application is the most effective means to control weeds in broad-acre cropping (Amare 2014) and the intense selection pressure imposed by herbicides has driven the evolution of resistance in multiple weed species. According to the International Survey of Herbicide Resistant Weeds as of 2018, there are 495 unique reported cases of herbicide resistance spanning 254 species (Heap 2018). This trend shows that herbicide application strategies over the past four decades is not sustainable and will eventually result in the complete loss of herbicide efficacy.

Maintaining the efficacy of herbicides requires active management of resistance in weed populations. The emergence of herbicide resistance in the field must be detected early and its development monitored. This is particularly critical as early detection of resistance can trigger the deployment of efficient pre-emptive strategies instead of reactive ones. The most widely adopted resistance assays measure mortality by scoring the number of dead and surviving plants after herbicide application. These methods generate useful and unambiguous binary data characterising levels of resistance; however, mortality assays may not be sensitive enough to reveal early signs of resistance that could be helpful for pre-emptive strategies. Resistance is increasingly found to be polygenic (Powles and Yu 2010), due to the combination of multiple target-site resistance (TSR) mechanisms (e.g. mutation or over-expression of the target protein) together with non-target-site resistance (NTSR) mechanisms (e.g. detoxification and sequestration mechanisms) (Délye 2013; Délye et al. 2013; Baucom 2016; Babineau 2017). Thus, measuring herbicide resistance of individual plants on a quantitative scale can provide the precision needed to finely monitor the development of resistance. This also enables a better capture of the underlying genetic nature of resistance by increasing the statistical power to detect small effects and eventually a better prediction of resistance in field populations.

Quantifying herbicide resistance and overall plant health can be approached in different ways, e.g. visual inspection, visible light imaging, chlorophyll fluorescence imaging and hyperspectral imaging (Bock et al. 2010; Mutka and Bart 2014). Visual inspection is the least costly but can be highly biased, inaccurate and low throughput. Image-based approaches are objective and therefore more accurate and potentially high-throughput.

Visual inspection at the population-level described in Powles et al (1998), Neve and Powles (2005), Kaiser et al (2013), Han et al (2016), Yang et al (2016), Babineau (2017), and Wang et al (2017) do not generate the level of precision required to monitor individual plant resistance. Furthermore, these require considerably larger areas to replicate whole populations compared with individual plant assays. In addition, using heterogeneous groups of individuals as the experimental unit does not allow for one-to-one genotype comparisons. Individual-level phenotyping through visual inspection (Liu et al. 2014) can be highly subjective and non-transferable. Assays performed *ex situ* i.e. in agar (Kaundun et al. 2011; Brosnan et al. 2017) and excised leaf assays (Shaner et al. 2005; Ma et al. 2015) minimise the confounding effects of environmental factors and reduces the amount of space required but at the expense of not accounting for the effects of edaphic factors and whole plant response in the case of excised leaf assays.

Hyperspectral imaging techniques (e.g. normalised difference vegetation index (NDVI), chlorophyll fluorescence and thermal imaging, as well as near-infrared (NIR) and Raman spectroscopy (Bock et al. 2010; Kaiser et al. 2013; Walter et al. 2015; Zhang et al. 2016; Altangerel 2017; Jin et al. 2017; Wang et al. 2017; Wang et al. 2017a) are capable of highly objective and transferable characterisation of individual plants. Raman spectroscopy has been used to distinguish abiotic stresses in plants (Altangerel 2017). Imaging pulse amplitude modulated (PAM) fluorometry with the purposely built Weed PAM (Kaiser et al. 2013; Zhang et al. 2116; Wang P et al. 2017; Wang P et al. 2017a) have been used to asses herbicide resistance of weed seedlings in greenhouse conditions (Wang et al. 2018), and in the field (Wang P et al. 2017). However, hyperspectral imaging techniques require substantial investment to purchase highly specialised equipment which may not be economical for small laboratories and farmers interested in simple fine-level monitoring of herbicide resistance.

The scale and magnitude of the herbicide resistance crisis motivate the need for accurate and repeatable bioassays with high accuracy and repeatability, yet relying on basic equipment available to the broadest possible user base (R4P Network 2016). Hence, visible light imaging

can provide a balance between cost, accuracy and throughput. The xenobiotic effects of herbicides cause leaf senescence which progresses from chlorosis to necrosis, as well as growth stunting and eventually death of susceptible plants, while resistant plants remain green, growing, and healthy (Sass et al. 2012; Hörtensteiner 2014). These effects can be captured simply and objectively using digital imaging in the visible light spectrum. Leaf senescence and overall plant health have been accurately measured and monitored in real-time using coloured images of plant leaves under the visible light spectrum (Contreras-Medina et al. 2012; Cai et al. 2016). This implies it is possible to assess plant health from coloured photographs through colour correction, deblurring for low resolution images (Cai et al. 2016), and followed by image segmentation to isolate the leaves from the background. Morphological characterisation and colour-based metrics can then be calculated using either RGB (red, green, blue) or HSV (hue, saturation, value) colour models (Bock et al. 2010; Sass et al. 2012; Contreras-Medina et al. 2012; Cai et al. 2016).

Here we develop a novel RGB-based colorimetric method to compute multiple quantitative measures of herbicide resistance for individual plants. Photographs taken at discrete time points and the rates of change over time were used to perform optimal feature selection using *elastic-net* to train a model for predicting individual plant mortality to herbicides. We highlight a potential application of the quantitative resistance estimates by generating predictions about resistance development under four herbicide rotation strategies.

3.2.3. Materials and Methods

Plant materials and herbicide application

Ryegrass population samples were collected from two cropping areas located 369 km apart in South-East Australia. The first population was sampled from different field crops (canola, wheat and barley) on a single farm in Inverleigh, Victoria; while the second population was from a barley field in Urana, New South Wales. Seeds were stored at room temperature for six months to release dormancy and germinated in 0.8% agar at 15°C/25°C night/day temperatures with 12-hour photoperiod and 50% relative humidity. Each seedling was transferred to individual pots with soil media composed of 55% peat moss, 15% perlite, 5% vermiculite and 25% sand (dry volume). For each population, approximately 500 plants were maintained. Thirty days after transplanting, each plant was cloned by separating four tillers, taking care to include roots or root primordia during transplanting. Fourteen days after cloning through tiller separation, each set of clones was exposed to one of the four herbicides (Table 3.1). Herbicides were applied at the recommended field rates for a total volume corresponding to 200L·ha^{-1} to simulate field application.

Table 3.1. List of herbicides used to characterise herbicide resistance in two *Lolium rigidum* populations.

HRAC Group	Commercial Product Name	Active Ingredient (a.i.)	Recommended Field Rate (g a.i. ha^{-1})
B	Apparent Sulfometuron 750WG	Sulfometuron	400
C	Nufarm Terbazine 875WG	Terbuthylazine	1,050
G	Yates Weedkiller 490SC	Glyphosate	720
K1	Nufarm TriflurX 480EC	Trifluralin	816

Phenotype acquisition and image analysis

For each individual plant, above-ground plant tissues were photographed 0, 7, and 14 days after herbicide application from 4 perpendicular angles (each angle was treated as a technical replicate) under well and uniformly lit condition in a light box (dimensions: 61×61×61 cm; 2 light bands with 30 light emitting diodes each for a total of 13,000 lumens) with a fixed reference white background using a Nikon D7000 camera at 4,928 × 3,264 pixels resolution in “Auto” mode without camera-flash. Survival was scored 21 days after herbicide application. The photographs were colour-corrected by scaling the pixel values in the red, green and blue channels with the correction factor, w_0/μ_i where w_0 is the expected pixel value of the white background and μ_i is the observed mean pixel value of the white background in the i^{th} colour

channel. The colour-corrected images were transformed into grayscale and then into a binary image using the user-defined threshold value (threshold value was set at 75% of w_0). Morphological transformations with a 5x5 kernel were used to improve the detection of the thin leaves. Object detection using topological structural analysis[35] was used to isolate the above-ground plant tissues from the background. Leaf area, 15 leaf colour-based metrics, and change over pairs of measurements or deltas (Table 3.2) were calculated from pixel information of the colour-corrected and background-subtracted photographs. This workflow is depicted in Figure 3.1.

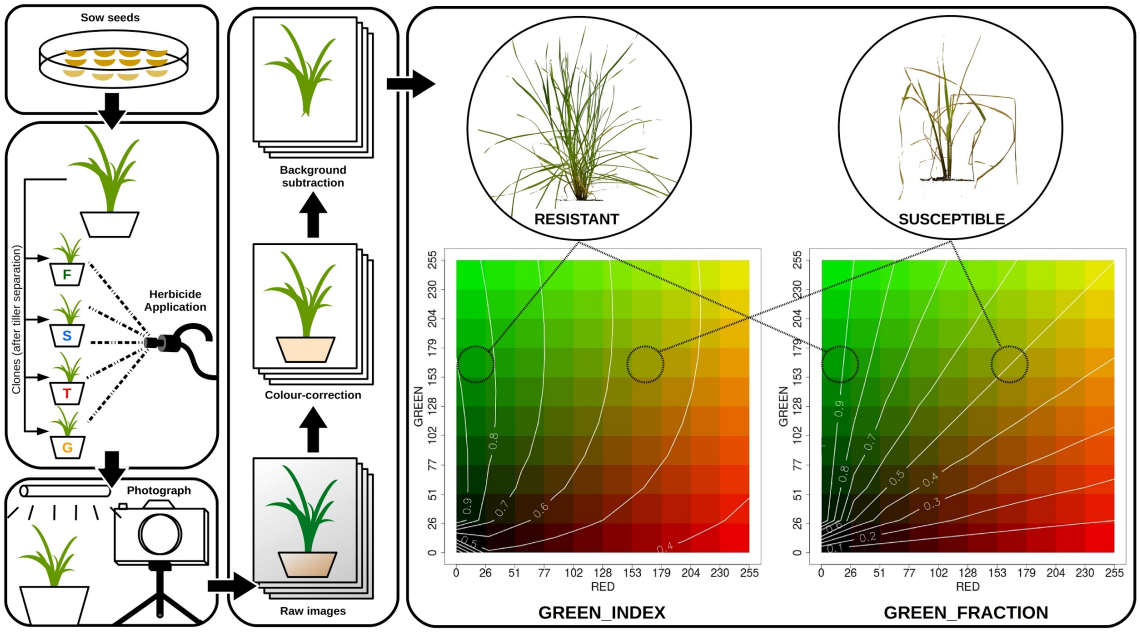


Figure 3.1. Workflow using instaGraminoid to extract the quantitative herbicide resistance estimates from plant photographs

Model construction and statistical analysis

To estimate herbicide resistance quantitatively, a logistic regression model for herbicide mortality was built on the 102 features (51 individual time points plus 51 deltas) using the R/glmnet package. The binary survival data, $y_i \in \{0,1\}$ was modelled using a logistic function,

$$p(y_i) = \frac{1}{1+e^{-(\beta_0 + x'_i \beta)}}, \qquad \text{(eq. 1)}$$

where $\boldsymbol{x}_i$ are the metrics extracted from the plant photographs and $\boldsymbol{\beta}$ are the regression coefficient associated with each metric. This logistic regression model was fitted to the mortality data using elastic-net penalisation (Friedman et al. 2010),

$$-\left[\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + x'_i\beta) - \log\left(1+e^{(\beta_0 + x'_i\beta)}\right)\right] + \lambda\left(\frac{1-\alpha}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1\right), \qquad \text{(eq. 2)}$$

where the first term is the loss function of the logistic model given the observed mortality data, $\boldsymbol{y}$; while the second term is the elastic-net penalty controlled by α and the tuning parameter λ. The α value was set to 0.5. The λ value was selected from a range of 100 values with λ_{min} = 0.0001 λ_{max} at all coefficients equal zero, by minimising the residual variance in the training set. The quantitative resistance, $\hat{y}_i$ was computed as the standardised negative exponent of e, $\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_i$ from the logistic equation. Metrics that were given non-zero $\boldsymbol{\beta}$ coefficients are hereafter termed features, consistent with the machine learning literature.

The performance of the *instaGraminoid* method in assessing plant resistance was measured as

the proportion of correctly classified individuals as resistant if $\frac{1}{1+e^{-\hat{y}_i}} \geq 0.5$ or as susceptible if $\frac{1}{1+e^{-\hat{y}_i}} < 0.5$. Transferability across populations and herbicide was assessed through 10-fold cross-validation by training the logistic model on a subset of the data (i.e. one population across herbicides, one herbicide across populations and one population for one herbicide) and testing the predictive ability of the model on a different subset. Each validation set was randomly divided into 10 subsets and the predictive ability of the trained model was estimated on each of these subsets. This was repeated for 10 iterations to get 100 performance estimations for each training-validation set pair. Correlations between resistances to herbicide treatments using Pearson's correlation coefficient (ρ) were calculated for each pair of herbicides. The receiver operating curve (ROC) plot was used to visualise the true positive discovery rate as the false positive rate increases.

Table 3.2. Plant health metrics derived from photographs of above-ground tissues of *Lolium rigidum* plants.

Metric	Description
a	leaf area visible from one angle (1 angle: 1 technical replicate)
μ_B	mean pixel value in the blue channel (range 0-255)
μ_G	mean pixel value in the green channel (range 0-255)
μ_R	mean pixel value in the red channel (range 0-255)
m_B	median pixel value in the blue channel
m_G	median pixel value in the green channel
m_R	median pixel value in the red channel
GREEN_INDEX	$\left[\left(1+\frac{\mu_G-\mu_R}{\mu_G+\mu_R}\right)+\left(1-\frac{\mu_B+\mu_G+\mu_R}{3\times 255}\right)\right]$
GREEN_FRACTION	$\frac{\mu_G}{\mu_B+\mu_G+\mu_R}$
$P(H_B)$	density-derived frequency of blue pixel values between 0 to 25
$P(H_G)$	density-derived frequency of green pixel values between 100 to 125
$P(H_R)$	density-derived frequency of red pixel values between 66 to 80
T_{G1}	proportion of green pixel values greater than 100
T_{G2}	number of green pixel values greater than 100 divided by the number of green pixel values less than 100
T_{R1}	proportion of red pixel values greater than 150
T_{R2}	number of red pixel values greater than 150 divided by the number of green pixel values less than 150
Δ_i	change in the above metric values between two measurement times i.e. Day14-Day7, Day14-Day0, and Day7-Day0

The resistance characterisation workflow (Figure 3.1), including the software (Figure 3.2) designed to extract the quantitative herbicide resistance trait are packaged into an open source project called *instaGraminoid* (https://gitlab.com/jeffersonfparil/instaGraminoid.git. Standalone executables for Mac OS (tested in OS X High Sierra), Microsoft Windows (tested in Windows 10), and Linux (tested in Ubuntu 18.04) can be found in https://gitlab.com/jeffersonfparil/instaGraminoid_executables.git). These programs were written in Bash 4.3 (Free Software Foundation 2016) using GNU-parallel version 2016 (Tange 2011), Python 2.7 (Python Software Foundation 2018) using the Open Source Computer Vision (OpenCV 3.4) (OpenCV 2017) and Scientific Python libraries (numpy 1.14 and scipy 1.0) (Jones et al. 2018), R 3.4 (R Core Team 2018) using the glmnet 2.0 (Friedman et al. 2010) and ROCR 1.0 (Sing et al. 2005) packages.

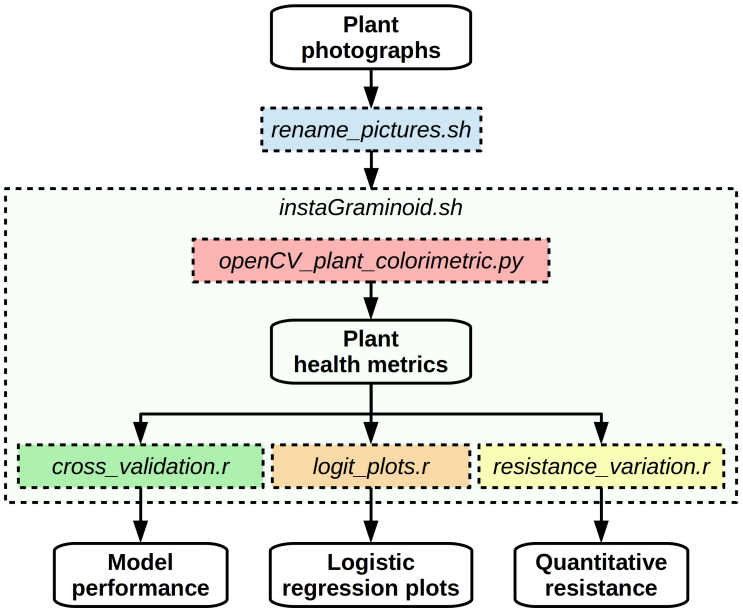


Figure 3.2. Architecture of the instaGraminoid package.

Herbicide resistance development simulation based on resistance correlation

The development of herbicide resistance under different scenarios was simulated based on the observed resistance levels and the correlations in resistance between the different herbicides using a quantitative genetics framework. The quantitative resistance estimates determined experimentally were scaled to range from 0 for maximum susceptibility to 1 for maximum resistance. Additive genetic variance (σ^2_a) was modelled as a parabolic function of resistance level, with maximum variance at the intermediate resistance level of 0.5 and minimum at the extreme resistance levels of 0 and 1. The evolution of resistance was simulated based on phenotypic variance (σ^2_p), additive genetic variance with the resulting heritability (σ^2_a/σ^2_p) set to a maximum of 80% and assuming an infinite population size. Selection for herbicide resistance was applied every generation for 12 generations using the multivariate breeder's equation (Lande and Arnold 1983):

$$\Delta z = G\beta, \qquad \text{(eq. 3)}$$

where Δz is the vector of response to selection; G is the variance-covariance matrix of resistance between any pair (i,j) of herbicide treatments, $\frac{\sigma(ij)}{\sigma(i)\sigma(j)}$ based on the observed correlations, and β is the vector of selection differentials imposed by a herbicide treatment ($\mu_{selected}$ - μ). Truncating selection was applied by selecting individuals with resistance value one unit of standard deviation greater than the mean. The mean resistance level of selected individuals was defined as:

$$\mu_{selected} = \frac{f(a) - f(b)}{\int_a^b f(x)\,dx}, \qquad \text{(eq. 4)}$$

where *f* is the probability density function of the standard normal distribution, *a* is the phenotypic value at $+1\sigma_p$ from the mean (corresponding to the truncating selection threshold) and *b* is 100% resistance. Three herbicide application scenarios were simulated: (1) continuous single herbicide, (2) herbicide rotation (sulfometuron - terbuthylazine - glyphosate - trifluralin) assuming no correlations between herbicide responses, and (3) herbicide rotation using the observed herbicide response correlations.

Resistance development was expressed as percent change in resistance from the initial resistance level,

$$\Delta\phi = \frac{\phi_t - \phi_0}{\phi_0} \times 100, \quad \text{(eq. 5)}$$

where Φ_t is the level of resistance at time *t*. The resulting resistance development curves were modelled using the logarithmic function,

$$\phi_t = a(t), \quad \text{(eq. 6)}$$

The scaling parameter, *a* was transformed from logarithmic scale back into linear scale so that the unit becomes change in resistance per unit time,

$$\Delta_{\frac{\phi}{t}} = \frac{a}{\exp(1)}. \quad \text{(eq. 7)}$$

This transformed metric was used to compare the rates of resistance development under the three different herbicide application scenarios.

3.2.4. Results

Herbicide resistance based on binary dead/alive data

Based on binary survival data, the survival rates to the different herbicides differed significantly between populations (Table 3.3). All the plants were resistant to trifluralin. All the plants were susceptible to glyphosate except for a single plant in the Urana population. Both populations showed variation in resistance to sulfometuron and terbuthylazine with the Inverleigh population being significantly more resistant.

Table 3.3. Proportion of resistant plants to the four herbicides. Chi-square test (df=1) was applied to test for independence between the two populations.

Herbicide	INVERLEIGH	URANA	χ^2 test p-value
Trifluralin	1.0000	1.0000	NA
Sulfometuron	0.1826	0.0567	9.21×10^{-9}
Terbuthylazine	0.1585	0.0452	3.13×10^{-8}
Glyphosate	0.0000	0.0023	9.65×10^{-1}

Performance of the colorimetric method, instaGraminoid

The run time for image processing and image-based metrics extraction from 1,000 jpeg images (3,264 × 4,928 pixels for a total size of 4GB) on a computer using 12 cores at 2.3GHz each, and 47GB of random access memory was ~3 hours with parallelisation and ~7 hours without parallelisation.

Qualitative assessment of the image processing showed that colour-correction and background subtraction using OpenCV functions were effective at detecting the above-ground plant tissues and removing the background (Figure 3.3). A threshold value of 75% w_0 (75% of expected white background pixel value) enabled effective background subtraction while minimizing artefacts compared with lower (50%) and higher (95%) threshold values.

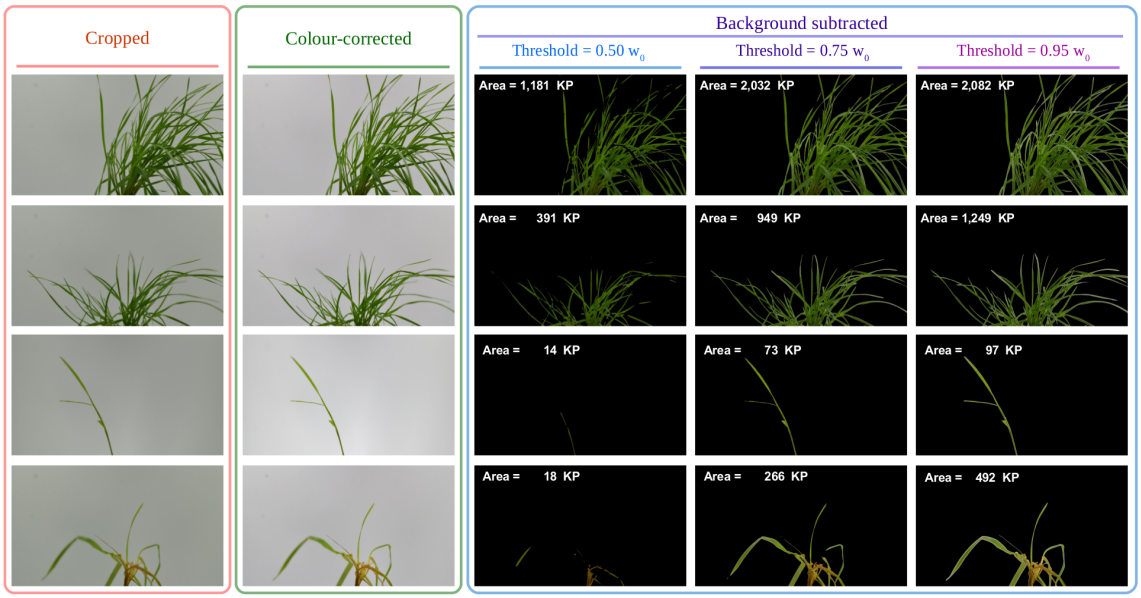


Figure 3.3. Image processing results for four sample plant photographs. (w_0 refers to the expected white background pixel value, and KP refers to ×1000 pixels).

All the metrics extracted from the photographs were highly repeatable (low variation among technical replicates taken from different angles, ANOVA F-test p-values ≤ 7.67×10^{-23}). The effects of most metrics were centred on zero and only a few metrics had large effects which indicates an efficient sparse feature selection using elastic-net penalisation (Figure 3.4).

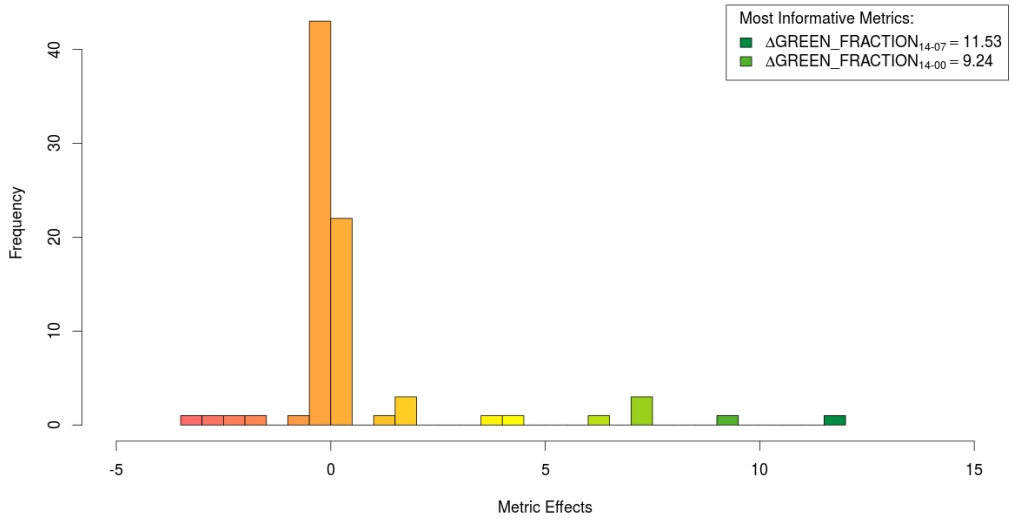


Figure 3.4. Histogram of the metric effects used for measuring the quantitative resistance of individual plants. (Top 2 metrics are labelled: ΔGREEN_FRACTION$_{14\text{-}07}$ - change in green fraction from day 14 to day 7, and ΔGREEN_FRACTION$_{14\text{-}00}$ - change in green fraction from day 14 to day 0; and uninformative features with standard deviations < 0.02 were omitted).

The *instaGraminoid* method correctly classified 95% of the plants as resistant or susceptible (Figure 3.5). The transferability of the phenotyping method was at least 92% across populations for a given herbicide (Table 3.4) and ranged from 23% to 100% across herbicide treatments (Table 3.5). This shows that a model trained in a single population and potentially a single herbicide can be transferred across different populations and different herbicide treatments despite different modes of action. However, models trained using glyphosate treatment data performed poorly in predicting trifluralin resistance.

Table 3.4. Proportion of plants correctly classified using the instaGraminoid colorimetric models trained on the populations in the row and validated in populations in the column (all these correlations are different based on Tukey's honest significant difference grouping at α=0.05).

Training Population	Validation Population	
	Inverleigh	Urana
Inverleigh	0.9347	0.9700
Urana	0.9201	0.9732

Table 3.5. Fraction of entries correctly classified using the instaGraminoid colorimetric models trained on herbicide treatments in the row and validated in herbicide treatments in the column (Letters correspond to Tukey's honest significant difference grouping at α=0.05).

Training Herbicide Treatment	Validation Herbicide Treatment			
	Trifluralin	Sulfometuron	Terbuthylazine	Glyphosate
Trifluralin*	-	-	-	-
Sulfometuron	0.8732^{e}	0.9378^{b}	0.8755^{e}	0.9989^{a}
Terbuthylazine	0.8892^{d}	0.9016^{c}	0.9082^{c}	0.9989^{a}
Glyphosate	0.2314^{g}	0.8665^{f}	0.8858^{d}	1.0000^{a}

* Model training using response variable data without any variation is not possible i.e. all the plants were unaffected by the trifluralin treatment.

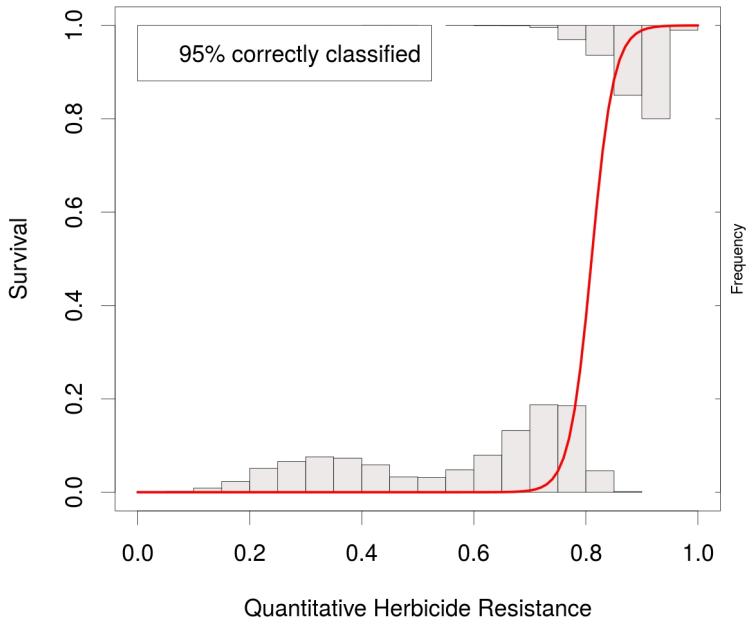


Figure 3.5. Capacity of the colorimetric model to determine resistance using logistic regression. (The histograms show the distribution of the quantitative resistance estimates in the susceptible [bottom] and resistant [top] plants. The red line is the logistic regression function relating the quantitative herbicide resistance to the observed survival data).

Correlated resistance to herbicides

Measuring herbicide resistance as a quantitative trait increased the sensitivity of detecting patterns of cross-resistance. Relationships of the herbicide resistance traits were assessed as percentage of overlap using the binary survival data, as well as correlation coefficients which can be more powerfully tested (Figure 3.6). A number of small yet significant positive correlations were identified between herbicide pairs: trifluralin-terbuthylazine, terbuthylazine-glyphosate, glyphosate-sulfometuron, and sulfometuron-terbuthylazine, with Pearson's correlation coefficients 0.11 (p-value=0.8%), 0.19 (<0.001%), 0.18 (0.001%) and 0.15 (<0.001%), respectively.

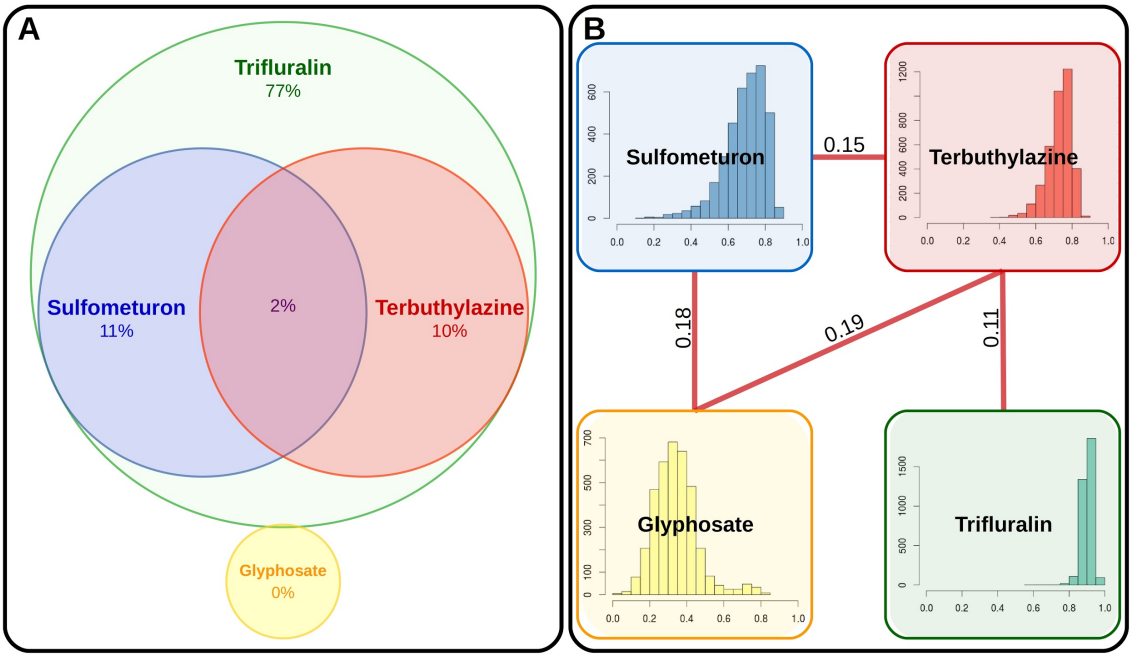


Figure 3.6. Patterns of herbicide resistance based on (A) binary survival data and (B) quantitative resistance estimates showing the significant correlations.

The consequence of these subtle positive correlations in herbicide resistance on the development of resistance was simulated using a response to selection model under different herbicide rotation strategies (Figure 3.7). The level of resistance to a specific herbicide is predicted to increase faster under continuous application of a single herbicide compared to herbicide rotation. Resistance could fixate as early as 3 generations of continuous herbicide application for populations with high initial resistance levels as in the trifluralin treatment or as late as 7 generations for populations with low initial resistance levels as in the glyphosate treatment. The rotation strategy is predicted to be less effective in the presence of correlated resistance, with the percent change in resistance per generation ($\Delta_{\varphi/t}$) ranging from 1.30% to 11.87% in absence of correlation and from 1.35% to 15.34% in the presence of these subtle positive correlations. For a population with high initial level of resistance, there is only 0.05% difference in the change in resistance per generation between correlated and uncorrelated resistance traits (i.e. trifluralin treatment); for a population with a low starting resistance level, there is 3.47% difference in the change in resistance per generation between correlated and uncorrelated resistance traits (i.e. glyphosate treatment).

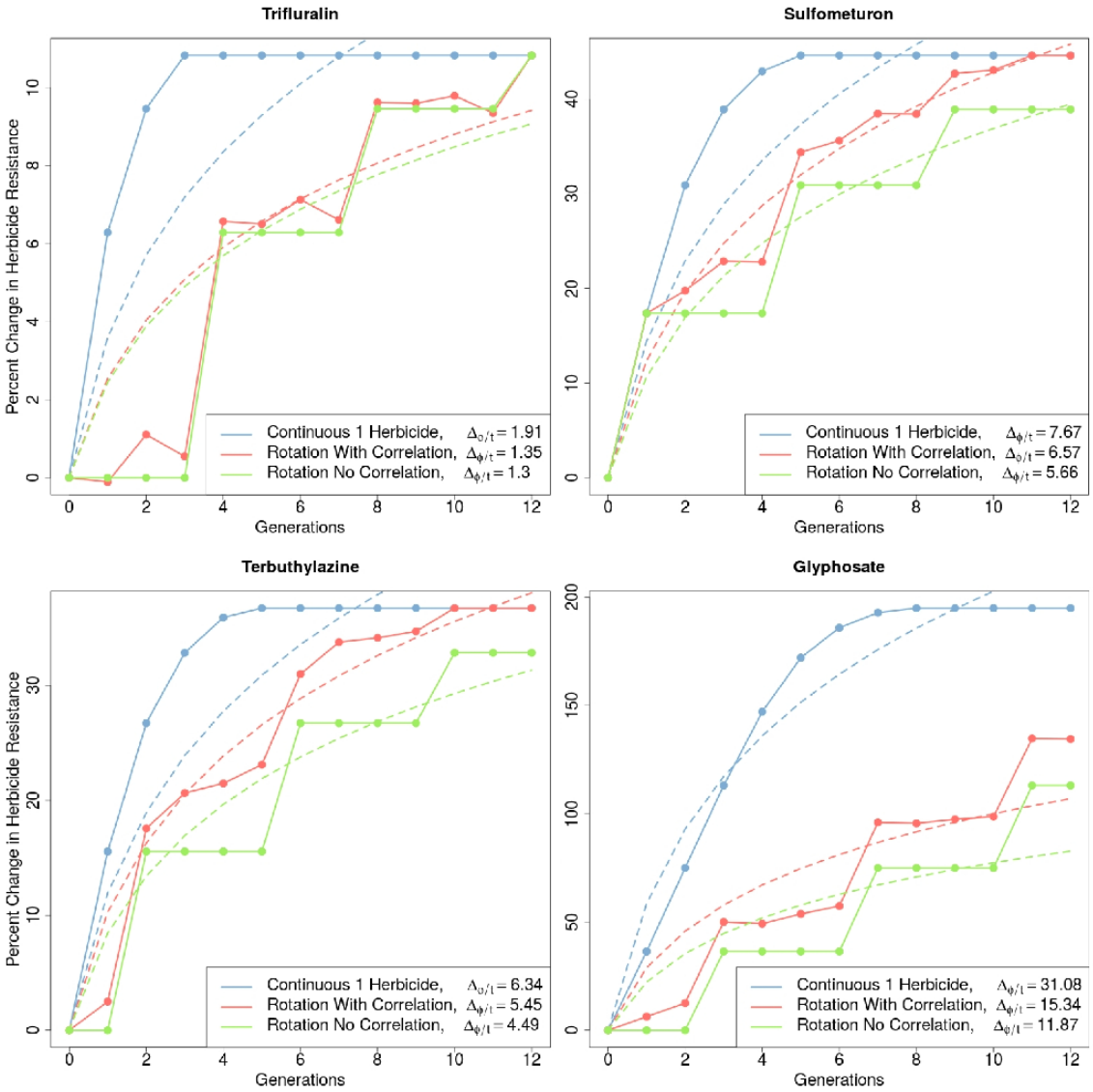


Figure 3.7. Simulated change in herbicide resistance as affected by herbicide rotation and resistance correlations. (Herbicide rotation: sulfometuron, terbuthylazine, glyphosate and trifluralin).

3.2.5. Discussion

Colorimetry as a robust approach to measure plant health

The instaGraminoid method accurately reflects plant survival after herbicide application and expresses herbicide resistance as a quantitative trait. It captures the essential image-based metrics explaining plant resistance and susceptibility to herbicides. Under stress and senescence, plants remobilise nitrogen from the photosynthetic tissues which results in chlorophyll breakdown (Hörtensteiner 2006; Hörtensteiner 2014; Wang and Blumwald 2014). More susceptible plants undergo greater stress and experience more rapid chlorosis and necrosis. This physiological change was captured in photographs and quantified as colorimetric features or metrics. The changes in green fraction from day 7 to 14 ($\Delta GREEN_FRACTION_{14\text{-}07}$) and day 0 to 14 ($\Delta GREEN_FRACTION_{14\text{-}00}$) are the most informative metrics after excluding the metrics with very low variances. In more complex context, deep learning could be used to expand the set of features from the photographs (Pound et a. 2017; Taghavi et al. 2018) used in the machine learning (glmnet) algorithm to predict resistance. However, this might generate biologically uninterpretable metrics and may result in parameter overfitting thus decreasing transferability (Taghavi et al. 2018). In addition, we aimed to select as model features the metrics that capture the progression of chlorosis and necrosis over time.

The rate at which chlorosis, necrosis and eventual plant death progressed differed not only between resistant and susceptible plants but also within each of these groups. This provides a strong justification for our colorimetric approach capable of measuring quantitatively these subtle yet significant differences in resistance. This was accomplished in a resource-effective, accurate and straightforward framework. However, when dealing with hundreds to thousands of plants, the task of photographing each individual plant is time consuming. Thus the next challenge is to extend the capability of *instaGraminoid* to identify and analyse photographs with multiple plants. In addition to the currently built-in functionalities and metrics, users can add their own metrics and use their own data to calibrate the model to specific weed species, populations and herbicides. The script "*openCV_plant_colorimetric.py*" can be modified to add new metrics and the script "*cross_validation.r*" can be used to build, train and test models using new input data. While the present study aimed to identify the most informative and biologically relevant metrics, user-specified metrics can be extracted to explain the traits of interest instead of applying the machine-learning procedure for feature selection to explain a target trait. This default model is nonetheless highly efficient for the current application of the method and successful at defining a quantitative resistance estimate by optimising the fit to the experimental data.

We showed through cross-validation that a model trained for one herbicide is highly transferable to other herbicide treatments with different modes of action. This demonstrates that the rates of chlorosis and necrosis are generally good indicators of plant health regardless of the type of herbicide used and possibly regardless of the type of stress, biotic or abiotic. However we observed poor transferability when the model was trained on a dataset with low phenotypic variation, i.e. glyphosate treatment where most plants were susceptible. This suggests that high levels of phenotype variability in the training dataset is critical for accurate prediction. Optimal transferability of the model can be achieved by training on a dataset representing the maximum range of phenotypic variation (Yates et al. 2018) – this applies to both the metrics and the response variable, be it herbicide resistance or any other plant health-related trait. The main recommendation for accurate prediction and early plant health assessment is to ensure that maximally variable populations are sampled for model training.

Relevance of the method to identify the molecular basis of herbicide resistance

Expressing herbicide resistance of individual plants as a quantitative trait using predictive models will enable fine-level and near real-time monitoring of plant health. Using our colorimetric method *instaGraminoid* and pre-trained *glmnet* predictive models, herbicide resistance of individual plants can be detected early using photographs.

Of the four herbicides used in this study, two resulted in low resistance variability in the two populations tested. These are glyphosate and trifluralin. Most of the plants did not survive 21 days after glyphosate application, most likely due to the low frequency of resistance-conferring alleles in the two populations. In comparison, all the plants survived trifluralin application. This apparent resistance of all the plants to trifluralin may be attributed to insufficient dose and application via spraying instead of the recommended soil-incorporation method (Epp et al. 2018). This probably resulted into high volatilisation of the herbicide and low absorption efficiency by the root and shoot apical meristems (Congreve and Cameron 2014). Trifluralin is used as a pre-emergence soil-incorporated herbicide and the recommended dosage used was optimised for germinating seeds not plants at mid-vegetative or tillering stage used in this study. However, the use of mid-vegetative plants was imperative to compare clones and test cross- and multiple resistance and similar post-emergence application of trifluralin has been shown to cause growth retardation in cotton plants when the terminal buds were exposed (Fischer 1966). An alternative would be to use families (half-sibs, full-sibs, early selfed or populations) but the one-to-one correspondence will be confounded by the heterogeneity of genotypes sampled across replications and treatments thus requiring larger sample sizes to minimise variance within group.

The quantitative herbicide resistance estimates will enable the application of powerful statistical analysis of the genetic basis of resistance, the accurate modelling of herbicide resistance evolution in agroecosystems and ultimately the improvement of herbicide resistance

management strategies. Binary survival data has been sufficient for studying the genetic basis of target site mutations and has been done extensively for many species and herbicides (Anderson and Gronwald 1991; Baerson 2002; Patzoldt et al. 2006; Yu et al. 2008; Hashim et al. 2012; Yang et al. 216). However, the search space for the causal genetic variants has typically been limited to the target site genes and well known detoxifying genes such as cytochrome P450 monooxygenases and glutathione S-transferases. Direct genetic studies without prior assumptions of the genetic elements conferring resistance such as genome-wide association studies are yet to be implemented in weeds. Our approach paves the way for harnessing high-throughput phenomics to next generation genomics in weeds.

Applications to field management of resistant weeds

The colorimetric method for herbicide resistance estimation presented here is more resource-effective compared to existing alternatives. A simple RGB camera using visible light sensors with resolutions of at least 5 megapixels can be used. In its current implementation, the method has limited portability; however it can be used in the field with pictures of the plants taken *in situ* by adapting the background subtraction function to isolate the plant tissues from the complex background. Any plant stage should be amenable to our protocol as long as the plants are not yet senescent as a result of natural aging and the level of stress is enough to induce a response in susceptible plants (e.g. the herbicide can be absorbed by the plant and the dosage is enough to induce plant stress responses). However hyperspectral imaging may out-perform our method because of the massive amounts of predictors that can be acquired, but the initial investment for acquiring the equipment is high. These can be used in a wide range of applications requiring non-invasive and non-destructive measurements. Abiotic and biotic stresses in plants have been measured using NIR spectroscopy (Jin et al. 2017). Furthermore, because of the massive amount of data that can be gathered with the different imaging techniques and the availability of easily accessible tools, deep learning techniques are becoming popular in plant phenomics (Ubbens et al. 2017; Singh et al. 2018). Deep learning techniques are being used to improve the performance of high-throughput image-based phenotyping workflows (Pound et al. 2017), for example in plant disease diagnosis (Ghosal et al. 2018; Toda and Okura 2019).

Nonetheless, the *instaGraminoid* machine-learning method enabled a direct estimation of cross-resistance. Pairwise resistance correlations between trifluralin-terbuthylazine, terbuthylazine-glyphosate, glyphosate-sulfometuron, and sulfometuron-terbuthylazine treatments were observed. These significant positive correlations in herbicide resistance are indicative of the absence of trade-offs in resistance across herbicides with different modes of action. The correlated resistance development shown by the herbicide rotation simulations depends on the assumption that the major genetic basis of resistance is NTSR. This needs to be validated through GWAS and functional genetics studies to confirm the genes/regulatory elements controlling resistance.

Heterogeneous herbicide applications across time and space (i.e. combination and rotation of different herbicides) is recommended to delay or manage the evolution of herbicide resistance (REX Consortium 2013; Hicks et al. 2018; Kreiner et al. 2018). However, the effectiveness of this strategy assumes independent, herbicide-specific TSR genes. In the presence and predominance of cross-herbicide NTSR genes, this strategy may be ineffective or even counter-productive. Under this strategy, the development of target site resistance would be delayed but the development of non-target site resistance that can induce resistance to multiple herbicides may be hastened which will be more problematic. Weed populations may develop resistance to herbicides they have not been previously exposed to. Integrated Weed Management (IWM) has been designed as a systems approach to minimise not only yield losses but also the negative impacts on the environment and health by managing weed populations (Swanton and Murphy 1996). IWM promotes the heterogeneity of weed control strategies, not just chemical control but also, mechanical, physical and biological weed control methods. The over-reliance on herbicides has promoted the evolution of herbicide resistant weed populations; however herbicides remain to be the cornerstone of modern weed control. In this context, timely monitoring of resistance levels through phenomics assays or genomic prediction (i.e. prediction of resistance levels based on genomic information) will boost the effectiveness, efficiency and sustainability of integrated weed control strategies.

3.2.6. Conclusion

The *instaGraminoid* method (https://gitlab.com/jeffersonfparil/instaGraminoid.git and https://gitlab.com/jeffersonfparil/instaGraminoid_executables.git), will provide end-users with a new, accurate, repeatable and transferable herbicide resistance assay that can be generalised for other types of plant health assessment. This method is able to quantify herbicide resistance of individual plants efficiently and accurately using visible light imaging and image analysis. Leaf greenness metrics were the most informative colorimetric features across the four herbicide treatments. The calibration of the model using *elastic-net* penalisation in a generalised linear model framework enables high transferability across populations and herbicide treatments. Using the quantitative herbicide resistance estimates, significant positive correlations were observed across herbicide treatments which may exacerbate the evolution of cross-resistance.

3.3. Materials and Methods

3.3.1. Weed populations and herbicides

Lolium rigidum populations were collected across SE Australia from crop and pasture fields (field populations) and from roads, lake sides, and an airstrip (feral populations) in 2016 and 2018 (Table 3.6). In 2016, two field populations from Inverleigh, Victoria and Urana, New South Wales were collected. In September 2018, five field populations were collected as live plants from Lexton, Rokewood and Skipton, Victoria. In November 2018, 60 field and feral populations were collected across Victoria, New South Wales and South Australia.

Table 3.6. List of the 67 *Lolium rigidum* populations sampled from South-East Australia in 2016 and 2018.

Dataset*	Name	Location	Type	Dataset*	Name	Location	Type
Intra-population	Inverleigh	Inverleigh	Field		ACC33	Dimboola	Field
	Urana	Urana	Field		ACC34	Dimboola	Field
Live collection	Lexton	Lexton	Field		ACC35	Pimpinio	Field
	Rokewood1	Rokewood	Field		ACC36	Pimpinio	Field
	Rokewood2	Rokewood	Field		ACC37	Dimboola	Field
	Skipton1	Skipton	Field		ACC38	Wilkur	Field
	Skipton2	Skipton	Field		ACC39	St. Arnaud	Field
Inter-population	ACC01	Glenrowan	Feral		ACC40	St. Arnaud	Field
	ACC02	Urana	Field		ACC41	Gringegalgona	Field
	ACC03	Lockhart	Field		ACC42	Gatum	Field
	ACC04	Lockhart	Field		ACC43	Gatum	Field
	ACC05	Lockhart	Field		ACC44	Gatum	Field
	ACC06	Lockhart	Field	Inter-population	ACC45	Cavendish	Field
	ACC07	Wagga wagga	Field		ACC46	Lake Bolac	Field
	ACC08	Wagga wagga	Field		ACC47	Moorlort	Field
	ACC10	Glenalla	Field		ACC48	Moorlort	Field
	ACC12	Glenalla	Field		ACC50	Moorlort	Field
	ACC14	Marrar	Field		ACC51	Moorlort	Field
	ACC15	Morongla	Field		ACC52	Elmore	Field
	ACC16	Morongla	Field		ACC53	Elmore	Field
	ACC17	Murrumbidgee	Field		ACC55	Cornella	Field
	ACC18	Waddi	Field		ACC57	Lang Lang	Feral
	ACC19	Hay	Feral		ACC58	Koo Wee Rup	Feral
	ACC20	Mildura	Feral		ACC60	Lake Bolac	Field
	ACC21	Cullulleraine	Feral		ACC61	Lake Bolac	Field
	ACC22	Hanson	Feral	Intra- & Inter-population	ACC09	Wagga wagga	Feral
	ACC23	Hart	Field		ACC11	Glenalla	Field
	ACC24	Hart	Field		ACC13	Glenalla	Field
	ACC26	Auburn	Field		ACC31	Kaniva	Field
	ACC27	Roseworhy	Field		ACC49	Moorlort	Field
	ACC28	Palmer	Field		ACC54	Cornella	Field
	ACC29	Coomandook	Field		ACC59	Lake Bolac	Field
	ACC30	Keith	Field		ACC62	Lake Bolac	Field
	ACC32	Dimboola	Feral	-	-	-	-

* - **Intra-population dataset** includes populations divided into 5 pools each using instaGraminoid; **Inter-population** dataset includes whole population data without subdivisions.

A total of five herbicide resistance traits were measured. The herbicides used and the respective dosages are listed in Table 3.7. The recommended application rates for Victoria, Australia were used. Clethodim is an acetyl-CoA carboxylase (ACCase) inhibitor, which disrupts fatty acid biosynthesis (Rendina and Felts 1988). Sulfometuron is an acetolactate synthase (ALS) or acetohydroxy acid synthase (AHAS) inhibitor, which disrupts the synthesis of branched amino acids - valine, leucine and isoleucine (Brown 1990). Terbuthylazine is a photosystem II (PSII) inhibitor (Simoneaux and Gould 2008). Glyphosate is a 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) inhibitor, which disrupts the synthesis of aromatic amino acids - tyrosine, tryptophan and phenylalanine (Steinrücken and Amrhein 1984). Trifluralin induces

tubulin depolymerisation, which inhibits cell division (Parka and Soper 1977). These are post-emergent herbicides except for trifluralin which is a strictly pre-emergent and sulfometuron which can be used as a pre- and post-emergent herbicide.

Table 3.7. List of herbicides used to characterise herbicide resistance in *Lolium rigidum* populations.

HRAC Group	Product Name	Active Ingredient (a.i.)	Recommended Field Rate (g a.i. ha^{-1})
A	Status Clethodim	Clethodim	120
B	Apparent Sulfometuron	Sulfometuron	400
C	Nufarm Terbazine	Terbuthylazine	1,050
G	Yates Weedkiller	Glyphosate	720
K1	Nufarm TriflurX	Trifluralin	816

3.3.2. Validation of instaGraminoid on field populations collected as live plants

To assess the viability of instaGraminoid as an easily accessible near-real-time herbicide resistance assessment, live plants collected from the field were used. *L. rigidum* weeds at varying vegetative stages from five field populations in Lexton (1 population with 220 plants), Rokewood (2 populations with 204 and 240 plants), and Skipton (2 populations with 330 and 191 plants) were collected in September 2018. The plants were transplanted from the field into trays with soil media described in Paril and Fournier-Level (2019). Sulfometuron was applied 14 days after collection. Binary and quantitative herbicide resistance data for each individual plant were collected following the instaGraminoid workflow.

3.3.3. South-East Australian populations characterisation

The distribution of herbicide resistances across the SE Australian landscape was assessed using the inter-population dataset, i.e. 60 populations which consisted of 51 field and 9 feral populations (refer to Table 3.6). Seed germination, growing condition and tiller separation methods were described in Paril and Fournier-Level (2019). At least 42 seeds from each population were sowed. Two replicates were made. Each population occupied at least one tray with a capacity for 42 individual plants. As dictated by the availability and viability of the seeds, as well as by the tillering capacity of the plants, some of the populations were phenotyped for up to 4 herbicide resistance traits. This resulted in an unbalanced experimental design where some populations were replicated more and tested under more herbicides than others. For clethodim and glyphosate resistances, all 60 populations were phenotyped. For sulfometuron and terbuthylazine, 47 and 26 populations were phenotyped, respectively. Trifluralin was not included because the pre-emergent nature of this herbicide resulted in suboptimal phenotypic variation as observed in the Inverleigh and Urana populations. Binary and quantitative herbicide resistance traits were measured at the population level. The instaGraminoid workflow was adapted for tray- or population-level measurements. The image cropping parameters were modified to include whole trays, and the survival rates per population were used instead of the individual plant survival data. Finally, the correlations between herbicide resistance traits at the population level were calculated.

3.3.4. Intra-population characterisation of 8 SE Australian populations

Eight populations observed to be maximally variable (≥10% and ≤75% survival rates) to at least one of the four herbicides (clethodim, glyphosate, sulfometuron, and terbuthylazine), were further characterised (refer to the intra- & inter-population dataset or the bottom-right rows of Table 3.6). For each herbicide resistance trait, two populations were selected: ACC13 and ACC49 for clethodim resistance, ACC31 and ACC62 for glyphosate resistance, ACC11 and ACC59 for sulfometuron resistance, and ACC09 and ACC54 for terbuthylazine resistance. Approximately 500 individual plants were individually characterised per population following the

instaGraminoid workflow.

3.4. Results

3.4.1. Validation of instaGraminoid on field populations collected as live plants

The quantitative herbicide resistance of individual plants from standing weed populations in the field were accurately measured with instaGraminoid. The population from Lexton, two populations from Rokewood, and two populations from Skipton exhibited 78.18%, 52.94%, 38.33%, 80.61% and 77.49% survival after sulfometuron application, respectively. The logistic function was 90% accurate in relating the instaGraminoid-derived quantitative sulfometuron resistance to the observed survival data (Figure 3.8).

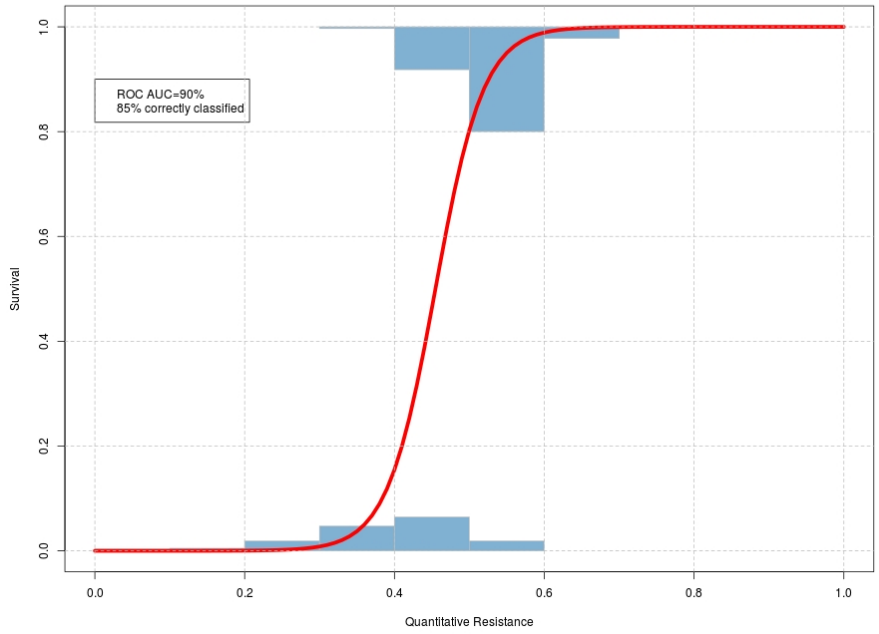


Figure 3.8. Logistic regression model for sulfometuron resistance of live plants collected from Lexton, Rokewood, and Skipton. (The histograms show the distribution of the quantitative sulfometuron resistance estimates in the susceptible [bottom] and resistant [top] plants. The red line is the logistic function relating the quantitative herbicide resistance to the observed survival data).

3.4.2. South-East Australian populations characterisation

Variations in resistance to clethodim, glyphosate, sulfometuron, and terbuthylazine exist across the SE Australian landscape (Figure 3.9). There are more populations with low resistances than highly resistant populations. All feral populations exhibited low resistances, except for one population collected from an agricultural airstrip that has been constantly sprayed with herbicides. Across all four herbicide resistance traits the feral populations had significantly lower resistances than field populations (based on Tukey's honest significant difference at α=5%), where 10.17% and 24.30% mean survival rates were observed for feral and field populations, respectively.

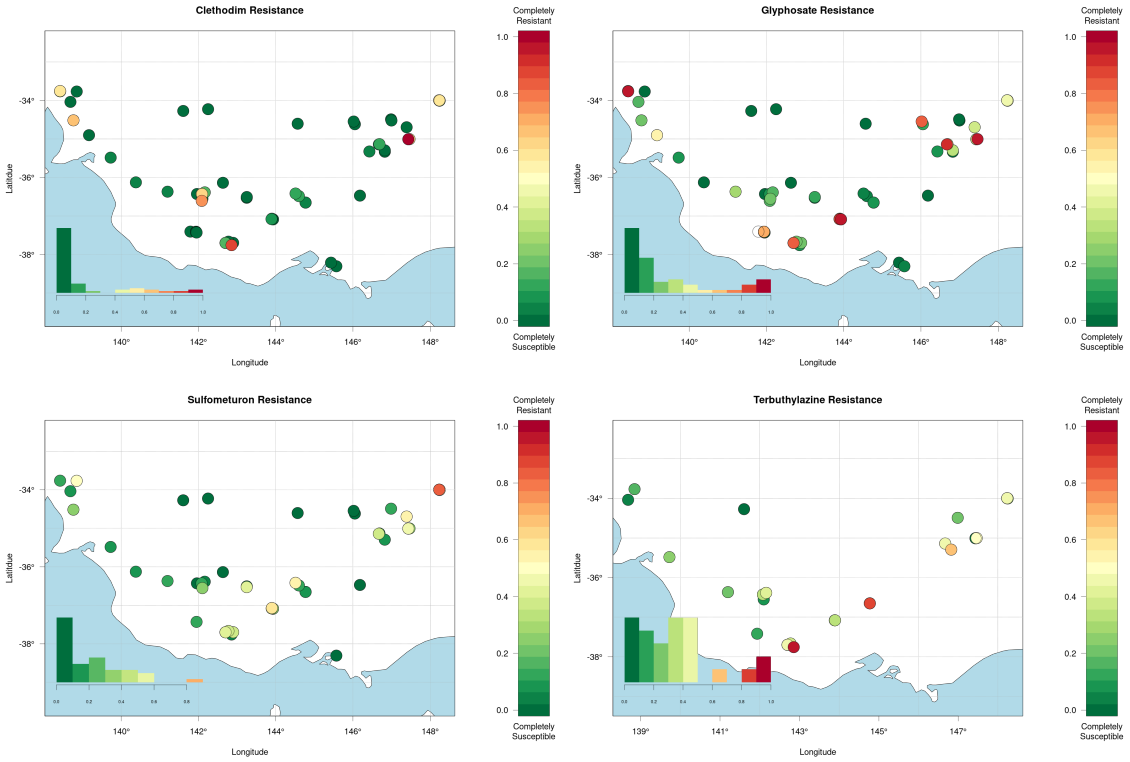


Figure 3.9. Herbicide resistance variation in *Lolium rigidum* populations across South-East Australia collected in November 2018. Herbicide resistances were measured as whole population survival rates. The points are the phenotyped populations. The histogram on the bottom left of every map shows the distribution of herbicide resistance based on the survival rate 28 days after herbicide application.

In terms of the population or tray level colorimetric measurements, the use of elastic-net penalisation under the generalised linear model framework allowed the straightforward use of the continuous percent survival data instead of binary survival data. This resulted in a maximum mean deviance ratio (explained deviance over total observed deviance) of only 29% with an elastic-net penalty parameter of zero or ridge regression penalty.

Using the survival rates of these populations, significant positive correlations between herbicide resistance traits were detected (Figure 3.10). These corroborate the previous findings using the quantitative herbicide resistances of Inverleigh and Urana populations (see Figure 3.6 Panel B). These significant positive correlations imply either the stacking of multiple herbicide-specific resistance alleles or the presence of broad-effect cross-resistance genes.

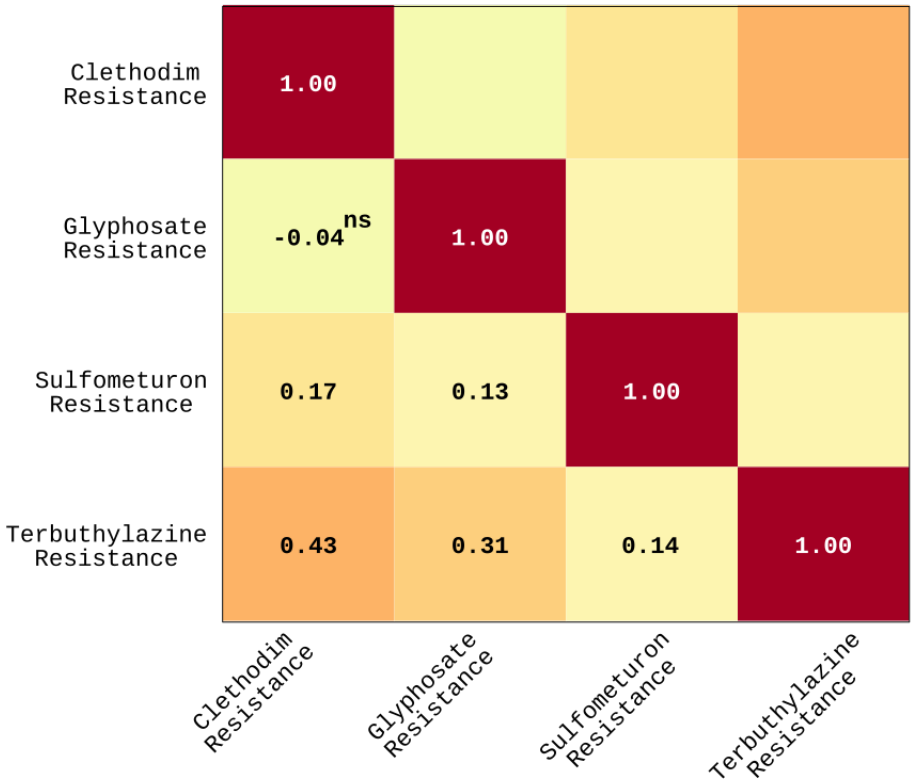


Figure 3.10. Correlations between herbicide resistance traits. The term “ns” refers to non-significant correlation at p<0.05.

3.4.3. Intra-population characterisation of 8 SE Australian populations

Among the 60 SE Australian populations characterised at the population level, 8 populations were further characterised at the individual plant level. The survival rates were 54% (ACC13) and 10% (ACC49) for clethodim, 26% (ACC31) and 75% (ACC62) for glyphosate, 46% (ACC11) and 41% (ACC59) for sulfometuron, and 50% (ACC09) and 32% (ACC54) for terbuthylazine. Seven of which were field populations and one was a feral population (ACC09) the one collected from an agricultural airstrip. All eight populations exhibited variations for herbicide resistance as measured by instaGraminoid (Figure 3.11).

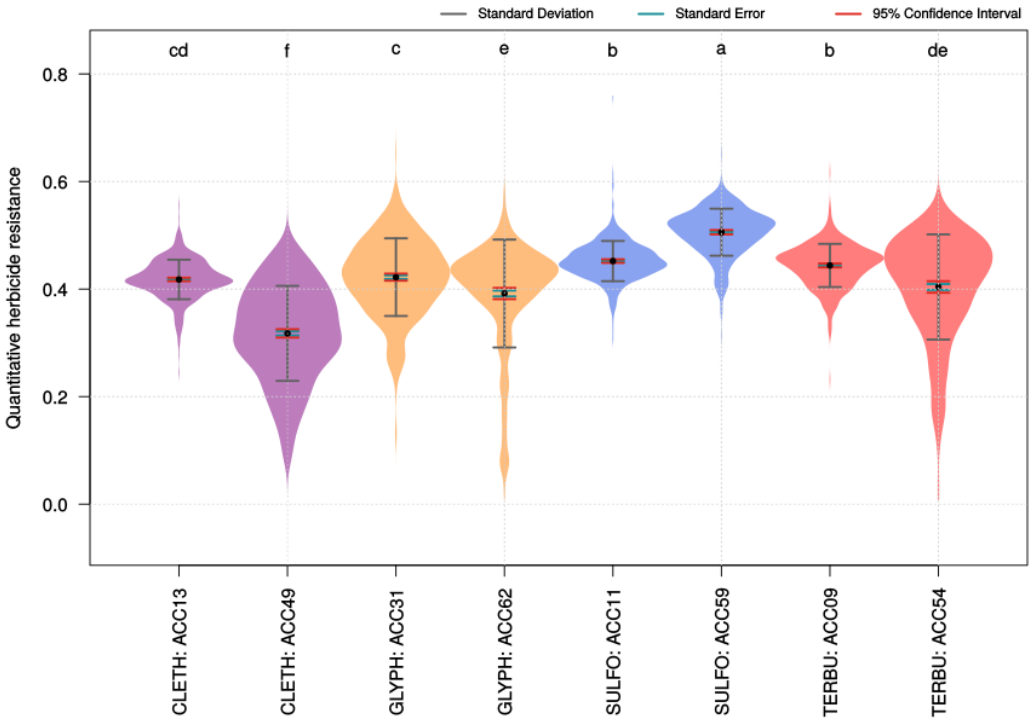


Figure 3.11. Intra-population quantitative herbicide resistance variations in 8 South-East Australian *Lolium rigidum* populations. Letters above each population-herbicide class refer to Tukey's honest significant grouping at α=0.05 (CLETH, GLYPH, SULFO and TERBU refers to clethodim, glyphosate, sulfometuron and terbuthylazine resistances, respectively).

3.5. Discussion

3.5.1. Quantitative herbicide resistance measurement using instaGraminoid

We developed an accessible and reliable method of measuring herbicide resistance of individual plants as a continuous quantitative trait, instaGraminoid (Paril and Fournier-Level 2019). The ubiquity of cameras and their low cost make instaGraminoid easily accessible to any laboratory requiring quantitative measures of herbicide resistance of individual plants and plant health in general. It is an unbiased alternative to manual visual rating of the degree of greenness and brownness of plant tissues. Plants remobilise nitrogen from photosynthetic tissues when under stress which results in chlorophyll breakdown or chlorosis (Hörtensteiner 2006; Christ and Hörtensteiner 2014; Wang and Blumwald 2014). As the plant's health deteriorates necrosis or the death of tissues occurs. The rate at which chlorosis, necrosis and eventual plant death progress vary not only between resistant and susceptible plants but also among resistant plants and among susceptible plants. The instaGraminoid method was able to capture these variations. This enabled the elucidation of small but significant positive correlations between herbicide resistance traits, indicative of either multiple- or cross-resistance.

The use of the recommended herbicide application rates, instead of performing a series of dose-response experiments was sufficient in characterising the *L. rigidum* populations across SE Australia. Dose-response experiments are resource-intensive and impractical for large sample sizes (Beckie et al. 2000). Using a single dose, specifically the recommended field dose for each herbicide, is valid since herbicide resistance as we've defined it is the ability of weeds to survive application rates at the recommended or commercial doses.

Binary survival data has been sufficient in studying the genetic basis of target site herbicide resistances (Anderson and Gronwald 1991; Baerson 2002; Patzoldt et al. 2006; Yu et al. 2008; Hashim et al. 2012; Délye 2013). However, genetic studies which do not assume target-site

resistance such as genome-wide association studies are lacking. Our phenotyping method, instaGraminoid paves the way for this by enabling the use of more powerful statistical analyses to associate genetic variants with quantitative herbicide resistance traits.

In terms of adapting instaGraminoid for population- or tray-level herbicide resistance measurements, it did not perform well (i.e. low deviance ratio) most probably because the plants at the centre of the tray were not well-represented since they were constantly blocked by the plants in front. This could be remedied by taking overhead photos in addition to the 4 perpendicular angles. However, this will require an image segmentation approach that can reliably distinguish between necrotic plant tissues and the soil.

3.5.2. Intra-population herbicide resistance

Using instaGraminoid for intra-population herbicide resistance assays, we observed variations in individual plant response, even when there was very low to no variation in the binary survival data. These quantitative measurements of individual plant resistance will enable fine-level and near real-time monitoring of herbicide resistance in the field. This will also allow the use of powerful statistical techniques to dissect the genetic basis of herbicide resistance beyond tests on contingency tables. However, one issue with the experimental design for the intra-population herbicide resistance assays was the absence of replicates. Replications will improve the precision of herbicide resistance estimates. Our approach is nonetheless valid as some plants may not be replicable due to intrinsic and resource limitations (Oksanen 2001), and replication studies can be performed in the future when more resources become available.

3.5.3. Herbicide resistance in SE Australia

Similar to a survey of herbicide resistance in annual ryegrass across Southeast Australia (Boutsalis et al. 2012), we found several herbicide-resistant populations across this landscape. The hotspots of highly resistant populations qualitatively coincide with regions of intense cultivation. Feral populations that are far away from intense agricultural activities and herbicide treatments exhibited significantly lower herbicide resistance than field populations. The variations within and between these populations can be exploited to infer likely models of herbicide resistance evolution.

3.6. Conclusion

The instaGraminoid method of expressing herbicide resistance of individual plants as a quantitative trait is highly accessible and reliable. This method can be used with seeds as starting material or with live plants collected from the field, and for individual plant and whole population characterisations. The resulting quantitative herbicide resistance measurements enable the use of powerful statistical techniques which would not be possible with binary survival data. Herbicide resistance variation across SE Australia is informative of the suitability of the sampled populations for quantitative and population genetics studies to dissect the genetic architecture of herbicide resistance and to infer the likely evolutionary paths towards herbicide resistance.

Chapter 4

Accurate Allele Frequencies from Pool-ddRADseq

4. Accurate Allele Frequencies from Pool-ddRADseq

4.1. Introduction

High-quality marker data is fundamental to quantitative trait loci (QTL) mapping, genome-wide association (GWAS), genomic prediction (GP), and population genetics inferences. Gathering genome-wide marker data from large (≥1×10^9 bases) and complex (large introns and abundant repetitive elements (Lynch and Conery 2003)) genomes is challenging. Genome complexity reduction techniques offer ways to tackle this problem. Specific subsets of large and complex genomes can be targeted through exome capture, transcriptome sequencing (RNAseq), and restriction site-associated DNA (RAD) characterisation. Exome capture reduces genome complexity by sequencing only the protein-coding regions (Albert et al. 2007; Warr et al. 2015). Although exome capture-derived genomic information is highly reproducible, the initial investment in developing the probes is high and the variation in non-coding regions is not represented (Hirsch et al. 2014). Similarly, RNAseq represents only the transcribed genomic regions and under-represents regulatory sequences (Zhao et al. 2019). In addition, RNAseq data is tissue-specific and highly affected by environmental conditions (Hirsch et al. 2014; Lopez-Maestre et al. 2016). Restriction site-associated DNA characterisation can be array-based, e.g. diversity arrays technology (DArT; Wenzl et al. 2004), or sequencing-based, i.e. restriction site-associated DNA sequencing (RADseq; Baird et al. 2008). Among these genome complexity reduction methods, RADseq is the most attractive due to the lower investment required and lower bias in genome representation (Andrews et al. 2016).

RADseq is a family of reduced-representation genome sequencing techniques for single nucleotide polymorphism (SNP) discovery and genotyping which includes the original RADseq protocol (Baird et al. 2008), and genotyping-by-sequencing (GBS; Elshire et al. 2011). They both use a single restriction enzyme with the latter involving indirect fragment size selection through multiplex PCR. Double digest RADseq (ddRADseq) improves the original RADseq by using a pair of restriction enzymes followed by direct fragment size selection which improves coverage and reduces the variance in coverage (Peterson et al. 2012; Toonen et al. 2013).

At the population level, RADseq techniques in tandem with pool sequencing (Pool-seq) have enabled the use of next-generation sequencing technologies for high-throughput landscape genomics studies in large groups of individuals and populations (Baird et al. 2008; Davey et al. 2011; Elshire et al. 2011; Andrews et al. 2016). The high level of detail captured by individual genotyping is not always biologically relevant or cost-effective for the genomic characterisation of large groups of individuals as commonly performed in ecological research. Pool-seq can generate more accurate allele frequency estimates than individual sequencing because of the larger sample sizes that can be sequenced for the same cost (Futschik and Schlötterer 2010). However, there are several issues with Pool-seq coupled with genome complexity reduction via RADseq, GBS or ddRADseq. The coverage may be insufficient to saturate large complex genomes with enough markers to capture all important loci (Lowry et al. 2017), and allele frequency estimates can be inaccurate for small pool sizes (Bansal 2010; Schlötterer et al. 2014). To maximise the accuracy of variant calling and allele frequency estimation, Pool-seq guidelines and recommendations have been proposed (Schlötterer et al. 2014; Fracassetti et al. 2015; Anand et al. 2016). Among these guidelines, large pool size and high sequencing coverage appear to be the most important factors. Pool sizes of at least 25 (Fracassetti et al. 2015; Gautier et al. 2013) or 40 individuals (Schlötterer et al. 2014); and sequencing depths greater than 50X (Zhu et al. 2012) and 65X (Gautier et al. 2013) have been recommended.

Despite these issues, the popularity of Pool-ddRADseq and similar Pool-seq coupled with reduced representation techniques is increasing (Pukk et al. 2015; Bélanger et al. 2016; Jiang et al. 2016; Kahnt et al. 2018; Nielsen et al. 2018). This family of techniques is used to characterise the genomic diversity of populations for studying adaptive evolution (Ellegren 2014; Storfer et al. 2018). It has also been used for QTL mapping using bulk segregant analysis in controlled crosses (Magwene et al. 2011; Dong et al. 2018) and GWAS in natural populations (Gaj et al. 2012; Yang and Kim 2015; Endler et al. 2016; Fournier-Level et al. 2017). However,

this has not yet been applied to non-model plant species with large complex genomes. *Lolium rigidum* is an important weed species with a large complex genome (2n=14 and an assumed genome size of 2Gb similar to *Lolium perenne* a closely related species (Byrne et al. 2015)) and a remarkable capacity to evolve resistance to many different herbicides. The evolution of resistance in this weed species can be studied using allele frequency data generated from Pool-ddRADseq.

In this chapter, we investigated the power of Pool-ddRADseq to accurately capture genome-wide allele frequency data in *L. rigidum* populations. The challenge is two-fold: how can ddRADseq reduce genome complexity while generating an unbiased representation of the genome? And how can Pool-seq yield accurate allele frequency estimates? Specifically, this chapter aims to:

- select the most appropriate restriction enzyme pair and fragment sizes to maximise the breadth and depth of coverage of ddRADseq;
- demonstrate an effective genome complexity reduction resulting in low heterogeneity of depth and breadth of sequencing coverage across the genome; and
- generate an optimal set of SNP filtering parameters to retain a genome-saturating number of SNPs with accurate allele frequency estimates from Pool-ddRADseq.

4.2. Materials and Methods

4.2.1. Genetic materials and pooling

Plant materials, growing conditions and experimental design were described in Chapter 3 Section 3.2 or in Paril and Fournier-Level (2019). Briefly, two *L. rigidum* populations were collected in Inverleigh, Victoria and Urana, New South Wales. Seeds were sown in Petri-dishes with 0.8% agar and germinated for 14 days inside a plant growth chamber set at 15°C/25°C night/day temperatures with a 12-hour photoperiod and 50% relative humidity. Seedlings were transplanted into 42-cell trays (4.5×4.5×4.5 cm cell dimensions) with soil media composed of 55% peat moss, 15% perlite, 5% vermiculite, and 25% sand by dry volume.

In each population, individual plants were arranged in order of increasing resistance levels and equally divided into 5 pools, where pool 1 has the lowest resistance and pool 5 has the highest, in preparation for genome-wide association studies. In cases where the population sizes do not equally divide by five, the remaining one to four individuals were allocated to the middle pool, i.e. pool 3. However, for the Pool-ddRADseq SNP filtering optimisation in this chapter, we disregard the pooling per population. Instead, we used each herbicide treatment as replication. Hence, the populations were replicated 4 times, each representing an herbicide treatment.

4.2.2. DNA extraction and sequencing

Leaf tissues were sampled prior to herbicide application and stored at -80°C. DNA was extracted from each pool. Approximately 100 mg of leaf tissue per plant were pooled and CTAB (cetyltrimethyl ammonium bromide) protocol (Clarke 2009) was used to extract the DNA. DNA was digested by pairs of restriction enzymes. DNA fragments were size-selected using magnetic bead-based kit (NucleoMag™ NGS Clean-up and Size Select).

For MinION nanopore-based nucleic acid sequencing (Oxford Nanopore Technologies Ltd.; ONT), the DNA fragments were end-repaired, A-tailed, barcoded per pool and ligated with MinION-specific adapters using ONT Ligation Sequencing Kit 1D (SQK-LSK109), ONT Native Barcoding Expansion 1-12 Kit (EXP-NBD103), NEB Blunt/TA Ligase Master Mix (M0367S), and NEBNext Quick Ligation Module (E6056). The quality and concentration of the resulting libraries were assessed using NanoDrop 1000 spectrophotometer and Qubit Fluorometer. The libraries were pooled and sequenced using the portable MinION device with the FLO-MIN106D flow cell.

For Illumina sequencing, a modified version of the ezRAD protocol of Toonen et al (2013) was used. The DNA fragments were size-selected (200-700 bp), end-repaired, A-tailed, barcoded per pool, ligated with Illumina-specific adapters, and PCR-enriched following the Illumina TruSeq DNA Kit protocol but modified for 1/3 reactions. The quality and concentration of each

library were assessed using Agilent 2200 TapeStation system. The DNA libraries were pooled and sequenced using one lane of the Illumina HiSeq X platform in 150 bp paired-end reads mode. The sequencing service was provided by GENEWIZ, Suzhou, China.

4.2.3. Reference genome and sequence alignment

In the absence of a *Lolium rigidum* genome assembly, we used as reference the genome of a closely related species, *Lolium perenne*. This draft genome (version 1.0; Byrne et al. (2015); Figure 1.9) with an assembly size of 1.128 Gb for an estimated whole genome size of 2.068 Gb, was used as reference genome for aligning *L. rigidum* sequencing reads. This has been successfully used as a reference genome in a pilot study for *L. rigidum* sequence alignment (Camm 2018).

For MinION sequence data, basecalling or translating the raw *fast5* electrical signals into *fastq* sequence files was performed using *albacore* (Wick et al. 2019). The reads were aligned to the reference genome using *minimap2 -ax map-ont* (Li 2018) with default parameters.

For Illumina sequence data, the *fastq* sequence files were aligned to the reference genome using *bwa mem* (Li 2013) with the default parameters listed in Appendix Table S4.1. Read pairs with mapping quality less than 20 phred score, i.e. >1% error rate, were filtered out. Base qualities were assessed with *fastQC* (Andrews 2010). Mapping summary statistics were computed using *samtools stats* (Li 2011).

4.2.4. Unbiased reduced-representation of the Lolium genome through ddRADseq

Genome complexity reduction experiments through ddRADseq were simulated *in silico*; validated using MinION sequencing for rapid quality assessment and protocol optimisation; and applied to the main sequencing experiment using the Illumina HiSeq X platform. The parameters tested were the restriction enzyme pairs and fragment sizes. The metrics used to assess the performance of ddRADseq were the breadth and depth of sequencing coverage. These metrics were calculated using non-overlapping sliding windows of 7Mbp, where the 48,415 scaffolds of the *L. perenne* reference genome were concatenated into a single linear sequence and divided into equally-sized non-overlapping windows, with each window covering at least 14 scaffolds. Depth of coverage per window was calculated as the sum of depths across all positions divided by the window size:

$$DEPTH = \frac{1}{w}\sum_{i=1}^{w} d_i \,,$$

where w is the window size in base pairs, and d_i is the depth at the i^{th} position. Breadth of coverage per window was calculated as the number of bases that was covered at least once divided by the window size:

$$BREADTH = \frac{1}{w}\sum_{i=1}^{w} c_i \,,$$

where $c_i = 1$, if $d_i \geq 1$ and $c_i = 0$, otherwise. Breadth and depth of coverage were calculated from the aligned, compressed and sorted bam files using *samtools depth -aa* (Li 2011). Two summary statistics, D_{50} and B_{50} were formulated to describe coverage as a function of the number of sequencing reads. D_{50} is the 50th percentile of *DEPTH* across windows. Similarly, B_{50} is the 50th percentile of the *BREADTH* across windows. The behaviour of these two summary statistics in response to the number of sequencing reads was used as an index of coverage saturation.

ddRADseq simulation

DDRADSEQTOOLS (Mora-Márquez et al. 2017) was used to simulate ddRADseq on the *L. perenne* genome. A set of 50 restriction enzymes equating to 1,225 enzyme pairs was tested *in silico* on approximately 10% of the genome composed of randomly selected scaffolds. The top

nine performing enzymes based on breadth and depth of coverage were identified. These top nine enzymes in tandem with 16 fragment size ranges (from 50bp to 400bp) were used to simulate ddRADseq on the whole *L. perenne* genome. The enzyme pairs and fragment sizes that resulted in the highest breadth and depth of coverage were identified.

ddRADseq MinION validation

The simulation results were experimentally validated using MinION sequencing on a test DNA sample. The predicted top-performing enzyme pair MseI-NsiI was compared to pairs with the commonly used enzyme PstI (Byrne et al. 2013; Gautier et al. 2013; Pukk et al. 2015; Ruttink et al. 2018), i.e. PstI-MseI and PstI-NsiI. Fragments between 200 and 700 bp long were selected during library preparation. Breadth and depth of coverage were averaged across the genome. Bootstrapping with 10 pseudo-replicates was performed to allow for analysis of variance and mean comparison using Tukey's honest significant difference (HSD).

Pool-ddRADseq on Illumina platform

After validation with MinION sequencing, the pooled DNA samples of the two populations were sequenced on Illumina HiSeq X platform in 150bp paired-end reads mode. The validated restriction enzyme pair MseI-NsiI and the Illumina-recommended fragment sizes ranging from 200 to 700 bp were used. Mean breadth and mean depth per window as well as D_{50} and B_{50} were calculated.

4.2.5. Accurate allele frequency estimation from Pool-ddRADseq

The effects of the SNP filtering parameters: sequencing depth, allele frequency, and basecalling quality on the accuracy of allele frequency estimates were assessed using simulated and empirical Pool-ddRADseq data.

Pool-ddRADseq simulation

The aim of simulating Pool-seq allele frequency estimation is to determine the relationship between depth of coverage and allele frequency on the accuracy of the estimated allele frequencies. It involved sampling biallelic genotypes from a binomial distribution with trial size of one and probability of success corresponding to a range of expected allele frequencies from 0 to 1. The number of samples represent the sequencing depth. Accuracy was measured as the absolute difference between expected and observed allele frequencies.

Pool-ddRADseq empirical data

Simulation results were validated using the empirical Illumina sequence data, SNP data subsets were generated using every combination of the three filtering parameters:

- minimum SNP depth (1X, 5X 10X, 20X, 30X, 50X, 75X, 100X and 1000X)
- minimum SNP quality score (Phred; 10, 15, 20, 30, and 40)
- minimum allele frequency (MAF; 0.1, 0.05, 0.01, and 0.001)

Sequencing read alignments were piled-up for each genomic position using *samtools mpileup* (Li 2011), and the number of SNP alleles were summarised into a synchronised pileup file using *mpileup2sync.jar* script of the Popoolation2 package (Kofler et al. 2011).

SNPs were filtered by minimum read depth and minimum allele frequency (MAF) with the aim of retaining a genome-saturating number of SNPs with accurate allele frequency estimates. To obtain empirical-based threshold values for the SNP filtering parameters, the number of SNPs (**SNP coun**t) and allele frequency accuracy (**accuracy**) were modelled as functions of the minimum depth threshold. Second degree polynomial regression models which explain a significant fraction of the variation without overfitting were developed. Accuracy per locus was expressed as the absolute pairwise difference in allele frequency estimates across replications. Since this accuracy ranged from 0 to 1, its distribution was modelled using a Beta distribution [$\beta(\alpha_1, \alpha_2)$]. The two shape parameters of the distribution were estimated using maximum

likelihood, and the resulting mean of the distribution, $\mu_\beta = \alpha_1/(\alpha_1+\alpha_2)$ was used to describe the overall allele frequency difference across the genome. SNP count was rescaled using the function:

$$C'_i = ((c_i - min_C) \times CF) + min_A$$

where C'_i is the transformed SNP count, c_i is the observed SNP count, $CF = \frac{max_A - min_A}{max_C - min_C}$ is the correction factor, max_C, min_C, max_A, and min_A are the maximum and minimum observed values of SNP count (C subscript) and accuracy (A subscript), respectively. Since minimum depth is inversely proportional to SNP count but proportional to accuracy, the minimum depth that defines a compromise between SNP count and accuracy was analytically calculated. The point of intersection between the accuracy and SNP count curves defines this *analytical point of compromise*. Using the expected number of SNPs defined at this point, the MAF threshold expected to pass the Bonferroni threshold with a risk α=5% for a causal loci when performing Pool-GWAS was calculated as:

$$MAF = \frac{x_{opt}}{2sp}$$

where:

$$x_{opt} = minimum(x \mid (1/p)^{x_i} < 0.05/l;\ \ for\ x \in [1, 2s]),$$

x_{opt} is the minimum number of copies of an allele in a pool required to pass the Bonferroni threshold, x is a vector containing different numbers of copies of an allele in a pool, s is the pool size, p is the number of pools, $(1/p)^{x_i}$ is the probability that x_i copies of an allele is found in a pool by chance, and l is the expected number of SNPs retained.

Additionally, for purposes other than Pool-GWAS (e.g. population genetics studies), the accuracy of allele frequency estimates are of prime importance. Accuracy and SNP count were simultaneously modelled as functions of depth and MAF (the two most important filtering parameters). The surfaces of the accuracy and SNP count models across the minimum depth and MAF axes were plotted. Three methods to find the optimum depth and MAF were used. The *observed point of compromise* is the highest point where the observed accuracy and the SNP count are least different, which was calculated as $argmax_{min.depth,\,MAF}\left(\frac{\mu_{Accuracy,\,SNP\,count}}{\sigma_{Accuracy,\,SNP\,count}}\right)$. On the other hand, the *analytical point of compromise* is defined as the point along the intersection of the two surfaces that maximised both accuracy and SNP count. Finally, the *line segment of weighted optimisation* is defined as the line segment intersecting the two surfaces where accuracy has 80% of the weight and SNP count has the remaining 20%, optimised using bounded limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS-B; (Byrd et al. 1995)).

4.3. Results

4.3.1. Reference genome and sequence alignment

The proportions of MinION and Illumina sequences that mapped to the *L. perenne* reference genome were assessed to validate the suitability of this genome assembly. The ddRADseq experiment using MinION sequencing produced a total of 910,604 sequences. Approximately 69% to 83% of these sequences mapped to the *L. perenne* reference genome. The Pool-ddRADseq experiment using the MseI-NsiI restriction enzyme pair under the Illumina HiSeq X sequencing platform generated ~798 million reads where 95% to 98% of which mapped to the *L. perenne* reference genome (Appendix Figure S4.1). This confirmed that the *L. perenne* genome assembly can be used effectively as reference genome for aligning *L. rigidum* sequences.

4.3.2. Unbiased reduced representation of the Lolium genome through ddRADseq

In order to generate an unbiased reduced representation of the *L. rigidum* genome, ddRADseq was simulated *in silico* using different restriction enzyme pairs and fragment size selection schemes. The results were validated with MinION sequencing and applied to the full-scale Illumina sequencing experiment.

ddRADseq simulation

In silico ddRADseq simulations showed that among the 50 restriction enzymes tested on 10% of the genome, AseI, BclI, BglII, MseI, NcoI, NdeI, NsiI, PciI and PstI were the nine enzymes that maximised breadth and depth of coverage. Among these top nine enzymes tested on the whole reference genome, the pair MseI-NsiI maximised the breadth and depth of coverage (Figure 4.1 Panels A and B) and MseI generally resulted in high sequencing depth (Appendix Figure S4.2). For fragment size selection, the top three longest fragment sizes, 50-400 bp, 100-400 bp and 50-300 bp maximised the depth of sequencing (Figure 4.1 Panel C). Hence, MseI and NsiI in tandem with large fragments sizes are expected to maximise the breadth and depth of coverage.

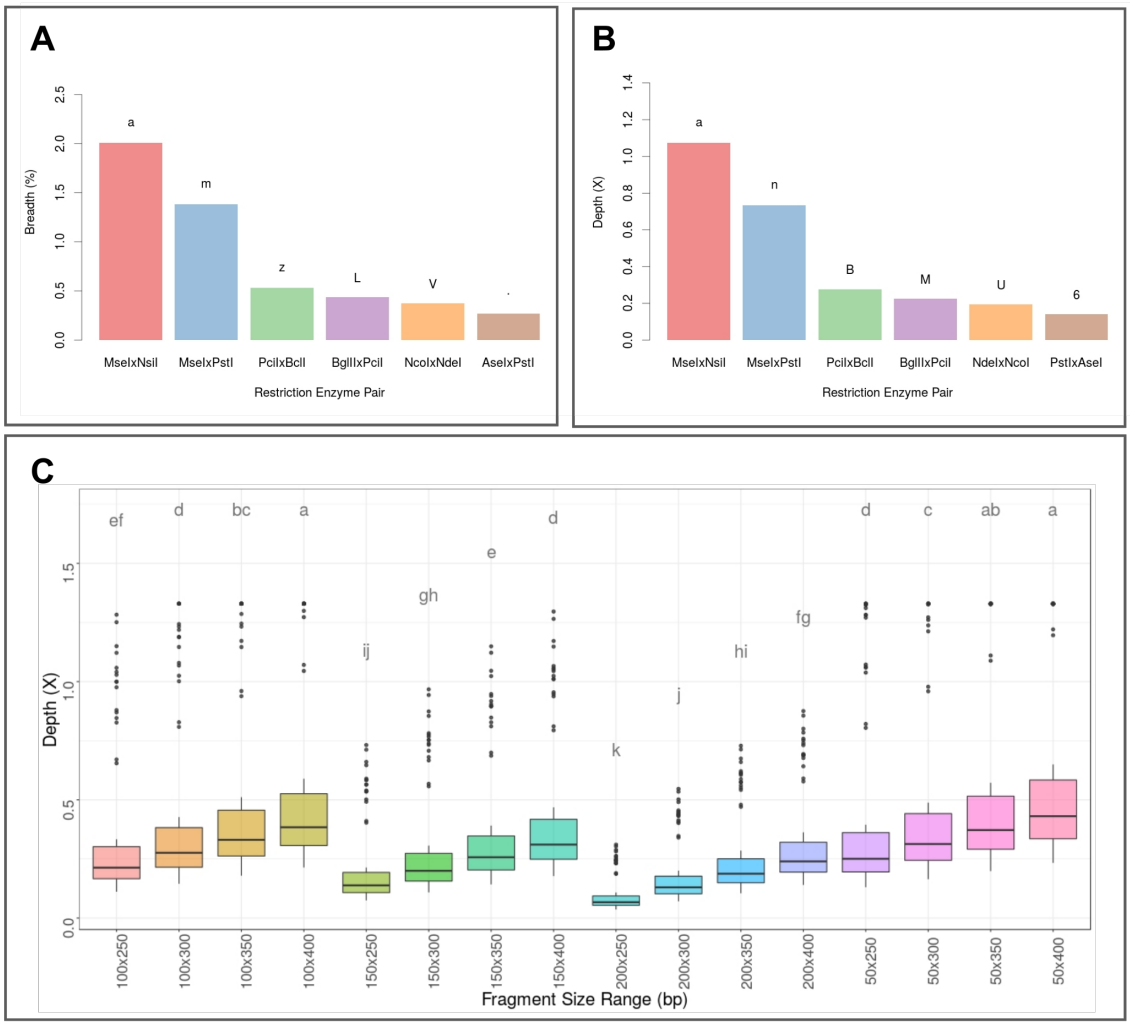


Figure 4.1. Breadth and depth of sequencing across the Lolium perenne genome based on simulated ddRADseq using different enzyme pairs (only a small subset is shown) and at varying fragment size ranges (letters refer to Tukey's HSD grouping at α=0.05).

ddRADseq MinION validation

The *in silico* results were validated empirically using MinION sequencing. Genome complexity reduction using the MseI-NsiI enzyme pair out-performed PstI-NsiI and PstI-MseI pairs in terms of breadth and depth of sequencing coverage (Figure 4.2).

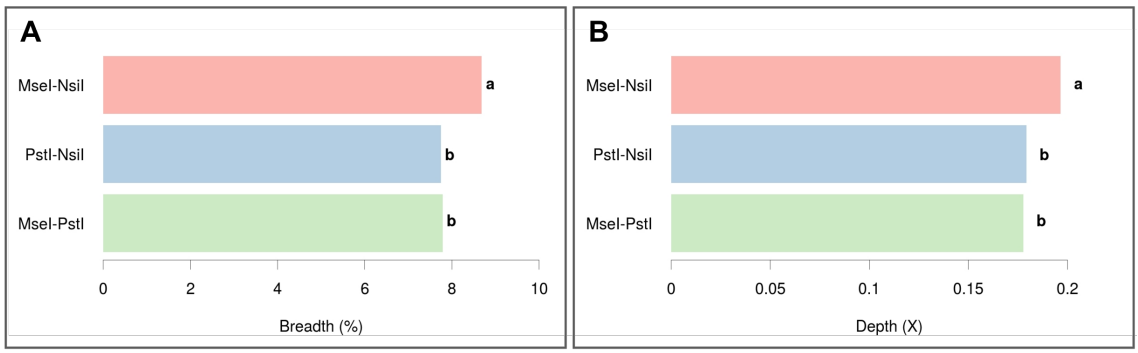


Figure 4.2. Breadth and depth of MinIon sequencing coverage using MseI-NsiI, PstI-NsiI and MseI-PstI restriction enzyme pairs (letters refer to Tukey's HSD grouping at α=0.05).

Pool-ddRADseq on Illumina platform

The results of the *in silico* simulation and MinION sequencing validation were applied to the high throughput Illumina sequencing experiment. On average 16% to 19% of the reference genome was covered with genome-wide depths ranging from 1.16X to 2.63X. This means that for the 17.5% of the genome that was actually covered by sequencing reads, the average depth was effectively $\frac{1.895\,X}{17.5\,\%} \approx 11\,X$. The distributions of breadth and depth across the genome averaged across the libraries sequenced are shown in Appendix Figures S4.3 and S4.4, respectively. The variation in breadth was uniform although the standard deviations are substantial. One region exhibited greater variation in depth than other regions. D_{50} increased linearly as the number of reads increases while B_{50} showed a sign of possible plateauing (Figure 4.3).

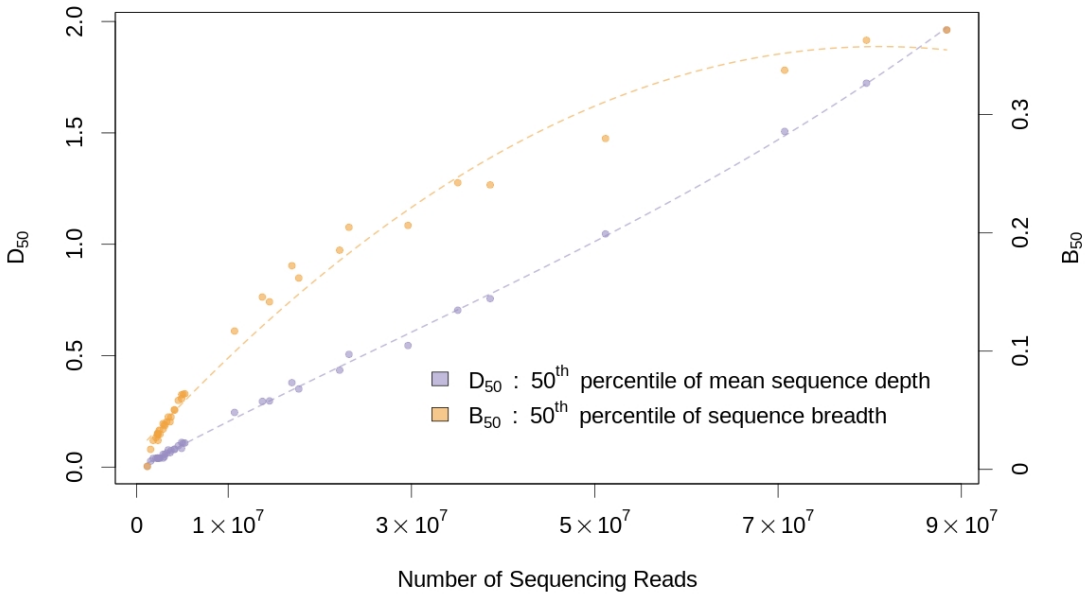


Figure 4.3. Breadth and depth of sequencing coverage as the number of reads increases.

4.3.3. Accurate allele frequency estimated from Pool-ddRADseq

To retain high quality SNPs which generate accurate allele frequency estimates, the search space for the optimum SNP filtering parameter thresholds (minimum depth and minimum allele frequency) was explored using simulated and observed Pool-ddRADseq data.

Pool-ddRADseq simulation

Pool-ddRADseq simulation showed that read depths of at least 100X will result in accurate and precise estimates of allele frequencies (Figure 4.4). The squared deviation of estimated vs. expected allele frequencies and the corresponding variance approached zero as read depth increases. On the other hand, this square deviation is maximised at intermediate allele frequency.

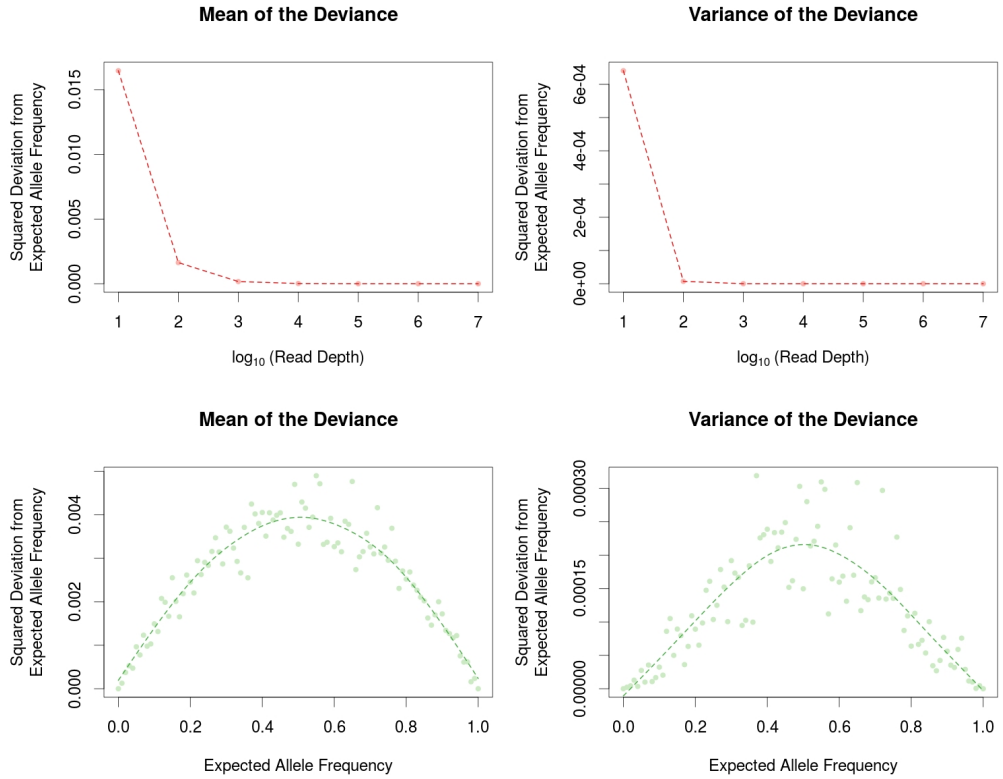


Figure 4.4. Allele frequency accuracy and precision simulations.

Given the empirically observed LD block sizes in *L. perenne* ranging from 0.5kb to 3kb (r^2 dips below 0.2) (Ponting et al. 2007), and assuming a genome size of 2Gb, the theoretical minimum number of SNPs required to saturate the genome as a function of LD block size as shown in Appendix Figure S4.5, ranged from 700,000 to 4 million SNPs.

Pool-ddRADseq empirical data

To validate the Pool-ddRADseq simulation result and the recommendations of Pool-seq review papers (Schlötterer et al. 2014; Fracassetti et al. 2015; Anand et al. 2016), the effects of the three SNP filtering parameters (minimum depth, minimum base quality and minimum allele frequency) on allele frequency accuracy and SNP count were investigated using empirical Pool-ddRADseq data.

The number of SNPs retained was reduced by a factor of 10 when the minimum depth cut-off was increased by a factor of 10 from 10X to 100X, and 100X to 1000X (Appendix Figure S4.6). The behaviours of allele frequency accuracy and SNP count in response to filtering for minimum SNP depth, minimum SNP quality, and minimum allele frequency using the Illumina sequence data are presented in Appendix Figure S4.7. Allele frequency accuracy improved as depth increased while the number of SNPs retained decreased. Both accuracy and SNP count decreased as the minimum allele frequency increased. Interestingly, accuracy and SNP count were unaffected by base quality since the quality of basecalling was already very high with expected error rates ranging from 0.014% to 0.028% (Appendix Figure S4.8). Under the assumption that basecalling quality remains in this range, it is justifiable to exclude base quality from the parameters that need to be optimised.

Accuracy and SNP count responded to MAF in the same direction but had opposite responses to minimum sequencing depth (Appendix Figure S4.7). The relationship between allele frequency accuracy and SNP count in the sets of SNPs filtered with varying levels of minimum depths and MAF are shown in Figure 4.5. We aimed to find a combination of minimum depth and MAF that minimises the error in allele frequency estimation and maximises the number of SNPs (a point in the upper left quadrant in Figure 4.5). As minimum depth increases, MAF

decreases in importance in terms of SNP count. This is shown by the decrease in the slope of the regression lines as minimum depth increases.

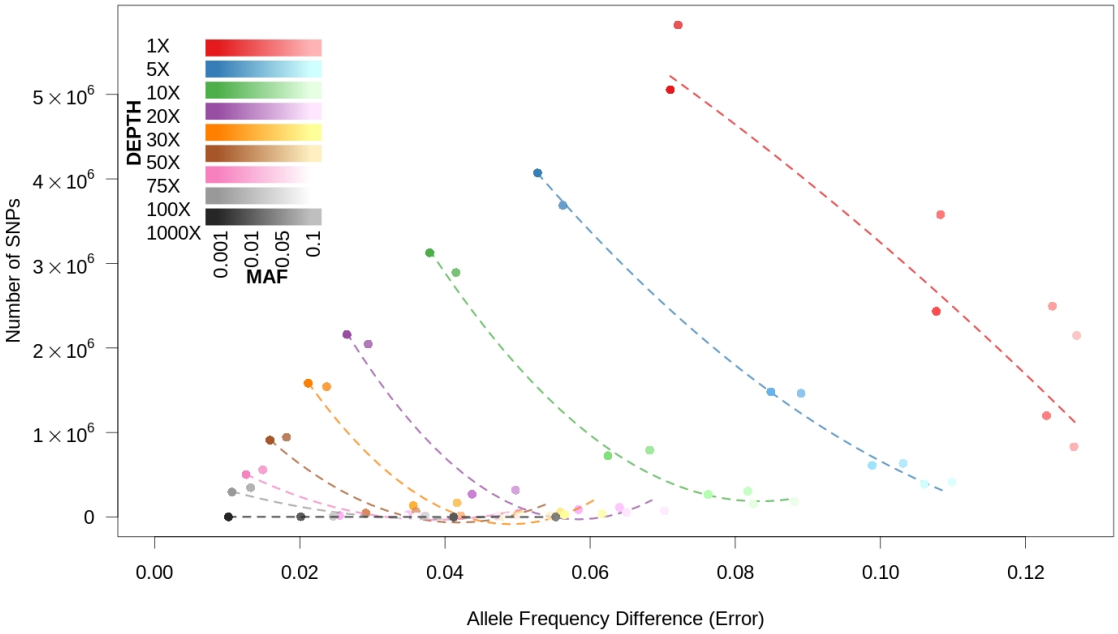


Figure 4.5. Effect of different sequencing depth and minimum allele frequency filtering parameters on allele frequency estimation accuracy and the number of SNPs retained.

The regression model that best explained the accuracy and SNP count as functions of depth and MAF appears to be the polynomial model of the 2nd degree (Appendix Figure S4.9). Model fit improved as complexity increased from linear to 4th degree polynomial. However, increasing model complexity from the 3rd to the 4th polynomial degree failed to shrink the residual variance, AIC, and BIC considerably. The residual variance is essentially zero at the 3rd and 4th degrees which suggest over-fitting above the 2nd degree polynomial model. Therefore, the polynomial model of the 2rd degree was used to model both accuracy and SNP count.

The minimum depth threshold of 8X defined the analytical point of compromise between the mean allele frequency accuracy and mean SNP count. At this point of intersection, the expected allele frequency deviance is at 0.07 and the expected number of SNPs retained is at 1.2 million (Figure 4.6).

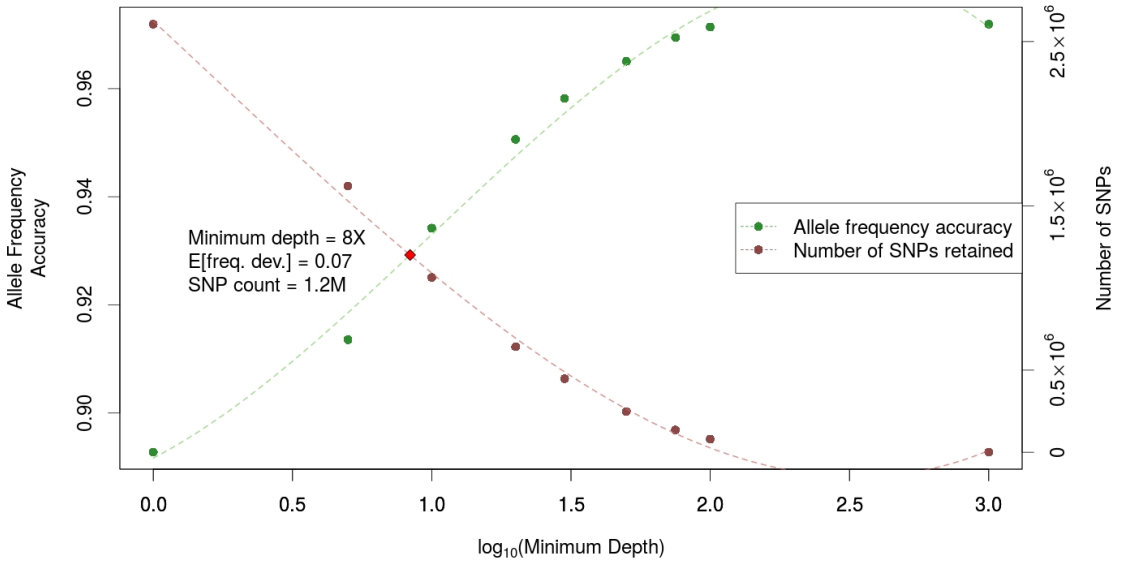


Figure 4.6. Optimising the compromise between allele frequency accuracy and number of SNPs as functions of minimum depth alone (E[Freq. Dev.]: expected deviation from actual allele frequency, SNP count: number of SNPs).

Using the 1.2 million SNPs expected to be retained, the Bonferroni threshold at α=5% was calculated to be 4.17×10^{-8}. Assuming that 5 pools will be used in Pool-GWAS, one random copy of an allele has a probability of being sampled in a particular pool by chance equal to ⅕, and the probability of x copies of this allele to be sampled in this pool by chance is equal to $(⅕)^x$. The minimum number of copies of an allele to pass the Bonferroni threshold is x_{opt} = 11, that is $(⅕)^{11} = 2.05 \times 10^{-8}$. Given this minimum of 11 copies of a causal allele, and assuming that 1.2 million SNPs, 100 individuals per pool and 5 pools will be used in Pool-GWAS, then the minimum allele frequency threshold was calculated to be $MAF = \frac{11}{2 \times 100 \times 5} = 0.011$.

For purposes other than Pool-GWAS, the accuracy and SNP count surfaces across the minimum depth and MAF axes were plotted (Figure 4.7). The *observed point of compromise* was found to be at 5X depth, and 0.001 MAF resulting in a deviance of 0.06 for allele frequency and in 3.69 million SNPs retained. On the other hand, the *analytical point of compromise* was determined to be at 3X depth and 0.001 MAF with expected deviance of allele frequency of 0.0706 and 2.81 million SNPs retained. In addition, the *line segment of weighted optimisation* intersects the two surfaces at 227X depth and 0.001 MAF with the lowest expected deviance of allele frequency of 0.01 and 600,000 SNPs retained.

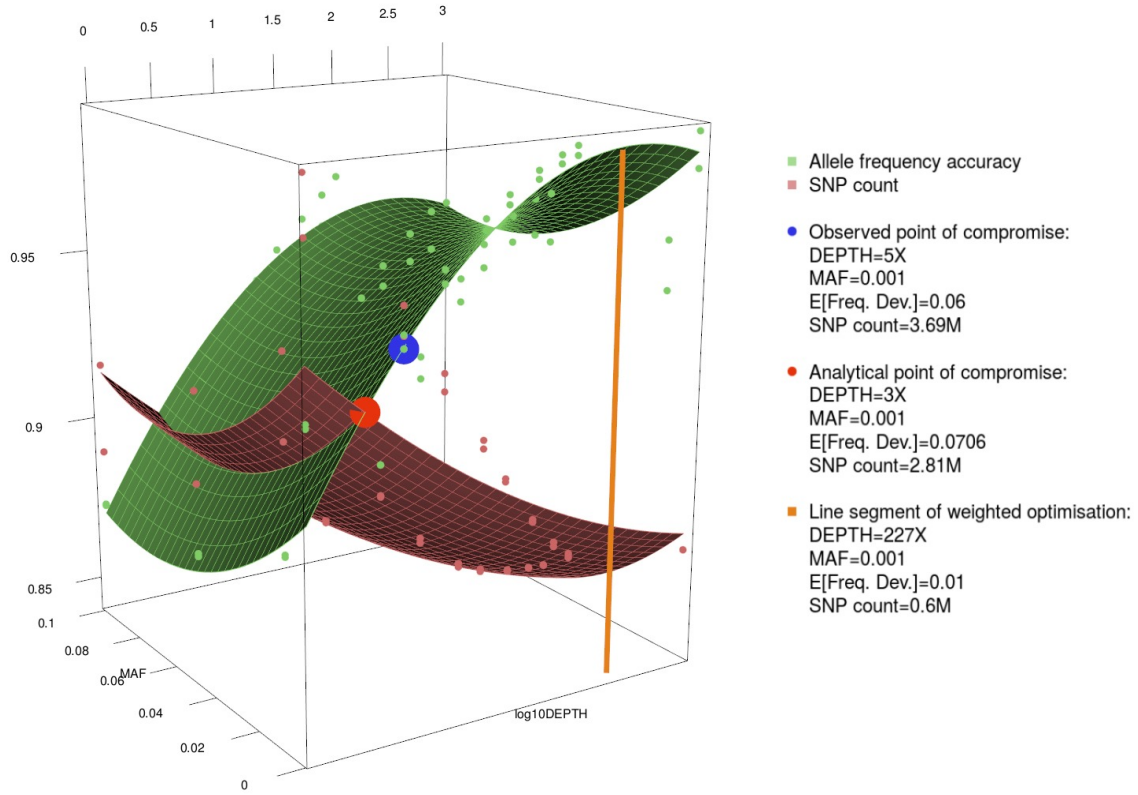


Figure 4.7. Optimising the compromise between allele frequency accuracy and number of SNPs as functions of minimum depth and minimum allele frequency (DEPTH: minimum sequence depth, MAF: minimum allele frequency, E[Freq. Dev.]: expected deviation from actual allele frequency, SNP count: number of SNPs).

The minimum depth of 227X and 0.001 MAF which resulted in the highest allele frequency accuracy while retaining 600,000 SNPs, were selected as the filtering parameter threshold values.

4.4. Discussion

To map the genetic basis of herbicide resistance using cost-effective Pool-GWAS in natural populations, and study the development of herbicide resistance *in vivo* using landscape genomics; we studied pool sequencing in tandem with double digest restriction site-associated DNA complexity reduction or Pool-ddRADseq. We have investigated the power of Pool-ddRADseq to capture accurate and genome-wide allele frequency data in *L. rigidum* populations. Allele frequency accuracy is affected by depth of coverage, and minimum allele frequency (MAF) thresholds. On the other hand, the number of SNPs required to saturate the genome is not only affected by depth, and MAF thresholds; but also by the breadth of coverage, genome size and linkage disequilibrium (LD) block size. The resulting large genome-wide SNP dataset was filtered using minimum depth, and minimum allele frequency thresholds in order to retain only the SNPs which were predicted to generate highly accurate allele frequency estimates while retaining enough SNPs to saturate the genome.

4.4.1. Genome complexity reduction for high and unbiased genome coverage

Genome complexity reduction through ddRADseq aims to find the best pair of restriction enzymes and fragment sizes to maximise genome coverage and minimise coverage heterogeneity. *L. rigidum* added an additional layer of complexity to this problem, because of the absence of a reference genome. However, the *L. perenne* draft genome (Byrne et al. 2015) was shown to be suitable as a reference genome for genotyping *L. rigidum* populations through Pool-ddRADseq, enabling the alignment of at least 95% of the *L. rigidum* sequences generated using Illumina technology. This is because these two *Lolium* species are closely related, are cross-compatible (Naylor 1960; Terrell 1966; Kloot 1983), and have a low degree of genetic differentiation based on Nei's genetic distance and F_{ST} (Guan et al. 2017). However, for more fine-detailed genomic characterisation i.e. comparative genomic and structural variation analyses of the *Lolium* genus, a species-specific reference genome would be required.

In terms of restriction enzyme pairs, results from ddRADseq simulations were consistent with the results from previous empirical and simulation studies. MseI is a frequent cutter, while NsiI, PciI and BclI are rare cutters. Previous studies have shown that pairs of restriction enzymes composed of a rare cutter in ddRADseq (e.g. EcoRI-G*AATTC and PstI-CTGCA*G) and a frequent cutter (e.g. MseI-T*TAA and MspI-C*CGG) resulted in high genome coverage across many different plant species (Peterson et al. 2012; Yang et al. 2016). Comparisons between ddRADseq and RNAseq for genome complexity reduction in a relatively recent study showed that in bees, ddRADseq generated to 3×10^7 reads and ~25,000 SNPs (15X depth threshold) which may be insufficiently dense for mapping large outbreeding populations; while RNAseq generated 3.6 times more SNPs (2×10^8 and ~90,000 SNPs) and better depth (66X) (Lozier et al. 2016). However, their methodology did not account for the suitability of other restriction enzyme pairs and the inflated bias and under-representation of non-transcribed genomic regions from RNAseq.

In terms of DNA fragment size selection, high sequencing coverage was observed by selecting the largest range of fragment sizes. This suggests that there may be no need for size selection as performed in the original RAD. However, the absence of size selection can also increase the heterogeneity in sequencing coverage (Lowry et al. 2017) and negatively affect the accuracy of allele frequency estimates. In addition, library preparation for the Illumina HiSeq X system recommends median fragment insert size range of 350 to 450 bp for 150 bp paired-end reads for efficient library clustering and sequencing (Illumina 2016).

We have validated the performance of MseI-NsiI restriction enzyme pair to generate high sequencing coverage using MinION nanopore sequencing. The MinION sequencing system enabled in-house sequencing experiments for rapid sequence returns, analyses and optimisations. The considerably higher sequencing error rate of 5% to 15% (Rang et al. 2018) compared with short-read Illumina sequencing of 0.24% (Pfeiffer et al. 2018) did not affect the accuracy of assessing the breadth and depth of coverage as functions of the restriction enzyme

pairs tested. This however prevented testing the accuracy of allele frequency estimates.

Applying the results of the ddRADseq simulation and MinION sequencing to the full-scale Pool-ddRADseq Illumina sequencing experiment, showed plateauing of B_{50} (50^{th} percentile of mean breadth of coverage). This means that increasing the number of reads does not further increase the fraction of the genome covered. This confirms that ddRADseq using the MseI-NsiI pair followed by fragment size selection is effective at reducing the complexity of the genome. Also, D_{50} (50^{th} percentile of mean depth of coverage) increased linearly as the number of sequencing reads increased. This is because additional reads simply pile-up to sites previously covered. The observed high variability in breadth and depth of coverage may be due to technical issues during library preparation. This includes variation in starting DNA quantity (Cumer et al. 2018), DNA fragment length distribution (Peterson et al. 2012), and amplification bias during PCR enrichment (Cumer et al. 2018; Davey et al. 2013). This variability in coverage can be reduced by performing more precise size selection and bead purification, which should also reduce PCR enrichment bias for smaller fragments. In addition, this variability is tolerable as long as it is possible to extract a genome-saturating number of loci with acceptable depths, base qualities and minimum allele frequencies to estimate accurate allele frequencies.

4.4.2. SNP filtering to retain a genome-saturating number of SNPs with accurate allele frequency estimates

The minimum SNP depth should be low enough to retain a large number of SNPs to saturate the genome evenly, but not too low so that allele frequency estimates are accurate. On the other hand, the MAF threshold should be high enough to retain loci with accurate allele frequencies, but not too high so that enough SNPs are retained to saturate the genome.

Biased representation of individuals, sequencing error rate, and the sequencing depth are the factors affecting the accuracy of allele frequency estimates. It is adversely affected by unequal representation of individual genotypes in Pool-seq experiments when pool sizes are small. But this is negligible for large pool sizes (Futschik and Schlötterer 2010; Zhu et al. 2012; Gautier et al. 2013), e.g. at least 25 individuals as suggested by Gautier et al (2013) and Fracassetti et al (2015) or at least 40 individuals as suggested by Schlötterer et al (2014). In addition, a recent study showed that estimating allele frequency without individual genotype calling, improved allele frequency accuracy (Warmuth and Ellegren 2019). These justified our method by which DNA samples were extracted from pools where pool sizes ranged from 66 to 95 individual plants. Although Illumina sequencing errors rates are typically below 0.001 (Kwon et al. 2013; Goodwin et al. 2016) read quality decreases towards the 3' end (Dohm et al. 2008) and for paired-end sequencing, the second read has lower than average quality with error rates increasing as fragment length increases (Tan et al. 2019). The low quality reads can be filtered out immediately after sequencing.

Genome and LD block sizes dictate the minimum number of DNA markers required to saturate the genome. We have approximated that a range of 0.7M to 4M SNPs is required to saturate the *L. rigidum* genome. A relatively recent study concluded that, given that LD block sizes for many species are orders of magnitude less than the typical density of RADseq markers, it is likely that many relevant loci may be missed (Lowry et al. 2017). However, this study failed to account for improvements in the breadth of coverage when using the best enzyme pair for ddRADseq. In addition, the breadth of coverage of Pool-ddRADseq can be improved by using more than one pair of RE and merging the sequencing results. The resulting redundant SNPs can be used as technical replicates to assess allele frequency accuracy while the non-redundant set of SNPs will improve the breadth of coverage.

Maximising the accuracy of allele frequency estimates and the number of available SNPs simultaneously is non-trivial. The depth and MAF filtering parameters as well as the corresponding expected accuracy and SNP count were different depending on which of the three optimisation methods was used. The *observed point of compromise* does not rely on any analytical model. This method accepts that the observed data perfectly captures the underlying relationships between accuracy, SNP count, depth and MAF and requires no interpolation. Filtering SNPs on other datasets using the depth and MAF defined by this point may not result in the expected accuracy and SNP count. On the other hand, the *analytical point of compromise*

can be thought of as assigning equal weights to both accuracy and SNP counts as a result of the direct superimposition of the accuracy and SNP count axes onto each other. Whereas, the *line segment of weighted optimisation* assigned more weight to accuracy compared with SNP count in order to have better allele frequency estimates. This increased the minimum depth by orders of magnitude but shrunk SNP count considerably. The set of parameters resulting from the weighted optimisation is preferred because it generated the lowest expected allele frequency deviation; however this frequency deviation (E[Freq. Dev.]=0.01) is 10 times larger than MAF (MAF=0.001). In addition, the number of SNPs retained is just below the theoretical genome-saturating SNP count which ranges from 0.7M to 4M SNPs. The simulation and empirical Pool-ddRADseq results showed that the MAF threshold was inversely proportional to accuracy, which suggests that the MAF threshold can go as low as $\frac{1}{2 \times poolsize}$pending that base calling accuracy remains high. Also, increasing the number of sequencing reads should increase depth and boost the number of SNPs available after filtering using a 227X minimum depth and 0.001 MAF thresholds. This can be accomplished by reducing the number of libraries pooled for each sequencing experiment or by using higher throughput sequencers, e.g. Illumina Novaseq 6000 system.

SNP filtering aims to retain informative loci at sufficiently high density to represent the genome. For the specific application of Pool-ddRADseq on Pool-GWAS with 5 pools, 100 individuals per pool and Bonferroni threshold at 5%; the *optimum point of compromise* at a minimum depth of 8X and the separately calculated MAF of 0.011 are preferred. This high MAF threshold should exclude loci with insufficient power to detect QTL and make the association analysis more time-efficient. In Pool-GWAS, the focus is to filter-out loci with allele frequencies below the theoretical minimum number of copies of a causal allele in a pool to pass the Bonferroni threshold. Alternatively, for population genetics studies, we recommend filtering the SNPs by 227X minimum depth and 0.001 MAF to maximise allele frequency accuracy. In population genetics studies, e.g. landscape genomics, allele frequency accuracy is of main importance. The validity of interpretations of population genetics summary statistics e.g. Fst, Watterson's estimator, and Tajima's D, rely on accurate estimates of allele frequency.

4.5. Conclusion

In conclusion, a genome complexity reduction-based pool sequencing (Pool-ddRADseq) methodology was developed to generate high quality allele frequency data that saturates the *Lolium rigidum* genome. The restriction enzyme pair MseI-NsiI was shown to result in high sequencing breadth and depth of coverage *in silico*, confirmed with nanopore sequencing and applied in Illumina high-throughput sequencing. The *Lolium perenne* draft genome proved to be a satisfactory reference genome for mapping *L. rigidum* sequences. The substantial variation in breadth and depth of sequencing coverage across the genome and across samples were offset by SNP filtering. SNP filtering parameter thresholds of 227X minimum depth and 0.001 MAF were predicted to result in accurate allele frequency estimates and a genome-saturating number of 600,000 SNPs amenable to quantitative and population genetics studies.

Chapter 5

Evolutionary quantitative genetics of herbicide resistance in weeds

5. Evolutionary quantitative genetics of herbicide resistance in weeds

5.1. Introduction

Convergent evolution in response to the same environmental challenges implies either a shared ancestry or a constrained evolutionary response due to the narrow biochemical basis of fitness. The repeated rapid evolution of herbicide resistance in weeds is an exemplary illustration of convergent evolution (Baucom 2019) but it is not clear whether this proceeds from shared standing genetic variation between populations, or from the limited mutational possibilities to endow resistance. Are herbicide resistance traits controlled by a few large-effect loci (monogenic or oligogenic) or by many loci with small effects (polygenic)? And are the corresponding resistance-conferring alleles present prior to the selection pressure from herbicide as standing genetic variation or are these the product of recent *de novo* mutations?

Addressing these questions is the key to understanding the evolution of herbicide resistance at the whole species level. Many herbicides are designed to target a specific biochemical pathway shared by most plant species. Additionally, stress response pathways like reactive oxygen species scavenging are shared across all plant species (Breusegem et al. 2008). Due to these commonalities, the solution to the stress induced by an herbicide is likely to be similar across different species. The genetic bases of herbicide resistance can be elucidated using genome-wide association studies (GWAS), a powerful gene mapping approach to detect putative quantitative trait loci (QTL). Herbicide resistance QTL can then be tracked across multiple populations with varying levels of resistance to infer the most likely evolutionary path taken by the different weed populations. Improved understanding of the genetic basis of herbicide resistance will also enable rapid monitoring of resistance in the field using genomic prediction (GP). Together with the knowledge of how herbicide resistance evolves, GWAS and GP will pave the way towards improved herbicide resistance control strategies.

Most studies which aimed to elucidate the genetic basis of herbicide resistance focused on target-site resistance (TSR) genes, whose effects are assumed *a priori*. However, TSR genes have been insufficient to explain the vast majority of the observed variation in herbicide resistance (Beckie and Tardif 2012; Délye et al. 2013; Délye 2013; Shaner and Beckie 2014). Many experiments have failed to recapture known TSR mutations and have in turn revealed an abundance of non-target-site resistance (NTSR) genes (Yu et al. 2009; Délye et al. 2011; Yu et al. 2013; Scarabel et al. 2015; Han et al. 2016; Rey-Caballero et al. 2017; Bai et al. 2019). NTSR genes include cytochrome P450, glutathione S-transferase, and glycosyltransferase gene families which are integral to the breakdown of herbicide molecules (Yuan et al. 2007). These NTSR genes can affect a broad range of herbicides with different modes of action and therefore can promote cross-resistance, i.e. cytochrome P450 conferring resistance to multiple modes of action (Han et al. 2020). Additionally, most of the previous studies on uncovering the genetic basis of herbicide resistance are under-powered because only two phenotype groups were involved, the susceptible and the resistant without consideration for the intermediate phenotypes. Therefore there is a need to perform a quantitative, genome-wide approach without assuming TSR. Knowledge of the genetic architecture of herbicide resistance traits and the identity of the resistance-associated loci can be used to infer the likely evolutionary path to herbicide resistance.

Herbicide resistance-conferring alleles may arise from standing genetic variation, or from *de novo* mutations, and may have been distributed across the landscape through gene flow, and pushed towards fixation by selection pressure from herbicides and random genetic drift. These evolutionary forces leave behind genomic signatures which can be detected. The coalescent theory (Kingman 1982) offers a coherent statistical framework to analyse these genomic signatures (Rosenberg and Nordborg 2002; Lee and Coop 2017). Variation in neutral allele frequency decreases as we get closer to the beneficial allele, this loss in variation is more pronounced if the beneficial allele is recent (i.e. product of new mutation) since less recombinations have the chance to occur compared with older alleles (Hermisson and Pennings 2005; Maynard and Haigh 2007; Boitard et al. 2009; Pritchard et al. 2010). If standing genetic

variation was the origin of the resistance alleles, then we expect soft sweeps and low fitness cost of the resistance alleles. If on the other hand, resistance-conferring mutations arose independently across populations and in parallel with the selection pressure from herbicides, we expect high effective population sizes and high genetic differentiation between resistant populations, low gene flow and hard sweeps. Furthermore, if gene flow is high then there is an increased chance of a single or few source populations from which resistance alleles originated.

The rapid evolution of herbicide resistance is more consistent with standing genetic variation as the source of resistance alleles rather than *de novo* mutations. Resistance to the first widely used herbicide 2,4-D, took 13 years to evolve (Hilton 1957). Triazine resistance in common groundsel (*Senecio vulgaris*) evolved within 10 years (Ryan 1970). Glyphosate resistance in *L. rigidum* evolved within 15 years (Powles et al. 1998). Assuming a high mutation rate of $1x10^{-7}$ per genome per generation with a high probability of 1% of all mutations will confer resistance, and a large effective population size of $5x10^{5}$; then at least 2000 generations are required to give rise to one resistance mutation (1/2000 chance that a population will acquire a resistance allele per generation). This is far from the observed 10 to 15 years it took to evolve herbicide resistance in the field. Empirical evidence for standing genetic variation as the source of resistance alleles in field populations is scant. In field populations of the weed, *Ipomoea purpurea* and the fruit fly, *Drosophila melanogaster*, herbicide and insecticide resistances, respectively, were found to be polygenic (Fournier-Level et al. 2019; Etten et al. 2020). These polygenic resistance traits support standing genetic variation as the primary source of resistance alleles. This is because adaptive evolution from standing genetic variation leads to the fixation of more alleles with small effects, i.e. polygenic adaptive traits (Barrett and Schluter 2008). In Canadian populations of the broadleaf weed, *Amaranthus tuberculatus*, standing genetic variation was the source of resistance alleles in introduced populations from the US; while in recently established agricultural populations new mutations are the source of resistance (Kreiner et al. 2018).

Insights that will be gained attempting to answer the following questions will improve our basic understanding of adaptive evolution, and guide the development of sustainable weed and herbicide resistance management strategies. What are the genetic and evolutionary bases of herbicide resistance in *Lolium rigidum* from South-East Australia? Are the genetic architectures truly polygenic or do major genes dominate? Are the resistance alleles shared across populations? What is the predominant source of herbicide resistance alleles, standing genetic variation or *de novo* mutations?

5.2. Materials and Methods

5.2.1. Plant material

We used the 67 *L. rigidum* populations sampled across SE Australia listed in Table 5.1 (see Chapter 3, Sub-section 3.3.1, Table 3.6 for details). The intra-population datasets were used in genome-wide association studies using GWAlpha (Fournier-Level, Robin, and Balding 2017). This included the Inverleigh and Urana populations, as well as the eight maximally variable populations with survival rates of 54% (ACC13) and 10% (ACC49) for clethodim, 26% (ACC31) and 75% (ACC62) for glyphosate, 46% (ACC11) and 41% (ACC59) for sulfometuron, and 50% (ACC09) and 32% (ACC54) for terbuthylazine. The merged intra- and inter-population datasets were used to build the genomic prediction and association studies (GPAS) models. Specific subsets of the November 2018 collection were used to infer the likely models of herbicide resistance evolution (Table 5.2).

Table 5.1. List populations used in the genomic analysis of herbicide resistance of *Lolium rigidum* across South-East Australia, and the corresponding herbicide resistance traits measured (details of the populations are found in Chapter 3 Table 3.6).

Dataset	Populations	Herbicide Resistance Traits*				
		C	G	S	T	F
Intra-population	Inverleigh		G	S	T	F
Intra-population	Urana		G	S	T	F
Intra-population	ACC09				T	
Intra-population	ACC11			S		
Intra-population	ACC13	C				
Intra-population	ACC31		G			
Intra-population	ACC49	C				
Intra-population	ACC54				T	
Intra-population	ACC59			S		
Intra-population	ACC62		G			
Merged-Clethodim	60 SE Australian populations (November 2018 collection), including 5 pools each from ACC13, and ACC49	C				
Merged-Glyphosate	60 SE Australian populations (November 2018 collection), including 5 pools each from ACC31, and ACC62		G			
Merged-Sulfometuron	47 SE Australian populations (November 2018 collection), including 5 pools each from ACC11, and ACC59			S		
Merged-Terbuthylazine	26 SE Australian populations (November 2018 collection), including 5 pools each from ACC09, and ACC54				T	
Live collection	Lexton, Rokewood, and Skipton			S		

* - Resistance to C for clethodim, G for glyphosate, S for sulfometuron, T for terbuthylazine, and F for trifluralin

5.2.2. Phenotyping and genotyping

Two types of continuous quantitative phenotype data were measured. The first was the instaGraminoid-derived individual plant resistance for the intra-population datasets. The second was the population or pool survival rate. For the intra-population datasets, unreplicated individual-level resistances were measured following the procedure described in Paril and Fournier-Level (2019) or in Chapter 3 Section 3.2. For the inter-population datasets, population-level survival rates were measured, and the experimental designs were unbalanced with some populations replicated more than others.

Genome-wide allele frequency data were obtained using Pool-ddRADseq as described in Chapter 4; however, DNeasy® Plant Mini Kit was used to conveniently extract the DNA instead of the CTAB protocol. Additionally, the minimum read depth was relaxed from 227X to the published recommendation of 50X (Schlötterer et al. 2014). Read depth was very variable across libraries and resulted in very low numbers of loci under the very stringent threshold of 227X determined in Chapter 4. Genotype data were extracted from at least 72 individuals per pool per population for the intra-population dataset and at least 21 individuals per population for the inter-population dataset. For the intra-population dataset, each population was divided into 5

pools arranged in order of increasing individual quantitative herbicide resistance levels, where pool 1 has the lowest resistance and pool 5 has the highest resistance.

5.2.3. GWAS

GWAS or QTL detection experiments were performed using univariate (iterative or one SNP at a time) model, and multiple regression (simultaneous allelic effects estimation) models. GWAlpha (Fournier-Level et al. 2017) was used to fit the univariate models. Elastic-net regularisation (Friedman et al. 2010) and linear mixed modelling (Henderson 1953; Henderson 1975) were used in multiple regression. We used the glmnet package for elastic-net regularised regression (Friedman et al. 2010), where the elastic-net penalty was set to 0 and 1, for ridge regression (RR) and least absolute shrinkage and selection operator (LASSO) models, respectively. The tuning parameter, λ was heuristically determined as the value that minimized the cross-validation error. The mixed model (MIXED-REML) can be expressed as,

$$y = X\beta + Z\mu + \varepsilon,$$

with

$$\mu \sim N(0, \sigma_\mu^2 I) \text{ and } \varepsilon \sim N(0, \sigma_\varepsilon^2 I),$$

where X is the $n \times p$ matrix of allele frequencies for n pools and p predictor alleles across loci, $\beta$is the corresponding vector of fixed effects of the alleles, Z is the $n \times n$ matrix of pairwise F_{ST} computed using the method-of-moments estimation from Pool-seq data described by Hivert et al (2018), μ is the vector of random effects of F_{ST}, ε is the vector of random residual effects, μ follows a multivariate normal distribution with a vector of zero means and a variance-covariance matrix $\sigma_\mu^2 I$, and ε also follows a multivariate normal distribution with a vector of zero means and a variance-covariance matrix $\sigma_\varepsilon^2 I$. The fixed effects were estimated using the less computationally taxing least squares solution, $\hat{\beta} = X^T (X X^T)^{-1} y$ instead of the usual $\hat{\beta} = (X^T X)^{-1} X^T y$. The variances ($\sigma_\mu^2$ and σ_ε^2) were estimated using restricted maximum likelihood (REML). Computations were performed using R statistical software (R Core Team 2018) and Julia (Zappa Nardelli et al. 2018).

GWAS were performed using intra-population datasets and merged intra- and inter-population datasets. Intra-population datasets were used in the univariate model: GWAlpha. Merged intra- and inter-population datasets were used in the linear-mixed model: MIXED-REML. The Bonferroni threshold was set at α=0.01/p. Manhattan plots were generated and GWAS peaks or the herbicide resistance QTL were identified.

5.2.4. Gene ontology analysis

Genes based on the *Lolium perenne* draft genome (Byrne et al. 2015) within ±1kb of the Pool-ddRADseq-captured loci and GWAS-detected QTL were mined for functional annotations based on orthology to the *Oryza sativa* gene sets (ENSEMBLE Oryza_sativa.IRGSP-1.0). Orthology was determined using BLAST (McGinnis and Madden 2004). BLAST hits with a minimum identity of 90%, and minimum query coverage of 1% were retained. The *O. sativa* genes corresponding to these BLAST hits were used in gene ontology (GO) term enrichment analysis. The web application, ShinyGO (Ge et al. 2020) was used to extract the GO terms and *Oryza sativa* Japonica group was the species selected. Using the GO terms from the Pool-ddRADseq loci as the null expectations, the GO terms associated with the GWAS peaks were tested for significant enrichment using Fisher's exact test (Fisher 1922).

Additionally, the candidate genes were used to identify the shared genetic bases of herbicide resistance between populations. The genes shared between at least two herbicide resistance traits were identified as candidate cross-resistance genes. Gene list overlaps were visualised using Venn diagrams implemented using the R package VennDiagram (Chen and Boutros 2011).

5.2.5. Composite likelihood of convergent evolutionary models

To test if the source of favourable alleles was primarily standing genetic variation or *de novo* mutations, we used the inference method of Lee and Coop (2017) based on a coalescent framework. This required estimates of the effective population sizes and recombination rates. The effective population sizes (N_e) were estimated using Tajima's π, where $\pi = 4 N_e \mu$. Poopolation (Kofler et al. 2011) was used to compute Tajima's π using a set of 16 neutral scaffolds with maximum -$\log_{10}$(p-value) < 1, across loci and herbicide resistance traits (Appendix Table S5.1). The mutation rate μ, was assumed to be a conservative 1.0 x 10^{-9} mutations per individual genome per generation (Kovalchuk et al. 2000; Wang et al. 2019). This mutation rate was used only in estimating the effective population size. The scaffold-specific recombination rates were estimated using the putatively non-neutral herbicide resistance-associated scaffolds in Table 5.2. Linkage disequilibrium (LD) was computed using LDx (Feder et al. 2012) based on individual sequencing read through the maximum likelihood approach (Clarke and Cardon 2005):

$$argmax_r \left[\prod_{i=1}^{2} \prod_{j=1}^{2} \left(\hat{p}_{ij} - \hat{D}\, r \right)^{n_{ij}} \right],$$

where *r* is the recombination rate between two biallelic loci, $\hat{p}_{ij}$is the estimated frequency of the *ij* read-haplotype, $n_{ij}$is the observed number of *ij* read-haplotypes, and $\hat{D} = \hat{p}_{11}\hat{p}_{22} - \hat{p}_{12}\hat{p}_{21}$. The per base-pair recombination rate is estimated as $r_{bp} = r / l$, where l is the distance in base-pairs between the two loci.

The composite likelihoods of the following models were calculated:

- model 1 (independent mutation model): the resistant populations acquired the resistance allele through independent *de novo* mutations after the onset of herbicide application,
- model 2 (gene flow model): the resistant populations share the resistance allele through gene flow,
- model 3 (standing variation model): the resistant populations share the resistance allele from a common ancestor,
- model 4 (gene flow and independent mutation model): some of the resistant populations share the resistance allele through gene flow; while the other population independently acquired the resistance allele through mutation, and
- model 5 (standing variation and independent mutation model): some of the resistant populations share the resistance allele from a common ancestor; while the other population independently acquired the resistance allele through mutation.

The probabilities of these evolutionary models were assessed using a coalescence framework implemented in R which is available in the github repository, dmc (https://github.com/kristinmlee/dmc/) (Lee and Coop 2017). The null hypothesis or the neutral expectations which the five models described above were compared with, were simulated using Hudson's program for generating samples under neutral models, ms (Hudson 2002). The selection coefficient, age of standing genetic variant, initial allele frequency, and migration rate were determined using maximum likelihood estimation. The search space for selection coefficient ranged from 0.0001 to 0.6. The search space for the age of the standing genetic variant, i.e. favourable allele was composed of 0, 5, 25, 100, 500, 1000, 1x10^4, and 1x10^6 generations. The search space for the initial frequency of this allele consisted of 1/2N_e, 1x10^{-5}, 1x10^{-4}, 1x10^{-3}, and 1x10^{-2}. Finally, the search space for migration rate was composed of 1x10^{-5}, 1x10^{-3}, 1x10^{-1}, 0.5, and 1 migrants per generation.

Under the dmc analysis framework, we assume that the resistance alleles were recently fixed in the resistant populations. Additionally, we assume that only the resistant populations are under herbicide selective pressure. For each herbicide resistance trait, two sets of populations (each set is composed of 3 susceptible and 3 resistant populations), and one scaffold with at least one herbicide resistance QTL were selected for dmc analysis (Table 5.2). We selected the scaffolds where the QTL were identified in both GWAS and dmc analysis. Since most of the resistant

populations we included in Table 5.2 were not completely resistant or completely susceptible, then the dmc models were under-powered. Despite this, the control of false positive rate remained high which effectively ruled-out the possibility of increased type I error stemming from an under-powered design.

Table 5.2. List of putatively non-neutral scaffolds for each herbicide resistance trait and the sets of populations used to estimate the likelihoods of different coalescent models of convergent evolution.

Herbicide	Scaffold	Susceptible Populations (Resistance)	Resistant Populations[c] (Resistance)
Clethodim	scaffold_3626|ref0003919	ACC01,ACC05,ACC06 (0%,0%,0%)	[ACC27,ACC34] & ACC36 (65%,61%,73%)
		ACC06,ACC38,ACC50 (0%,0%,0%)	[ACC08,ACC09],ACC11 (97%,59%,100%)
Glyphosate	scaffold_198|ref0030880[a]	ACC01,ACC02,ACC15 (0%,7%,1%)	ACC18,[ACC47,ACC50] (81%,90%,84%)
		ACC38,ACC39,ACC55 (0%,0%,0%)	[ACC04,ACC09],ACC24 (88%,93%,94%)
Sulfometuron	scaffold_6912|ref0027973[b]	ACC01,ACC19,ACC21 (0%,0%,0%)	ACC14,[ACC51,ACC53] (52%,56%,52%)
		ACC03,ACC20,ACC57 (0%,0%,0%)	[ACC11,ACC12],ACC53 (46%,81%,52%)
Terbuthylazine	scaffold_6912|ref0027973[b]	ACC07,ACC21,ACC26 (2%,0%,5%)	[ACC04,ACC06],ACC55 (45%,65%,86%)
		ACC07,ACC21,ACC26 (2%,0%,5%)	ACC09,[ACC60,ACC62] (50%,92%,42%)

[a] Associated with Os04g0115650 or Bph33 - brown planthopper resistance gene
[b] Associated with Os10g0480500 or DRB6 - double-stranded RNA-binding protein 6
[c] Populations inside the brackets share the resistance alleles through gene flow or through a common ancestor depending on the model being tested.

The scripts used in the composite likelihood estimation, as well as the log-likelihood profiles and maximum likelihood estimates of the population genetics parameters for other scaffolds not listed in Table 5.2 are available in https://gitlab.com/jeffersonfparil/popgen-models.

5.2.6. GP accuracy

The insights into how herbicide resistance evolves across the landscape need to be complemented with efficient and timely monitoring techniques to effectively manage its negative impacts on agriculture. Genomic prediction (GP) using the same statistical framework as GWAS, can be used to rapidly monitor herbicide resistance without the need for time- and resource-intensive herbicide resistance assays.

GP models were built using merged intra- and inter-population datasets. There were four models used. The first one is the univariate model, i.e. GWAlpha (Fournier-Level et al. 2017). The next three are multiple linear regression models: RR, LASSO, and MIXED-REML. The model fitting procedure for GWAlpha was the same as in GWAS and described in detail in Fournier-Level et al. (2017). The estimation of the predicted phenotypes for this model is two-fold. First, the polygenic scores of the training set (S_{train}) were calculated as:

$$s_{train} = X_{train}\beta,$$

where X_{train} is the $n \times p$ matrix of allele frequencies for n pools and p predictor alleles across loci, $\beta$is the vector of GWApha-derived allele effects, both are from the training dataset. These polygenic scores were regressed against the observed phenotypes of the training set (y_{train}),

$$y_{train} = \alpha_0 + \alpha_1 s_{train},$$

where α_0 is the intercept, and $\alpha_1$is the slope. Finally, the polygenic scores of the validation set,

$$s_{valid} = X_{valid}\beta,$$

were transformed into the predicted phenotype values ($y_{predicted}$) using the linear function,

$$y_{predicted} = \alpha_0 + \alpha_1 s_{valid}.$$

The model fitting procedures for the RR, LASSO and MIXED-REML were the same as described for GWAS. Phenotype predictions ($y_{predicted}$) were computed as,

$$y_{predicted} = X_{valid}\beta,$$

where X_{valid} is the matrix of allele frequencies of the validation set, and $\beta$is the vector of allele effects estimated using the training set.

Genomic prediction accuracy was measured using the root mean square error between observed and predicted phenotype values,

$$RMSE = \sqrt{\frac{\sum (y_{observed} - y_{predicted})^2}{n}},$$

where $y_{observed}$is the observed herbicide resistance, $y_{predicted}$ is the predicted herbicide resistance, and n is the number of pools or populations.

To assess the prediction accuracies of RR, LASSO and MIXED-REML, we used the merged inter- and intra-population datasets. The phenotype data used were the survival rates instead of the instaGraminoid-derived phenotype data. Leave-one-out cross-validation was performed instead of 10-fold cross validation due to the small number of observations available which ranged from 46 to 77 populations and pools.

To assess the feasibility of GP for the rapid monitoring of herbicide resistance in the field, we predicted the sulfometuron resistance of five populations collected as live plants from Lexton, Rokewood and Skipton in Victoria. We used the instaGraminoid-derived phenotype data and the GWAlpha-derived allele effects. We trained the Pool-GPAS models on the populations from Inverleigh, Victoria and Urana, New South Wales.

5.3. Results

5.3.1. GWAS and gene ontology enrichment analysis

GWAS identified a total of 2,763 SNPs associated with the four herbicides tested (Appendix Figures S5.1 to S5.18). For each herbicide resistance trait, the number of associations above the Bonferroni threshold ranged from 482 to 1,011, suggesting the predominance of loci of small effects and only a few loci with large effects as illustrated in Figure 5.1. There is no explicit approach to compute the amount of variance explained by each QTL detected by GWAlpha; however, the top 10% of the identified QTL accounted for 10% to 18% of the cumulative QTL effects.

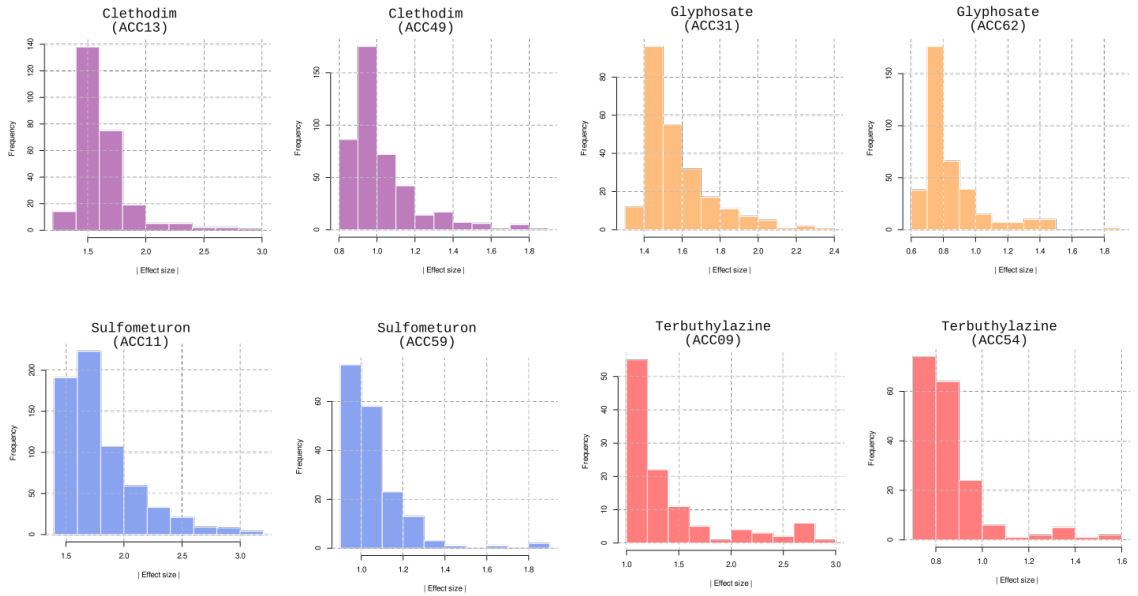


Figure 5.1. Distributions of the absolute values of the allele effects of loci above the Bonferroni threshold from 8 intra-population GWAS experiments.

Candidate gene lists were significantly enriched for stress-related response genes (Table 5.3). However, none of the resistance-associated loci were linked to target-site resistance (TSR) genes since our Pool-ddRADseq genotyping did not capture any of the TSR genes (i.e. ACCase gene for clethodim, EPSPS gene for glyphosate, ALS gene for sulfometuron, and psbA for terbuthylazine). Additionally, some of the significantly enriched GO terms are broad in scope, e.g. biological regulation and regulation of metabolic processes. Nonetheless, the total number of associated SNP loci outnumbered the set of linked and confidently annotated genes that were used in the analysis. Hence, there are potentially more genes that have not been included in the GO term enrichment analysis.

Table 5.3. Summary of GWAS results and gene ontology enrichment analysis.

Herbicide resistance trait	Population	Number of loci tested	Number of loci (and scaffolds) above the Bonferroni threshold at α=0.01/p**		Significantly enriched gene ontology terms (Fisher's exact test p-values and the total number of genes used in the analysis)
Clethodim Resistance	Merged*	46,408	605 (398)	1,011 (600)	Anatomical structure morphogenesis (0.0153), biosynthetic process (0.0680), & response to abiotic stimulus (0.0010) across 59 genes
	ACC13	115,141	160 (124)		
	ACC49	245,705	257 (144)		
Glyphosate Resistance	Merged*	11,765	57 (51)	482 (320)	Biological regulation (0.0275), regulation of growth (0.0191), response to stimulus (0.0598), & signalling (0.0527) across 25 genes
	ACC31	154,233	153 (112)		
	ACC62	118,129	212 (121)		
	Inverleigh	46,092	55 (47)		
	Urana	21,765	8 (8)		
Sulfometuron Resistance	Merged*	23,591	108 (79)	778 (424)	Cellular localisation (0.0560), & regulation of metabolic process (0.0958) across 28 genes
	ACC11	84,199	383 (195)		
	ACC59	95,001	106 (77)		
	Inverleigh	263,794	154 (94)		
	Urana	58,752	31 (24)		
Terbuthylazine Resistance	Merged*	20,174	261 (180)	602 (379)	Biological regulation (0.0377), cell wall organisation (0.0582), regulation of cellular process (0.0411), & regulation of metabolic process (0.0231) across 23 genes
	ACC09	124,778	63 (48)		
	ACC54	139,037	104 (74)		
	Inverleigh	33,759	104 (45)		
	Urana	107,323	83 (68)		

* - Merged intra- and inter-population datasets resulted in lower numbers of loci tested because each genotype data was a consensus across multiple datasets.
** - p is the number of loci

For each herbicide resistance trait, there were more genes specific to each population than shared between populations (Appendix Figures S5.19 to S5.22). We identified candidate cross-resistance genes (Figure 5.2 and Table 5.4). The candidate gene common to all four herbicide resistance traits was the double-stranded RNA-binding protein 6 (DRB6). It is involved in gene and chromatin silencing. Many of these candidate cross-resistance genes are poorly characterised and often not directly related to response to xenobiotics with the noticeable exception of 15 detoxification genes.

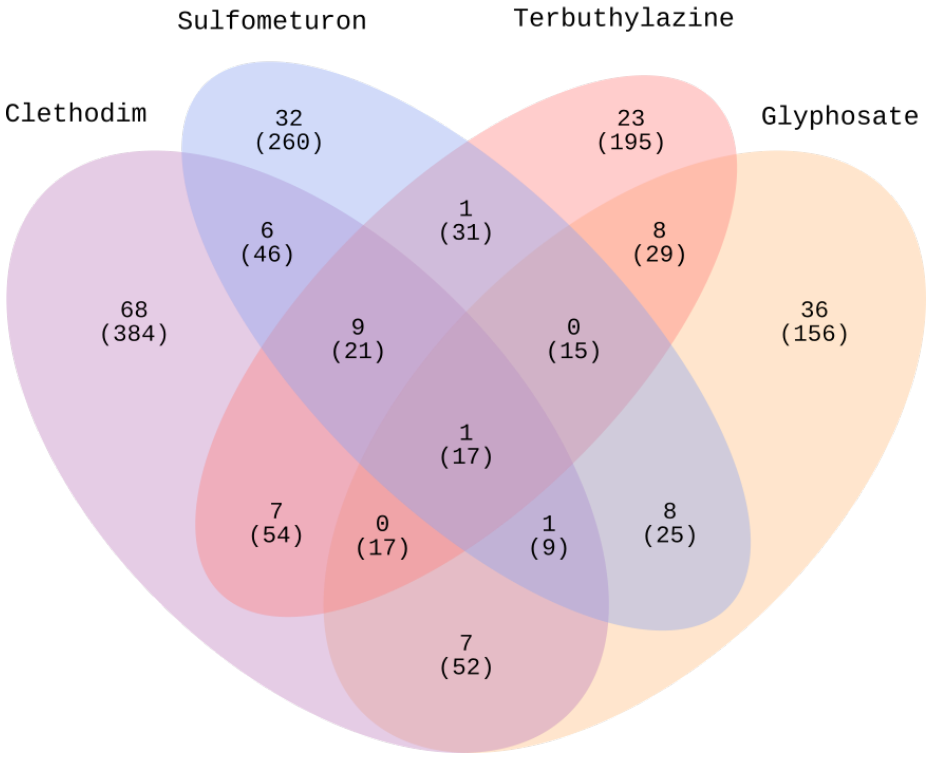


Figure 5.2. Overlap between candidate gene lists and scaffolds containing at least one herbicide resistance-associated locus. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.

Table 5.4. List of candidate cross-resistance genes with descriptions obtained from The UniProt Consortium (2019) (continued in the next two pages).

Herbicide Resistance Trait Combinations	*L. perenne* annotations (scaffold: positions)	*O. sativa* gene name	Description*
Clethodim - Glyphosate - Sulfometuron - Terbuthylazine	6912: 32663-38188	DRB6 (Os10g0480500)	Double-stranded RNA-binding protein 6
Clethodim - Glyphosate - Sulfometuron	11087: 10146-21116	Os03g0112101	Intracellular protein transport
Clethodim - Sulfometuron - Terbuthylazine	210: 221558-223725	ENSRNA049446581	-
	2843: 54929-58509	Os01g0340801	-
	210: 221558-223725	Os02g0810900	Transcription regulation
	2843: 60055-81855	Os03g0102333	Negative regulation of translation
	210: 221558-223725	Os04g0619000	Transcription regulation
	210: 221558-223725	Os08g0200600	Transcription regulation
	210: 221558-223725	Os08g0200750	-
	2843: 54929-58509	Os10g0415600	Acylamino-acid-releasing enzyme 1 (serine-type endopeptidase activity)
	210: 221558-223725	Os12g0610600	Transcription regulation
Clethodim - Glyphosate	1520: 122398-125786	Os02g0280400	Peptidyl-serine phosphorylation
	1573: 83618-95551	RAD54 (Os02g0762800)	DNA repair and recombination protein
	324: 64991-70219	Os03g0239000	Xyloglucan endotransglucosylase/hydrolase (cell wall organization)
	198: 98183-98971	Os04g0115600	-
	198: 98183-98971	Os04g0115650	-
	198: 98183-98971	Os04g0115800	-
	2865: 87064-89686	RER1 (Os06g0708300)	Retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum
Clethodim - Sulfometuron	820: 71681-77293	Os02g0266100	U1 small nuclear ribonucleoprotein C
	820: 71681-77293	Os05g0466900	Protein serine/threonine kinase activity
	2721: 63509-64198	Os07g0493800	Protein serine/threonine kinase activity

* - The absence of information regarding gene function is encoded as “-”.

Table 5.4. (continuation: 2 of 3).

Herbicide Resistance Trait Combinations	*L. perenne* annotations (scaffold: positions)	*O. sativa* gene name	Description*
Clethodim - Sulfometuron	1952: 41723-46059	Os07g0544900	Structural constituent of ribosome
	820: 71681-77293	Os12g0517800	-
	820: 71681-77293	Os12g0518000	Integral component of membrane
Clethodim - Terbuthylazine	4210: 66918-69208	Os03g0273800	Haloacid dehalogenase-like hydrolase family protein
	4210: 33363-36260	Os03g0372650	Integral component of membrane
	580: 129058-136122	Os03g0747800	Cysteine synthase (cysteine biosynthetic process from serine)
	185: 177310-179822	Os04g0581100	Dioxygenase activity
	580: 129058-136122	Os06g0602900	Oxidoreductase activity
	580: 129058-136122	Os06g0602950	-
	580: 129058-136122	Os11g0644000	ATP-binding
Glyphosate - Sulfometuron	167: 212566-229006	Os02g0224400	Aminopeptidase activity and manganese ion binding
	167: 175374-180716	Os02g0224800	Integral component of membrane
	167: 175374-180716	Os02g0224850	-
	348: 106683-111913	Os05g0196500	Lysine-specific demethylase
	1148: 71299-75643	Os05g0378800	Integral component of membrane
	167: 175374-180716	Os06g0563000	Methyltransferase (methylation)
	167: 175374-180716	Os06g0563125	-
	140: 211839-220334	Os10g0188100	Transcription initiation factor IIF subunit alpha
Glyphosate - Terbuthylazine	1711: 94584-96380	Os03g0332500	Structural constituent of ribosome
	4788: 59083-65420	HMA5 (Os04g0556000)	Copper-transporting ATPase (detoxification of copper ion)
	4788: 71909-74014	Os04g0556300	Glutathione peroxidase (response to oxidative stress)
	123: 204712-214195	Os06g0707100	Xenobiotic transmembrane transporter activity
	123: 27495-32392	Os06g0708000	Mitogen-activated protein kinase 12 (defence response)

* - The absence of information regarding gene function is encoded as “-”.

Table 5.4. (continuation: 3 of 3).

Herbicide Resistance Trait Combinations	*L. perenne* annotations (scaffold: positions)	*O. sativa* gene name	Description*
Glyphosate - Terbuthylazine	123: 27495-32392	Os06g0708075	-
	123: 198627-200788	CYP71P1 (Os12g0268000)	Tryptamine 5-hydroxylase (heme binding defence response)
	123: 198627-200788	Os12g0268100	-
Sulfometuron - Terbuthylazine	7864: 37646-40326	Os05g0553800	Chromatin assembly or disassembly

* - The absence of information regarding gene function is encoded as "-".

5.3.2. Composite likelihood of convergent evolutionary models

The analysis of the composite likelihoods suggested that standing genetic variation is the more likely source of resistance alleles than *de novo* mutations emerging independently in different populations (Figure 5.3). The most likely models for the two sets of resistant and susceptible populations were generally consistent across all herbicide resistance traits.

For clethodim, the observed variation in frequency for the resistance allele and for the resistance trait at the landscape-level were better explained by standing genetic variation than independent mutations as the source of the resistance alleles (Figure 5.3, Clethodim resistance panel). The standing genetic variation model (model 3) exhibited the highest peak near the resistance QTL for the first set of populations (Figure 5.3, Clethodim resistance panel, left plot). On the other hand, the standing genetic variation with an independent mutation in one population model (model 5) showed the highest peak near the putative QTL for the other set of populations (Figure 5.3, Clethodim resistance panel, right plot). These two models exhibited the highest mean composite log-likelihoods for both sets of populations. Additionally, despite having the estimated age of the standing genetic variant at zero generations for model 3 in Appendix Table S5.2 and models 3 and 5 in Appendix Table S5.3, the initial allele frequencies are too high, i.e. 0.01 and 0.1, to overlap with the mutation model (model 1).

For glyphosate, the relative composite log-likelihood profile was similar to that of clethodim (Figure 5.3, Glyphosate resistance panel). Standing genetic variation was the most likely source of the resistance allele. For the first set of populations, the age of the resistance allele was zero generation with an initial frequency of 0.1 which rules out model 1 (Appendix Table S5.4). For the other set of populations, the resistance allele was estimated to be 25 or 50 generations old for model 3 and model 5, respectively (Appendix Table S5.5).

For sulfometuron and terbuthylazine using scaffold_6912|ref0027973 harbouring the putative cross-resistance gene, DRB6, standing genetic variation was found to be the most likely source of the resistance alleles (Figure 5.3, Sulfometuron and Terbuthylazine resistance panels). Similar to the findings for clethodim and glyphosate, the zero generation-old resistance alleles at 0.1 initial frequencies ruled out model 1 (Appendix Tables S5.6 to S5.9).

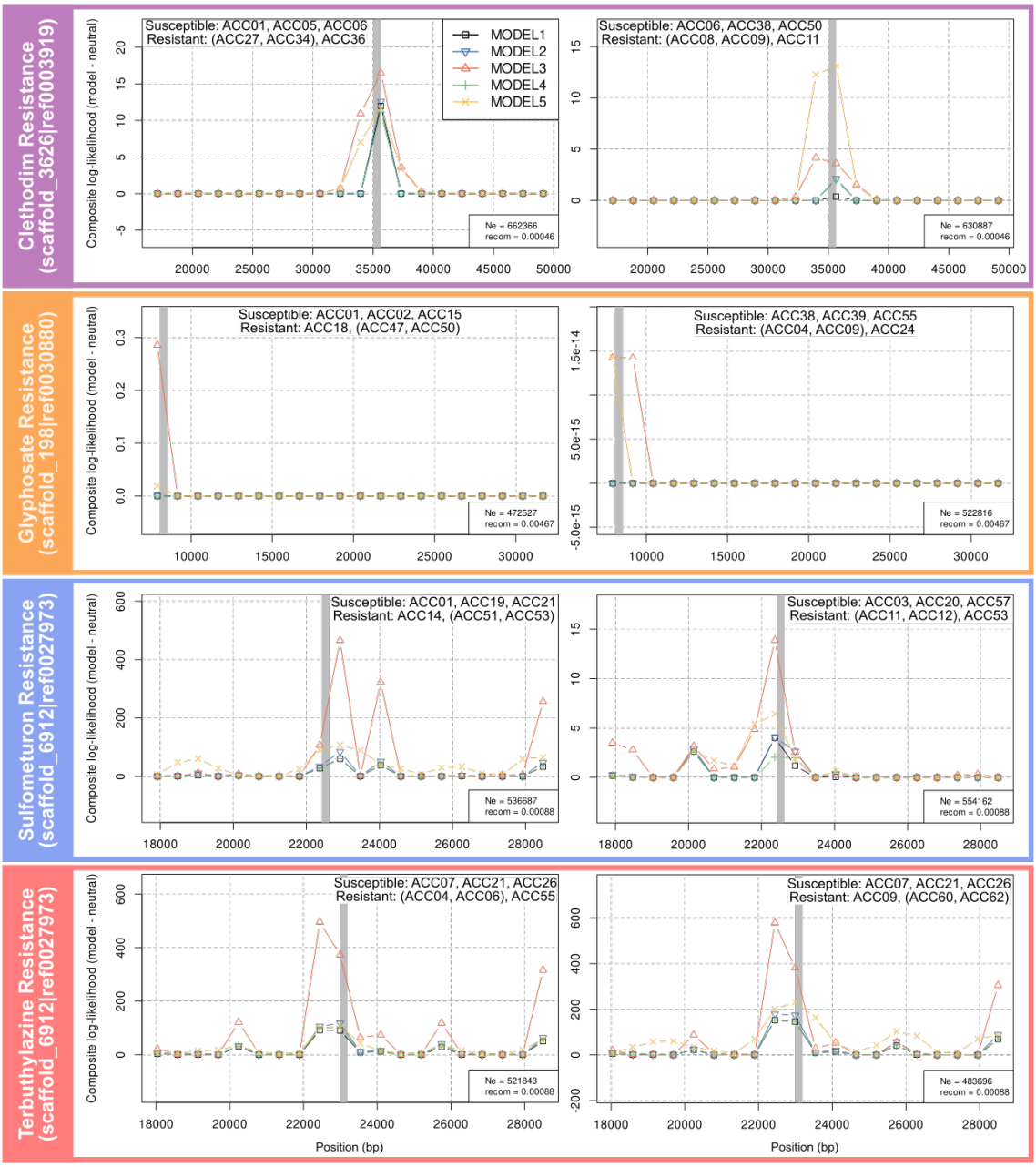


Figure 5.3. Relative composite likelihood profile across scaffolds per herbicide resistance trait where the strongest genetic effects were mapped. **Model 1 (black)**: all selected populations have independent mutations of the beneficial allele. **Model 2 (blue)**: all selected populations share the beneficial allele via migration. **Model 3 (red)**: the beneficial allele was standing in the ancestor of all selected populations. **Model 4 (green)**: the two resistant populations inside the parenthesis share beneficial allele via migration, while the other population has an independent mutation at the same locus. **Model 5 (yellow)**: the beneficial allele was standing in the ancestor of the two resistant populations inside the parenthesis, while the other population has an independent mutation at the same locus. **Gray vertical line** marks the position of the resistance QTL.

5.3.3. GP accuracy

Genomic predictions using the merged intra- and inter-population datasets resulted in ±20.11% to ±26.21% deviations from the observed resistances across the four herbicide resistance traits (Figure 5.4). In terms of RMSE, MIXED-REML performed better than RR and LASSO for glyphosate and terbuthylazine resistance traits; while LASSO performed better for clethodim and sulfometuron resistances. The highest prediction accuracy was achieved for terbuthylazine resistance using MIXED-REML.

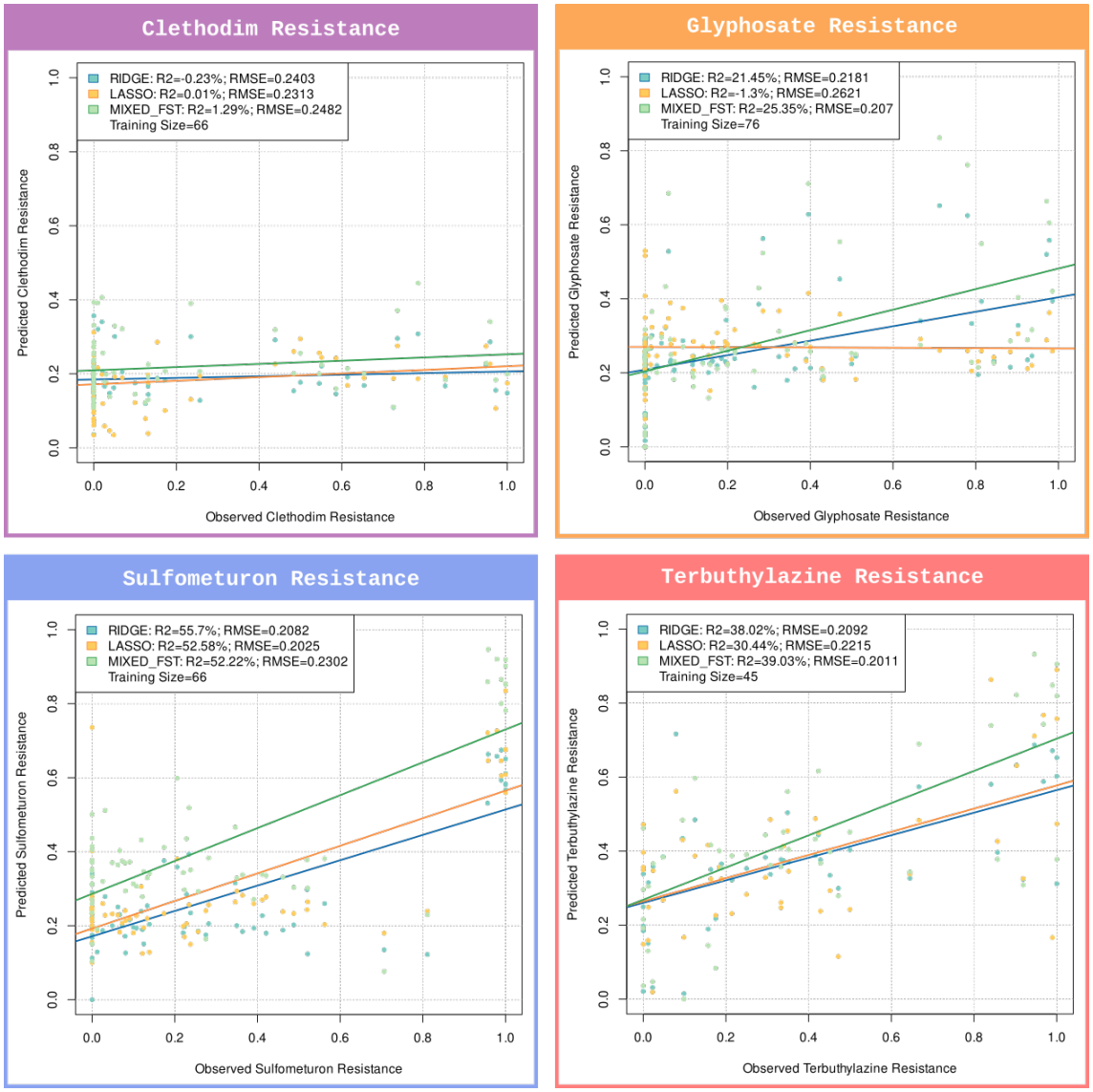


Figure 5.4. Genomic prediction accuracies per herbicide resistance trait using leave-one-out cross-validation. The models tested are RIDGE for ridge regression, LASSO for least absolute shrinkage and selection operator regression, and MIXED_FST is linear mixed model regression using F_{ST} as the random covariate and restricted maximum likelihood estimation of variance components. R2 is the adjusted coefficient of determination, and RMSE is the root mean square deviation of the prediction from the observation.

The accurate genomic predictions of sulfometuron resistances using the five populations collected as live plants demonstrated the feasibility of GP for the rapid monitoring of herbicide resistance in the field (Figure 5.5). However, the range of observed mean quantitative resistances (57.3% to 60.6%) were narrow.

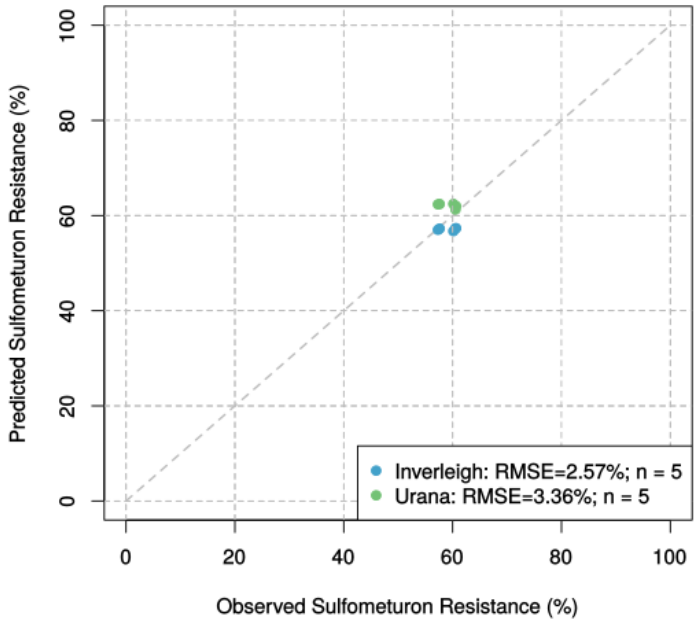


Figure 5.5. Accuracy of predicting the mean quantitative sulfometuron resistance of Lexton, Rokewood, and Skipton populations using the GWAlpha-derived allelic effects from Inverleigh and Urana populations. *RMSE* is the root mean square error between predicted and observed values and *n* is the number of populations: one for Lexton and two each for Rokewood and Skipton.

5.4. Discussion

5.4.1. Polygenicity of herbicide resistance traits and the genetic basis of cross-resistance

We have performed the first GWAS experiments on weed herbicide resistance. The genome-wide approach and the use of multiple populations in GWAS are in sharp contrast with the majority of previous studies which focused narrowly on target site resistance (TSR) mechanisms using a few contrasting populations. A small number of non-TSR-centric approaches have been performed in the past, including artificial selection, bulk-segregant analysis of parental, F1, F2 and backcross families and population genomic screening (Busi and Powles 2009; Busi et al 2011; Busi et al. 2013; Etten et al. 2020). The lack of consideration for polygenic resistance and the narrow focus on TSR in past studies has been detrimental to controlling herbicide resistance in weeds (Manalil 2015). Indeed, TSR genes can have large effects and induce rapid evolution as a consequence of high selection pressure in the form of high herbicide dosage (Preston et al. 2009; Powles and Yu 2010), but lowering herbicide dose to reduce the evolution of TSR (Gorddard et al. 1995) is inappropriate. It does not take into account polygenic resistance which can be more robust and can induce cross-resistance.

The predominance of small effect QTL implies that selection pressure imposed by herbicides acts on multiple loci simultaneously and low dose or lower selection pressure experienced by populations at field margins can exacerbate resistance development. A similar result was found in a study exposing *L. rigidum* populations to low dose of herbicide, i.e. sub-optimal herbicide application rates can induce the rapid adaptation to multiple herbicides by the accumulation of adaptive alleles with small effects (Busi et al. 2013). Therefore, resistance monitoring should not only include samples within the field but also along the borders and neighbouring weed patches.

The majority of the herbicide resistance-associated loci were not close to known genes. Additionally, these candidate loci do not include any of the expected TSR genes, i.e. ACCase EPSPS, ALS, and pbsA genes for clethodim, glyphosate, sulfometuron and terbuthylazine resistances, respectively. We suggest four reasons to explain this. First is the failure to include

scaffolds containing the TSR during Pool-ddRADseq; second is the poor and fragmented state of the reference genome; third is the suboptimal genome annotations; and fourth is the possible absence of TSR variants in the populations used. Our Pool-ddRADseq protocol may or may not have included SNP markers linked to putative TSR genes, but the fragmented nature of the reference genome used (N50=70kb) prevented us from recognising the linkage when markers and the TSR genes are located in different scaffolds. The genome annotations of the *Lolium perenne* genome assembly available to us is rudimentary. Finally, the TSR variants may simply have been absent in the populations we used. However, there are certainly many more genes yet to be characterised that would have small but significant effects on resistance.

The candidate genes for herbicide resistance include stress response and detoxification genes, as well as many poorly characterised genes. This is in line with the results from a recent population genomics screen of glyphosate-resistant *Ipomoea purpurea* populations where the expected target-site resistance variants (i.e. over-expression) were not found; instead detoxification genes were identified (Etten et al. 2020). We found that the significantly enriched functions among our clethodim and glyphosate resistance genes were abiotic and chemical stress response. The same cannot be said for sulfometuron or terbuthylazine resistance, where the significantly enriched GO terms are more general, i.e. regulation of cellular and metabolic processes. This supports the omnigenic hypothesis of polygenic traits, which states that most of the genetic effects reside outside the genes directly related to the core pathways controlling the trait - this include regulatory pathways and seemingly unrelated genes (Boyle et al. 2017).

We showed evidence of cross-resistance genes (Figure 5.2 and Appendix Figures S5.19 to S5.22), and population-specific herbicide resistance-associated loci. The presence of common loci and genes between at least two herbicide resistance traits is evidence for cross-resistance. These are the putative genetic bases of the correlated resistance traits we reported in the instaGraminoid article (Paril and Fournier-Level 2019; i.e. Figure 3.6) and Figure 3.10. DRB6, the candidate gene common to the four herbicide resistance traits (clethodim, glyphosate, sulfometuron, and terbuthylazine resistances), is poorly characterised. Experimental evidence at the transcript level suggests it belongs to the group of plant double-stranded RNA binding proteins (DRBs). This family of proteins is a key factor involved in the biogenesis and function of small RNAs (Clavel et al. 2016). Furthermore, the population-specific herbicide resistance-associated loci and candidate genes highlight the need for characterising multiple populations to improve QTL detection power, which supports findings from previous multi-population GWAS studies on plants (J. Yu et al. 2008; Bouchet et al. 2017), cattle (Gebreyesus et al. 2019) and humans (Wijmenga and Zhernakova 2018). Additionally, the relatively small number of associated SNPs shared among populations (Appendix Figures S5.19 to S5.22) is in contrast with the results of Duhoux et al (2017) where a common set of NTSR genes were significantly upregulated in geographically distinct populations. This suggests that different populations may exploit different regulatory pathways to control the same herbicide.

Potentially more genes underlie the resistance-associated loci, however due to the poor state of annotations and the fragmented state of the reference genome, most of the associated loci did not collocate near annotated genes. GWAS are nonetheless powerful hypothesis generators ultimately requiring cross-validation. The herbicide resistance QTL detected need to be functionally validated in future experiments. Nonetheless, our results can be directly used in genomic prediction for rapid monitoring and effective management of herbicide resistance.

5.4.2. Herbicide resistance evolution in *Lolium rigidum* across South-East Australia

Herbicide resistance evolution in *L. rigidum* populations across SE Australia is more likely the result of co-opting existing variation to confer resistance rather than the emergence of new mutations. This is despite their similar selection signatures in large populations under the classical view of genetic variation (Barton 2010). Standing genetic variation as the more likely origin of resistance alleles is further supported by our observations that herbicide resistance traits are highly polygenic. A recent mutational origin for these hundreds of resistance-conferring alleles is unlikely given the rarity of favourable mutations (Elena et al. 1996; Orr 2010) and the rapid rate at which herbicide resistances have been evolving (Baucom 2016). Standing genetic variation facilitates rapid adaptive evolution since existing favourable alleles have higher

frequency before selection than new mutations (Colosimo et al. 2005; Barrett and Schluter 2008; Domingues et al. 2012).

Pre-existing alleles prior to being selected for by herbicide application may have been (1) neutral and maintained in the population by chance, (2) favourable, or (3) slightly deleterious but cryptically by recessiveness (Yanchukov and Proulx 2012) or hitch-hiking when linked to favourable alleles (Hill and Robertson 1966; Brown et al. 1989; McVean and Charlesworth 2000; Loewe and Hill 2010). These pre-existing alleles endow weed populations with the inherent capacity to rapidly overcome herbicide exposure.

There are fundamental biochemical and physiological constraints to the fixation or loss of adaptive alleles (Olson-Manning et al. 2012). These constraints imply the existence of optimal phenotypes, and a limit to evolution (Stern 2013; Stayton 2015; Sackton and Clark 2019). The search space for biological solutions to survive herbicide application is likely to be large and riddled with local optima and potentially a narrow global optimum. Historical mutations observed as standing genetic variation and new mutations represent a random-walk across such space. Suboptimal solutions typically include resistance alleles which incur fitness costs. Better solutions are cross-resistances with or without fitness costs. The optimal solution is cross-resistance which improves fitness even in the absence of herbicides. This was observed in an experimental evolution of herbicide resistance in *Chlamydomonas reinhardtii*, i.e. the most resistant populations were not only cross-resistant but were also more fit in the ancestral herbicide-free environment (Vogwill et al. 2012).

The rapid evolution of herbicide resistance from standing genetic variation stresses the fact that resistance evolution in weeds is inevitable. Hence, this evolution can only be slowed-down to a manageable pace, allowing for new herbicides to be developed and more sustainable solutions such as integrated weed management or transformational technologies such as gene drives to be adopted (Hamilton 1967; Curtis 1968; Champer et al. 2016). Due to the lack of appreciation for standing genetic variation as a source of resistance, early recommendations to reduce herbicide dosage (Gorddard et al. 1995) may have accelerated resistance evolution instead of slowing it. Under low selection intensities it has been predicted that adaptations from standing genetic variation is more likely than new mutations (Hermisson and Pennings 2005). A promising strategy to slow this evolution is to use a mixture of different herbicides at high enough doses to be inhibitory (Lagator et al. 2013).

Our findings using a coalescent-based composite likelihood approach can be complemented by analysing the full extent of resistance variation at the genome and landscape levels. More general approaches without the assumptions of recent sweeps and complete resistance or susceptibility of populations, while accounting for the whole resistance landscape and multiple resistance loci simultaneously, are needed to corroborate or invalidate our findings. Approximate Bayesian computation (ABC; (Beaumont et al. 2002) is one such approach. ABC generates likelihood-free posterior probability distributions of population genetics parameters using Markov Chain Monte Carlo simulations contingent upon the complex observed datasets consisting of phenomic and genomic data distilled into informative summary statistics (Pritchard et al. 1999; Beaumont et al. 2002; Beaumont 2019).

5.4.3. Genomic prediction using Pool-seq data for rapid and cost-effective herbicide resistance assessment

A linear-mixed modelling approach was integrated into the GWAlpha package for GWAS and GP using Pool-seq data. This approach accounts for population structure and enables multi-population Pool-GWAS, where each pool represents a population. A linear-mixed model accounting for multiple sources of variation in *Lolium perenne* F2 families has been demonstrated to yield predictive abilities ranging from 0.34 to 0.77 correlation between observed and predicted phenotypes (Guo et al. 2018). Thus genomic prediction (GP) using population- and pool-level genotype data to predict phenotype has the potential to improve the management of herbicide-resistant weeds by identifying likely resistant populations at early stages of resistance evolution. The utility of GP lies in its exclusive reliance on genotype data over time-consuming and resource-intensive phenotyping assays. In plant and animal breeding, it was also shown to yield higher genetic gain per year compared to conventional breeding

methods (Hickey et al. 2017).

At the whole population scale, measuring the population-level proportion of resistance is informative enough without the need for individual plant-level quantitative herbicide resistance estimates. Additionally, farmers are more interested in the state of resistance of populations rather than individual plants. Accurate genomic predictions were made using the GWAlpha-inferred allelic effects and the mean instaGraminoid-derived individual plant herbicide resistance in the Lexton, Rokewood and Skipton populations. However, the range of values for these populations is very narrow and the prediction accuracies may change drastically at the extreme values we failed to observe using these populations. In addition, the tray-level instaGraminoid protocol requires further development to improve reliability, despite the high accuracy of the individual-level phenotyping protocol.

The prediction accuracies presented here are similar to those of previous genomic prediction and selection experiments on plants and animals (Technow et al. 2013; Zhao et al. 2013; Cerrudo et al. 2018). In the closely-related *L. perenne* whose genome we used as reference, the genomic prediction for crown rust resistance yielded a maximum predictive ability measured as Pearson's correlation coefficient of 52% (Arojju et al. 2018). GP accuracies greater than 0.90 adjusted R^2 are rare. Accuracies ranging from 0.79 to 0.92 R^2 have been obtained on wood phenotypes of Douglas-fir trees using exome capture but the authors attribute these high accuracies to pedigree effects rather than individual causal QTL (Thistlethwaite et al. 2017).

Our approach to GP using Pool-seq data on the non-model, poorly genomically characterised, yet agroecologically important weed species *Lolium rigidum* has a lot of room for improvement. To improve GP accuracy, the number of training populations, replications, and consensus SNPs across populations need to increase, and the control of population structure needs to improve. Better prediction accuracy is expected with more populations used for model training (Asoro et al. 2011; Lorenz et al. 2011; Akdemir and Isidro-Sánchez 2019). The intra-population datasets are unreplicated, while the inter-population datasets are unbalanced where some populations have been replicated more than others. More reliable phenotype data generated from replicated and balanced experimental designs is straightforward to attain.

Inconsistent Pool-ddRADseq genome coverage resulted in a reduced number of consensus loci genotyped across populations and pools. Improving coverage consistency will boost the number of SNPs by capturing genomic regions more closely linked to causal variants. There are two ways of achieving this with Pool-ddRADseq: protocol optimisation, and missing data imputation. Optimising the Pool-ddRADseq protocol involves finding better restriction enzyme pairs which will improve genome coverage, and better DNA fragment selection which will boost the number of consensus loci through size selection and adapter design. Adapters designed to be specific to each one of the enzyme cut sites (Peterson et al. 2012) may be used instead of non-specific Illumina adapters (Toonen et al. 2013). Imputing missing genotype data requires linkage disequilibrium (LD) information (Li et al. 2009; Schurz et al. 2019). Since individual haplotype data are not available with Pool-seq, LD information can be extracted from individual sequencing reads instead of individual genotypes (Feder et al. 2012). Long-read sequencing technologies like MinION can overcome the limits imposed by short-read sequencing thereby allowing the estimation of long-range LD.

Finally, accounting for population structure can remove nuisance or non-causal genetic effects, i.e. pedigree/relatedness effects and improve the transferability of GP models. This is very challenging with Pool-seq data since individual genotypes are unavailable. To circumvent this problem, training datasets can be designed to better represent the validation or test datasets (Hoffstetter et al. 2016; Akdemir and Isidro-Sánchez 2019; Edwards et al. 2019). Additionally, further development toward controlling for population structure like genetic similarity matrices adapted to pool-seq should be considered.

5.5. Conclusion

The herbicide resistance traits studied, i.e. clethodim, glyphosate, sulfometuron, and terbuthylazine resistances are polygenic with the predominance of associated SNPs with small effects. In each herbicide resistance trait, different sets of candidate QTL and genes contribute

to resistance in different populations, however some QTL and candidate genes are shared between populations. Evidence of cross-resistance QTL and genes were found. These herbicide resistance alleles are more likely to have originated from standing genetic variation than from *de novo* mutations. Genomic prediction using population-level data can be useful in herbicide resistance monitoring and management by identifying potentially resistant populations at the early stages of resistance evolution.

Chapter 6

General discussion and the evolution of herbicide resistance in SE Australian *Lolium rigidum* populations

6. General discussion and the evolution of herbicide resistance in SE Australian *Lolium rigidum* populations

The remarkable propensity of life to adapt to challenging conditions is fueled by genetic variation. The mechanism through which these variations originate through random mutations is fundamentally stochastic (Nei 1980; Barton 2010). The forces of selection determine which of these genetic variants increase increase in frequency (Darwin 1859). There are fundamental biochemical and physiological constraints to the fixation or loss of adaptive alleles (Olson-Manning et al. 2012). These constraints imply the existence of optimal phenotypes, and a limit to evolution (Stayton 2015; Stern 2013; Sackton and Clark 2019).

The search space for biological solutions to challenging conditions is likely to be very large, with multiple local optima. The accumulation of beneficial genetic mutations represent a random-walk across this search space. Suboptimal solutions may incur fitness costs; while better solutions have positive pleiotropic effects, improving fitness across multiple environmental conditions. A parsimonious genetic basis for the phenotypes resulting in an optimal fitness is more likely to originate from the rich reservoir of standing genetic variation rather than from new mutations with low likelihoods of conferring parallel benefits to multiple traits. Standing genetic variation presents a large reservoir of potentially beneficial variants across many loci which selective forces can readily act on. This is what we have inferred and observed as the more likely source for the adaptive evolution of *Lolium rigidum* across South-East Australia to four different herbicides, clethodim, glyphosate, sulfometuron and terbuthylazine. We observed that herbicide resistance traits were highly polygenic and we identified numerous putative cross-resistance loci. We also inferred that the causal alleles were likely to have originated from standing genetic variation rather than from new mutations.

6.1. A model of herbicide resistance evolution in *Lolium rigidum* across SE Australia

The observed polygenicity of the herbicide resistance traits is more consistent with standing genetic variation as the origin of resistance alleles than new mutations. This is because at the start of selection from herbicides, the initial frequencies of beneficial alleles from standing genetic variation are higher than from new mutations allowing for the observed rapid evolution (Baucom 2016; Baucom 2019) to be more likely. Additionally, this origin explains the abundance of putative cross-resistance loci, since even before the onset of selection from herbicides, these loci may have general beneficial effects on fitness. If we assume a pleiotropic allele which positively affects multiple traits is more favourable than an allele which only affects a single trait, then cross-resistance alleles are more beneficial than herbicide-specific resistance alleles. Hence, it is within our expectations to observe polygenic herbicide resistance traits with an abundance of cross-resistance loci.

Variability is the fuel through which evolution proceeds. Standing genetic variation represents historical mutations that have accumulated and maintained in populations due to chance (drift), selective advantage, or linkage with beneficial variants. This variability is positively correlated with effective population size (N_e) (Kimura 1991; Hague and Routman 2016), and widespread distribution (Caley and Schluter 1997) which are hallmarks of outcrossing species (Holsinger 2000; Pannell and Labouche 2013) such as *L. rigidum*. The *L. rigidum* populations in SE Australia descended from multiple populations introduced directly or indirectly from the Mediterranean, which explains the abundant standing genetic variation through shared ancestry (Kloot 1983).

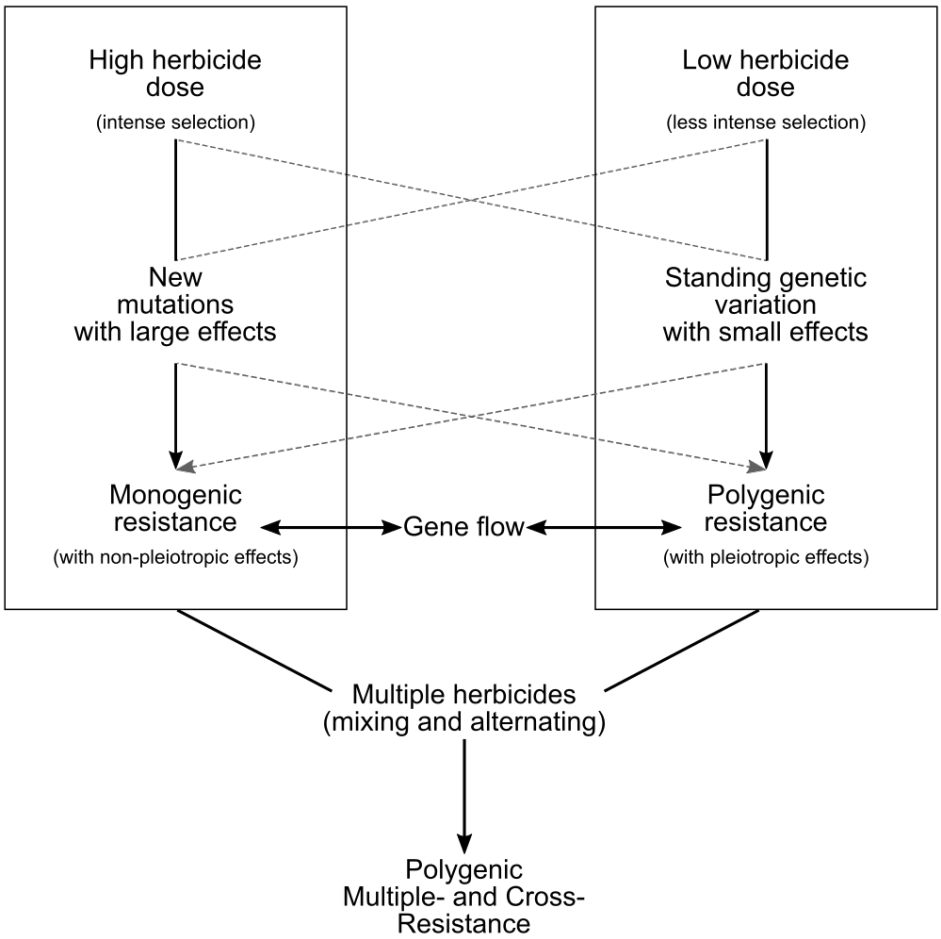


Figure 6.1. Model of the adaptive evolution of weeds to herbicides in agroecosystems.

My model for resistance evolution is depicted in Figure 6.1. Succinctly, this model states that adaptive evolution favours parsimony through pleiotropy. Recruiting existing favourable variants to achieve a co-adaptation to multiple challenging conditions is a parsimonious solution. Sustained intense selection pressure on populations with large N_e will favour monogenic resistance evolution; however, neighbouring populations may be inadvertently exposed to suboptimal or low herbicide dose and hence to less intense selection which favours polygenic resistance evolution. This is an extension and application of the findings in *Chlamydomonas reinhardtii*, *L. rigidum*, and is emphasised in some of the reviews of herbicide resistance in weeds (Vogwill et al. 2012; Lagator et al. 2013; Busi et al. 2013; Délye et al. 2013). Gene flow between these populations is likely, thereby exacerbating herbicide resistance evolution by generating more robust genetic bases. The resulting polygenic resistance, controlled by a mixture of large and small effects loci, is robust since the loss of one to a few resistance-conferring alleles will not lead to a complete breakdown of resistance. These resistance alleles are driven towards fixation with sustained selection pressure from herbicides. Fitness costs and linkage drag can slow down this process. Application of different herbicides with different modes of action repeats this cycle, where pleiotropic alleles which confer resistance to multiple herbicides and even improved fitness in the absence of herbicides have the most fitness advantage. Parsimony through pleiotropy is supported by the omnigenic view of complex traits, i.e. gene networks must be sufficiently interconnected such that a single gene affects multiple traits (Boyle et al. 2017). Hence, alternating and/or mixing herbicides with different modes of action has the potential to accelerate the evolution of super-resistant weed populations.

This stresses the importance of using a variety of non-chemical weed control strategies, i.e. mechanical, physical, and biological methods, in tandem with herbicides. This is to keep weed population sizes small to reduce the chances of selecting for cross-resistant individuals. This can slow down the evolution of both monogenic (as expected for TSR) and polygenic (when both NTSR and TSR are involved) herbicide resistance traits.

To test this model of parsimonious adaptive evolution through pleiotropy, I propose two research approaches. The first one is to perform multivariate or multi-trait GWAS to dissect the genetic bases of multiple correlated traits within a single framework. In addition to mapping the genetic basis of herbicide cross-resistance, the correlation between resistance and other fitness-related traits should also be mapped, e.g. other abiotic and biotic stress resistances. The abundance of detected cross-resistance QTL from these multi-trait GWAS will provide support for this model of adaptive evolution. The second one is to study the correlation between genome complexity including the number of genes and organism complexity. The motivating argument is that genome complexity based on the number of genes in the genome is not positively correlated with species complexity, e.g. the single-celled human pathogen *Trichomonas vaginalis* has more than twice the number of protein-coding genes than humans despite being biologically simpler (Van Straalen and Roelofs 2006). I recommend seeking answers to the following questions. Does adaptation to many varied conditions favour large complex genomes and more genes? An affirmative answer would imply the stacking of environment- or condition-specific genes. On the other hand, a negative answer implies the co-adaptation of genes or the predominance of pleiotropic genes.

6.2. Polygenic herbicide resistance

How does the observation of putatively polygenic herbicide resistance traits support the herbicide resistance evolution model of parsimony through pleiotropy? The rate at which weed populations have been evolving resistance to herbicide is too fast to be compatible with a model where new mutations are the primary source of resistance-conferring alleles. Standing genetic variation is the more likely source. Evolution of resistance from standing genetic variation favours polygenic resistance traits because of the abundance of historical mutations which may have pre-existing positive effects on fitness and stress-response. These beneficial standing genetic variants may be adapted to confer herbicide resistance as well. Therefore polygenicity is likely to further favour pleiotropic alleles.

The speed at which resistance to different herbicides has evolved in multiple weed species has been striking. 2,4-D (2,4-dichlorophenoxyacetic acid) started being used extensively as a herbicide to control broadleaf weeds in 1944 (Peterson 1967). Within 13 years, in 1957, the first report of 2,4-D resistance in spreading dayflower (*Commelina diffusa*) was recorded (Hilton 1957). Triazine resistance in common groundsel (*Senecio vulgaris*) evolved within 10 years of applying simazine and atrazine once or twice annually from 1958 to 1968 (Ryan 1970). Glyphosate resistance in *L. rigidum evolved* within 15 years of extensive and recurrent glyphosate application in Australia from 1981 to 1996 (Powles et al. 1998). Adaptive evolution from standing genetic variation is faster than from *de novo* mutations, and it leads to the fixation of more alleles with small effects (i.e. polygenic adaptive traits) (Barrett and Schluter 2008). Polygenic resistance can be argued to be more stable, i.e. the multitude of loci involved acts as a buffer for when one or several alleles fail to confer resistance, overall resistance does not totally breakdown with them.

The agroecosystems these weed populations occupy are spatially and temporally heterogeneous in terms of the selection pressure imposed, i.e. different herbicides singly or in combination may be used in different areas and times, and fallow periods may occur where herbicides are not applied at all. Hence, alleles which confer resistance to multiple herbicides as well as contribute to improved fitness in the absence of herbicides are the most favourable. A multitude of these cross-resistance, pleiotropic alleles may exist as standing genetic variation, ready for selective forces to act upon. An experimental evolution study in 2012 of *Chlamydomonas reinhardtii* showed that a mixture of different herbicides works on slowing down resistance evolution if the doses are high enough to be inhibitory; however if doses were suboptimal, it resulted in the evolution of cross-resistance (Vogwill et al. 2012). A year later in 2013, a different experimental evolution of *C. reinhardtii* from the same group showed that pleiotropic effects of NTSR genes can improve resistance to multiple herbicides and even improve fitness in herbicide-free environments (Lagator et al. 2013).

6.3. Rapid monitoring of herbicide resistance with genomic prediction and Pool-seq

Assuming herbicide resistance traits are polygenic and are evolving under the principle of parsimony through pleiotropy, how do we rapidly and accurately monitor the adaptive evolution of weeds to herbicides across agroecosystems? Tracking the causal loci across a landscape is more challenging for polygenic traits than for monogenic and oligogenic traits. Polymerase chain reaction genotyping experiments can suffice for mono- and oligogenic traits, while a genome-wide approach is required for polygenic traits. We used Pool-seq in tandem with a genome complexity reduction technique, Pool-ddRADseq, to characterise *L. rigidum* populations across SE Australia. We have shown the potential of using these genome-wide Pool-seq allele frequency data to rapidly monitor the levels of resistance of populations through genomic prediction.

We demonstrated through a simulation experiment that Pool-seq is more cost-effective than individual sequencing (Indi-seq). For the same budgetary constraint, more populations can be characterised with Pool-seq than with Indi-seq. Since sample sizes should be sufficiently large to accurately represent a population, this can be prohibitively expensive with Indi-seq. On the other hand, this does not limit the number of individuals that can be included in each pool with Pool-seq. In fact, increasing the number of individuals included in each pool improves the accuracy of allele frequency estimates (Ferretti et al. 2013). Pool-seq can accurately capture genome-wide allele frequency information (Zhu et al. 2012; Gautier et al. 2013; Rellstab et al. 2013; Fracassetti et al. 2015). It is more cost-effective than Indi-seq, and it can include more individuals, grouped into one or a few pools and sequenced at high depth (Futschik and Schlötterer 2010; Gautier et al. 2013). However, it is prone to biases in genome representation when sample size is below 40 individuals and depth of coverage is below 50X (Cutler and Jensen 2010; Schlötterer et al. 2014). It also loses linkage disequilibrium information (Fariello et al. 2017); however this presents an opportunity to develop novel analysis methods (Cutler and Jensen 2010).

Using Pool-seq data to build genomic prediction (GP) models at the population level has the potential to improve the management of herbicide-resistant weeds by identifying resistant populations at early stages of their evolution. The success of an intervention strategy to reduce the spread and development of resistance is more effective when executed earlier rather than later (Tomson and Vlad 2014; Derde et al. 2014; Baquero et al. 2015). The advantage of GP over conventional phenotyping is not its accuracy (at least not yet) but the speed at which it can provide estimates of the phenotype value. Additionally, there is potential for cost reduction with GP where a single genotyping experiment can yield predictions for multiple herbicides, simultaneously. In plant and animal breeding, this translates to higher genetic gain per unit of time compared with conventional breeding methods (Hickey et al. 2017).

We can extend the scope of the GP models investigated in this thesis to account for the specificity of the adaptive evolution of weeds. Weediness characteristics (e.g. seed dormancy and prolificacy) can be monitored in addition to multiple herbicide resistance traits. Multivariate GP models can be used to predict resistance to multiple herbicides and other fitness-related traits, simultaneously. This can be effectively realised with a training dataset that is larger and more representative than the one we currently have for *L. rigidum* populations from SE Australia. Amassing these data will involve sampling and characterising more populations, and improving the Pool-ddRADseq protocol to yield more consistent genomic data.

6.4. Recommendations to improve GWAS power and GP accuracy

Populations need to be collected from more locations and characterised more effectively to improve the power of GWAS and the accuracy of GP. Larger sample sizes, more replications and more balanced experimental designs during phenotyping will improve the accuracy of the measured resistance levels. Technical improvements in Pool-ddRADseq for better consistency of genome coverage across samples, coupled with imputation techniques will boost the number

of informative SNPs that can be used in GWAS and GP.

6.4.1. Improvements in phenotyping

The first step of improvement that can be proposed in *L. rigidum* herbicide resistance phenotyping is to increase the level of replication. Given more resources to work with, each plant can be replicated at least three times using clonal propagation through tiller separation. At the population level, replicating and representing each population equally will balance the experimental design. This will result in higher statistical power and less susceptibility to departures from the assumption of error homoscedasticity than unbalanced designs (Shaw and Mitchell-Olds 1993).

The cost-effective colorimetric phenotyping method we developed (instaGraminoid) was sufficient for the GWAS and GP experiments performed. However, there is room to improve this method with more sophisticated imaging instruments like hyperspectral imagers. Hyperspectral imaging will provide high-dimensional data which can be used to model not only the physical and colorimetric plant responses to herbicides but also their internal physiological and biochemical responses. Hyperspectral imaging will enable non-destructive measurements of plant biomass as well as water and protein contents (Mishra et al. 2017).

6.4.2. Improvements in genotyping

Improving the consistency of Pool-ddRADseq genome coverage will boost the number of SNPs by capturing genomic regions more closely linked to causal variants. Better restriction enzyme pairs can help improve genome coverage. Adapters can be designed to be specific to each of the enzyme cut sites as in the original ddRADseq protocol (Peterson et al. 2012). This will improve the consistency of the genomic data but at the cost of reducing the breadth of coverage. Missing data can also be imputed. This requires linkage disequilibrium (LD) information (Li et al. 2009; Schurz et al. 2019) which is not straightforward to obtain from Pool-seq data. LD information can be extracted from individual sequencing reads instead of individual genotypes (Feder et al. 2012). Additionally, long-read sequencing technologies like MinION can overcome the limits imposed by short-read sequencing thereby allowing the estimation of long-range LD.

Despite mapping at least 95% of the Pool-ddRADseq (Illumina reads) to the *L. perenne* genome assembly (Byrne et al. 2015), the highly fragmented nature and suboptimal state of the genome annotations made it difficult to analyse the GWAS results. Drafting a *L. rigidum* reference genome and developing a good annotation will allow us to better identify herbicide resistance-associated loci and test their validity.

6.4.3. Improvements in sampling and genotype-to-phenotype mapping

Expanding the training data will improve the representation of the landscape which in turn will increase GP accuracy (Akdemir and Isidro-Sánchez 2019; Edwards et al. 2019; Hoffstetter et al. 2016). This will also increase the chances of detecting QTL that were missed in our current collection but present in other populations over the landscape.

As the amount of “omics” data available increases (Yang et al. 2020; Reuter et al. 2015), machine learning techniques such as neural networks are becoming more feasible options to generate genotype-to-phenotype maps for GWAS and GP. These model-free non-parametric or black box approaches can effectively estimate genetic effects beyond additive effects to include non-additive effects most notably epistasis. The linear models we used, only accounted for the additive genetic effects of the SNPs, while the dominance and epistatic effects were lumped together in the error term. Incorporating these non-additive genetic effects into a linear model is intractable because the SNP interaction matrix will be massive and likely to be highly multicollinear. Random forest (Li et al. 2018; Abdollahi-Arpanahi et al. 2020; Ogutu et al. 2011), and deep learning (Abdollahi-Arpanahi et al. 2020; Ma et al. 2018; Zingaretti et al. 2020; González-Camacho et al. 2012; Montesinos-López et al. 2018; Waldmann 2018) have been

used to perform GWAS and GP.

6.5. Recommendation to control herbicide resistance in weeds

Under our adaptive evolution model, over-reliance on herbicide will lead to super-resistant weed populations regardless of the herbicide application strategies we employ. Continuous application of an herbicide with a single mode of action will likely results in the rapid evolution of resistance to that specific herbicide (Jasieniuk et al. 1996) forcing the use of another mode of action which repeats the cycle. Mixing and rotating herbicides with different modes of action will likely slow down the evolution of resistance but at the cost of eventually generating cross-resistant superweeds (Vogwill et al. 2012; Lagator et al. 2013). Hence, recalcitrance to diversifying weed control methods beyond herbicide application entails a trade-off between speed of resistance evolution and resistance severity until ultimately super-resistant weeds evolve. My data validate the recommendation of intensifying non-chemical weed control as proposed since Gorddard et al in 1995 but with the major amendment of not reducing herbicide doses.

I propose a genomically-informed rotation of chemical and non-chemical weed control strategies (Figure 6.2). Economically-detrimental, crop- and weed-specific herbicide resistance levels will be determined. These are the levels of resistance above which weeds incur significant crop yield losses. Resistance of weed populations will be regularly and rapidly monitored using GP. The most variable populations or the best representatives of the landscape will need to be phenotypically and genotypically characterised regularly to re-train or update the GP models. Levels of resistance in the other populations will be predicted using these updated GP models. In addition, weed populations on the margin of agroecosystems may be developing resistances and have the potential to migrate to field populations, and need to be included. Paddock-specific herbicide mixtures will be recommended. The recommendation will be based on which herbicides the weed populations in the paddocks are susceptible to. The herbicide mixtures will be applied until the resistance threshold is reached. At that point, non-chemical methods will be recommended until resistance levels go below the threshold again. The cycle is repeated and other or new herbicides may be recommended.

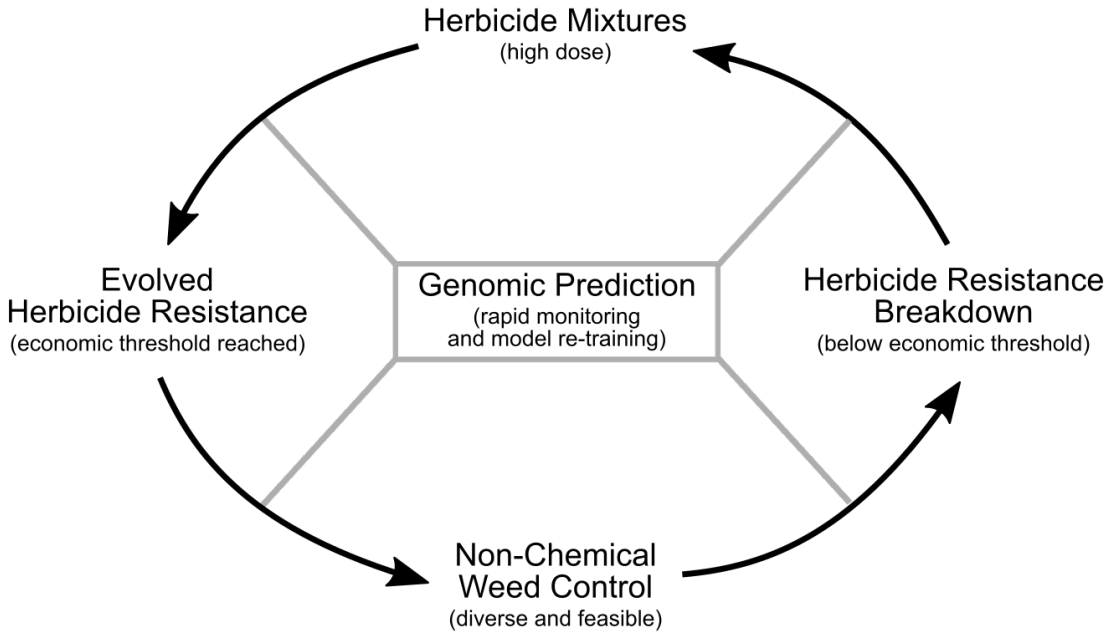


Figure 6.2. Genomically-informed rotation of chemical and non-chemical control strategies.

I am proposing a weed control rotation strategy that I hope will ease farmers into transitioning towards more integrated weed management (IWM). The reluctance of farmers to practise more IWM strategies is due to the convenience and maintained cost-effectiveness of herbicides.

Farmers and agronomists are aware and concerned about the increasing problem of herbicide resistance in weeds (Llewellyn et al. 2007; Wilson et al. 2016; Moss 2018). However, they are constantly confronted with relatively short-term business decisions to maximise output, and herbicide application remains to be the easily accessible and cost-effective approach (Jussaume and Dentzman 2016). Moss (2018) alluded to the possibility of government intervention to improve the adoption of IWM.

In SE Australia, sheep grazing is usually practised in combination with crop cutting for silage before weed seed set and cutting low for hay. Hot water or steam treatments are gaining traction in Australia as an alternative to chemical control (Banks and Sandral 2007). This approach can be very effective in high-value horticultural cropping systems like orchards; however its large-scale applications can be very costly. Therefore, there is a need to further develop non-chemical weed control methods for cost-effective broad-acre applications.

6.6. Conclusion

Herbicide resistance in weeds is an outstanding example of rapid adaptive and convergent evolution. My research into this phenomenon have yielded insights into the genetic basis of herbicide resistance, how adaptive evolution proceeds, and helped us build a genomically-informed herbicide resistance management strategy. Herbicide resistance in weeds is likely to be polygenic. Cross-resistances were observed and evidence of pleiotropic genetic architectures were found. The origin of these resistance-conferring alleles are more likely to be standing genetic variation (i.e. historical mutations maintained in the populations via chance, selection, or linkage with favourable alleles) than new mutations. This line of evidence supports our adaptive evolution model of parsimony through pleiotropy, i.e. alleles favourably affecting resistance to multiple herbicides as well as fitness in herbicide-free environments are the most favourable. With these findings, we have demonstrated that big funding is not required to generate meaningful and insightful results using the increasingly accessible genomics tools and powerful statistical frameworks. Cost-effective and accurate phenotyping and population-level genotyping methods (i.e. instaGraminoid and Pool-ddRADseq, respectively) can be used to build genomic prediction models. This has the potential to improve herbicide resistance management with the rapid and accurate monitoring of resistance evolution in multiple agroecosystems across vast areas. Furthermore, these genomics and statistical tools can help wean farmers from the over-reliance on herbicides and ease them into more integrated weed management for more sustainable agricultural production.

Bibliography

7. Bibliography

1001 Genomes Consortium, Stefan R. Henz, Svante Holm, Matthew Horton, Mike Jarsulic, Randall A. Kerstetter, Arthur Korte, et al. 2016. “1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis Thaliana.” Cell 166 (2): 481–491. https://doi.org/10.1016/j.cell.2016.05.063.

Aalto, Juha, Pentti Pirinen, Juha Heikkinen, and Ari Venäläinen. 2013. “Spatial Interpolation of Monthly Climate Data for Finland: Comparing the Performance of Kriging and Generalized Additive Models.” Theoretical and Applied Climatology 112 (1): 99–111. https://doi.org/10.1007/s00704-012-0716-9.

Abdollahi-Arpanahi, Rostam, Daniel Gianola, and Francisco Peñagaricano. 2020. “Deep Learning versus Parametric and Ensemble Methods for Genomic Prediction of Complex Phenotypes.” Genetics Selection Evolution 52 (1): 12. https://doi.org/10.1186/s12711-020-00531-z.

Abraham, Gad, and Michael Inouye. 2015. “Genomic Risk Prediction of Complex Human Disease and Its Clinical Application.” Current Opinion in Genetics & Development, Molecular and genetic bases of disease, 33 (August): 10–16. https://doi.org/10.1016/j.gde.2015.06.005.

Adegoke, Anthony, Adekunle Faleye, Gulshan Singh, and Thor Stenström. 2016. “Antibiotic Resistant Superbugs: Assessment of the Interrelationship of Occurrence in Clinical Settings and Environmental Niches.” Molecules 22 (1): 29. https://doi.org/10.3390/molecules22010029.

Akdemir, Deniz, and Julio Isidro-Sánchez. 2019. “Design of Training Populations for Selective Phenotyping in Genomic Prediction.” Scientific Reports 9 (1): 1446. https://doi.org/10.1038/s41598-018-38081-6.

Albert, Thomas J, Michael N Molla, Donna M Muzny, Lynne Nazareth, David Wheeler, Xingzhi Song, Todd A Richmond, et al. 2007. “Direct Selection of Human Genomic Loci by Microarray Hybridization.” Nature Methods 4 (11): 903–905. https://doi.org/10.1038/nmeth1111.

Alqudah, Ahmad M., Ahmed Sallam, P. Stephen Baenziger, and Andreas Börner. 2020. “GWAS: Fast-Forwarding Gene Identification and Characterization in Temperate Cereals: Lessons from Barley – A Review.” Journal of Advanced Research 22 (March): 119–35. https://doi.org/10.1016/j.jare.2019.10.013.

Amare, Tesfay. 2014. “Effect of Weed Control Methods on Weeds and Wheat (Triticum Aestivum L.) Yield.” *African Journal of Agricultural Research* 2 (3): 124–128. https://doi.org/10.5897/AJAR2013.8288.

Anand, Santosh, Eleonora Mangano, Nadia Barizzone, Roberta Bordoni, Melissa Sorosina, Ferdinando Clarelli, Lucia Corrado, Filippo Martinelli Boneschi, Sandra D ’alfonso, and Gianluca De Bellis. 2016. “Next Generation Sequencing of Pooled Samples: Guideline for Variants’ Filtering.” Nature Publishing Group. https://doi.org/10.1038/srep33735.

Anderson, M P, and J W Gronwald. 1991. “Atrazine Resistance in a Velvetleaf (Abutilon Theophrasti) Biotype Due to Enhanced Glutathione S-Transferase Activity.” Plant Physiol 96 (1): 104–109. https://doi.org/10.1104/pp.96.1.104.

Andrews, Kimberly R, Jeffrey M Good, Michael R Miller, Gordon Luikart, and Paul A Hohenlohe. 2016. “Harnessing the Power of RADseq for Ecological and Evolutionary Genomics.” Nature Reviews Genetics 17 (2): 81–92. https://doi.org/10.1038/nrg.2015.28.

Andrews. 2010. FastQC A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ansorge, Wilhelm, Brian Sproat, Josef Stegemann, Christian Schwager, and Martin Zenke. 1987. “Automated DNA Sequencing: Ultrasensitive Detection of Fluorescent Bands during Electrophoresis.” Nucleic Acids Research 15 (11): 4593–4602. https://doi.org/10.1093/nar/15.11.4593.

Ansorge, Wilhelm, Brian S. Sproat, Josef Stegemann, and Christian Schwager. 1986. “A Non-Radioactive Automated Method for DNA Sequence Determination.” Journal of Biochemical and Biophysical Methods 13 (6): 315–23. https://doi.org/10.1016/0165-022X(86)90038-2.

Anthony, Richard G., Teresa R. Waldin, John A. Ray, SimonW. J. Bright, and Patrick J. Hussey. 1998. “Herbicide Resistance Caused by Spontaneous Mutation of the Cytoskeletal Protein Tubulin.” Letters To Nature 393 (May): 6–9.

Aquadro, Charles F., Susan F. Desse, Molly M. Bland, Charles H. Langley, and Cathy C. Laurie-Ahlberg. 1986. “Molecular Population Genetics of the Alcohol Dehydrogenase Gene Region of DROSOPHILA MELANOGASTER.” Genetics 114 (4): 1165–90.

Arojju, Sai Krishna, Patrick Conaghan, Susanne Barth, Dan Milbourne, Michael D. Casler, Trevor R. Hodkinson, Thibauld Michel, and Stephen L. Byrne. 2018. “Genomic Prediction of Crown Rust Resistance in Lolium Perenne.” BMC Genetics 19 (1): 35. https://doi.org/10.1186/s12863-018-0613-z.

Ashworth, Mark, Patrick White, and David Armstrong. 2014. “Dangerous Ideas: Resistant Superbugs and the Potential of the Microbiome.” The British Journal of General Practice 64 (629): 637. https://doi.org/10.3399/bjgp14X682921.

Asoro, Franco G., Mark A. Newell, William D. Beavis, M. Paul Scott, and Jean-Luc Jannink. 2011. “Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats.” The Plant Genome 4 (2): 132–44. https://doi.org/10.3835/plantgenome2011.02.0007.

Atwell, Susanna, Yu S. Huang, Bjarni J. Vilhjálmsson, Glenda Willems, Matthew Horton, Yan Li, Dazhe Meng, et al. (2010). “Genome-Wide Association Study of 107 Phenotypes in Arabidopsis Thaliana Inbred Lines.” *Nature* 465, no. 7298: 627–31. https://doi.org/10.1038/nature08800.

Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty. 1944. “Studies on the chemical nature of the substance inducing transformation of Pneumococcal types.” The Journal of Experimental Medicine 79 (2): 137–58.

Babineau, Marielle, Solvejg K Mathiassen, Michael Kristensen, and Per Kudsk. 2017. “Fitness of ALS-Inhibitors Herbicide Resistant Population of Loose Silky Bentgrass (Apera Spica-Venti).” *Frontiers in Plant Science* 8: 1660. https://doi.org/10.3389/fpls.2017.01660.

Baerson, Scott R. 2002. “Glyphosate-Resistant Goosegrass. Identification of a Mutation in the Target Enzyme 5-Enolpyruvylshikimate-3-Phosphate Synthase.” Plant Physiology 129 (3): 1265–1275. https://doi.org/10.1104/pp.001560.

Baerson, Scott R, Damian J Rodriguez, Nancy A Biest, Minhtien Tran, Roger W Kreuger, Gerald M Dill, James E Pratley, et al. 2002. “Investigating the Mechanism of Glyphosate Resistance in Rigid Ryegrass (Lolium Ridigum) Published by : Weed Science Society of America Stable URL : Http://Www.Jstor.Org/Stable/4046644 REFERENCES Linked References Are Available on JSTOR for This Article.” Weed Science 50 (6): 721–730.

Bailey, Dale L., David W. Townsend, Peter E. Valk, and Michael N. Maisey, eds. 2005. Positron Emission Tomography: Basic Sciences. London: Springer-Verlag. https://doi.org/10.1007/b136169.

Bailey-Serres, Julia, and Ron Mittler. 2006. “The Roles of Reactive Oxygen Species in Plant Cells.” Plant Physiology 141 (2): 311. https://doi.org/10.1104/pp.104.900191.

Baird, Nathan A., Paul D. Etter, Tressa S. Atwood, Mark C. Currey, Anthony L. Shiver, Zachary A. Lewis, Eric U. Selker, William A. Cresko, and Eric A. Johnson. 2008. “Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers.” Edited by Justin C. Fay. PLoS ONE 3 (10): e3376. https://doi.org/10.1371/journal.pone.0003376.

Bai, Shuang, Fengwen Zhang, Zuren Li, Hengzhi Wang, Qian Wang, Jinxin Wang, Weitang Liu, and Lianyang Bai. 2019. “Target-Site and Non-Target-Site-Based Resistance to Tribenuron-Methyl in Multiply-Resistant Myosoton Aquaticum L.” Pesticide Biochemistry and Physiology 155 (March): 8–14. https://doi.org/10.1016/j.pestbp.2018.12.004.

Baker, H. G. 1965. “Characteristics and Modes of Origin of Weeds.” Characteristics and Modes of Origin of Weeds., 147–72.

Ball, Roderick D. 2013. “Designing a GWAS: Power, Sample Size, and Data Structure.” In Genome-Wide Association Studies and Genomic Prediction, edited by Cedric Gondro, Julius van der Werf, and Ben Hayes, 37–98. Methods in Molecular Biology. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-62703-447-0_3.

Banks, John, and Graeme Sandral. 2007. “Report on weed control using hot water / steam and herbicides in the city of Joondalup.” http://api.joondalup.wa.gov.au/files/councilmeetings/2007/attach11brf210807.pdf.

Bansal, Vikas. 2010. “A Statistical Method for the Detection of Variants from Next-Generation Resequencing of DNA Pools.” Bioinformatics (Oxford, England) 26 (12): i318–24. https://doi.org/10.1093/bioinformatics/btq214.

Baquero, Fernando, Val F Lanza, Rafael Cantón, and Teresa M Coque. 2015. “Public Health Evolutionary Biology of Antimicrobial Resistance: Priorities for Intervention.” Evolutionary Applications 8 (3): 223–39. https://doi.org/10.1111/eva.12235.

Barbitoff, Yury A., Dmitrii E. Polev, Andrey S. Glotov, Elena A. Serebryakova, Irina V. Shcherbakova, Artem M. Kiselev, Anna A. Kostareva, Oleg S. Glotov, and Alexander V. Predeus. 2020. “Systematic Dissection of Biases in Whole-Exome and Whole-Genome Sequencing Reveals Major Determinants of Coding Sequence Coverage.” Scientific Reports 10 (1): 1–13. https://doi.org/10.1038/s41598-020-59026-y.

Barrett, Rowan D. H., and Dolph Schluter. 2008. “Adaptation from Standing Genetic Variation.” Trends in Ecology & Evolution 23 (1): 38–44. https://doi.org/10.1016/j.tree.2007.09.008.

Barton, N. H. 2010. “Mutation and the Evolution of Recombination.” Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1544): 1281–94. https://doi.org/10.1098/rstb.2009.0320.

Bastide, Héloïse, Andrea Betancourt, Viola Nolte, Raymond Tobler, Petra Stöbe, Andreas Futschik, and Christian Schlötterer. 2013. “A Genome-Wide, Fine-Scale Map of Natural Pigmentation Variation in Drosophila Melanogaster.” Edited by Patricia Wittkopp. PLoS Genetics 9 (6): e1003534. https://doi.org/10.1371/journal.pgen.1003534.

Battlay, Paul, Joshua M Schmidt, Alexandre Fournier-Level, and Charles Robin. 2016. “Genomic and Transcriptomic Associations Identify a New Insecticide Resistance Phenotype for the Selective Sweep at the Cyp6g1 Locus of Drosophila Melanogaster.” G3 (Bethesda, Md.). https://doi.org/10.1534/g3.116.031054.

Baucom, Regina S. 2016. “The Remarkable Repeated Evolution of Herbicide Resistance.” American Journal of Botany 103 (2): 181–183. https://doi.org/10.3732/ajb.1500510.

Baucom, Regina S. 2019. Evolutionary and Ecological Insights from Herbicide-Resistant Weeds: What Have We Learned about Plant Adaptation, and What Is Left to Uncover? https://doi.org/10.1111/nph.15723.

Bayes, Mr., and Mr. Price. 1763. “An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.” Philosophical Transactions (1683-1775) 53: 370–418.

Beaumont, Mark A. 2019. “Approximate Bayesian Computation.” Annual Review of Statistics and Its Application 6 (1): 379–403. https://doi.org/10.1146/annurev-statistics-030718-105212.

Beaumont, Mark A, Wenyang Zhang, and David J Balding. 2002. “Approximate Bayesian Computation in Population Genetics.” Genetics 162 (4): 2025–2035.

Beckie, Hugh J., and François J. Tardif. 2012. “Herbicide Cross Resistance in Weeds.” Crop Protection 35 (May): 15–28. https://doi.org/10.1016/J.CROPRO.2011.12.018.

Beckie, Hugh J., Ian M. Heap, Reid J. Smeda, and Linda M. Hall. 2000. “Screening for Herbicide Resistance in Weeds.” Weed Technology 14 (2): 428–45. https://doi.org/10.1614/0890-037X(2000)014[0428:SFHRIW]2.0.CO;2.

Beckie, Hugh J., Roberto Busi, Muthukumar V. Bagavathiannan, and Sara L. Martin. 2019. “Herbicide Resistance Gene Flow in Weeds: Under-Estimated and under-Appreciated.” Agriculture, Ecosystems and Environment 283 (June): 1–11. https://doi.org/10.1016/j.agee.2019.06.005.

Begum, Hasina, Jennifer E. Spindel, Antonio Lalusin, Teresita Borromeo, Glenn Gregorio, Jose Hernandez, Parminder Virk, Bertrand Collard, and Susan R. McCouch. 2015. “Genome-Wide Association Mapping for Yield and Other Agronomic Traits in an Elite Breeding Population of Tropical Rice (Oryza Sativa).” PLoS ONE 10 (3). https://doi.org/10.1371/journal.pone.0119873.

Beissinger, Timothy M, Candice N Hirsch, Brieanne Vaillancourt, Shweta Deshpande, Kerrie Barry, C Robin Buell, Shawn M Kaeppler, Daniel Gianola, and Natalia de Leon. 2014. “A Genome-Wide Scan for Evidence of Selection in a Maize Population under Long-Term Artificial Selection for Ear Number.” Genetics 196 (3): 829–40. https://doi.org/10.1534/genetics.113.160655.

Bélanger, Sébastien, Patricio Esteves, Isabelle Clermont, Martine Jean, and François Belzile. 2016. “Genotyping-by-Sequencing on Pooled Samples and Its Use in Measuring Segregation Bias during the Course of Androgenesis in Barley.” The Plant Genome 9 (1): 0. https://doi.org/10.3835/plantgenome2014.10.0073.

Belz, Regina G. 2007. “Allelopathy in Crop/Weed Interactions — an Update.” Pest Management Science 63 (4): 308–26. https://doi.org/10.1002/ps.1320.

Benirschke, Kurt. 2004. “Francis Galton: Pioneer of Heredity and Biometry.” Journal of Heredity 95 (3): 273–273. https://doi.org/10.1093/jhered/esh039.

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Controlling the False Discovery Rate: A Practical and

Powerful Approach to Multiple Testing." Source Journal of the Royal Statistical Society. Series B (Methodological) Journal of the Royal Statistical Society. Series B J. R. Statist. Soc. B 57 (1): 289–300.

Bennett, O. L., D. A. Ashley, and B. D. Doss. 1966. "Cotton Responses to Black Plastic Mulch and Irrigation1." Agronomy Journal 58 (1): 57–60. https://doi.org/10.2134/agronj1966.00021962005800010019x.

Bentley, David R., Shankar Balasubramanian, Harold P. Swerdlow, Geoffrey P. Smith, John Milton, Clive G. Brown, Kevin P. Hall, et al. 2008. "Accurate Whole Human Genome Sequencing Using Reversible Terminator Chemistry." Nature 456 (7218): 53–59. https://doi.org/10.1038/nature07517.

Berger, Abi. 2002. "Magnetic Resonance Imaging." BMJ : British Medical Journal 324 (7328): 35.

Bernal-Vasquez, AM., Möhring, J., Schmidt, M., Schönleben, M., Schön, S.-C., & Piepho, H.-P. (2014). The importance of phenotypic data analysis for genomic prediction - a case study comparing different spatial models in rye. *BMC Genomics 15, 646* . https://doi.org/10.1186/1471-2164-15-646

Bertholdsson, N.-O. 2005. "Early Vigour and Allelopathy – Two Useful Traits for Enhanced Barley and Wheat Competitiveness against Weeds." Weed Research 45 (2): 94–102. https://doi.org/10.1111/j.1365-3180.2004.00442.x.

Boake, Christine R. B., Stevan J. Arnold, Felix Breden, Lisa M. Meffert, Michael G. Ritchie, Barbara J. Taylor, Jason B. Wolf, and Allen J. Moore. 2002. "Genetic Tools for Studying Adaptation and the Evolution of Behavior." The American Naturalist 160 (S6): S143–59. https://doi.org/10.1086/342902.

Bock, C. H., G. H. Poole, P. E. Parker, and T. R. Gottwald. 2010. "Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging." Critical Reviews in Plant Sciences 29 (2): 59–107. https://doi.org/10.1080/07352681003617285.

Boitard, S., C. Schlotterer, V. Nolte, R. V. Pandey, and A. Futschik. 2012. "Detecting Selective Sweeps from Pooled Next-Generation Sequencing Samples." Molecular Biology and Evolution 29 (9): 2177–2186. https://doi.org/10.1093/molbev/mss090.

Bonferroni, Carlo Emilio. 1936. "Teoria Statistica Delle Classi e Calcolo Delle Probabilità." Pubblicazioni Del R Istituto Superiore Di Scienze Economiche e Commerciali Di Firenze 8: 3–62. https://doi.org/10.4135/9781412961288.n455.

Botstein, D, R L White, M Skolnick, and R W Davis. 1980. "Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms." American Journal of Human Genetics 32 (3): 314–31.

Bouchet, Sophie, Marcus O. Olatoye, Sandeep R. Marla, Ramasamy Perumal, Tesfaye Tesso, Jianming Yu, Mitch Tuinstra, and Geoffrey P. Morris. 2017. "Increased Power To Dissect Adaptive Traits in Global Sorghum Diversity Using a Nested Association Mapping Population." Genetics 206 (2): 573–85. https://doi.org/10.1534/genetics.116.198499.

Boutsalis, Peter, Gurjeet S. Gill, and Christopher Preston. 2012. "Incidence of Herbicide Resistance in Rigid Ryegrass (*Lolium rigidum*) across Southeastern Australia." *Weed Technology* 26, no. 3 (July 2012): 391–98. https://doi.org/10.1614/WT-D-11-00150.1.

Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. 2017. "An Expanded View of Complex Traits: From Polygenic to Omnigenic." Cell 169 (7): 1177–86. https://doi.org/10.1016/j.cell.2017.05.038.

Breusegem, Frank Van, Julia Bailey-Serres, and Ron Mittler. 2008. "Unraveling the Tapestry of Networks Involving Reactive Oxygen Species in Plants." Plant Physiology 147 (3): 978–84. https://doi.org/10.1104/pp.108.122325.

Bridges, David C. 1992. Crop Losses Due to Weeds in the United States, 1992. Weed Science Society of America. http://agris.fao.org/agris-search/search.do?recordID=US9521216.

Bridle, Jon R., & Vines, Timothy H., (2007). Limits to evolution at range margins: when and why does adaptation fail?. *Trends in Ecology & Evolution. Volume 22, Issue 3*, pp 140-147. ISSN 0169-5347. https://doi.org/10.1016/j.tree.2006.11.002.

Brodersen, Jakob, P. Anders Nilsson, Lars-Anders Hansson, Christian Skov, & Christer Brönmark. (2008). "Condition-Dependent Individual Decision-Making Determines Cyprinid Partial Migration." *Ecology* 89, no. 5: 1195–1200. https://doi.org/10.1890/07-1318.1.

Brosnan, James T., Jose J. Vargas, Eric H. Reasor, Roberto Viggiani, Gregory K. Breeden, and John M. Zobel. 2017. "A Diagnostic Assay to Detect Herbicide Resistance in Annual

Bluegrass (Poa Annua).” *Weed Technology* 31 (04): 609–616. https://doi.org/10.1017/wet.2017.26.

Brouard, Jean-Simon, Brian Boyle, Eveline M. Ibeagha-Awemu, and Nathalie Bissonnette. 2017. “Low-Depth Genotyping-by-Sequencing (GBS) in a Bovine Population: Strategies to Maximize the Selection of High Quality Genotypes and the Accuracy of Imputation.” BMC Genetics 18 (April). https://doi.org/10.1186/s12863-017-0501-y.

Brown, A. H. D., G. J. Lawrence, M. Jenkin, J. Douglass, and E. Gregory. 1989. “Linkage Drag in Backcross Breeding in Barley.” Journal of Heredity 80 (3): 234–39. https://doi.org/10.1093/oxfordjournals.jhered.a110841.

Brown, Hugh M. 1990. “Mode of Action, Crop Selectivity, and Soil Relations of the Sulfonylurea Herbicides.” Pesticide Science 29 (3): 263–81. https://doi.org/10.1002/ps.2780290304.

Bulmer, M.G. 1971. “The Effect of Selection on Genetic Variability.” The American Naturalist. https://www.journals.uchicago.edu/doi/10.1086/282718.

Burgos, Nilda R. 2015. “Whole-Plant and Seed Bioassays for Resistance Confirmation.” Weed Science, no. Special Issue: 152–165. https://doi.org/10.1614/WS-D-14-00019.1.

Busi, R, and S B Powles. 2009. “Evolution of Glyphosate Resistance in a Lolium Rigidum Population by Glyphosate Selection at Sublethal Doses.” Heredity 103 (4): 318–325. https://doi.org/10.1038/hdy.2009.64.

Busi, R, M M Vila-Aiub, and S B Powles. 2011. “Genetic Control of a Cytochrome P450 Metabolism-Based Herbicide Resistance Mechanism in Lolium Rigidum.” Heredity 106 (5): 817–824. https://doi.org/10.1038/hdy.2010.124.

Busi, Roberto, Marcelo Girotto, and Stephen B. Powles. 2016. “Response to Low-Dose Herbicide Selection in Self-Pollinated Avena Fatua.” Pest Management Science 72 (3). https://doi.org/10.1002/ps.4032.

Busi, Roberto, Paul Neve, and Stephen Powles. 2013. “Evolved Polygenic Herbicide Resistance in Lolium Rigidum by Low-Dose Herbicide Selection within Standing Genetic Variation.” Evolutionary Applications 6 (2): 231–242. https://doi.org/10.1111/j.1752-4571.2012.00282.x.

Bustos-Korts, D., Malosetti, M., Chapman S., Biddulph, B., & van Eeuwijk, F. (2016). Improvement of Predictive Ability by Uniform Coverage of the Target Genetic Space. *G3 Genes|Genomes|Genetics, Volume 6, Issue 11,* 1 November 2016, Pages 3733–3747, https://doi.org/10.1534/g3.116.035410

Byrd, Richard H., Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. 1995. “A Limited Memory Algorithm for Bound Constrained Optimization.” SIAM Journal on Scientific Computing 16 (5): 1190–1208. https://doi.org/10.1137/0916069.

Byrne, Stephen, Adrian Czaban, Bruno Studer, Frank Panitz, Christian Bendixen, and Torben Asp. 2013. “Genome Wide Allele Frequency Fingerprints (GWAFFs) of Populations via Genotyping by Sequencing.” Edited by Frederick CC. Leung. PLoS ONE 8 (3): e57438. https://doi.org/10.1371/journal.pone.0057438.

Byrne, Stephen L., Istvan Nagy, Matthias Pfeifer, Ian Armstead, Suresh Swain, Bruno Studer, Klaus Mayer, et al. 2015. “A Synteny-Based Draft Genome Sequence of the Forage Grass Lolium Perenne.” The Plant Journal 84 (4): 816–826. https://doi.org/10.1111/tpj.13037.

CABI. 2017. Lolium Rigidum. In: Invasive Species Compendium. www.cabi.org/isc.

Cai, Jinhai, Mamoru Okamoto, Judith Atieno, Tim Sutton, Yongle Li, and Stanley J. Miklavcic. 2016. “Quantifying the Onset and Progression of Plant Senescence by Color Image Analysis for High Throughput Applications.” Edited by Panagiotis Kalaitzis. *PLOS ONE* 11 (6): e0157102. https://doi.org/10.1371/journal.pone.0157102.

Caley, M. Julian, and Dolph Schluter. 1997. “The Relationship Between Local and Regional Diversity.” Ecology 78 (1): 70–80. https://doi.org/10.1890/0012-9658(1997)078[0070:TRBLAR]2.0.CO;2.

Callaway, Ewen. 2017. “New Concerns Raised over Value of Genome-Wide Disease Studies.” Nature News 546 (7659): 463. https://doi.org/10.1038/nature.2017.22152.

Camm, Benjamin. 2018. “The Feasibility of a Gene Drive in Lolium Rigidum to Remove Resistance to the Herbicide Sulfometuron.” Master’s Thesis, The University of Melbourne.

Cammen, Kristina M., Thomas F. Schultz, W. Don Bowen, Michael O. Hammill, Wendy B. Puryear, Jonathan Runstadler, Frederick W. Wenzel, Stephanie A. Wood, and Michael Kinnison. 2018. “Genomic Signatures of Population Bottleneck and Recovery in Northwest Atlantic Pinnipeds.” Ecology and Evolution 8 (13): 6599–6614. https://doi.org/10.1002/ece3.4143.

Capinera, John L. 2005. “Relationships between Insect Pests and Weeds: An Evolutionary Perspective.” Weed Science 53 (6): 892–901.

Cappa, Eduardo P., Yousry A. El-Kassaby, Martín N. Garcia, Cintia Acuña, Nuno M. G. Borralho, Dario Grattapaglia, and Susana N. Marcucci Poltri. 2013. “Impacts of Population Structure and Analytical Models in Genome-Wide Association Studies of Complex Traits in Forest Trees: A Case Study in Eucalyptus Globulus.” PLOS ONE 8 (11): e81267. https://doi.org/10.1371/journal.pone.0081267.

Carter, L. M., R. F. Colwick, and J. R. Tavernetti. 1960. “Evaluating flame-burner design for weed control in cotton.” Trans. Amer. Soc. Agric. Eng. 3 (2): 125–28.

Castle, W. E. 1921. “An Improved Method of Estimating the Number of Genetic Factors Concerned in Cases of Blending Inheritance.” Science 54 (1393): 223–223. https://doi.org/10.1126/science.54.1393.223.

Cerrudo, Diego, Shiliang Cao, Yibing Yuan, Carlos Martinez, Edgar Antonio Suarez, Raman Babu, Xuecai Zhang, and Samuel Trachsel. 2018. “Genomic Selection Outperforms Marker Assisted Selection for Grain Yield and Physiological Traits in a Maize Doubled Haploid Population Across Water Treatments.” Frontiers in Plant Science 9. https://doi.org/10.3389/fpls.2018.00366.

Champer, Jackson, Anna Buchman, and Omar S. Akbari. 2016. “Cheating Evolution: Engineering Gene Drives to Manipulate the Fate of Wild Populations.” Nature Reviews Genetics 17 (3): 146–59. https://doi.org/10.1038/nrg.2015.34.

Chao, Ruey-Ming, Hsien-Chu Wu, and Zi-Chun Chen. 2009. “Image Segmentation by Automatic Histogram Thresholding.” In Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, 136–141. ICIS ’09. Seoul, Korea: Association for Computing Machinery. https://doi.org/10.1145/1655925.1655951.

Charlesworth, B, and D Charlesworth. 2017. “Population Genetics from 1966 to 2016.” Heredity 118 (1): 2–9. https://doi.org/10.1038/hdy.2016.55.

Charlesworth, B., D. Charlesworth, and M. T. Morgan. 1990. “Genetic Loads and Estimates of Mutation Rates in Highly Inbred Plant Populations.” Nature 347 (6291): 380–382. https://doi.org/10.1038/347380a0.

Charlesworth, Brian. 2012. “The Effects of Deleterious Mutations on Evolution at Linked Sites.” Genetics 190 (1): 5–22. https://doi.org/10.1534/genetics.111.134288.

Charng, Jason, Mark Simcoe, Paul G. Sanfilippo, R. Rand Allingham, Alex W. Hewitt, Chris J. Hammond, David A. Mackey, and Seyhan Yazar. 2020. “Age-Dependent Regional Retinal Nerve Fibre Changes in SIX1/SIX6 Polymorphism.” Scientific Reports 10 (July). https://doi.org/10.1038/s41598-020-69524-8.

Charudattan, Raghavan, and Amos Dinoor. 2000. “Biological Control of Weeds Using Plant Pathogens: Accomplishments and Limitations.” Crop Protection, XIVth International Plant Protection Congress, 19 (8): 691–95. https://doi.org/10.1016/S0261-2194(00)00092-2.

Cheng, Changde, Bradley J. White, Colince Kamdem, Keithanne Mockaitis, Carlo Costantini, Matthew W. Hahn, and Nora J. Besansky. 2012. “Ecological Genomics of Anopheles Gambiae along a Latitudinal Cline: A Population-Resequencing Approach.” Genetics 190 (4): 1417–1432. https://doi.org/10.1534/genetics.111.137794.

Cheng, H. D., X. H. Jiang, Y. Sun, and Jingli Wang. 2001. “Color Image Segmentation: Advances and Prospects.” Pattern Recognition 34 (12): 2259–81. https://doi.org/10.1016/S0031-3203(00)00149-7.

Chen, Hanbo, and Paul C. Boutros. 2011. “VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R.” BMC Bioinformatics 12 (1): 35. https://doi.org/10.1186/1471-2105-12-35.

Chen, Junjie, and Xinghua Shi. 2019. “Sparse Convolutional Denoising Autoencoders for Genotype Imputation.” Genes 10 (9): 652. https://doi.org/10.3390/genes10090652.

Christ, Bastien, and Stefan Hörtensteiner. 2014. “Mechanism and Significance of Chlorophyll Breakdown.” Journal of Plant Growth Regulation 33 (1): 4–20. https://doi.org/10.1007/s00344-013-9392-y.

Christoffers, Michael J, Michelle L Berg, and Calvin G Messersmith. 2002. “An Isoleucine to Leucine Mutation in Acetyl-CoA Carboxylase Confers Herbicide Resistance in Wild Oat.” Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada 45: 1049–1056. https://doi.org/10.1139/g02-080.

Clarke, Geraldine M., and Lon R. Cardon. 2005. “Disentangling Linkage Disequilibrium and Linkage From Dense Single-Nucleotide Polymorphism Trio Data.” Genetics 171 (4):

2085–95. https://doi.org/10.1534/genetics.105.047431.
Clarke, Joseph D. 2009. “Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation.” Cold Spring Harbor Protocols 2009 (3): pdb.prot5177. https://doi.org/10.1101/pdb.prot5177.
Clavel, Marion, Thierry Pélissier, Thomas Montavon, Marie-Aude Tschopp, Marie-Noëlle Pouch-Pélissier, Julie Descombin, Viviane Jean, Patrice Dunoyer, Cécile Bousquet-Antonelli, and Jean-Marc Deragon. 2016. “Evolutionary History of Double-Stranded RNA Binding Proteins in Plants: Identification of New Cofactors Involved in EasiRNA Biogenesis.” Plant Molecular Biology 91 (1–2): 131–47. https://doi.org/10.1007/s11103-016-0448-9.
Cochran, William G. 1954. “Some Methods for Strengthening the Common X2 Tests.” Biometrics 10 (4): 417–51. https://doi.org/10.2307/3001616.
Colosimo, Pamela F., Kim E. Hosemann, Sarita Balabhadra, Guadalupe Villarreal, Mark Dickson, Jane Grimwood, Jeremy Schmutz, Richard M. Myers, Dolph Schluter, and David M. Kingsley. 2005. “Widespread Parallel Evolution in Sticklebacks by Repeated Fixation of Ectodysplasin Alleles.” Science (New York, N.Y.) 307 (5717): 1928–33. https://doi.org/10.1126/science.1107239.
Condon, Greg. 2020. “New Chemistry – What’s New, What’s Coming and How to Keep Them for Longer.” New chemistry – what’s new, what’s coming and how to keep them for longer. Grains Research and Development Corporation. Grains Research and Development Corporation. February 25, 2020. https://grdc.com.au/resources-and-publications/grdc-update-papers/tab-content/grdc-update-papers/2020/02/new-chemistry-whats-new,-whats-coming-and-how-to-keep-them-for-longer.
Congreve, Mark, and John Cameron. 2014. *Soil Behaviour of Pre-Emergent Herbicides in Australian Farming Systems – Reference Manual for Advisers*. Edited by M Congreve and J Cameron. GRDC publication, Australia.
Contreras-Medina, Luis M, Roque A Osornio-Rios, Irineo Torres-Pacheco, Rene de J Romero-Troncoso, Ramon G Guevara-González, and Jesus R Millan-Almaraz. 2012. “Smart Sensor for Real-Time Quantification of Common Symptoms Present in Unhealthy Plants.” *Sensors (Basel, Switzerland)* 12 (1): 784–805. https://doi.org/10.3390/s120100784.
Cotto, O., Wessely, J., Georges, D., Klonner, G., Schmid, M., Dullinger, S., Thuiller, W., & Guillaume, F.. (2017). A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. *Nat Commun* **8,** 15399. https://doi.org/10.1038/ncomms15399
Croom, Edward. 2012. “Chapter Three - Metabolism of Xenobiotics of Human Environments.” In Progress in Molecular Biology and Translational Science, edited by Ernest Hodgson, 112:31–88. Toxicology and Human Environments. Academic Press. https://doi.org/10.1016/B978-0-12-415813-9.00003-9.
Crossa, José, Paulino Pérez-Rodríguez, Jaime Cuevas, Osval Montesinos-López, Diego Jarquín, Gustavo de los Campos, Juan Burgueño, et al. 2017. “Genomic Selection in Plant Breeding: Methods, Models, and Perspectives.” Trends in Plant Science 22 (11): 961–75. https://doi.org/10.1016/j.tplants.2017.08.011.
Crow, James F. 1997. “The High Spontaneous Mutation Rate: Is It a Health Risk?” Proceedings of the National Academy of Sciences of the United States of America 94 (16): 8380–86.
Crow, James F., and Motoo Kimura. 1970. An Introduction to Population Genetics Theory. https://doi.org/10.2307/1529706.
Crutchfield, Donald A., Gail A. Wicks, and Orvin C. Burnside. 1986. “Effect of Winter Wheat (Triticum Aestivum) Straw Mulch Level on Weed Control.” Weed Science 34 (1): 110–14. https://doi.org/10.1017/S0043174500026564.
Cui, Feng, Mei-Xia Li, Hai-Jing Chang, Yun Mao, Han-Ying Zhang, Li-Xia Lu, Shuai-Guo Yan, Ming-Lin Lang, Li Liu, and Chuan-Ling Qiao. 2015. “Carboxylesterase-Mediated Insecticide Resistance: Quantitative Increase Induces Broader Metabolic Resistance than Qualitative Change.” Pesticide Biochemistry and Physiology 121 (June): 88–96. https://doi.org/10.1016/j.pestbp.2014.12.016.
Cumer, Tristan, Charles Pouchon, Frédéric Boyer, Glenn Yannic, Delphine Rioux, Aurélie Bonin, and Thibaut Capblancq. 2018. “Double-Digest RAD-Sequencing: Do Wet and Dry Protocol 1 Parameters Impact Biological Results?” BioRxiv 33 (0). https://doi.org/10.1101/445122.
Curtis, C. F. 1968. “Possible Use of Translocations to Fix Desirable Genes in Insect Pest Populations.” Nature 218 (5139): 368–69. https://doi.org/10.1038/218368a0.
Cutler, David J, and Jeffrey D Jensen. 2010. “To Pool, or Not to Pool?” Genetics 186 (1): 41–3.

https://doi.org/10.1534/genetics.110.121012.
Dai, Xiaowen, Libin Jin, Anqi Shi, and Lei Shi. 2016. “Outlier Detection and Accommodation in General Spatial Models.” Statistical Methods & Applications 25 (3): 453–75. https://doi.org/10.1007/s10260-015-0348-1.
Darwin, Charles. 1859. On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life. D. Appleton and Company. http://darwin-online.org.uk/converted/pdf/1861_OriginNY_F382.pdf.
Davey, John W., Paul A. Hohenlohe, Paul D. Etter, Jason Q. Boone, Julian M. Catchen, and Mark L. Blaxter. 2011. “Genome-Wide Genetic Marker Discovery and Genotyping Using next-Generation Sequencing.” Nature Reviews Genetics 12 (7): 499–510. https://doi.org/10.1038/nrg3012.
Davey, John W., Timothée Cezard, Pablo Fuentes-Utrilla, Cathlene Eland, Karim Gharbi, and Mark L. Blaxter. 2013. “Special Features of RAD Sequencing Data: Implications for Genotyping.” Molecular Ecology 22 (11): 3151–3164. https://doi.org/10.1111/mec.12084.
Davies, Julian, and Dorothy Davies. 2010. “Origins and Evolution of Antibiotic Resistance.” Microbiology and Molecular Biology Reviews 74 (3): 417–33. https://doi.org/10.1128/MMBR.00016-10.
Délye, Christophe. 2013. “Unravelling the Genetic Bases of Non-Target-Site-Based Resistance (NTSR) to Herbicides: A Major Challenge for Weed Science in the Forthcoming Decade.” Pest Management Science 69 (2): 176–187. https://doi.org/10.1002/ps.3318.
Délye, Christophe, Annick Matéjicek, and Séverine Michel. 2008. “Cross-Resistance Patterns to ACCase-Inhibiting Herbicides Conferred by Mutant ACCase Isoforms in Alopecurus Myosuroides Huds. (Black-Grass), Re-Examined at the Recommended Herbicide Field Rate.” Pest Management Science 64 (11): 1179–1186. https://doi.org/10.1002/ps.1614.
Délye, Christophe, K Boucansaud, F Pernin, and V Le Corre. 2009. “Variation in the Gene Encoding Acetolactate-Synthase in Lolium Species and Proactive Detection of Mutant, Herbicide-Resistant Alleles.” Weed Research 49 (3): 326–336. https://doi.org/10.1111/j.1365-3180.2009.00700.x.
Délye, Christophe, Marie Jasieniuk, and Valérie Le Corre. 2013. “Deciphering the Evolution of Herbicide Resistance in Weeds.” Trends in Genetics 29 (11): 649–658. https://doi.org/10.1016/j.tig.2013.06.001.
Délye, Christophe, Yosra Menchari, Séverine Michel, and Henri Darmency. 2004. “Molecular Bases for Sensitivity to Tubulin-Binding Herbicides in Green Foxtail.” Plant Physiology 136 (4): 3920. https://doi.org/10.1104/pp.103.037432.
Délye, C., J. a. C. Gardin, K. Boucansaud, B. Chauvel, and C. Petit. 2011. “Non-Target-Site-Based Resistance Should Be the Centre of Attention for Herbicide Resistance Research: Alopecurus Myosuroides as an Illustration.” Weed Research 51 (5): 433–37. https://doi.org/10.1111/j.1365-3180.2011.00864.x.
D’Emden, Francis H., Rick S. Llewellyn, and Michael P. Burton. 2006. “Adoption of Conservation Tillage in Australian Cropping Regions: An Application of Duration Analysis.” Technological Forecasting and Social Change 73 (6): 630–47. https://doi.org/10.1016/j.techfore.2005.07.003.
De Prado, R, J Jorrin, and L Garcia-Torres, eds. 1997. Weed and Crop Resistance. SPRINGER SCIENCE+BUSINESS MEDIA, B.V.
Derde, Lennie P G, Ben S Cooper, Herman Goossens, Surbhi Malhotra-Kumar, Rob J L Willems, Marek Gniadkowski, Waleria Hryniewicz, et al. 2014. “Interventions to Reduce Colonisation and Transmission of Antimicrobial-Resistant Bacteria in Intensive Care Units: An Interrupted Time Series Study and Cluster Randomised Trial.” The Lancet Infectious Diseases 14 (1): 31–39. https://doi.org/10.1016/S1473-3099(13)70295-0.
Devlin, B., and Kathryn Roeder. 1999. “Genomic Control for Association Studies.” Biometrics 55 (4): 997–1004. https://doi.org/10.1111/j.0006-341X.1999.00997.x.
Dexter, John S. 1914. “The Analysis of a Case of Continuous Variation in Drosophila by a Study of Its Linkage Relations.” The American Naturalist 48 (576): 712–58. https://doi.org/10.1086/279446.
Dinelli, G, I Marotti, P Catizone, A Bonetti, J M Urbano, and J Barnes. 2008. “Physiological and Molecular Basis of Glyphosate Resistance in C. Bonariensis (L.) Cronq. Biotypes from Spain.” Weed Research 48 (January): 257–265.
Dohm, Juliane C, Claudio Lottaz, Tatiana Borodina, and Heinz Himmelbauer. 2008. “Substantial Biases in Ultra-Short Read Data Sets from High-Throughput DNA Sequencing.” Nucleic Acids Research 36 (16): e105. https://doi.org/10.1093/nar/gkn425.

Domingues, Vera S., Yu-Ping Poh, Brant K. Peterson, Pleuni S. Pennings, Jeffrey D. Jensen, and Hopi E. Hoekstra. 2012. “Evidence of Adaptation from Ancestral Variation in Young Populations of Beach Mice.” Evolution; International Journal of Organic Evolution 66 (10): 3209–23. https://doi.org/10.1111/j.1558-5646.2012.01669.x.

Dong, Wei, Defeng Wu, Guoshen Li, Dewei Wu, and Zicheng Wang. 2018. “Next-Generation Sequencing from Bulked Segregant Analysis Identifies a Dwarfism Gene in Watermelon.” Scientific Reports 8 (1): 2908. https://doi.org/10.1038/s41598-018-21293-1.

Dormann, Carsten F., Stanislaus J. Schymanski, Juliano Cabral, Isabelle Chuine, Catherine Graham, Florian Hartig, Michael Kearney, et al. 2012. “Correlation and Process in Species Distribution Models: Bridging a Dichotomy.” Journal of Biogeography 39 (12): 2119–31. https://doi.org/10.1111/j.1365-2699.2011.02659.x.

Drummond, Alexei J, Geoff K Nicholls, Allen G Rodrigo, and Wiremu Solomon. 2002. “Estimating Mutation Parameters, Population History and Genealogy Simultaneously From Temporally Spaced Sequence Data.” Genetics 161 (July): 1307–1320.

Dudbridge, Frank. 2013. “Power and Predictive Accuracy of Polygenic Risk Scores.” PLOS Genetics 9 (3): e1003348. https://doi.org/10.1371/journal.pgen.1003348.

Duforet-Frebourg, Nicolas, Keurcien Luu, Guillaume Laval, Eric Bazin, and Michael G. B. Blum. 2016. “Detecting Genomic Signatures of Natural Selection with Principal Component Analysis: Application to the 1000 Genomes Data.” Molecular Biology and Evolution 33 (4): 1082–93. https://doi.org/10.1093/molbev/msv334.

Duhoux, Arnaud, Sébastien Carrère, Alexis Duhoux, and Christophe Délye. 2017. “Transcriptional Markers Enable Identification of Rye-Grass (Lolium Sp.) Plants with Non-Target-Site-Based Resistance to Herbicides Inhibiting Acetolactate-Synthase.” *Plant Science* 257 (April 1, 2017): 22–36. https://doi.org/10.1016/j.plantsci.2017.01.009.

Dusfour, Isabelle, John Vontas, Jean-Philippe David, David Weetman, Dina M. Fonseca, Vincent Corbel, Kamaraju Raghavendra, et al. 2019. “Management of Insecticide Resistance in the Major Aedes Vectors of Arboviruses: Advances and Challenges.” PLOS Neglected Tropical Diseases 13 (10): e0007615. https://doi.org/10.1371/journal.pntd.0007615.

Eckert, Andrew J., and Rodney J. Dyer. 2012. “Defining the Landscape of Adaptive Genetic Diversity.” Molecular Ecology 21 (12): 2836–38. https://doi.org/10.1111/j.1365-294X.2012.05615.x.

Edwards, R., and M. Hannah. 2014. “Focus on Weed Control.” PLANT PHYSIOLOGY 166 (3): 1087–89. https://doi.org/10.1104/pp.114.900496.

Edwards, Stefan McKinnon, Jaap B. Buntjer, Robert Jackson, Alison R. Bentley, Jacob Lage, Ed Byrne, Chris Burt, et al. 2019. “The Effects of Training Population Design on Genomic Prediction Accuracy in Wheat.” Theoretical and Applied Genetics 132 (7): 1943–52. https://doi.org/10.1007/s00122-019-03327-y.

Egley, G. H., and J. M. Chandler. 1983. “Longevity of Weed Seeds After 5.5 Years in the Stoneville 50-Year Buried-Seed Study.” Weed Science 31 (2): 264–70. https://doi.org/10.1017/S0043174500068958.

Ekblom, R., and J. Galindo. 2011. “Applications of next Generation Sequencing in Molecular Ecology of Non-Model Organisms.” Heredity 107 (1): 1–15. https://doi.org/10.1038/hdy.2010.152.

Elena, Santiago F., Vaughn S. Cooper, and Richard E. Lenski. 1996. “Punctuated Evolution Caused by Selection of Rare Beneficial Mutations.” Science 272 (5269): 1802–4. https://doi.org/10.1126/science.272.5269.1802.

Ellegren, Hans. 2014. “Genome Sequencing and Population Genomics in Non-Model Organisms.” Trends in Ecology & Evolution 29 (1): 51–63. https://doi.org/10.1016/J.TREE.2013.09.008.

Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler, and Sharon E. Mitchell. 2011. “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.” Edited by Laszlo Orban. PLoS ONE 6 (5): e19379. https://doi.org/10.1371/journal.pone.0019379.

Endelman, Jeffrey B. 2011. “Ridge Regression and Other Kernels for Genomic Selection with R Package RrBLUP.” The Plant Genome 4 (3): 250–55. https://doi.org/10.3835/plantgenome2011.08.0024.

Endler, Lukas, Andrea J Betancourt, Viola Nolte, and Christian Schlötterer. 2016. “Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in Drosophila Melanogaster.” Genetics 202 (2): 843–55.

https://doi.org/10.1534/genetics.115.183376.
Epp, Jeffrey B, Paul R Schmitzer, and Gary D Crouse. 2018. “Fifty Years of Herbicide Research: Comparing the Discovery of Trifluralin and Halauxifen-Methyl.” *Pest Management Science* 74 (1): 9–16. https://doi.org/10.1002/ps.4657.
Es, H. M. van, and C. L. van Es. 1993. “Spatial Nature of Randomization and Its Effect on the Outcome of Field Experiments.” Agronomy Journal 85 (2): 420–28. https://doi.org/10.2134/agronj1993.00021962008500020046x.
Etcheverry, Angela Virginia, Maria Mercedes Alemán, and Trinidad Figueroa Fleming. (2008). “Flower Morphology, Pollination Biology and Mating System of the Complex Flower of *Vigna caracalla* (Fabaceae: Papilionoideae).” *Annals of Botany* 102, no. 3 (September 1, 2008): 305–16. https://doi.org/10.1093/aob/mcn106.
Etcheverry, A. V., J. J. Protomastro, and C. Westerkamp. (2003). “Delayed Autonomous Self-Pollination in the Colonizer Crotalaria Micans (Fabaceae: Papilionoideae): Structural and Functional Aspects.” *Plant Systematics and Evolution* 239, no. 1/2 (2003): 15–28.
Etten, Megan Van, Kristin M. Lee, Shu-Mei Chang, and Regina S. Baucom. 2020. “Parallel and Nonparallel Genomic Responses Contribute to Herbicide Resistance in Ipomoea Purpurea, a Common Agricultural Weed.” PLOS Genetics 16 (2): e1008593. https://doi.org/10.1371/journal.pgen.1008593.
Evans, Tim J. 2013. “Chapter 2 - Toxicokinetics and Toxicodynamics.” In Small Animal Toxicology (Third Edition), edited by Michael E. Peterson and Patricia A. Talcott, 13–19. Saint Louis: W.B. Saunders. https://doi.org/10.1016/B978-1-4557-0717-1.00002-8.
Ewing, Brent, and Phil Green. 1998. “Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities.” Genome Research 8 (3): 186–94. https://doi.org/10.1101/gr.8.3.186.
Exposito-Alonso, Moises, François Vasseur, Wei Ding, George Wang, Hernán A. Burbano, and Detlef Weigel. 2018. “Genomic Basis and Evolutionary Potential for Extreme Drought Adaptation in Arabidopsis Thaliana.” Nature Ecology & Evolution 2 (2): 352–58. https://doi.org/10.1038/s41559-017-0423-0.
Fadista, João, Alisa K. Manning, Jose C. Florez, and Leif Groop. 2016. “The (in)Famous GWAS P -Value Threshold Revisited and Updated for Low-Frequency Variants.” European Journal of Human Genetics 24 (8): 1202–5. https://doi.org/10.1038/ejhg.2015.269.
Fan, Ruzong, and Momiao Xiong. 2002. “High Resolution Mapping of Quantitative Trait Loci by Linkage Disequilibrium Analysis.” European Journal of Human Genetics 10 (10): 607–15. https://doi.org/10.1038/sj.ejhg.5200843.
Fardisi, Mahsa, Ameya D. Gondhalekar, Aaron R. Ashbrook, and Michael E. Scharf. 2019. “Rapid Evolutionary Responses to Insecticide Resistance Management Interventions by the German Cockroach (Blattella Germanica L.).” Scientific Reports 9 (1): 1–10. https://doi.org/10.1038/s41598-019-44296-y.
Fariello, María Inés, Simon Boitard, Sabine Mercier, David Robelin, Thomas Faraut, Cécile Arnould, Julien Recoquillay, et al. 2017. “Accounting for Linkage Disequilibrium in Genome Scans for Selection without Individual Genotypes: The Local Score Approach.” Molecular Ecology 26 (14): 3700–3714. https://doi.org/10.1111/mec.14141.
Fatumo, Segun, Tommy Carstensen, Oyekanmi Nashiru, Deepti Gurdasani, Manjinder Sandhu, and Pontiano Kaleebu. 2019. “Complimentary Methods for Multivariate Genome-Wide Association Study Identify New Susceptibility Genes for Blood Cell Traits.” Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00334.
Fawcett, Tom. 2006. “An Introduction to ROC Analysis.” Pattern Recognition Letters 27 (8): 861–74. https://doi.org/10.1016/j.patrec.2005.10.010.
Feder, Alison F., Dmitri A. Petrov, and Alan O. Bergland. 2012. “LDx: Estimation of Linkage Disequilibrium from High-Throughput Pooled Resequencing Data.” PLOS ONE 7 (11): e48588. https://doi.org/10.1371/journal.pone.0048588.
Fedurco, Milan, Anthony Romieu, Scott Williams, Isabelle Lawrence, and Gerardo Turcatti. 2006. “BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies.” Nucleic Acids Research 34 (3): e22. https://doi.org/10.1093/nar/gnj023.
Feng, Sheng, Shengchu Wang, Chia-Cheng Chen, and Lan Lan. 2011. “GWAPower: A Statistical Power Calculation Software for Genome-Wide Association Studies with Quantitative Traits.” BMC Genetics 12 (1): 12. https://doi.org/10.1186/1471-2156-12-12.
Ferretti, Luca, Sebastián E. Ramos-Onsins, and Miguel Pérez-Enciso. 2013. “Population Genomics from Pool Sequencing.” Molecular Ecology 22 (22): 5561–5576. https://doi.org/

10.1111/mec.12522.
Finch-Savage, William E., and Gerhard Leubner-Metzger. 2006. “Seed Dormancy and the Control of Germination.” New Phytologist 171 (3): 501–23. https://doi.org/10.1111/j.1469-8137.2006.01787.x.
Fischer, Bill B. 1966. “Effects of Post-Emergence Application of Trifluralin for Selective Weed Control in Cotton.” *California Agriculture*, no. March 1966: 10–11.
Fisher, R. A. 1922. “On the Interpretation of X2 from Contingency Tables, and the Calculation of P.” Journal of the Royal Statistical Society 85 (1): 87–94. https://doi.org/10.2307/2340521.
Fisher, Ronald A. 1918. “The Correlation Between Relatives on the Supposition of Mendelian Inheritance.” Transactions of the Royal Society of Edinburgh 52: 399–433.
Flatt, Thomas. 2020. “Life-History Evolution and the Genetics of Fitness Components in Drosophila Melanogaster.” Genetics 214 (1): 3–48. https://doi.org/10.1534/genetics.119.300160.
FAO (Food and Agriculture Organization of the United Nations). 2020. “FAOSTAT Statistical Database.” 2020. http://www.fao.org/faostat/en/#data/EP/visualize.
Forouzesh, A, E Zand, S Soufizadeh, and S Samadi Foroushani. 2015. “Classification of Herbicides According to Chemical Family for Weed Resistance Management Strategies - an Update.” Edited by Enrique Barriuso. Weed Research 55 (4): 334–358. https://doi.org/10.1111/wre.12153.
Fournier-Level, A., A. Korte, M. D. Cooper, M. Nordborg, J. Schmitt, and A. M. Wilczek. 2011. “A Map of Local Adaptation in Arabidopsis Thaliana.” Science 334 (6052): 86–89. https://doi.org/10.1126/science.1209271.
Fournier-Level, Alexandre, Charles Robin, and David J. Balding. 2017. “GWAlpha: Genome-Wide Estimation of Additive Effects (Alpha) Based on Trait Quantile Distribution from Pool-Sequencing Experiments.” Bioinformatics. https://doi.org/10.1093/bioinformatics/btw805.
Fournier-Level, Alexandre, Robert T. Good, Stephen A. Wilcox, Rahul V. Rane, Michelle Schiffer, Wei Chen, Paul Battlay, et al. 2019. “The Spread of Resistance to Imidacloprid Is Restricted by Thermotolerance in Natural Populations of Drosophila Melanogaster.” Nature Ecology & Evolution 3 (4): 647–56. https://doi.org/10.1038/s41559-019-0837-y.
Fracassetti, Marco, Philippa C Griffin, and Yvonne Willi. 2015. “Validation of Pooled Whole-Genome Re-Sequencing in Arabidopsis Lyrata.” PloS One 10 (10): e0140462. https://doi.org/10.1371/journal.pone.0140462.
Free Software Foundation. 2016. GNU Bash. https://www.gnu.org/software/bash/.
Frey, Steven K., Edward Topp, Izhar U.H. Khan, Bonnie R. Ball, Mark Edwards, Natalie Gottschall, Mark Sunohara, and David R. Lapen. 2015. “Quantitative Campylobacter Spp., Antibiotic Resistance Genes, and Veterinary Antibiotics in Surface and Ground Water Following Manure Application: Influence of Tile Drainage Control.” Science of The Total Environment 532 (November): 138–53. https://doi.org/10.1016/j.scitotenv.2015.03.114.
Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. “Regularization Paths for Generalized Linear Models via Coordinate Descent.” Journal of Statistical Software 33 (1): 1–22.
Friend, E. J., and D. A. Hopwood. 1971. “The Linkage Map of Streptomyces Rimosus.” Microbiology 68 (2): 187–97. https://doi.org/10.1099/00221287-68-2-187.
Fuller, Zachary L., Veronique J. L. Mocellin, Luke A. Morris, Neal Cantin, Jihanne Shepherd, Luke Sarre, Julie Peng, et al. (2020) “Population Genetics of the Coral Acropora Millepora: Toward Genomic Prediction of Bleaching.” *Science* 369, no. 6501 (July 17, 2020). https://doi.org/10.1126/science.aba4674.
Futschik, Andreas, and Christian Schlötterer. 2010. “The next Generation of Molecular Markers from Massively Parallel Sequencing of Pooled DNA Samples.” Genetics 186 (1): 207–18. https://doi.org/10.1534/genetics.110.114397.
Fu, Yun-Xin, and Haying Huai. 2003. “Estimating Mutation Rate: How to Count Mutations?” Genetics 164 (2): 797–805.
Gaines, Todd A., Lothar Lorentz, Andrea Figge, Johannes Herrmann, Frank Maiwald, Mark-Christoph Ott, Heping Han, et al. 2014. “RNA-Seq Transcriptome Analysis to Identify Genes Involved in Metabolism-Based Diclofop Resistance in Lolium Rigidum.” *The Plant Journal* 78, no. 5 (2014): 865–76. https://doi.org/10.1111/tpj.12514.
Gaines, T. A., W. Zhang, D. Wang, B. Bukun, S. T. Chisholm, D. L. Shaner, S. J. Nissen, et al.

2010. “Gene Amplification Confers Glyphosate Resistance in Amaranthus Palmeri.” Proceedings of the National Academy of Sciences 107 (3): 1029–1034. https://doi.org/10.1073/pnas.0906649107.

Gaines, Todd A., Lothar Lorentz, Andrea Figge, Johannes Herrmann, Frank Maiwald, Mark-Christoph Christoph Ott, Heping Han, et al. 2014. “RNA-Seq Transcriptome Analysis to Identify Genes Involved in Metabolism-Based Diclofop Resistance in Lolium Rigidum.” Plant Journal 78 (5): 865–876. https://doi.org/10.1111/tpj.12514.

Gaj, Pawel, Natalia Maryan, Ewa E. Hennig, Joanna K. Ledwon, Agnieszka Paziewska, Aneta Majewska, Jakub Karczmarski, et al. 2012. “Pooled Sample-Based GWAS: A Cost-Effective Alternative for Identifying Colorectal and Prostate Cancer Risk Variants in the Polish Population.” Edited by Kin Mang Lau. PLoS ONE 7 (4): e35307. https://doi.org/10.1371/journal.pone.0035307.

Gao, Xiaoyi, Joshua Starmer, and Eden R. Martin. 2008. “A Multiple Testing Correction Method for Genetic Association Studies Using Correlated Single Nucleotide Polymorphisms.” Genetic Epidemiology 32 (4): 361–69. https://doi.org/10.1002/gepi.20310.

Garbout, Amin, Lars J. Munkholm, Søren B. Hansen, Bjørn M. Petersen, Ole L. Munk, and Radoslaw Pajor. 2012. “The Use of PET/CT Scanning Technique for 3D Visualization and Quantification of Real-Time Soil/Plant Interactions.” Plant and Soil 352 (1): 113–27. https://doi.org/10.1007/s11104-011-0983-8.

García, Maria J., Candelario Palma-Bautista, Antonia M. Rojano-Delgado, Enzo Bracamonte, João Portugal, Ricardo Alcántara-de la Cruz, and Rafael De Prado. 2019. “The Triple Amino Acid Substitution TAP-IVS in the EPSPS Gene Confers High Glyphosate Resistance to the Superweed Amaranthus Hybridus.” *International Journal of Molecular Sciences* 20, no. 10 (May 15, 2019). https://doi.org/10.3390/ijms20102396.

García-Ramos, Gisela, and Mark Kirkpatrick. 1997. “Genetic Models of Adaptation and Gene Flow in Peripheral Populations.” Evolution 51 (1): 21–28. https://doi.org/10.1111/j.1558-5646.1997.tb02384.x.

García-Sánchez, Francisco, Luis Galvez-Sola, Juan J. Martínez-Nicolás, Raquel Muelas-Domingo, and Manuel Nieves. 2017. “Using Near-Infrared Spectroscopy in Agricultural Systems.” Developments in Near-Infrared Spectroscopy, March. https://doi.org/10.5772/67236.

Garg, N., and G. Manchanda. 2009. “ROS Generation in Plants: Boon or Bane?” Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology 143 (1): 81–96. https://doi.org/10.1080/11263500802633626.

Garland, Theodore Jr., and Michael Rose, eds. 2010. Experimental Evolution. Concepts, Methods, and Applications of Selection Experiments. 1st ed. Berkeley and Los Angeles, California: University of California Press. https://academic.oup.com/icb/article/50/5/909/642777.

Gautier, Mathieu, Julien Foucaud, Karim Gharbi, Timothée Cézard, Maxime Galan, Anne Loiseau, Marian Thomson, Pierre Pudlo, Carole Kerdelhué, and Arnaud Estoup. 2013. “Estimation of Population Allele Frequencies from Next-Generation Sequencing Data: Pool-versus Individual-Based Genotyping.” Molecular Ecology 22 (14): 3766–3779. https://doi.org/10.1111/mec.12360.

Geber, M.A. & Eckhart, V.M. (2005). Experimental studies of adaptation in *Clarkia xantiana*: II. Fitness variation across a subspecies border. *Evolution, 59: 521-531.* https://doi.org/10.1111/j.0014-3820.2005.tb01012.x

Gebreyesus, G., A. J. Buitenhuis, N. A. Poulsen, M. H. P. W. Visker, Q. Zhang, H. J. F. van Valenberg, D. Sun, and H. Bovenhuis. 2019. “Multi-Population GWAS and Enrichment Analyses Reveal Novel Genomic Regions and Promising Candidate Genes Underlying Bovine Milk Fatty Acid Composition.” BMC Genomics 20 (1): 178. https://doi.org/10.1186/s12864-019-5573-9.

Geman, S., and D. Geman. 1984. “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.” IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-6 (6): 721–41. https://doi.org/10.1109/TPAMI.1984.4767596.

Georges, Michel, Carole Charlier, and Ben Hayes. 2019. “Harnessing Genomic Information for Livestock Improvement.” Nature Reviews Genetics 20 (3): 135–56. https://doi.org/10.1038/s41576-018-0082-2.

Ge, Steven Xijin, Dongmin Jung, and Runan Yao. 2020. “ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants.” Edited by Alfonso Valencia. Bioinformatics 36 (8): 2628–29. https://doi.org/10.1093/bioinformatics/btz931.

Ge, Yufeng, Abbas Atefi, Huichun Zhang, Chenyong Miao, Raghuprakash Kastoori Ramamurthy, Brandi Sigmon, Jinliang Yang, and James C. Schnable. 2019. “High-Throughput Analysis of Leaf Physiological and Chemical Traits with VIS–NIR–SWIR Spectroscopy: A Case Study with a Maize Diversity Panel.” Plant Methods 15 (1): 66. https://doi.org/10.1186/s13007-019-0450-8.

Ghosal, Sambuddha, David Blystone, Asheesh K Singh, Baskar Ganapathysubramanian, Arti Singh, and Soumik Sarkar. 2018. “An Explainable Deep Machine Vision Framework for Plant Stress Phenotyping.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (18): 4613–4618. https://doi.org/10.1073/pnas.1716999115.

Gijsenij, Arjan, and Theo Gevers. 2011. “Color Constancy Using Natural Image Statistics and Scene Semantics.” IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (4): 687–98. https://doi.org/10.1109/TPAMI.2010.93.

Gillespie, John H. 1994. The Causes of Molecular Evolution. Oxford University Press.

Gill, Gurjeet. 1996. “Why Annual Ryegrass Is a Problem in Australian Agriculture.” Plant Protection Quarterly 11 (Supplement 1): 193–195.

Giovagnoli, Alessandra, and Paola Sebastiani. 1989. “Experimental Designs for Mean and Variance Estimation in Variance Components Models.” Computational Statistics & Data Analysis 8 (1): 21–28. https://doi.org/10.1016/0167-9473(89)90062-5.

Goggin, Danica E., Stephen B. Powles, and Kathryn J. Steadman. 2012. “Understanding Lolium Rigidum Seeds: The Key to Managing a Problem Weed?” Agronomy 2 (3): 222–239. https://doi.org/10.3390/agronomy2030222.

Gondro, Cedric, Julius van der Werf, and Ben Hayes, eds. 2013. Genome-Wide Association Studies and Genomic Prediction. 1st ed. Vol. 1019. Methods in Molecular Biology. Humana Press. https://doi.org/10.1007/978-1-62703-447-0_20.

González-Camacho, J. M., G. de los Campos, P. Pérez, D. Gianola, J. E. Cairns, G. Mahuku, R. Babu, and J. Crossa. 2012. “Genome-Enabled Prediction of Genetic Values Using Radial Basis Function Neural Networks.” Theoretical and Applied Genetics. https://doi.org/10.1007/s00122-012-1868-9.

Goodwin, Sara, John D. McPherson, and W. Richard McCombie. 2016. “Coming of Age: Ten Years of next-Generation Sequencing Technologies.” Nature Reviews Genetics 17 (6): 333–351. https://doi.org/10.1038/nrg.2016.49.

Gorddard, Russell J., David J. Pannell, and Greg Hertzler. 1995. “An Optimal Control Model for Integrated Weed Management Under Herbicide Resistance*.” Australian Journal of Agricultural Economics 39 (1): 71–87. https://doi.org/10.1111/j.1467-8489.1995.tb00544.x.

Goudet, J., & Büchi. (2006). The effects of dominance, regular inbreeding and sampling design on QST, an estimator of population differentiation for quantitative traits. *Genetics 172*, 1337-1347. https://doi.org/10.1534/genetics.105.050583.

Greenbury, Sam F., Mauricio Barahona, and Iain G. Johnston. 2020. “HyperTraPS: Inferring Probabilistic Patterns of Trait Acquisition in Evolutionary and Disease Progression Pathways.” Cell Systems 10 (1): 39-51.e10. https://doi.org/10.1016/j.cels.2019.10.009.

Griffin, H. G., and A. M. Griffin. 1993. “DNA Sequencing.” Applied Biochemistry and Biotechnology 38 (1): 147. https://doi.org/10.1007/BF02916418.

Griswold, C K. 2006. “Gene Flow’s Effect on the Genetic Architecture of a Local Adaptation and Its Consequences for QTL Analyses.” Heredity 96 (6): 445–53. https://doi.org/10.1038/sj.hdy.6800822.

Guan, Y., Hohn, S., & Merico, A. (2015). Suitable environmental ranges for potential coral reef habitats in the tropical ocean. *PloS one*, *10*(6), e0128831. https://doi.org/10.1371/journal.pone.0128831

Guan, Xuanli, Nana Yuyama, Alan Stewart, Chenglong Ding, Nengxiang Xu, Takako Kiyoshi, and Hongwei Cai. 2017. “Genetic Diversity and Structure of Lolium Species Surveyed on Nuclear Simple Sequence Repeat and Cytoplasmic Markers.” Frontiers in Plant Science 8: 584. https://doi.org/10.3389/fpls.2017.00584.

Gunsolus, Jeffrey L. 1990. “Mechanical and Cultural Weed Control in Corn and Soybeans.” American Journal of Alternative Agriculture 5 (3): 114–19.

Guo, Xiangyu, Fabio Cericola, Dario Fè, Morten G Pedersen, Ingo Lenk, Christian S Jensen, Just Jensen, and Lucas L Janss. 2018. “Genomic Prediction in Tetraploid Ryegrass Using Allele Frequencies Based on Genotyping by Sequencing.” Frontiers in Plant Science 9: 1165. https://doi.org/10.3389/fpls.2018.01165.

Gurdasani, D., Barroso, I., Zeggini, E. et al. (2019). Genomics of disease risk in globally diverse

populations. *Nat Rev Genet 20,* 520–535. https://doi.org/10.1038/s41576-019-0144-0
Hague, M T J, and E J Routman. 2016. “Does Population Size Affect Genetic Diversity? A Test with Sympatric Lizard Species.” Heredity 116 (1): 92–98. https://doi.org/10.1038/hdy.2015.76.
Haldane, J. B. S. 1924. “A Mathematical Theory of Natural and Artificial Selection-I.” Transactions of the Cambridge Philosophical Society 23: 19–41. https://doi.org/10.1007/BF02459574.
Hallauer, Arnel R. 2007. “History, Contribution, and Future of Quantitative Genetics in Plant Breeding: Lessons From Maize.” Crop Science 47 (S3): S-4-S-19. https://doi.org/10.2135/cropsci2007.04.0002IPBS.
Hall, L M, S R Moss, and S B Powles. 1995. Mechanism of Resistance to Chlorotoluron in Two Biotypes of the Grass Weed Alopecurus Myosuroides. Vol. 53. https://doi.org/10.1006/pest.1995.1066.
Hamilton, W. D. 1967. “Extraordinary Sex Ratios.” Science 156 (3774): 477–88. https://doi.org/10.1126/science.156.3774.477.
Han, Heping, Qin Yu, Michael J. Widderick, and Stephen B. Powles. 2016. “Target-Site EPSPS Pro-106 Mutations: Sufficient to Endow Glyphosate Resistance in Polyploid Echinochloa Colona?” Pest Management Science. https://doi.org/10.1002/ps.4038.
Han, H., Yu, Q., Beffa, R., González, S., Maiwald, F., Wang, J. and Powles, S.B. 2020. “Cytochrome P450 CYP81A10v7 in Lolium rigidum confers metabolic resistance to herbicides across at least five modes of action.” Plant Journal. https://doi.org/10.1111/tpj.15040
Hansson, Bengt, Hanna Sigeman, Martin Stervander, Maja Tarka, Suvi Ponnikas, Maria Strandh, Helena Westerdahl, and Dennis Hasselquist. 2018. “Contrasting Results from GWAS and QTL Mapping on Wing Length in Great Reed Warblers.” Molecular Ecology Resources 18 (4): 867–76. https://doi.org/10.1111/1755-0998.12785.
Harding, Dylan P., and Manish N. Raizada. 2015. “Controlling Weeds with Fungi, Bacteria and Viruses: A Review.” Frontiers in Plant Science 6 (August). https://doi.org/10.3389/fpls.2015.00659.
Hartley, H. O., and J. N. K. Rao. 1967. “Maximum-Likelihood Estimation for the Mixed Analysis of Variance Model.” Biometrika 54 (1/2): 93–108. https://doi.org/10.2307/2333854.
Hasenbein, Simone, Jade Peralta, Sharon P. Lawler, and Richard E. Connon. 2017. “Environmentally Relevant Concentrations of Herbicides Impact Non-Target Species at Multiple Sublethal Endpoints.” Science of The Total Environment 607–608 (December): 733–43. https://doi.org/10.1016/j.scitotenv.2017.06.270.
Hashim, Saima, Asad Jan, Yukari Sunohara, Mayumi Hachinohe, Hideki Ohdan, and Hiroshi Matsumoto. 2012. “Mutation of Alpha-Tubulin Genes in Trifluralin-Resistant Water Foxtail (Alopecurus Aequalis).” Pest Management Science 68 (3): 422–429. https://doi.org/10.1002/ps.2284.
Hästbacka, Johanna, Albert de la Chapelle, Ilkka Kaitila, Pertti Sistonen, Alix Weaver, and Eric Lander. 1992. “Linkage Disequilibrium Mapping in Isolated Founder Populations: Diastrophic Dysplasia in Finland.” Nature Genetics 2 (3): 204–11. https://doi.org/10.1038/ng1192-204.
Hastie, T. J., and R. J. Tibshirani. 1990. Generalized Additive Models. CRC Press.
Hastings, W. K. 1970. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” Biometrika 57 (1): 97–109. https://doi.org/10.2307/2334940.
Häusler, Thomas. 2006. Viruses vs. Superbugs. London: Palgrave Macmillan UK. https://doi.org/10.1007/978-0-230-55228-9.
Hawkins, J.P., Roberts, C.M. & Clark, V. (2000), The threatened status of restricted-range coral reef fish species. *Animal Conservation*, 3: 81-88. https://doi.org/10.1111/j.1469-1795.2000.tb00089.x
Heap, Ian M. 2014. Global Perspective of Herbicide-Resistant Weeds. Vol. 70. 9. https://doi.org/10.1002/ps.3696.
Heap, Ian M. 2020. “The International Survey of Herbicide Resistant Weeds.” The International Survey of Herbicide Resistant Weeds. 2020. www.weedscience.org.
Heather, James M., and Benjamin Chain. 2016. “The Sequence of Sequencers: The History of Sequencing DNA.” Genomics 107 (1): 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003.
Henderson, C R. 1953. “Estimation of Variance and Covariance Components.” Biometrics 9 (2): 226–52.
Henderson, C. R. 1975. “Best Linear Unbiased Estimation and Prediction under a Selection

Model." Biometrics 31 (2): 423. https://doi.org/10.2307/2529430.
Henderson, C. R., Oscar Kempthorne, S. R. Searle, and C. M. von Krosigk. 1959. "The Estimation of Environmental and Genetic Trends from Records Subject to Culling." Biometrics 15 (2): 192–218. https://doi.org/10.2307/2527669.
Hendry, G. A. F., K. Thompson, C. J. Moss, E. Edwards, and P. C. Thorpe. 1994. "Seed Persistence: A Correlation Between Seed Longevity in the Soil and Ortho-Dihydroxyphenol Concentration." Functional Ecology 8 (5): 658–64. https://doi.org/10.2307/2389929.
Hermisson, Joachim, and Pleuni S. Pennings. 2005. "Soft Sweeps: Molecular Population Genetics of Adaptation From Standing Genetic Variation." Genetics 169 (4): 2335–52. https://doi.org/10.1534/genetics.104.036947.
Hermsen, Rutger, J. Barrett Deris, and Terence Hwa. 2012. "On the Rapidity of Antibiotic Resistance Evolution Facilitated by a Concentration Gradient." Proceedings of the National Academy of Sciences of the United States of America 109 (27): 10775–80. https://doi.org/10.1073/pnas.1117716109.
Hickey, John M., Tinashe Chiurugwi, Ian Mackay, and Wayne Powell. 2017. "Genomic Prediction Unifies Animal and Plant Breeding Programs to Form Platforms for Biological Discovery." Nature Genetics 49 (9): 1297–1303. https://doi.org/10.1038/ng.3920.
Hicks, Helen L., David Comont, Shaun R. Coutts, Laura Crook, Richard Hull, Ken Norris, Paul Neve, Dylan Z. Childs, and Robert P. Freckleton. 2018. "The Factors Driving Evolved Herbicide Resistance at a National Scale." *Nature Ecology & Evolution* 2 (3): 529–536. https://doi.org/10.1038/s41559-018-0470-1.
Hill, W. G., and Alan Robertson. 1966. "The Effect of Linkage on Limits to Artificial Selection." Genetics Research 8 (3): 269–94. https://doi.org/10.1017/S0016672300010156.
Hilton, H.W. 1957. "Herbicide Tolerant Strains of Weeds." Hawaiian Sugar Planters Assoc. Ann. Rept., 69.
Hirschberg, J, and L McIntosh. 1983. "Molecular Basis of Herbicide Resistance in Amaranthus Hybridus." Science 222 (4630): 1346–1349. https://doi.org/10.1126/science.222.4630.1346.
Hirsch, C. D., J. Evans, C. R. Buell, and C. N. Hirsch. 2014. "Reduced Representation Approaches to Interrogate Genome Diversity in Large Repetitive Plant Genomes." Briefings in Functional Genomics 13 (4): 257–267. https://doi.org/10.1093/bfgp/elt051.
Hirschhorn, Joel N., and Mark J. Daly. 2005. "Genome-Wide Association Studies for Common Diseases and Complex Traits." Nature Reviews Genetics 6 (2): 95–108. https://doi.org/10.1038/nrg1521.
Hivert, Valentin, Raphaël Leblois, Eric J Petit, Mathieu Gautier, and Renaud Vitalis. 2018. "Measuring Genetic Differentiation from Pool-Seq Data." Genetics 210 (1): 315–330. https://doi.org/10.1534/genetics.118.300900.
Hochberg, O., M. Sibony, and B. Rubin. 2009. "The Response of ACCase-Resistant Phalaris Paradoxa Populations Involves Two Different Target Site Mutations." Weed Research 49 (1): 37–46. https://doi.org/10.1111/j.1365-3180.2008.00677.x.
Hoel, Paul. 1943. "The Accuracy of Sampling Methods in Ecology on JSTOR." The Annals of Mathematical Statistics 14 (3): 289–300.
Hoffmann, Ary A., and Loren H. Rieseberg. 2008. "Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?" Annual Review of Ecology, Evolution, and Systematics 39 (December): 21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532.
Hoffstetter, Amber, Antonio Cabrera, Mao Huang, and Clay Sneller. 2016. "Optimizing Training Population Data and Validation of Genomic Selection for Economic Traits in Soft Winter Wheat." G3: Genes, Genomes, Genetics 6 (9): 2919–28. https://doi.org/10.1534/g3.116.032532.
Holliday, Jason A., Kermit Ritland, and Sally N. Aitken. 2010. "Widespread, Ecologically Relevant Genetic Markers Developed from Association Mapping of Climate-Related Traits in Sitka Spruce (Picea Sitchensis)." New Phytologist 188 (2): 501–14. https://doi.org/10.1111/j.1469-8137.2010.03380.x.
Holsinger, Kent E. 2000. "Reproductive Systems and Evolution in Vascular Plants." Proceedings of the National Academy of Sciences 97 (13): 7037–42. https://doi.org/10.1073/pnas.97.13.7037.
Hornoy, Benjamin, Nathalie Pavy, Sébastien Gérardi, Jean Beaulieu, & Jean Bousquet. (2015). "Genetic Adaptation to Climate in White Spruce Involves Small to Moderate Allele

Frequency Shifts in Functionally Diverse Genes." *Genome Biology and Evolution* 7, no. 12: 3269–85. https://doi.org/10.1093/gbe/evv218.
Horowitz, Menashe, Yael Regev, and Geza Herzlinger. 1983. "Solarization for Weed Control." Weed Science 31 (2): 170–79. https://doi.org/10.1017/S0043174500068788.
Hörtensteiner, S. 2006. "Chlorophyll Degradation during Senescence." Annual Review of Plant Biology 57 (1): 55–77. https://doi.org/10.1146/annurev.arplant.57.032905.105212.
HRAC. 2018. Herbicide Resistance Action Committee. http://www.hracglobal.com/.
Huang, Honglin, Farhan Ullah, Dao-Xiu Zhou, Ming Yi, and Yu Zhao. 2019. "Mechanisms of ROS Regulation of Plant Development and Stress Responses." Frontiers in Plant Science 10. https://doi.org/10.3389/fpls.2019.00800.
Humphry, R. W., D. Blake, D. Fenlon, G. Horgan, J. C. Low, and G. J. Gunn. 2002. "The Quantitative Measurement of Antimicrobial Resistance in Escherichia Coli at the Meta-Population Level (Meta-Population Analysis)." Letters in Applied Microbiology 35 (4): 326–30. https://doi.org/10.1046/j.1472-765X.2002.01199.x.
Igoshin, Alexander V., Andrey A. Yurchenko, Nadezhda M. Belonogova, Dmitry V. Petrovsky, Ruslan B. Aitnazarov, Vladimir A. Soloshenko, Nikolay S. Yudin, and Denis M. Larkin. 2019. "Genome-Wide Association Study and Scan for Signatures of Selection Point to Candidate Genes for Body Temperature Maintenance under the Cold Stress in Siberian Cattle Populations." BMC Genetics 20 (1): 26. https://doi.org/10.1186/s12863-019-0725-0.
Ikegawa, Shiro. 2012. "A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going." Genomics & Informatics 10 (4): 220–5. https://doi.org/10.5808/GI.2012.10.4.220.
Illumina. 2016. "HiSeq X TM Series of Sequencing Systems," 1–4. https://doi.org/10.1186/s13073-017-0408-2.
Illumina Inc. 2020. "Maximum Read Length for Illumina Sequencing Platforms." November 11, 2020. https://sapac.support.illumina.com/bulletins/2020/04/maximum-read-length-for-illumina-sequencing-platforms.html.
International Schizophrenia Consortium. 2009. "Common Polygenic Variation Contributes to Risk of Schizophrenia That Overlaps with Bipolar Disorder." Nature 460 (7256): 748–52. https://doi.org/10.1038/nature08185.
Ito, Michio, Tetsuo Matsumoto, and Marco A. Quinones. 2007. "Conservation Tillage Practice in Sub-Saharan Africa: The Experience of Sasakawa Global 2000." Crop Protection, Weed Science in Time of Transition, 26 (3): 417–23. https://doi.org/10.1016/j.cropro.2006.06.017.
Jain, SK, and AD Bradshaw. 1966. "Evolutionary Divergence among Adjacent Plant Populations." Heredity 21: 407–441. https://doi.org/10.1038/hdy.1966.42.
Jannink, Jean Luc, Aaron J. Lorenz, and Hiroyoshi Iwata. 2010. "Genomic Selection in Plant Breeding: From Theory to Practice." Briefings in Functional Genomics and Proteomics 9 (2): 166–177. https://doi.org/10.1093/bfgp/elq001.
Jasieniuk, Marie, Anita L Brule-Babel, and Ian N Morrison. 1996. "The Evolution and Genetics of Herbicide Resistance in Weeds1." Weed Science 44: 176–193.
Jensen, Jeffrey D., Bret A. Payseur, Wolfgang Stephan, Charles F. Aquadro, Michael Lynch, Deborah Charlesworth, and Brian Charlesworth. 2019. "The Importance of the Neutral Theory in 1968 and 50 Years on: A Response to Kern and Hahn 2018." Evolution 73 (1): 111–14. https://doi.org/10.1111/evo.13650.
Jiang, Longda, Zhili Zheng, Ting Qi, Kathryn E. Kemper, Naomi R. Wray, Peter M. Visscher, and Jian Yang. 2019. "A Resource-Efficient Tool for Mixed Model Association Analysis of Large-Scale Data." Nature Genetics 51 (12): 1749–55. https://doi.org/10.1038/s41588-019-0530-8.
Jiang, Ning, Fengjun Zhang, Jinhua Wu, Yue Chen, Xiaohua Hu, Ou Fang, Lindsey J Leach, Di Wang, and Zewei Luo. 2016. "A Highly Robust and Optimized Sequence-Based Approach for Genetic Polymorphism Discovery and Genotyping in Large Plant Populations." TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik 129 (9): 1739–57. https://doi.org/10.1007/s00122-016-2736-9.
Jin, Shichao, Yanjun Su, Shilin Song, Kexin Xu, Tianyu Hu, Qiuli Yang, Fangfang Wu, et al. 2020. "Non-Destructive Estimation of Field Maize Biomass Using Terrestrial Lidar: An Evaluation from Plot Level to Individual Leaf Level." Plant Methods 16 (1): 69. https://doi.org/10.1186/s13007-020-00613-5.
Jin, Xiaoli, Chunhai Shi, Chang Yeon Yu, Toshihiko Yamada, and Erik J Sacks. 2017.

"Determination of Leaf Water Content by Visible and Near-Infrared Spectrometry and Multivariate Calibration in Miscanthus." *Frontiers in Plant Science* 8: 721. https://doi.org/10.3389/fpls.2017.00721.

Johnston, Mark, and Daniel J Schoen. 2017. "Mutation Rates and Dominance Levels of Genes Affecting Total Fitness in Two Angiosperm Species Author (s): Mark O . Johnston and Daniel J . Schoen Published by : American Association for the Advancement of Science Stable URL : Http://Www.Jstor.Org/Stabl" 267 (5195): 226–229.

Johnston, Susan E., John C. McEWAN, Natalie K. Pickering, James W. Kijas, Dario Beraldi, Jill G. Pilkington, Josephine M. Pemberton, and Jon Slate. 2011. "Genome-Wide Association Mapping Identifies the Genetic Basis of Discrete and Quantitative Variation in Sexual Weaponry in a Wild Sheep Population." Molecular Ecology 20 (12): 2555–66. https://doi.org/10.1111/j.1365-294X.2011.05076.x.

Jones, Eric, Travis Oliphant, Pearu Peterson, and Others. 2018. *SciPy: Open Source Scientific Tools for Python*. http://www.scipy.org/.

Jost, Lou. 2008. "GST and Its Relatives Do Not Measure Differentiation." Molecular Ecology 17 (18): 4015–26. https://doi.org/10.1111/j.1365-294X.2008.03887.x.

Juchau, Mont R., and Hao Chen. 1998. "Chapter 17 - Developmental Enzymology: Xenobiotic Biotransformation." In Handbook of Developmental Neurotoxicology, edited by William Slikker and Louis W. Chang, 321–37. San Diego: Academic Press. https://doi.org/10.1016/B978-012648860-9.50023-6.

Jung, Ho-Youl, Yun-Ju Park, Young-Jin Kim, Jung-Sun Park, Kuchan Kimm, and InSong Koh. 2007. "New Methods for Imputation of Missing Genotype Using Linkage Disequilibrium and Haplotype Information." Information Sciences 177 (3): 804–14. https://doi.org/10.1016/j.ins.2006.07.017.

Jussaume, Raymond, and Katherine Dentzman. 2016. "Farmers' Perspectives on Management Options for Herbicide-Resistant Weeds." Agricultural & Applied Economics Association 31 (4).

Jussaume, Raymond A., Katherine Dentzman, and Micheal D. K. Owen. 2019. "Producers, Weeds, and Society." Journal of Integrated Pest Management 10 (1). https://doi.org/10.1093/jipm/pmy017.

Kahnt, Belinda, Panagiotis Theodorou, Antonella Soro, Hilke Hollens-Kuhr, Michael Kuhlmann, Anton Pauw, and Robert J. Paxton. 2018. "Small and Genetically Highly Structured Populations in a Long-Legged Bee, Rediviva Longimanus, as Inferred by Pooled RAD-Seq." BMC Evolutionary Biology 18 (1): 196. https://doi.org/10.1186/s12862-018-1313-z.

Kalender, Willi A. 2011. Computed Tomography: Fundamentals, System Technology, Image Quality, Applications. John Wiley & Sons.

Kaler, Avjinder S., and Larry C. Purcell. 2019. "Estimation of a Significance Threshold for Genome-Wide Association Studies." BMC Genomics 20 (1): 618. https://doi.org/10.1186/s12864-019-5992-7.

Kaiser, Y. I., A. Menegat, and R. Gerhards. 2013. "Chlorophyll Fluorescence Imaging: A New Method for Rapid Detection of Herbicide Resistance in Alopecurus Myosuroides." *Weed Research* 53 (6): 399–406. https://doi.org/10.1111/wre.12043.

Kang, Hyun Min, Jae Hoon Sul, Susan K Service, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara Sabatti, and Eleazar Eskin. 2010. "Variance Component Model to Account for Sample Structure in Genome-Wide Association Studies." Nature Genetics 42 (4): 348–354. https://doi.org/10.1038/ng.548.

Kaundun, Shiv Shankhar. 2010. "An Aspartate to Glycine Change in the Carboxyl Transferase Domain of Acetyl CoA Carboxylase and Non-Target-Site Mechanism(s) Confer Resistance to ACCase Inhibitor Herbicides in a Lolium Multiflorum Population." Pest Management Science 66 (11): 1249–1256. https://doi.org/10.1002/ps.2003.

Kaundun, Shiv S., Sarah Jane Hutchings, Richard P. Dale, Anushka Howell, James A. Morris, Vance C. Kramer, Vinod K. Shivrain, and Eddie McIndoe. 2017. "Mechanism of Resistance to Mesotrione in an Amaranthus Tuberculatus Population from Nebraska, USA." PLoS ONE 12 (6). https://doi.org/10.1371/journal.pone.0180095.

Kazan, Kemal, and Rebecca Lyons. 2016. "The Link between Flowering Time and Stress Tolerance." Journal of Experimental Botany 67 (1): 47–60. https://doi.org/10.1093/jxb/erv441.

Keinan, Alon, and Andrew G. Clark. 2012. "Recent Explosive Human Population Growth Has Resulted in an Excess of Rare Genetic Variants." Science (New York, N.Y.) 336 (6082): 740–43. https://doi.org/10.1126/science.1217283.

Kelley, Joanna L., Jennifer Madeoy, John C. Calhoun, Willie Swanson, and Joshua M. Akey. 2006. “Genomic Signatures of Positive Selection in Humans and the Limits of Outlier Approaches.” Genome Research 16 (8): 980–89. https://doi.org/10.1101/gr.5157306.
Kern, Andrew D, and Matthew W Hahn. 2018. “The Neutral Theory in Light of Natural Selection.” Molecular Biology and Evolution 35 (6): 1366–71. https://doi.org/10.1093/molbev/msy092.
Khanna, Pragya. 2010. Cell and Molecular Biology. I. K. International Pvt Ltd.
Kimura, Motoo. 1968. “Evolutionary Rate at the Molecular Level.” Nature 217 (5129): 624–626. https://doi.org/10.1038/217624a0.
Kimura, Motoo. 1991. “The Neutral Theory of Molecular Evolution: A Review of Recent Evidence.” The Japanese Journal of Genetics 66 (4): 367–86. https://doi.org/10.1266/jjg.66.367.
Kimura, Motoo. 2020. “The Neutral Theory and Molecular Evolution.” In My Thoughts on Biological Evolution, edited by Motoo Kimura, 119–38. Evolutionary Studies. Singapore: Springer. https://doi.org/10.1007/978-981-15-6165-8_8.
Kimura, Motoo, and Tomoko Ohta. 1974. “On Some Principles Governing Molecular Evolution*.” Proceedings of the National Academy of Sciences of the United States of America 71 (7): 2848–52.
Kingman, J. F. C. 1982. “The Coalescent.” Stochastic Processes and Their Applications 13 (3): 235–48. https://doi.org/10.1016/0304-4149(82)90011-4.
Kirkpatrick, Mark, and Nick Barton. 2006. “Chromosome Inversions, Local Adaptation and Speciation.” Genetics 173 (1): 419–34. https://doi.org/10.1534/genetics.105.047985.
Kishikawa, Toshihiro, Yukihide Momozawa, Takeshi Ozeki, Taisei Mushiroda, Hidenori Inohara, Yoichiro Kamatani, Michiaki Kubo, and Yukinori Okada. 2019. “Empirical Evaluation of Variant Calling Accuracy Using Ultra-Deep Whole-Genome Sequencing Data.” Scientific Reports 9 (1): 1–10. https://doi.org/10.1038/s41598-018-38346-0.
Kloot, P. M. 1983. “The Genus Lolium in Australia.” Australian Journal of Botany 31 (4): 421–35. https://doi.org/10.1071/bt9830421.
Knight, Jo, Scott F. Saccone, Zhehao Zhang, Dennis G. Ballinger, and John P. Rice. 2009. “A Comparison of Association Statistics between Pooled and Individual Genotypes.” Human Heredity 67 (4): 219–25. https://doi.org/10.1159/000194975.
Kofler, Robert, R. V. Pandey, and Christian Schlötterer. 2011. “PoPoolation2: Identifying Differentiation between Populations Using Sequencing of Pooled DNA Samples (Pool-Seq).” Bioinformatics 27 (24): 3435–3436. https://doi.org/10.1093/bioinformatics/btr589.
Kojima, Kaname, Shu Tadaka, Fumiki Katsuoka, Gen Tamiya, Masayuki Yamamoto, and Kengo Kinoshita. 2020. “A Genotype Imputation Method for De-Identified Haplotype Reference Information by Using Recurrent Neural Network.” PLOS Computational Biology 16 (10): e1008207. https://doi.org/10.1371/journal.pcbi.1008207.
Kovalchuk, Igor, Olga Kovalchuk, and Barbara Hohn. 2000. “Genome-Wide Variation of the Somatic Mutation Frequency in Transgenic Plants.” The EMBO Journal 19 (17): 4431–38. https://doi.org/10.1093/emboj/19.17.4431.
Kreiner, Julia M, Darci Giacomini, Felix Bemm, Bridgit Waithaka, Julian Regalado, Christa Lanz, Julia Hildebrandt, et al. 2018. “Multiple Modes of Convergent Adaptation in the Spread of Glyphosate-Resistant Amaranthus Tuberculatus.” BioRxiv, no. 1: 1–17. https://doi.org/10.1101/498519.
Kreiner, Julia M, John R Stinchcombe, and Stephen I Wright. 2018. “Population Genomics of Herbicide Resistance: Adaptation via Evolutionary Rescue.” *Annu. Rev. Plant Biol* 69: 611–35. https://doi.org/10.1146/annurev-arplant-042817.
Kreitman, M. 1983. “Nucleotide Polymorphism at the Alcohol Dehydrogenase Locus of Drosophila Melanogaster.” Nature 304 (5925): 412–17. https://doi.org/10.1038/304412a0.
Kruglyak, L. 1999. “Prospects for Whole-Genome Linkage Disequilibrium Mapping of Common Disease Genes.” Nature Genetics 22 (2): 139–44. https://doi.org/10.1038/9642.
Kudsk, P., and J. C. Streibig. 2003. “Herbicides – a Two-Edged Sword*.” Weed Research 43 (2): 90–102. https://doi.org/10.1046/j.1365-3180.2003.00328.x.
Kumar, Gaurav, and Pradeep Kumar Bhatia. 2014. “A Detailed Review of Feature Extraction in Image Processing Systems.” In 2014 Fourth International Conference on Advanced Computing Communication Technologies, 5–12. https://doi.org/10.1109/ACCT.2014.74.
Kumar, Kishore R., Mark J. Cowley, and Ryan L. Davis. 2019. “Next-Generation Sequencing and Emerging Technologies.” Seminars in Thrombosis and Hemostasis 45 (7): 661–73. https://doi.org/10.1055/s-0039-1688446.

Kumar, Vipan, Prashant Jha, Darci Giacomini, Eric P. Westra, and Philip Westra. 2015. “Molecular Basis of Evolved Resistance to Glyphosate and Acetolactate Synthase-Inhibitor Herbicides in Kochia (Kochia Scoparia) Accessions from Montana.” Weed Science. https://doi.org/10.1614/WS-D-15-00021.1.

Kuritz, S J, J R Landis, and G G Koch. 1988. “A General Overview of Mantel-Haenszel Methods: Applications and Recent Developments.” Ann. Rev. Public Health 9: 123–60.

Kwon, Sunyoung, Seunghyun Park, Byunghan Lee, and Sungroh Yoon. 2013. “In-Depth Analysis of Interrelation between Quality Scores and Real Errors in Illumina Reads.” In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 635–638. IEEE. https://doi.org/10.1109/EMBC.2013.6609580.

Lafayette, Lev, and Bernd Wiebelt. 2017. “Spartan and NEMO: Two HPC-Cloud Hybrid Implementations.” In 2017 IEEE 13th International Conference on E-Science (e-Science), 458–59. Auckland: IEEE. https://doi.org/10.1109/eScience.2017.70.

Lafond, Jonathan A., Liwen Han, and Pierre Dutilleul. 2015. “Concepts and Analyses in the CT Scanning of Root Systems and Leaf Canopies: A Timely Summary.” Frontiers in Plant Science 6 (December). https://doi.org/10.3389/fpls.2015.01111.

Lagator, Mato, Tom Vogwill, Andrew Mead, Nick Colegrave, and Paul Neve. 2013. “Herbicide Mixtures at High Doses Slow the Evolution of Resistance in Experimentally Evolving Populations of Chlamydomonas Reinhardtii.” New Phytologist 198 (3): 938–945. https://doi.org/10.1111/nph.12195.

Lande, Russell, and Stevan J. Arnold. 1983. “The Measurement of Selection on Correlated Characters.” *Evolution* 37 (6): 1210. https://doi.org/10.2307/2408842.

Lander, E. S., and D. Botstein. 1986. “Mapping Complex Genetic Traits in Humans: New Methods Using a Complete RFLP Linkage Map.” Cold Spring Harbor Symposia on Quantitative Biology 51 (January): 49–62. https://doi.org/10.1101/SQB.1986.051.01.007.

Lander, E. S., and D. Botstein. 1989. “Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps.” Genetics 121 (1): 185–99.

Langley, C H, E Montgomery, and W F Quattlebaum. 1982. “Restriction Map Variation in the Adh Region of Drosophila.” Proceedings of the National Academy of Sciences of the United States of America 79 (18): 5631–35.

Laplante, Julie, Istvan Rajcan, and François J. Tardif. 2009. “Multiple Allelic Forms of Acetohydroxyacid Synthase Are Responsible for Herbicide Resistance in Setaria Viridis.” Theoretical and Applied Genetics 119 (4): 577–585. https://doi.org/10.1007/s00122-009-1067-5.

Le Roux P. (2009) Plants at the margin. Ecological limits and climate change. *Ann Bot. 104(7):ix*. doi:10.1093/aob/mcp220

Lee, Kristin M, and Graham Coop. 2017. “Distinguishing Among Modes of Convergent Adaptation Using Population Genomic Data.” https://doi.org/10.1534/genetics.117.300417.

Lemerle, D., B. Verbeek, and N. Coombes. 1995. “Losses in Grain Yield of Winter Crops from Lolium Rigidum Competition Depend on Crop Species, Cultivar and Season.” Weed Research 35 (6): 503–9. https://doi.org/10.1111/j.1365-3180.1995.tb01648.x.

Lemon, R. W. 1994. “Insecticide Resistance.” The Journal of Agricultural Science 122 (3): 329–33. https://doi.org/10.1017/S0021859600067241.

Lewontin, R. C., and J. L. Hubby. 1966. “A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of DROSOPHILA PSEUDOOBSCURA.” Genetics 54 (2): 595–609.

Li, Bo, Nanxi Zhang, You-Gan Wang, Andrew W. George, Antonio Reverter, and Yutao Li. 2018. “Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods.” Frontiers in Genetics 9. https://doi.org/10.3389/fgene.2018.00237.

Lichtfouse, Eric, ed. 2015. Sustainable Agriculture Reviews. Vol. 15. Sustainable Agriculture Reviews. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-09132-7.

Liggett, Wm M, Willet M Hays, Samuel B Green, Otto Lugger, and Harry Snyder. 1893. “Wheat.” University of Minnesota Agricultural Experiment Station.

Li, H. 2011. “A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data.” Bioinformatics 27 (21): 2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Li, Heng. 2013. “Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM,” March. http://arxiv.org/abs/1303.3997.
Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” Bioinformatics, May. https://doi.org/10.1093/bioinformatics/bty191.
Li, Junrui, Haipeng Li, Mattias Jakobsson, Sen Li, Per Sjödin, and Martin Lascoux. 2012. “Joint Analysis of Demography and Selection in Population Genetics: Where Do We Stand and Where Could We Go?” Molecular Ecology 21 (1): 28–44. https://doi.org/10.1111/j.1365-294X.2011.05308.x.
Li, Mingyao, Chun Li, and Weihua Guan. 2008. “Evaluation of Coverage Variation of SNP Chips for Genome-Wide Association Studies.” European Journal of Human Genetics 16 (5): 635–43. https://doi.org/10.1038/sj.ejhg.5202007.
Liepe, Katharina J., Andreas Hamann, Pia Smets, Connor R. Fitzpatrick, & Sally N. Aitken. (2016). “Adaptation of Lodgepole Pine and Interior Spruce to Climate: Implications for Reforestation in a Warming World.” *Evolutionary Applications* 9, no. 2: 409–19. https://doi.org/10.1111/eva.12345.
Linde, Celeste C., Leon M. Smith, and Rod Peakall. 2016. “Weeds, as Ancillary Hosts, Pose Disproportionate Risk for Virulent Pathogen Transfer to Crops.” BMC Evolutionary Biology 16 (1): 101. https://doi.org/10.1186/s12862-016-0680-6.
Lindsay, Jodi A., and Matthew T. G. Holden. 2004. “Staphylococcus Aureus: Superbug, Super Genome?” Trends in Microbiology 12 (8): 378–85. https://doi.org/10.1016/j.tim.2004.06.004.
Lippert, Christoph. 2003. “Linear Mixed Models for Genome-Wide Association Studies.” PhD Thesis.
Liu, J. G., K. J. Mahoney, P. H. Sikkema, and C. J. Swanton. 2009. “The Importance of Light Quality in Crop–Weed Competition.” Weed Research 49 (2): 217–24. https://doi.org/10.1111/j.1365-3180.2008.00687.x.
Liu, Li, Yayu Wang, Di Zhang, Zhuoxin Chen, Xiaoshu Chen, Zhijian Su, and Xionglei He. 2020. “The Origin of Additive Genetic Variance Driven by Positive Selection.” Molecular Biology and Evolution 37 (8): 2300–2308. https://doi.org/10.1093/molbev/msaa085.
Liu, Mingyang, Andrew G Hulting, and Carol A Mallory-Smith. 2014. “Characterization of Multiple-Herbicide-Resistant Italian Ryegrass (Lolium Perenne Spp. Multiflorum).” *Pest Management Science* 70 (7): 1145–1150. https://doi.org/10.1002/ps.3665.
Liu, Yidong, Dongtao Ren, Sharon Pike, Stephen Pallardy, Walter Gassmann, and Shuqun Zhang. 2007. “Chloroplast-Generated Reactive Oxygen Species Are Involved in Hypersensitive Response-like Cell Death Mediated by a Mitogen-Activated Protein Kinase Cascade.” The Plant Journal 51 (6): 941–54. https://doi.org/10.1111/j.1365-313X.2007.03191.x.
Li, Xinghua, Changsheng Nie, Yuchen Liu, Yu Chen, Xueze Lv, Liang Wang, Jianwei Zhang, et al. 2019. “A Genome-Wide Association Study Explores the Genetic Determinism of Host Resistance to Salmonella Pullorum Infection in Chickens.” Genetics, Selection, Evolution : GSE 51 (September). https://doi.org/10.1186/s12711-019-0492-4.
Li, Yong, Xue-Xia Zhang, Run-Li Mao, Jie Yang, Cai-Yun Miao, Zhuo Li, and Ying-Xiong Qiu. 2017. “Ten Years of Landscape Genomics: Challenges and Opportunities.” Frontiers in Plant Science 8: 2136. https://doi.org/10.3389/fpls.2017.02136.
Li, Yun, Cristen Willer, Serena Sanna, and Gonçalo Abecasis. 2009. “Genotype Imputation.” Annual Review of Genomics and Human Genetics 10: 387–406. https://doi.org/10.1146/annurev.genom.9.081307.164242.
Llewellyn, RS, D Ronning, J Ouzman, S Walker, A Mayfield, and M Clarke. 2016. “Impact of Weeds on Australian Grain Production: The Cost of Weeds to Australian Grain Growers and the Adoption of Weed Management and Tillage Practices Report for GRDC.” Grains Research and Development Corporation and the Commonwealth Scientific and Industrial Research Organisation.
Llewellyn, Rick S., Robert K. Lindner, David J. Pannell, and Stephen B. Powles. 2007. “Herbicide Resistance and the Adoption of Integrated Weed Management by Western Australian Grain Growers.” Agricultural Economics 36 (1): 123–30. https://doi.org/10.1111/j.1574-0862.2007.00182.x.
Loewe, Laurence, and William G. Hill. 2010. “The Population Genetics of Mutations: Good, Bad and Indifferent.” Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1544): 1153–67. https://doi.org/10.1098/rstb.2009.0317.
Long, Anthony D., and Charles H. Langley. 1999. “The Power of Association Studies to Detect

the Contribution of Candidate Genetic Loci to Variation in Complex Traits." Genome Research 9 (8): 720–31. https://doi.org/10.1101/gr.9.8.720.

Lopez-Maestre, Hélène, Lilia Brinza, Camille Marchet, Janice Kielbassa, Sylvère Bastien, Mathilde Boutigny, David Monnin, et al. 2016. "SNP Calling from RNA-Seq Data without a Reference Genome: Identification, Quantification, Differential Analysis and Impact on the Protein Sequence." Nucleic Acids Research 44 (19): gkw655. https://doi.org/10.1093/nar/gkw655.

Lorenz, Aaron J., Shiaoman Chao, Franco G. Asoro, Elliot L. Heffner, Takeshi Hayashi, Hiroyoshi Iwata, Kevin P. Smith, Mark E. Sorrells, and Jean Luc Jannink. 2011. Genomic Selection in Plant Breeding. Knowledge and Prospects. 1st ed. Vol. 110. C. Elsevier Inc. https://doi.org/10.1016/B978-0-12-385531-2.00002-5.

Lowry, David B, Sean Hoban, Joanna L Kelley, Katie E Lotterhos, Laura K Reed, Michael F Antolin, and Andrew Storfer. 2017. "Breaking RAD: An Evaluation of the Utility of Restriction Site-Associated DNA Sequencing for Genome Scans of Adaptation." Molecular Ecology Resources 17 (2): 142–152. https://doi.org/10.1111/1755-0998.12635.

Lozier, Jeffrey D., Jason M. Jackson, Michael E. Dillon, and James P. Strange. 2016. "Population Genomics of Divergence among Extreme and Intermediate Color Forms in a Polymorphic Insect." Ecology and Evolution 6 (4): 1075–1091. https://doi.org/10.1002/ece3.1928.

Lubke, GH, C Laurin, R Walters, N Eriksson, P Hysi, TD Spector, GW Montgomery, NG Martin, SE Medland, and DI Boomsma. 2013. "Gradient Boosting as a SNP Filter: An Evaluation Using Simulated and Hair Morphology Data." Journal of Data Mining in Genomics & Proteomics 4 (October). https://doi.org/10.4172/2153-0602.1000143.

Luckey, John A., Howard Drossman, Anthony J. Kostichka, David A. Mead, Jonathan D'Cunha, Tracy B. Norris, and Lloyd M. Smith. 1990. "High Speed DNA Sequencing by Capillary Electrophoresis." Nucleic Acids Research 18 (15): 4417–21. https://doi.org/10.1093/nar/18.15.4417.

Lukačišinová, Marta, and Tobias Bollenbach. 2017. "Toward a Quantitative Understanding of Antibiotic Resistance Evolution." Current Opinion in Biotechnology 46 (August): 90–97. https://doi.org/10.1016/j.copbio.2017.02.013.

Lynch, Michael, and John S Conery. 2003. "The Origins of Genome Complexity." Science (New York, N.Y.) 302 (5649): 1401–4. https://doi.org/10.1126/science.1089370.

Macgregor, Stuart, Peter M. Visscher, and Grant Montgomery. 2006. "Analysis of Pooled DNA Samples on High Density Arrays without Prior Knowledge of Differential Hybridization Rates." Nucleic Acids Research 34 (7). https://doi.org/10.1093/nar/gkl136.

Mackay, Trudy F. C., Stephen Richards, Eric A. Stone, Antonio Barbadilla, Julien F. Ayroles, Dianhui Zhu, Sònia Casillas, et al. 2012. "The Drosophila Melanogaster Genetic Reference Panel." Nature 482 (7384): 173–78. https://doi.org/10.1038/nature10811.

Magwene, Paul M., John H. Willis, and John K. Kelly. 2011. "The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing." Edited by Adam Siepel. PLoS Computational Biology 7 (11): e1002255. https://doi.org/10.1371/journal.pcbi.1002255.

Malecot, G. 1948. "Les Mathematiques de l'Heredite (Maison et Cie, Paris)." Les Mathématiques de l'hérédité. Masson and Cie, Paris. https://www.cabdirect.org/cabdirect/abstract/19490101200.

Manalil, Sudheesh. 2014. Evolution of Herbicide Resistance in Lolium Rigidum under Low Herbicide Rates: An Australian Experience. Vol. 54. 2. https://doi.org/10.2135/cropsci2013.04.0248.

Manalil, Sudheesh. 2015. "An Analysis of Polygenic Herbicide Resistance Evolution and Its Management Based on a Population Genetics Approach." Basic and Applied Ecology 16 (2): 104–11. https://doi.org/10.1016/j.baae.2015.01.004.

Manel, Stéphanie, Marco Andrello, Karine Henry, Daphné Verdelet, Aude Darracq, Pierre-Edouard Guerin, Bruno Desprez, and Pierre Devaux. 2018. "Predicting Genotype Environmental Range from Genome–Environment Associations." Molecular Ecology 27 (13): 2823–33. https://doi.org/10.1111/mec.14723.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, et al. 2009. "Finding the Missing Heritability of Complex Diseases." Nature 461 (7265): 747–53. https://doi.org/10.1038/nature08494.

Mantel, Nathan, and William Haenszel. 1959. "Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease." JNCI: Journal of the National Cancer Institute 22 (4): 719–48. https://doi.org/10.1093/jnci/22.4.719.

March, Ruth E. 1999. “Gene Mapping by Linkage and Association Analysis.” Molecular Biotechnology 13 (2): 113–22. https://doi.org/10.1385/MB:13:2:113.
Mardis, Elaine R. 2017. “DNA Sequencing Technologies: 2006–2016.” Nature Protocols 12 (2): 213–18. https://doi.org/10.1038/nprot.2016.182.
Marees, Andries T., Hilde de Kluiver, Sven Stringer, Florence Vorspan, Emmanuel Curis, Cynthia Marie-Claire, and Eske M. Derks. 2018. “A Tutorial on Conducting Genome-wide Association Studies: Quality Control and Statistical Analysis.” International Journal of Methods in Psychiatric Research 27 (2). https://doi.org/10.1002/mpr.1608.
Martin, Steven G., Rene C. Van Acker, and Lyle F. Friesen. 2001. “Critical Period of Weed Control in Spring Canola.” Weed Science 49 (3): 326–33. https://doi.org/10.1614/0043-1745(2001)049[0326:CPOWCI]2.0.CO;2.
Masabni, Joseph G, Bernard H Zandstra, and Joseph G Masabni. 1999. “A Serine-to-Threonine Mutation in Linuron-Resistant Portulaca Oleracea.” Weed Science 47 (4): 393–400.
Mason, H. S., J. C. North, and M. Vanneste. 1965. “Microsomal Mixed-Function Oxidations: The Metabolism of Xenobiotics.” Federation Proceedings 24 (5): 1172–80.
Ma, R., J. J. Skelton, and D. E. Riechers, “Measuring rates of herbicide metabolism in dicot weeds with an excised leaf assay.” 2015. Journal of Visualized Experiments, vol. 103.
Ma, Wenlong, Zhixu Qiu, Jie Song, Jiajia Li, Qian Cheng, Jingjing Zhai, and Chuang Ma. 2018. “A Deep Convolutional Neural Network Approach for Predicting Phenotypes from Genotypes.” Planta 248 (5): 1307–18. https://doi.org/10.1007/s00425-018-2976-9.
Maxam, A. M., and W. Gilbert. 1977. “A New Method for Sequencing DNA.” Proceedings of the National Academy of Sciences 74 (2): 560–64. https://doi.org/10.1073/pnas.74.2.560.
Maynard, John, and John Haigh. 2007. “The Hitch-Hiking Effect of a Favourable Gene.” Genetics Research 89 (5–6): 391–403. https://doi.org/10.1017/S0016672308009579.
McCraw, J M, W Spoor, and West Mains Road. 1983. “Self-Incompatibility in Lolium Species. 1. Lolium Rigidum Gaud. and L. Multiflorum L.” Heredity 50: 21–27.
McGinnis, S., and T. L. Madden. 2004. “BLAST: At the Core of a Powerful and Diverse Set of Sequence Analysis Tools.” Nucleic Acids Research 32 (Web Server): W20–W25. https://doi.org/10.1093/nar/gkh435.
McVean, Gilean A. T., and Brian Charlesworth. 2000. “The Effects of Hill-Robertson Interference Between Weakly Selected Mutations on Patterns of Molecular Evolution and Variation.” Genetics 155 (2): 929–44.
Mechant, E., T. De Marez, O. Hermann, R. Olsson, and R. Bulcke. 2008. Target Site Resistance to Metamitron in Chenopodium Album L. 21.
Medd, RW, BA Auld, DR Kemp, and D Murison. 1985. “The Influence of Wheat Density and Spatial Arrangement on Annual Ryegrass, Lolium Rigidum Gaudin, Competition.,” 3, Australian Journal of Agricultural Research (36): 361–71.
Mendes, Rafael R., Hudson K. Takano, Jéssica F. Leal, Amanda S. Souza, Sarah Morran, Rubem S. Oliveira Jr, Fernando S. Adegas, Todd A. Gaines, and Franck E. Dayan. 2020. “Evolution of EPSPS Double Mutation Imparting Glyphosate Resistance in Wild Poinsettia (Euphorbia Heterophylla L.).” *PLOS ONE* 15, no. 9 (September 10, 2020): e0238818. https://doi.org/10.1371/journal.pone.0238818.
Menges, R. M., and J. R. Wayland. 1974. “Uhf Electromagnetic Energy for Weed Control in Vegetables.” Weed Science 22 (6): 584–90. https://doi.org/10.1017/S0043174500038352.
Mengistu, Lemma W, George W Mueller-Warrant, Aaron Liston, and Reed E Barker. 2000. “PsbA Mutation (Valine 219 to Isoleucine) in \textlessi\textgreaterPoa Annua\textless/I\ textgreater Resistant to Metribuzin and Diuron.” Pest Management Science 56 (November 1999): 209–217.
Mengistu, Lemma W, Michael J. Christoffers, and Rodney G. Lym. 2005. “A PsbA Mutation in Kochia Scoparia (L) Schrad from Railroad Rights-of-Way with Resistance to Diuron, Tebuthiuron and Metribuzin.” Pest Management Science 61 (11): 1035–1042. https://doi.org/10.1002/ps.1079.
Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. “Equation of State Calculations by Fast Computing Machines.” The Journal of Chemical Physics 21 (6): 1087–92. https://doi.org/10.1063/1.1699114.
Meuwissen, T H, and M E Goddard. 2000. “Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria with Closely Linked Marker Loci.” Genetics 155 (1): 421–30.
Meuwissen, T H, B J Hayes, and M E Goddard. 2001. “Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.” Genetics 157 (4): 1819–29.

Meuwissen, Theo HE. 2009. “Accuracy of Breeding Values of ‘unrelated’ Individuals Predicted by Dense SNP Genotyping.” Genetics Selection Evolution 41 (1): 35. https://doi.org/10.1186/1297-9686-41-35.

Micheletti, Steven J., and Shawn R. Narum. 2018. “Utility of Pooled Sequencing for Association Mapping in Nonmodel Organisms.” Molecular Ecology Resources 18 (4): 825–837. https://doi.org/10.1111/1755-0998.12784.

Mishra, Puneet, Mohd Shahrimie Mohd Asaari, Ana Herrero-Langreo, Santosh Lohumi, Belén Diezma, and Paul Scheunders. 2017. “Close Range Hyperspectral Imaging of Plants: A Review.” Biosystems Engineering 164 (December): 49–67. https://doi.org/10.1016/j.biosystemseng.2017.09.009.

Moeinizade, Saba, Guiping Hu, Lizhi Wang, and Patrick S. Schnable. 2019. “Optimizing Selection and Mating in Genomic Selection with a Look-Ahead Approach: An Operations Research Framework.” G3: Genes, Genomes, Genetics 9 (7): 2123–33. https://doi.org/10.1534/g3.118.200842.

Monaco, Thomas J, Stephen C Weller, and Floyd M Ashton. 2002. WEED SCIENCE Principles and Practices. John Wiley & Sons, Inc., New York.

Monaghan, N. M. 1980. “The Biology and Control of Lolium Rigidum as a Weed of Wheat.” Weed Research 20 (2): 117–121. https://doi.org/10.1111/j.1365-3180.1980.tb00054.x.

Montesinos-López, Osval A., Abelardo Montesinos-López, José Crossa, José C. Montesinos-López, David Mota-Sanchez, Fermín Estrada-González, Jussi Gillberg, Ravi Singh, Suchismita Mondal, and Philomin Juliana. 2018. “Prediction of Multiple-Trait and Multiple-Environment Genomic Data Using Recommender Systems.” G3: Genes, Genomes, Genetics 8 (1): 131–47. https://doi.org/10.1534/g3.117.300309.

Mora-Márquez, F., V. García-Olivares, B. C. Emerson, and U. López de Heredia. 2017. “Ddradseqtools : A Software Package for in Silico Simulation and Testing of Double-Digest RADseq Experiments.” Molecular Ecology Resources 17 (2): 230–246. https://doi.org/10.1111/1755-0998.12550.

Morgan, T. H. 1910. “Sex Limited Inheritance in Drosophila.” Science 32 (812): 120–22. https://doi.org/10.1126/science.32.812.120.

Morgan, Thomas Hunt, Alfred Henry Sturtevant, Calvin Blackman Bridges, and Hermann Joseph Muller. 1915. The Mechanism of Mendelian Heredity. H. Holt.

Moss, Stephen. 2019. “Integrated Weed Management (IWM): Why Are Farmers Reluctant to Adopt Non-Chemical Alternatives to Herbicides?” Pest Management Science 75 (5): 1205–11. https://doi.org/10.1002/ps.5267.

Muller, Hermann J. 1916. “The Mechanism of Crossing-Over. II. IV. The Manner of Occurrence of Crossing-Over.” The American Naturalist 50 (593): 284–305. https://doi.org/10.1086/279541.

Mullineaux, Lauren S., Anna Metaxas, Stace E. Beaulieu, Monika Bright, Sabine Gollner, Benjamin M. Grupe, Santiago Herrera, et al. (2018). “Exploring the Ecology of Deep-Sea Hydrothermal Vents in a Metacommunity Framework.” *Frontiers in Marine Science* 5. https://doi.org/10.3389/fmars.2018.00049.

Mutka, Andrew M, and Rebecca S Bart. 2014. “Image-Based Phenotyping of Plant Disease Symptoms.” *Frontiers in Plant Science* 5: 734. https://doi.org/10.3389/fpls.2014.00734.

Murphy, Kevin P. 2007. “Conjugate Bayesian Analysis of the Gaussian Distribution.”

Najarro, Michael A., Jennifer L. Hackett, Brittny R. Smith, Chad A. Highfill, Elizabeth G. King, Anthony D. Long, and Stuart J. Macdonald. 2015. “Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila.” Edited by Nadia Singh. PLOS Genetics 11 (11): e1005663. https://doi.org/10.1371/journal.pgen.1005663.

Nakagawa, Shinichi, and Robert P. Freckleton. 2008. “Missing Inaction: The Dangers of Ignoring Missing Data.” Trends in Ecology & Evolution 23 (11): 592–96. https://doi.org/10.1016/j.tree.2008.06.014.

Nakib, A., H. Oulhadj, and P. Siarry. 2007. “Image Histogram Thresholding Based on Multiobjective Optimization.” Signal Processing 87 (11): 2516–34. https://doi.org/10.1016/j.sigpro.2007.04.001.

Naylor, Beryl. 1960. “Species Differentiation in the Genus Lolium.” Heredity 15 (2–3): 219–233. https://doi.org/10.1038/hdy.1960.78.

Nei, Masatoshi. 1972. “Genetic Distance between Populations.” The American Naturalist 106 (949): 283–92. https://doi.org/10.1086/282771.

Nei, Masatoshi. 1980. “Stochastic Theory of Population Genetics and Evolution.” In Vito Volterra Symposium on Mathematical Models in Biology, edited by Claudio Barigozzi, 17–47.

Lecture Notes in Biomathematics. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-93161-1_2.
Nei, Masatoshi, and Fumio Tajima. 1981. “DNA Polymorphism Detectable by Restriction Endonucleases.” Genetics 97 (1): 145–63.
Nelson, Matthew R., Daniel Wegmann, Margaret G. Ehm, Darren Kessner, Pamela St. Jean, Claudio Verzilli, Judong Shen, et al. 2012. “An Abundance of Rare Functional Variants in 202 Drug Target Genes Sequenced in 14,002 People.” Science (New York, N.Y.) 337 (6090): 100–104. https://doi.org/10.1126/science.1217876.
Neuenschwander, Samuel, Frédéric Michaud, Jérôme Goudet, and Oliver Stegle. 2018. “QuantiNemo 2: A Swiss Knife to Simulate Complex Demographic and Genetic Scenarios, Forward and Backward in Time.” Edited by Oliver Stegle. Bioinformatics, August. https://doi.org/10.1093/bioinformatics/bty737.
Neve, Paul, and S Powles. 2005. “High Survival Frequencies at Low Herbicide Use Rates in Populations of Lolium Rigidum Result in Rapid Evolution of Herbicide Resistance.” Heredity 95 (6): 485–492. https://doi.org/10.1038/sj.hdy.6800751.
Nielsen, Erica S., Romina Henriques, Robert J. Toonen, Ingrid S. S. Knapp, Baocheng Guo, and Sophie von der Heyden. 2018. “Complex Signatures of Genomic Variation of Two Non-Model Marine Species in a Homogeneous Environment.” BMC Genomics 19 (1): 347. https://doi.org/10.1186/s12864-018-4721-y.
Nitschke, Lutz, and Walter Schüssler. 1998. “Surface Water Pollution by Herbicides from Effluents of Waste Water Treatment Plants.” Chemosphere 36 (1): 35–41. https://doi.org/10.1016/S0045-6535(97)00286-5.
Oerke, E.C. 2006. “Crop Losses to Pests.” Journal of Agricultural Science 144: 31–43. https://doi.org/10.1017/ S0021859605005708.
Ogutu, Joseph O., Hans-Peter Piepho, and Torben Schulz-Streeck. 2011. “A Comparison of Random Forests, Boosting and Support Vector Machines for Genomic Selection.” BMC Proceedings 5 (3): S11. https://doi.org/10.1186/1753-6561-5-S3-S11.
Ohta, Tomoko, and Motoo Kimura. 1971. “On the Constancy of the Evolutionary Rate of Cistrons.” Journal of Molecular Evolution 1 (1): 18–25. https://doi.org/10.1007/BF01659391.
Oksanen, Lauri. 2001. “Logic of Experiments in Ecology: Is Pseudoreplication a Pseudoissue?” Oikos 94 (1): 27–38.
Olson-Manning, Carrie F., Maggie R. Wagner, and Thomas Mitchell-Olds. 2012. “Adaptive Evolution: Evaluating Empirical Support for Theoretical Predictions.” Nature Reviews Genetics 13 (12): 867–877. https://doi.org/10.1038/nrg3322.
OpenCV. 2017. Open Source Computer Vision Library. https://github.com/itseez/opencv.
Ophel, Kathy M., Alan F. Bird, and Allen Kerr. 1993. “Association of Bacteriophage Particles with Toxin Production by Clavibacter Toxicus, the Causal Agent of Annual Ryegrass Toxicity.” Phytopathology 83 (6): 676–81.
Orr, H. Allen. 2010. “The Population Genetics of Beneficial Mutations.” Philosophical Transactions of the Royal Society B: Biological Sciences 365 (1544): 1195–1201. https://doi.org/10.1098/rstb.2009.0282.
Orson, J. H. 1999. “The Cost to the Farmer of Herbicide Resistance.” Weed Technology 13 (3): 607–11.
Ott, Jurg, Jing Wang, and Suzanne M. Leal. 2015. “Genetic Linkage Analysis in the Age of Whole-Genome Sequencing.” Nature Reviews Genetics 16 (5): 275–84. https://doi.org/10.1038/nrg3908.
Ozaki, Kouichi, Yozo Ohnishi, Aritoshi Iida, Akihiko Sekine, Ryo Yamada, Tatsuhiko Tsunoda, Hiroshi Sato, et al. 2002. “Functional SNPs in the Lymphotoxin-α Gene That Are Associated with Susceptibility to Myocardial Infarction.” Nature Genetics 32 (4): 650–654. https://doi.org/10.1038/ng1047.
Pais, A.L., Whetten, R.W. & Xiang, Q.-Y. (2020). Population structure, landscape genomics, and genetic signatures of adaptation to exotic disease pressure in Cornus florida L. - Insights from GWAS and GBS data. *J. Syst. Evol., 58*: 546-570. https://doi.org/10.1111/jse.12592
Pallares, Luisa F., Bettina Harr, Leslie M. Turner, and Diethard Tautz. 2014. “Use of a Natural Hybrid Zone for Genomewide Association Mapping of Craniofacial Traits in the House Mouse.” Molecular Ecology 23 (23): 5756–70. https://doi.org/10.1111/mec.12968.
Pal, Nikhil R, and Sankar K Pal. 1993. “A Review on Image Segmentation Techniques.” Pattern Recognition 26 (9): 1277–94. https://doi.org/10.1016/0031-3203(93)90135-J.

Panjvani, Karim, Anh V. Dinh, and Khan A. Wahid. 2019. “LiDARPheno – A Low-Cost LiDAR-Based 3D Scanning System for Leaf Morphological Trait Extraction.” Frontiers in Plant Science 10. https://doi.org/10.3389/fpls.2019.00147.

Pankin, Artem, Janine Altmüller, Christian Becker, and Maria von Korff. 2018. “Targeted Resequencing Reveals Genomic Signatures of Barley Domestication.” New Phytologist 218 (3): 1247–1259. https://doi.org/10.1111/nph.15077.

Pannell, John R., and Anne-Marie Labouche. 2013. “The Incidence and Selection of Multiple Mating in Plants.” Philosophical Transactions of the Royal Society B: Biological Sciences 368 (1613): 20120051. https://doi.org/10.1098/rstb.2012.0051.

Paril, Jefferson F., and Alexandre J. Fournier-Level. 2019. “InstaGraminoid, a Novel Colorimetric Method to Assess Herbicide Resistance, Identifies Patterns of Cross-Resistance in Annual Ryegrass.” Plant Phenomics 2019 (April): 1–13. https://doi.org/10.34133/2019/7937156.

Parish, S. 1990. “The Flame Treatment of Weed Seedlings under Controlled Conditions.” The Flame Treatment of Weed Seedlings under Controlled Conditions., no. No. 45: 193–96.

Parka, S. J., and O. F. Soper. 1977. “The Physiology and Mode of Action of the Dinitroaniline Herbicides.” Weed Science 25 (1): 79–87. https://doi.org/10.1017/S0043174500032975.

Park, Kee Woong, and Carol A Mallory-Smith. 2006. “PsbA Mutation (Asn266 to Thr) in Senecio Vulgaris L. Confers Resistance to Several PS II-Inhibiting Herbicides.” Pest Management Science 62 (9): 880–885. https://doi.org/10.1002/ps.1252.

Park, Sang Tae, and Jayoung Kim. 2016. “Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing.” International Neurourology Journal 20 (Suppl 2): S76. https://doi.org/10.5213/inj.1632742.371.

Patterson, H. D., and R. Thompson. 1971. “Recovery of Inter-Block Information When Block Sizes Are Unequal.” Biometrika 58 (3): 545–54. https://doi.org/10.1093/biomet/58.3.545.

Patzoldt, William L, Aaron G Hager, Joel S McCormick, and Patrick J Tranel. 2006. “A Codon Deletion Confers Resistance to Herbicides Inhibiting Protoporphyrinogen Oxidase.” Proceedings of the National Academy of Sciences of the United States of America 103 (33): 12329–34. https://doi.org/10.1073/pnas.0603137103.

Patzoldt, William L, and Patrick J Tranel. 2007. “Multiple ALS Mutations Confer Herbicide Resistance in Waterhemp (Amaranthus Tuberculatus) [Electronic Resource].” Weed Science 55 (5): 421–428. https://doi.org/10.1614/WS-06-213.1.

Penone, Caterina, Ana D. Davidson, Kevin T. Shoemaker, Moreno Di Marco, Carlo Rondinini, Thomas M. Brooks, Bruce E. Young, Catherine H. Graham, and Gabriel C. Costa. 2014. “Imputation of Missing Data in Life-History Trait Datasets: Which Approach Performs the Best?” Methods in Ecology and Evolution 5 (9): 961–70. https://doi.org/10.1111/2041-210X.12232.

Pereira, Helcio Duarte, José Marcelo Soriano Viana, Andréa Carla Bastos Andrade, Fabyano Fonseca e Silva, and Geísa Pinheiro Paes. 2018. “Relevance of Genetic Relationship in GWAS and Genomic Prediction.” Journal of Applied Genetics 59 (1): 1–8. https://doi.org/10.1007/s13353-017-0417-2.

Perez-Jones, Alejandro, Suphannika Intanon, and Carol Mallory-Smith. 2009. “Molecular Analysis of Hexazinone-Resistant Shepherd’s-Purse (Capsella Bursa-Pastoris) Reveals a Novel PsbA Mutation.” Weed Science 57 (06): 574–578. https://doi.org/10.1614/WS-09-089.1.

Perrier, C., B. Delahaie, and A. Charmantier. 2018. “Heritability Estimates from Genomewide Relatedness Matrices in Wild Populations: Application to a Passerine, Using a Small Sample Size.” Molecular Ecology Resources 18 (4): 838–53. https://doi.org/10.1111/1755-0998.12886.

Peter, Mary Lou. 2016. “Left Uncontrolled, Weeds Would Cost Billions in Economic Losses Every Year.” 2016. https://www.ksre.k-state.edu/news/stories/2016/05/uncontrolled-weeds051216.html.

Peterson, Brant K., Jesse N. Weber, Emily H. Kay, Heidi S. Fisher, and Hopi E. Hoekstra. 2012. “Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species.” Edited by Ludovic Orlando. PLoS ONE 7 (5): e37135. https://doi.org/10.1371/journal.pone.0037135.

Peterson, Gale E. 1967. “The Discovery and Development of 2,4-D.” Agricultural History 41 (3): 243–54.

Petit, Cécile, Guillaume Bay, Fanny Pernin, and Christophe Délye. 2010. “Prevalence of Cross- or Multiple Resistance to the Acetyl-Coenzyme a Carboxylase Inhibitors Fenoxaprop,

Clodinafop and Pinoxaden in Black-Grass (Alopecurus Myosuroides Huds.) in France." Pest Management Science 66 (2): 168–177. https://doi.org/10.1002/ps.1851.

Pfeiffer, Franziska, Carsten Gröber, Michael Blank, Kristian Händler, Marc Beyer, Joachim L. Schultze, and Günter Mayer. 2018. "Systematic Evaluation of Error Rates and Causes in Short Samples in Next-Generation Sequencing." Scientific Reports 8 (1): 10950. https://doi.org/10.1038/s41598-018-29325-6.

Pflugfelder, Daniel, Ralf Metzner, Dagmar van Dusschoten, Rüdiger Reichel, Siegfried Jahnke, and Robert Koller. 2017. "Non-Invasive Imaging of Plant Roots in Different Soils Using Magnetic Resonance Imaging (MRI)." Plant Methods 13 (1): 102. https://doi.org/10.1186/s13007-017-0252-9.

Poland, Jesse A., and Trevor W. Rife. 2012. "Genotyping-by-Sequencing for Plant Breeding and Genetics." The Plant Genome 5 (3): 92–102. https://doi.org/10.3835/plantgenome2012.05.0005.

Ponting, Rebecca C., Michelle C. Drayton, Noel O. I. Cogan, Mark P. Dobrowolski, Germán C. Spangenberg, Kevin F. Smith, and John W. Forster. 2007. "SNP Discovery, Validation, Haplotype Structure and Linkage Disequilibrium in Full-Length Herbage Nutritive Quality Genes of Perennial Ryegrass (Lolium Perenne L.)." Molecular Genetics and Genomics 278 (5): 585–597. https://doi.org/10.1007/s00438-007-0275-4.

Pop, Mihai. 2009. "Genome Assembly Reborn: Recent Computational Challenges." Briefings in Bioinformatics 10 (4): 354–66. https://doi.org/10.1093/bib/bbp026.

Pound, Michael P., Jonathan A. Atkinson, Alexandra J. Townsend, Michael H. Wilson, Marcus Griffiths, Aaron S. Jackson, Adrian Bulat, et al. 2017. "Deep Machine Learning Provides State-of-the-Art Performance in Image-Based Plant Phenotyping." *GigaScience* 6 (10). https://doi.org/10.1093/gigascience/gix083.

Powles, Stephen B., and John M. Matthews. 1992. "Multiple Herbicide Resistance in Annual Ryegrass (Lolium Rigidum): A Driving Force for the Adoption of Integrated Weed Management." In Resistance '91: Achievements and Developments in Combating Pesticide Resistance, edited by Ian Denholm, Alan L. Devonshire, and Derek W. Hollomon, 75–87. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-2862-9_7.

Powles, Stephen B., and Qin Yu. 2010. "Evolution in Action: Plants Resistant to Herbicides." Annual Review of Plant Biology 61 (1): 317–347. https://doi.org/10.1146/annurev-arplant-042809-112119.

Powles, Stephen B, Debrah F Lorraine-Colwill, James J Dellow, and Christopher Preston. 1998. "Evolved Resistance to Glyphosate in Rigid Ryegrass (Lolium Rigidum) in Australia." Weed Science 46 (5): 604–607.

Prado, R., C. Rioboo, C. Herrero, and A. Cid. 2009. "The Herbicide Paraquat Induces Alterations in the Elemental and Biochemical Composition of Non-Target Microalgal Species." Chemosphere 76 (10): 1440–44. https://doi.org/10.1016/j.chemosphere.2009.06.003.

Preston, Christopher, Angela M. Wakelin, Fleur C. Dolman, Yazid Bostamam, and Peter Boutsalis. 2009. "A Decade of Glyphosate-Resistant Lolium around the World: Mechanisms, Genes, Fitness, and Agronomic Management." Weed Science 57 (4): 435–441. https://doi.org/10.1614/WS-08-.

Pritchard, J K, and N A Rosenberg. 1999. "Use of Unlinked Genetic Markers to Detect Population Stratification in Association Studies." American Journal of Human Genetics 65 (1): 220–28.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. "Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites." Molecular Biology and Evolution 16 (12): 1791–98. https://doi.org/10.1093/oxfordjournals.molbev.a026091.

Provine, William B. 2001. The Origins of Theoretical Population Genetics. 2nd ed., with A new afterword. The Chicago History of Science and Medicine. Chicago: University of Chicago Press.

Pukk, Lilian, Freed Ahmad, Shihab Hasan, Veljo Kisand, Riho Gross, and Anti Vasemägi. 2015. "Less Is More: Extreme Genome Complexity Reduction with DdRAD Using Ion Torrent Semiconductor Technology." Molecular Ecology Resources 15 (5): 1145–1152. https://doi.org/10.1111/1755-0998.12392.

Pulido, Francisco. (2007) "The Genetics and Evolution of Avian Migration." *BioScience* 57, no. 2: 165–74. https://doi.org/10.1641/B570211.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” The American Journal of Human Genetics 81 (3): 559–75. https://doi.org/10.1086/519795.

R4P Network. 2016. Trends and Challenges in Pesticide Resistance Detection. Vol. 21. 10.

Ralph, Peter L., and Graham Coop. 2015. “The Role of Standing Variation in Geographic Convergent Adaptation.” The American Naturalist 186 (Suppl 1): S5–23. https://doi.org/10.1086/682948.

Rang, Franka J., Wigard P. Kloosterman, and Jeroen de Ridder. 2018. “From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy.” Genome Biology 19 (1): 90. https://doi.org/10.1186/s13059-018-1462-9.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. Vienna, Austria. https://www.r-project.org/.

Reddy, K N, and V K Nandula. 2012. “Herbicide Resistant Crops: History, Development and Current Technologies.” Indian Journal of Agronomy 57 (1): 7.

Reeves, T. G. 1976. “Effect of Annual Ryegrass (Lolium Rigidum Gaud.) on Yield of Wheat.” Weed Research 16 (1): 57–63. https://doi.org/10.1111/j.1365-3180.1976.tb00379.x.

Rellstab, Christian, Stefan Zoller, Andrew Tedder, Felix Gugerli, and Martin C. Fischer. 2013. “Validation of SNP Allele Frequencies Determined by Pooled Next-Generation Sequencing in Natural Populations of a Non-Model Plant Species.” Edited by Gabriel AB Marais. PLoS ONE 8 (11): e80422. https://doi.org/10.1371/journal.pone.0080422.

Rendina, Alan R., and Judy M. Felts. 1988. “Cyclohexanedione Herbicides Are Selective and Potent Inhibitors of Acetyl-CoA Carboxylase from Grasses.” Plant Physiology 86 (4): 983–86. https://doi.org/10.1104/pp.86.4.983.

Reuter, Jason A, Damek V Spacek, and Michael P Snyder. 2015. “High-Throughput Sequencing Technologies.” Molecular Cell 58 (4): 586–97. https://doi.org/10.1016/j.molcel.2015.05.004.

REX Consortium. 2013. “Heterogeneity of Selection and the Evolution of Resistance.” *Trends in Ecology & Evolution* 28 (2): 110–118. https://doi.org/10.1016/j.tree.2012.09.001.

Rey-Caballero, Jordi, Julio Menéndez, Maria D. Osuna, Marisa Salas, and Joel Torra. 2017. “Target-Site and Non-Target-Site Resistance Mechanisms to ALS Inhibiting Herbicides in Papaver Rhoeas.” Pesticide Biochemistry and Physiology 138 (May): 57–65. https://doi.org/10.1016/j.pestbp.2017.03.001.

Richards, F. J. 1959. “A Flexible Growth Function for Empirical Use.” Journal of Experimental Botany 10 (2): 290–301. https://doi.org/10.1093/jxb/10.2.290.

Riginos, C., Crandall, E.D., Liggins, L., Bongaerts, P., & Treml, E.A. (2016). Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. *Current Zoology, Volume 62, Issue 6*, December 2016, Pages 581–601, https://doi.org/10.1093/cz/zow067

Riley, I. T., and A. C. McKay. 1991. “Inoculation of Lolium Rigidum with Clavibacter Sp., the Toxigenic Bacteria Associated with Annual Ryegrass Toxicity.” Journal of Applied Bacteriology 71 (4): 302–6. https://doi.org/10.1111/j.1365-2672.1991.tb03794.x.

Rincent, R., Charcosset, A., & Moreau, L. (2017). Predicting genomic selection efficiency to optimize calibration set and to assess prediction accuracy in highly structured populations. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, *130*(11), 2231–2247. https://doi.org/10.1007/s00122-017-2956-7

Roberts, H. A. 1964. “Emergence and Longevity in Cultivated Soil of Seeds of Some Annual Weeds.” Weed Research 4 (4): 296–307. https://doi.org/10.1111/j.1365-3180.1964.tb00304.x.

Robin, Charles, Paul Battlay, and Alexandre Fournier-Level. 2019. “What Can Genetic Association Panels Tell Us about Evolutionary Processes in Insects?” Current Opinion in Insect Science 31: 99–105. https://doi.org/10.1016/j.cois.2018.12.004.

Roff, Derek A. 1997. Evolutionary Quantitative Genetics. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-4080-9.

Rojas, Clemencia M., Muthappa Senthil-Kumar, Vered Tzin, and Kirankumar Mysore. 2014. “Regulation of Primary Plant Metabolism during Plant-Pathogen Interactions and Its Contribution to Plant Defense.” Frontiers in Plant Science 5. https://doi.org/10.3389/fpls.2014.00017.

Rollini, P., B. Mach, and J. Gorski. 1985. “Linkage Map of Three HLA-DR Beta-Chain Genes: Evidence for a Recent Duplication Event.” Proceedings of the National Academy of

Sciences 82 (21): 7197–7201. https://doi.org/10.1073/pnas.82.21.7197.
Roos, A P W de, B J Hayes, and M E Goddard. 2009. “Reliability of Genomic Predictions across Multiple Populations.” Genetics 183 (4): 1545–53. https://doi.org/10.1534/genetics.109.104935.
Rosenberg, Noah A., Lucy Huang, Ethan M. Jewett, Zachary A. Szpiech, Ivana Jankovic, and Michael Boehnke. 2010. “Genome-Wide Association Studies in Diverse Populations.” Nature Reviews Genetics 11 (5): 356–66. https://doi.org/10.1038/nrg2760.
Roser, Max. 2020. “Pesticides.” Our World in Data. 2020.
Ruttink, T, K Lamour, S van Glabeke, M Hegarty, C Verwimp, E Veeckman, H Muylle, et al. 2018. “Pooling Resources: Allele Frequency Fingerprinting in Lolium Perenne.” PAG XXVI Grasslands Workshop.
Ryan, G. F. 1970. “Resistance of Common Groundsel to Simazine and Atrazine.” Weed Science 18 (5): 614–16.
Sackton, Timothy B., and Nathan Clark. 2019. “Convergent Evolution in the Genomics Era: New Insights and Directions.” Philosophical Transactions of the Royal Society B: Biological Sciences 374 (1777): 20190102. https://doi.org/10.1098/rstb.2019.0102.
Sanger, F., and A. R. Coulson. 1975. “A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase.” Journal of Molecular Biology 94 (3): 441–48. https://doi.org/10.1016/0022-2836(75)90213-2.
Sanger, Frederick. 1998. “Sequences, Sequences, and Sequences.” Annual Review of Biochemistry, no. 57: 1–28.
Sanger, F., S. Nicklen, and A. R. Coulson. 1977. “DNA Sequencing with Chain-Terminating Inhibitors.” Proceedings of the National Academy of Sciences of the United States of America 74 (12): 5463–67.
Santangelo, James S., Marc T. J. Johnson, and Rob W. Ness. 2018. “Modern Spandrels: The Roles of Genetic Drift, Gene Flow and Natural Selection in the Evolution of Parallel Clines.” Proceedings of the Royal Society B: Biological Sciences 285 (1878): 20180230. https://doi.org/10.1098/rspb.2018.0230.
Santos, Alesandro Souza, and Fernanda Amato Gaiotto. 2020. “Knowledge Status and Sampling Strategies to Maximize Cost-Benefit Ratio of Studies in Landscape Genomics of Wild Plants.” Scientific Reports 10 (1): 1–9. https://doi.org/10.1038/s41598-020-60788-8.
Sass, László, Petra Majer, and Éva Hideg. 2012. “Leaf Hue Measurements: A High-Throughput Screening of Chlorophyll Content.” In *High-Throughput Phenotyping in Plants: Methods and Protocols, Methods in Molecular Biology*, edited by Jennifer Normanly, 61–69.
Sax, Karl. 1923. “The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS.” Genetics 8 (6): 552–60.
Scarabel, Laura, Fanny Pernin, and Christophe Délye. 2015. “Occurrence, Genetic Control and Evolution of Non-Target-Site Based Resistance to Herbicides Inhibiting Acetolactate Synthase (ALS) in the Dicot Weed Papaver Rhoeas.” Plant Science. https://doi.org/10.1016/j.plantsci.2015.06.008.
Schaffner, Urs, Martin Hill, Tom Dudley, and Carla D’Antonio. 2020. “Post-Release Monitoring in Classical Biological Control of Weeds: Assessing Impact and Testing Pre-Release Hypotheses.” Current Opinion in Insect Science, Ecology • Parasites/Parasitoids/Biological control, 38 (April): 99–106. https://doi.org/10.1016/j.cois.2020.02.008.
Scheben, Armin, Jacqueline Batley, and David Edwards. 2017. “Genotyping-by-Sequencing Approaches to Characterize Crop Genomes: Choosing the Right Tool for the Right Application.” Plant Biotechnology Journal 15 (2): 149–161. https://doi.org/10.1111/pbi.12645.
Schielzeth, Holger, and Shinichi Nakagawa. 2013. “Nested by Design: Model Fitting and Interpretation in a Mixed Model Era.” Methods in Ecology and Evolution 4 (1): 14–24. https://doi.org/10.1111/j.2041-210x.2012.00251.x.
Schielzeth, Holger, Niels J. Dingemanse, Shinichi Nakagawa, David F. Westneat, Hassen Allegue, Céline Teplitsky, Denis Réale, Ned A. Dochtermann, László Zsolt Garamszegi, and Yimen G. Araya-Ajoy. 2020. “Robustness of Linear Mixed-Effects Models to Violations of Distributional Assumptions.” Methods in Ecology and Evolution 11 (9): 1141–52. https://doi.org/10.1111/2041-210X.13434.
Schlötterer, Christian. 2004. “The Evolution of Molecular Markers — Just a Matter of Fashion?” Nature Reviews Genetics 5 (1): 63–69. https://doi.org/10.1038/nrg1249.

Schlötterer, Christian, Raymond Tobler, Robert Kofler, and Viola Nolte. 2014. “Sequencing Pools of Individuals - Mining Genome-Wide Polymorphism Data without Big Funding.” Nature Reviews 15: 749–765.

Schlötterer, Christian, Robert Kofler, E Versace, R Tobler, and S U Franssen. 2015. “Combining Experimental Evolution with Next-Generation Sequencing: A Powerful Tool to Study Adaptation from Standing Genetic Variation.” Heredity 114 (5): 431–440. https://doi.org/10.1038/hdy.2014.86.

Schneider, Dominik, Laura S. Lopez, Meng Li, Joseph D. Crawford, Helmut Kirchhoff, and Hans-Henning Kunz. 2019. “Fluctuating Light Experiments and Semi-Automated Plant Phenotyping Enabled by Self-Built Growth Racks and Simple Upgrades to the IMAGING-PAM.” Plant Methods 15 (1): 156. https://doi.org/10.1186/s13007-019-0546-1.

Schoen, Daniel J., Jacques L. David, and Thomas M. Bataillon. 1998. “Deleterious Mutation Accumulation and the Regeneration of Genetic Resources.” Proceedings of the National Academy of Sciences of the United States of America 95 (1): 394–99.

Schurz, Haiko, Stephanie J. Müller, Paul David van Helden, Gerard Tromp, Eileen G. Hoal, Craig J. Kinnear, and Marlo Möller. 2019. “Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population.” Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00034.

Schwass, D. E., and J. W. Finley. 1985. “Overview: The Influence of Nutrition on Xenobiotic Metabolism.” In Xenobiotic Metabolism: Nutritional Effects, edited by John W. Finley and Daniel E. Schwass, 277:1–10. ACS Symposium Series. Washington, D.C.: American Chemical Society. https://doi.org/10.1021/bk-1985-0277.ch001.

Searle, Shayle R., George Casella, and Charles E. McCulloch. 2009. Variance Components. John Wiley & Sons.

Selmoni, Oliver, Elia Vajana, Annie Guillaume, Estelle Rochat, and Stéphane Joost. 2020. “Sampling Strategy Optimization to Increase Statistical Power in Landscape Genomics: A Simulation-Based Approach.” Molecular Ecology Resources 20 (1): 154–69. https://doi.org/10.1111/1755-0998.13095.

Sexton, Jason P., Patrick J. McIntyre, Amy L. Angert, and Kevin J. Rice. 2009. “Evolution and Ecology of Species Range Limits.” Annual Review of Ecology, Evolution, and Systematics 40 (1): 415–36. https://doi.org/10.1146/annurev.ecolsys.110308.120317.

Shaner, Dale L., and Hugh J. Beckie. 2014. The Future for Weed Control and Technology. Vol. 70. 9. https://doi.org/10.1002/ps.3706.

Shaner, D.L., T. Nadler-Hassar, W. B. Henry, and C. H. Koger. 2005. “A rapid in vivo shikimate accumulation assay with excised leaf discs.” Weed Science, vol. 53, no. 6, pp. 769–774.

Sharma, Gaurav, and Raja Bala. 2017. Digital Color Imaging Handbook. CRC Press.

Shaw, Ruth G., and Thomas Mitchell-Olds. 1993. “Anova for Unbalanced Data: An Overview.” Ecology 74 (6): 1638–45. https://doi.org/10.2307/1939922.

Shi, Huwenbo, Gleb Kichaev, and Bogdan Pasaniuc. 2016. “Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data.” The American Journal of Human Genetics 99 (1): 139–53. https://doi.org/10.1016/j.ajhg.2016.05.013.

Shin, Jimin, and Chaeyoung Lee. 2015. “Statistical Power for Identifying Nucleotide Markers Associated with Quantitative Traits in Genome-Wide Association Analysis Using a Mixed Model.” Genomics 105 (1): 1–4. https://doi.org/10.1016/j.ygeno.2014.11.001.

Siang Tan, Khang, and Nor Ashidi Mat Isa. 2011. “Color Image Segmentation Using Histogram Thresholding – Fuzzy C-Means Hybrid Approach.” Pattern Recognition 44 (1): 1–15. https://doi.org/10.1016/j.patcog.2010.07.013.

Silvestri, Luciana, Alejandro Sosa, Fernando Mc Kay, Marcelo Diniz Vitorino, Martin Hill, Costas Zachariades, Stephen Hight, et al. 2020. “Implementation of Access and Benefit-Sharing Measures Has Consequences for Classical Biological Control of Weeds.” BioControl 65 (2): 125–41. https://doi.org/10.1007/s10526-019-09988-4.

Simoneaux, Bruce J., and Thomas J. Gould. 2008. “Chapter 7 - Plant Uptake and Metabolism of Triazine Herbicides.” In The Triazine Herbicides, edited by Homer M. LeBaron, Janis E. McFarland, and Orvin C. Burnside, 73–99. San Diego: Elsevier. https://doi.org/10.1016/B978-044451167-6.50010-6.

Sims, David, Ian Sudbery, Nicholas E. Ilott, Andreas Heger, and Chris P. Ponting. 2014. “Sequencing Depth and Coverage: Key Considerations in Genomic Analyses.” Nature Reviews Genetics 15 (2): 121–32. https://doi.org/10.1038/nrg3642.

Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer. 2005. “ROCR: Visualizing Classifier Performance in R.” *Bioinformatics* 21 (20): 3940–3941.

https://doi.org/10.1093/bioinformatics/bti623.
Singh, Asheesh Kumar, Baskar Ganapathysubramanian, Soumik Sarkar, and Arti Singh. 2018. “Deep Learning for Plant Stress Phenotyping: Trends and Future Perspectives.” *Trends in Plant Science* 23 (10): 883–898. https://doi.org/10.1016/j.tplants.2018.07.004.
Slaughter, D. C., D. K. Giles, and D. Downey. 2008. “Autonomous Robotic Weed Control Systems: A Review.” Computers and Electronics in Agriculture, Emerging Technologies For Real-time and Integrated Agriculture Decisions, 61 (1): 63–78. https://doi.org/10.1016/j.compag.2007.05.008.
Smith, Johanna L., Miranda L. Wilson, Sara M. Nilson, Troy N. Rowan, David L. Oldeschulte, Robert D. Schnabel, Jared E. Decker, and Christopher M. Seabury. 2019. “Genome-Wide Association and Genotype by Environment Interactions for Growth Traits in U.S. Gelbvieh Cattle.” BMC Genomics 20 (1): 926. https://doi.org/10.1186/s12864-019-6231-y.
Smith, Lloyd M., Steven Fung, Michael W. Hunkapiller, Tim J. Hunkapiller, and Leroy E. Hood. 1985. “The Synthesis of Oligonucleotides Containing an Aliphatic Amino Group at the 5′ Terminus: Synthesis of Fluorescent DNA Primers for Use in DNA Sequence Analysis.” Nucleic Acids Research 13 (7): 2399–2412. https://doi.org/10.1093/nar/13.7.2399.
Smythies, Carleton. 1834. “No. III. EXPERIMENTS ON THE GROWTH OF SWEDISH TURNIPS.” Transactions of the Society, Instituted at London, for the Encouragement of Arts, Manufactures, and Commerce 49 (2): 13–15.
Solberg, T. R., A. K. Sonesson, J. A. Woolliams, and T. H. E. Meuwissen. 2008. “Genomic Selection Using Different Marker Types and Densities.” Journal of Animal Science 86 (10): 2447–54. https://doi.org/10.2527/jas.2007-0010.
Song, Kai, Li Li, and Guofan Zhang. 2016. “Coverage Recommendation for Genotyping Analysis of Highly Heterologous Species Using Next-Generation Sequencing Technology.” Scientific Reports 6 (1): 1–7. https://doi.org/10.1038/srep35736.
South, Andy, and Ian M. Hastings. 2018. “Insecticide Resistance Evolution with Mixtures and Sequences: A Model-Based Explanation.” Malaria Journal 17 (1): 80. https://doi.org/10.1186/s12936-018-2203-y.
Stayton, C. Tristan. 2015. “What Does Convergent Evolution Mean? The Interpretation of Convergence and Its Implications in the Search for Limits to Evolution.” Interface Focus 5 (6). https://doi.org/10.1098/rsfs.2015.0039.
Steinrücken, Hans Christian, and Nikolaus Amrhein. 1984. “5-Enolpyruvylshikimate-3-Phosphate Synthase of Klebsiella Pneumoniae.” European Journal of Biochemistry 143 (2): 341–49. https://doi.org/10.1111/j.1432-1033.1984.tb08378.x.
Stern, David L. 2013. “The Genetic Causes of Convergent Evolution.” Nature Reviews Genetics 14 (11): 751–64. https://doi.org/10.1038/nrg3483.
Stevens, O. A. 1932. “The Number and Weight of Seeds Produced by Weeds.” American Journal of Botany 19 (9): 784–94. https://doi.org/10.2307/2436042.
Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." Journal of the Royal Statistical Society. Series B (Methodological) 36, no. 2: 111-47. Accessed December 8, 2020. http://www.jstor.org/stable/2984809.
Stopps, Gregory J., Robert E. Nurse, and Peter H. Sikkema. 2013. “The Effect of Time of Day on the Activity of Postemergence Soybean Herbicides.” Weed Technology 27 (4): 690–95. https://doi.org/10.1614/WT-D-13-00035.1.
Storfer, Andrew, Austin Patton, and Alexandra K Fraik. 2018. “Navigating the Interface Between Landscape Genetics and Landscape Genomics.” Frontiers in Genetics 9: 68. https://doi.org/10.3389/fgene.2018.00068.
Sturtevant, A. H. 1913. “The Linear Arrangement of Six Sex-Linked Factors in Drosophila, as Shown by Their Mode of Association.” Journal of Experimental Zoology 14 (1): 43–59. https://doi.org/10.1002/jez.1400140104.
Sul, Jae Hoon, Lana S Martin, and Eleazar Eskin. 2018. “Population Structure in Genetic Studies: Confounding Factors and Mixed Models.” PLoS Genetics 14 (12): e1007309. https://doi.org/10.1371/journal.pgen.1007309.
Swanton, Clarence J., and Stephen D. Murphy. 1996. “Weed Science beyond the Weeds: The Role of Integrated Weed Management (IWM) in Agroecosystem Health on JSTOR.” Weed Science 44 (2): 437–445.
Swerdlow, Harold, and Raymond Gesteland. 1990. “Capillary Gel Electrophoresis for Rapid, High Resolution DNA Sequencing.” Nucleic Acids Research 18 (6): 1415–19. https://doi.org/10.1093/nar/18.6.1415.
Tabashnik, Bruce E., and Nancy L. Cushing. 1989. “Quantitative Genetic Analysis of Insecticide

Resistance: Variation in Fenvalerate Tolerance in a Diamondback Moth (Lepidoptera: Plutellidae) Population." Journal of Economic Entomology 82 (1): 5–10. https://doi.org/10.1093/jee/82.1.5.

Tackenberg, Oliver. 2007. "A New Method for Non-Destructive Measurement of Biomass, Growth Rates, Vertical Biomass Distribution and Dry Matter Content Based on Digital Image Analysis." Annals of Botany 99 (4): 777–83. https://doi.org/10.1093/aob/mcm009.

Taghavi Namin, Sarah, Mohammad Esmaeilzadeh, Mohammad Najafi, Tim B. Brown, and Justin O. Borevitz. 2018. "Deep Phenotyping: Deep Learning for Temporal Phenotype/Genotype Classification." *Plant Methods* 14 (1): 66. https://doi.org/10.1186/s13007-018-0333-4.

Taiz, Lincoln, and Eduardo Zeiger. 2002. Plant Physiology. 3rd ed. Sinauer Associates. https://academic.oup.com/aob/article-lookup/doi/10.1093/aob/mcg079.

Tajima, Fumio. 1983. "Evolutionary Relationship of DNA Sequences in Finite Populations." Genetics 105 (2).

Tajima, Fumio. 1989. "Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism." Genetics 123: 585–595.

Takeno, Kiyotoshi. 2016. "Stress-Induced Flowering: The Third Category of Flowering Response." Journal of Experimental Botany 67 (17): 4925–34. https://doi.org/10.1093/jxb/erw272.

Talbot, Christopher J., Alison Nicod, Stacey S. Cherny, David W. Fulker, Allan C. Collins, and Jonathan Flint. 1999. "High-Resolution Mapping of Quantitative Trait Loci in Outbred Mice." Nature Genetics 21 (3): 305–8. https://doi.org/10.1038/6825.

Tan, Ge, Lennart Opitz, Ralph Schlapbach, and Hubert Rehrauer. 2019. "Long Fragments Achieve Lower Base Quality in Illumina Paired-End Sequencing." Scientific Reports 9 (1): 2856. https://doi.org/10.1038/s41598-019-39076-7.

Tange, O. 2011. GNU Parallel - The Command-Line Power Tool. The USENIX Magazine.

Tanksley, S. D., N. D. Young, A. H. Paterson, and M. W. Bonierbale. 1989. "RFLP Mapping in Plant Breeding: New Tools for an Old Science." Bio/Technology 7 (3): 257–64. https://doi.org/10.1038/nbt0389-257.

Taylor, A L. 1970. "Current Linkage Map of Escherichia Coli." Bacteriological Reviews 34 (2): 155–75.

Technow, Frank, Anna Bürger, and Albrecht E. Melchinger. 2013. "Genomic Prediction of Northern Corn Leaf Blight Resistance in Maize with Combined or Separated Training Sets for Heterotic Groups." G3: Genes|Genomes|Genetics 3 (2): 197–203. https://doi.org/10.1534/g3.112.004630.

Tennessen, Jacob A., and Joshua M. Akey. 2011. "Parallel Adaptive Divergence among Geographically Diverse Human Populations." PLoS Genetics 7 (6). https://doi.org/10.1371/journal.pgen.1002127.

Terrell, Edward E. 1966. "Taxonomic Implications of Genetics in Ryegrasses (Lolium)." The Botanical Review 32 (2): 138–64. https://doi.org/10.1007/BF02858658.

Terry, Michael. 1945. "Soil Erosion in Australia." The Geographical Journal 105 (3/4): 121. https://doi.org/10.2307/1789563.

Thapa, Suresh, Feiyu Zhu, Harkamal Walia, Hongfeng Yu, and Yufeng Ge. 2018. "A Novel LiDAR-Based Instrument for High-Throughput, 3D Measurement of Morphological Traits in Maize and Sorghum." Sensors (Basel, Switzerland) 18 (4). https://doi.org/10.3390/s18041187.

The UniProt Consortium. 2019. "UniProt: A Worldwide Hub of Protein Knowledge." Nucleic Acids Research 47 (D1): D506–15. https://doi.org/10.1093/nar/gky1049.

Thistlethwaite, Frances R., Blaise Ratcliffe, Jaroslav Klápště, Ilga Porth, Charles Chen, Michael U. Stoehr, and Yousry A. El-Kassaby. 2017. "Genomic Prediction Accuracies in Space and Time for Height and Wood Density of Douglas-Fir Using Exome Capture as the Genotyping Platform." BMC Genomics 18 (1): 930. https://doi.org/10.1186/s12864-017-4258-5.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society. Series B (Methodological) 58 (1): 267–88.

Tilk, Susanne, Alan Bergland, Aaron Goodman, Paul Schmidt, Dmitri Petrov, and Sharon Greenblum. 2019. "Accurate Allele Frequencies from Ultra-Low Coverage Pool-Seq Samples in Evolve-and-Resequence Experiments." G3: Genes, Genomes, Genetics 9 (12): 4159–68. https://doi.org/10.1534/g3.119.400755.

Timmons, F. L. 1970. "A History of Weed Control in the United States and Canada." Weed

Science 18 (2): 294–307.
Toda, Yosuke, and Fumio Okura. 2019. “How Convolutional Neural Networks Diagnose Plant Disease.” *Plant Phenomics* 2019 (March): 9237136. https://doi.org/10.1155/2019/9237136.
Tomson, Göran, and Ioana Vlad. 2014. “The Need to Look at Antibiotic Resistance from a Health Systems Perspective.” Upsala Journal of Medical Sciences 119 (2): 117–24. https://doi.org/10.3109/03009734.2014.902879.
Toonen, Robert J, Jonathan B Puritz, Zac H Forsman, Jonathan L Whitney, Iria Fernandez-Silva, Kimberly R Andrews, and Christopher E Bird. 2013. “EzRAD: A Simplified Method for Genomic Genotyping in Non-Model Organisms.” PeerJ 1: e203. https://doi.org/10.7717/peerj.203.
Torkamani, Ali, Nathan E. Wineinger, and Eric J. Topol. 2018. “The Personal and Clinical Utility of Polygenic Risk Scores.” Nature Reviews Genetics 19 (9): 581–90. https://doi.org/10.1038/s41576-018-0018-x.
Truong, Kevin. 2020. “Sequencing Giant Illumina Scraps $1.2 Billion PacBio Acquisition.” San Francisco Business Times, January 3, 2020. https://www.bizjournals.com/sanfrancisco/news/2020/01/03/sequencing-giant-illumina-scraps-1-2-billion.html.
Turrens, Julio F. 2003. “Mitochondrial Formation of Reactive Oxygen Species.” The Journal of Physiology 552 (Pt 2): 335–44. https://doi.org/10.1113/jphysiol.2003.049478.
Ubbens, Jordan R, and Ian Stavness. 2017. “Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks.” *Frontiers in Plant Science* 8: 1190. https://doi.org/10.3389/fpls.2017.01190.
Upadhyaya, Mahesh K., and Robert E. Blackshaw. 2007. Non-Chemical Weed Management: Principles, Concepts and Technology. CABI.
Van Straalen, N. I., and D. Roelofs. 2006. “Introduction to Ecological Genetics.” New York, Oxford University Press
Vignal, Alain, Denis Milan, Magali SanCristobal, and André Eggen. 2002. “A Review on SNP and Other Types of Molecular Markers and Their Use in Animal Genetics.” Genetics Selection Evolution 34 (3): 275. https://doi.org/10.1186/1297-9686-34-3-275.
Villalobos, Francisco J, and Elias Fereres. 2016. Principles of Agronomy for Sustainable Agriculture. https://doi.org/10.1007/978-3-319-46116-8.
Visscher, Peter M., Matthew A. Brown, Mark I. McCarthy, and Jian Yang. 2012. “Five Years of GWAS Discovery.” American Journal of Human Genetics 90 (1): 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029.
Visscher, Peter M., Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 2017. “10 Years of GWAS Discovery: Biology, Function, and Translation.” The American Journal of Human Genetics 101 (1): 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.
Vogwill, T., M. Lagator, N. Colegrave, and Paul Neve. 2012. “The Experimental Evolution of Herbicide Resistance in Chlamydomonas Reinhardtii Results in a Positive Correlation between Fitness in the Presence and Absence of Herbicides.” Journal of Evolutionary Biology 25 (10): 1955–1964. https://doi.org/10.1111/j.1420-9101.2012.02558.x.
Voss-Fels, Kai Peter, Mark Cooper, and Ben John Hayes. 2019. “Accelerating Crop Genetic Gains with Genomic Selection.” Theoretical and Applied Genetics 132 (3): 669–86. https://doi.org/10.1007/s00122-018-3270-8.
Wakeley, John. 2009. Coalescent Theory: An Introduction (Chapter 1). Greenwood Village, CO, USA: Roberts and Company.
Waldmann, Patrik. 2018. “Approximate Bayesian Neural Networks in Genomic Prediction.” Genetics Selection Evolution 50 (1): 70. https://doi.org/10.1186/s12711-018-0439-1.
Walsh, Bruce. 2006. Evolutionary Quantitative Genetics. June. https://books.google.es/books?id=G70GCAAAQBAJ.
Walter, Achim, Frank Liebisch, and Andreas Hund. 2015. “Plant Phenotyping: From Bean Weighing to Image Analysis.” *Plant Methods* 11: 14. https://doi.org/10.1186/s13007-015-0056-8.
Walsh, Bruce, and Michael Lynch. 2018. Evolution and Selection of Quantitative Traits. 1st ed. May. Oxford University Press. http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html#2B.
Wandinger, Ulla. 2005. “Introduction to Lidar.” In Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere, edited by Claus Weitkamp, 1–18. Springer Series in Optical

Sciences. New York, NY: Springer. https://doi.org/10.1007/0-387-25101-4_1.
Wang, Hongru, Xun Xu, Filipe Garrett Vieira, Yunhua Xiao, Zhikang Li, Jun Wang, Rasmus Nielsen, and Chengcai Chu. 2016. “The Power of Inbreeding: NGS-Based GWAS of Rice Reveals Convergent Evolution during Rice Domestication.” Molecular Plant 9 (7): 975–85. https://doi.org/10.1016/j.molp.2016.04.018.
Wang, Huanhuan, Yi Lin, Zeliang Wang, Yunjun Yao, Yuhu Zhang, and Ling Wu. 2017. “Validation of a Low-Cost 2D Laser Scanner in Development of a More-Affordable Mobile Terrestrial Proximal Sensing System for 3D Plant Structure Phenotyping in Indoor Environment.” Computers and Electronics in Agriculture 140 (August): 180–89. https://doi.org/10.1016/j.compag.2017.06.002.
Wang, Jiabo, Zhengkui Zhou, Zhe Zhang, Hui Li, Di Liu, Qin Zhang, Peter J. Bradbury, Edward S. Buckler, and Zhiwu Zhang. 2018. “Expanding the BLUP Alphabet for Genomic Prediction Adaptable to the Genetic Architectures of Complex Traits.” Heredity 121 (6): 648–662. https://doi.org/10.1038/s41437-018-0075-0.
Wang, Long, Yilun Ji, Yingwen Hu, Huaying Hu, Xianqin Jia, Mengmeng Jiang, Xiaohui Zhang, et al. 2019. “The Architecture of Intra-Organism Mutation Rate Variation in Plants.” PLOS Biology 17 (4): e3000191. https://doi.org/10.1371/journal.pbio.3000191.
Wang, Meiyue, and Shizhong Xu. 2019. “Statistical Power in Genome-Wide Association Studies and Quantitative Trait Locus Mapping.” Heredity 123 (3): 287–306. https://doi.org/10.1038/s41437-019-0205-3.
Wang, Pei, Gerassimos Peteinatos, Hui Li, Frank Brändle, Erhard Pfündel, Hans G. Drobny, and Roland Gerhards. 2017. “Rapid Monitoring of Herbicide-Resistant Alopecurus Myosuroides Huds. Using Chlorophyll Fluorescence Imaging Technology.” Journal of Plant Diseases and Protection, October, 1–9. https://doi.org/10.1007/s41348-017-0131-7.
Wang P, Peteinatos GG, Gerhards R. 2017a. “In field identification of herbicide resistant Apera spica-venti using chlorophyll fluorescence.” *Adv Anim Biosci*. 8(02):283-287.
Wang, Songhu, and Eduardo Blumwald. 2014. “Stress-Induced Chloroplast Degradation in Arabidopsis Is Regulated via a Process Independent of Autophagy and Senescence-Associated Vacuoles.” The Plant Cell 26 (12): 4875–88. https://doi.org/10.1105/tpc.114.133116.
Ward, Sarah M., & Marie Jasieniuk. (2009). Review: Sampling Weedy and Invasive Plant Populations for Genetic Diversity Analysis. *Weed Science 57, no. 6* : 593-602. http://www.jstor.org/stable/40586878.
Warmuth, Vera M., and Hans Ellegren. 2019. “Genotype-Free Estimation of Allele Frequencies Reduces Bias and Improves Demographic Inference from RADSeq Data.” Molecular Ecology Resources, January. https://doi.org/10.1111/1755-0998.12990.
Warr, Amanda, Christelle Robert, David Hume, Alan Archibald, Nader Deeb, and Mick Watson. 2015. “Exome Sequencing: Current and Future Perspectives.” G3 (Bethesda, Md.) 5 (8): 1543–50. https://doi.org/10.1534/g3.115.018564.
Watson, J. D., and F. H. C. Crick. 1953. “Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid.” Nature 171 (4356): 737–38. https://doi.org/10.1038/171737a0.
Watterson, G. A. 1975. “On the Number of Segregating Sites in Genetical Models without Recombination.” Theoretical Population Biology 276 (7): 256–276. https://doi.org/10.1039/b316709g.
Watterson, G. A. 1978. “The Homozygosity Test of Neutrality.” Genetics 88 (2): 405–17.
Weir, Bruce S., and C. Clark Cockerham. 1984. “Estimating F-Statistics for the Analysis of Population Structure Author (s): B . S . Weir and C . Clark Cockerham Reviewed Work (s): Published by : Society for the Study of Evolution Stable URL : Http://Www.Jstor.Org/ Stable/2408641 .” Evolution 38 (6): 1358–1370. https://doi.org/128.103.149.52.
Weiss, Matthew, Richard A. Sniezko, Daniela Puiu, Marc W. Crepeau, Kristian Stevens, Steven L. Salzberg, Charles H. Langley, David B. Neale, and Amanda R. De La Torre. (2020). “Genomic Basis of White Pine Blister Rust Quantitative Disease Resistance and Its Relationship with Qualitative Resistance.” *The Plant Journal* 104, no. 2: 365–76. https://doi.org/10.1111/tpj.14928.
Wenzl, Peter, Jason Carling, David Kudrna, Damian Jaccoud, Eric Huttner, Andris Kleinhofs, and Andrzej Kilian. 2004. “Diversity Arrays Technology (DArT) for Whole-Genome Profiling of Barley.” Proceedings of the National Academy of Sciences 101 (26): 9915–9920. https://doi.org/10.1073/pnas.0401076101.
West, Mike. 1984. “Outlier Models and Prior Distributions in Bayesian Linear Regression.”

Journal of the Royal Statistical Society: Series B (Methodological) 46 (3): 431–39. https://doi.org/10.1111/j.2517-6161.1984.tb01317.x.
Weston, Leslie A. 1990. “Cover Crop and Herbicide Influence on Row Crop Seedling Establishment in No-Tillage Culture.” Weed Science 38 (2): 166–71. https://doi.org/10.1017/S0043174500056320.
Wetterstrand, KA. 2020. “DNA Sequencing Costs: Data.” Genome.Gov. August 25, 2020. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
White, G. M., S. R. Moss, and A. Karp. 2005. “Differences in the Molecular Basis of Resistance to the Cyclohexanedione Herbicide Sethoxydim in Lolium Multiflorum.” Weed Research 45 (6): 440–448. https://doi.org/10.1111/j.1365-3180.2005.00478.x.
Whittaker, John C., Robin Thompson, and Mike C. Denham. 2000. “Marker-Assisted Selection Using Ridge Regression.” Genetical Research 75 (2): 249–52. https://doi.org/10.1017/S0016672399004462.
Wick, R.R., L.M Judd, and K.E. Holt. 2019. “Performance of neural network basecalling tools for Oxford Nanopore sequencing.” *Genome Biol* 20**,** 129. https://doi.org/10.1186/s13059-019-1727-y
Wijmenga, Cisca, and Alexandra Zhernakova. 2018. “The Importance of Cohort Studies in the Post-GWAS Era.” Nature Genetics 50 (3): 322–28. https://doi.org/10.1038/s41588-018-0066-3.
Williams, Byron K., and Eleanor D. Brown. 2019. “Sampling and Analysis Frameworks for Inference in Ecology.” Methods in Ecology and Evolution 10 (11): 1832–42. https://doi.org/10.1111/2041-210X.13279.
Willing, Eva-Maria, Paul Bentzen, COCK Van Oosterhout, Margarete Hoffmann, Joanne Cable, Felix Breden, Detlef Weigel, and Christine Dreyer. 2010. “Genome-Wide Single Nucleotide Polymorphisms Reveal Population History and Adaptive Divergence in Wild Guppies.” Molecular Ecology 19 (5): 968–84. https://doi.org/10.1111/j.1365-294X.2010.04528.x.
Willis, A. J., and J. Memmott. 2005. “The Potential for Indirect Effects between a Weed, One of Its Biocontrol Agents and Native Herbivores: A Food Web Approach.” Biological Control, Science and Decision Making in Biological Control of Weeds: Benefits and Risks of Biological Control, 35 (3): 299–306. https://doi.org/10.1016/j.biocontrol.2005.07.013.
Wilson, C L. 1969. “Use of Plant Pathogens in Weed Control.” Annual Review of Phytopathology 7 (1): 411–34. https://doi.org/10.1146/annurev.py.07.090169.002211.
Wilson, R. S., N. Hooker, M. Tucker, J. LeJeune, and D. Doohan. 2009. “Targeting the Farmer Decision Making Process: A Pathway to Increased Adoption of Integrated Weed Management.” Crop Protection 28 (9): 756–64. https://doi.org/10.1016/j.cropro.2009.05.013.
Wojcik, Genevieve L., Mariaelisa Graff, Katherine K. Nishimura, Ran Tao, Jeffrey Haessler, Christopher R. Gignoux, Heather M. Highland, et al. 2019. “Genetic Analyses of Diverse Populations Improves Discovery for Complex Traits.” Nature 570 (7762): 514–18. https://doi.org/10.1038/s41586-019-1310-4.
Wolf, Thomas M., S. Kent Harrison, Franklin R. Hall, and Jane Cooper. 2000. “Optimizing Postemergence Herbicide Deposition and Efficacy through Application Variables in No-till Systems.” Weed Science 48 (6): 761–68. https://doi.org/10.1614/0043-1745(2000)048[0761:OPHDAE]2.0.CO;2.
Wong, Tzu-Tsung. 2015. "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation." Pattern Recognition. Volume 48, Issue 9. pp 2839-2846. ISSN 0031-3203. https://doi.org/10.1016/j.patcog.2015.03.009.
Wordsworth, Sarah, Brett Doble, Katherine Payne, James Buchanan, Deborah A. Marshall, Christopher McCabe, and Dean A. Regier. 2018. “Using ‘Big Data’ in the Cost-Effectiveness Analysis of Next-Generation Sequencing Technologies: Challenges and Potential Solutions.” Value in Health 21 (9): 1048–53. https://doi.org/10.1016/j.jval.2018.06.016.
Worthington, Margaret, and Chris Reberg-Horton. 2013. “Breeding Cereal Crops for Enhanced Weed Suppression: Optimizing Allelopathy and Competitive Ability.” Journal of Chemical Ecology 39 (2): 213–31. https://doi.org/10.1007/s10886-013-0247-6.
Wright, S. 1938. “The Distribution of Gene Frequencies Under Irreversible Mutation.” Proceedings of the National Academy of Sciences 24 (7): 253–59. https://doi.org/10.1073/pnas.24.7.253.
Wright, Sewall. 1921. “Systems of Mating. V. General Considerations.” Genetics 6 (2): 167–78.

Wright, Sewall. 1931. “Evolution in Mendelian Populations.” Genetics 16 (2): 97–159.
Wright, Sewall. 1949. “The Genetical Structure of Populations.” Annals of Eugenics 15 (1): 323–54. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x.
Wright, Sewall. 1984. Evolution and the Genetics of Populations, Volume 1: Genetic and Biometric Foundations. University of Chicago Press.
WSSA. 2018. Weed Science Society of America. http://wssa.net/.
Wu, Pingxian, Kai Wang, Jie Zhou, Qiang Yang, Xidi Yang, Anan Jiang, Yanzhi Jiang, et al. 2019. “A Genome Wide Association Study for the Number of Animals Born Dead in Domestic Pigs.” BMC Genetics 20 (1): 4. https://doi.org/10.1186/s12863-018-0692-x.
Xiao, Yingjie, Haijun Liu, Liuji Wu, Marilyn Warburton, and Jianbing Yan. 2017. “Genome-Wide Association Studies in Maize: Praise and Stargaze.” Molecular Plant 10 (3): 359–74. https://doi.org/10.1016/j.molp.2016.12.008.
Xu, Hongyan, and Yun-Xin Fu. 2004. “Estimating Effective Population Size or Mutation Rate with Microsatellites.” Genetics 166 (1): 555–63.
Yaduraju, Nanjapur, and A.N. Rao. 2013. “Implications of Weeds and Weed Management on Food Security and Safety in the Asia-Pacific Region.” In .
Yamamoto, E. 1998. “A-Tubulin Missense Mutations Correlate with Antimicrotubule Drug Resistance in Eleusine Indica.” The Plant Cell 10 (2): 297–308. https://doi.org/10.1105/tpc.10.2.297.
Yanchukov, Alexey, and Stephen Proulx. 2012. “Invasion of Gene Duplication Through Masking for Maladaptive Gene Flow.” Evolution 66 (5): 1543–55. https://doi.org/10.1111/j.1558-5646.2011.01551.x.
Yang, Guo-Qian, Yun-Mei Chen, Jin-Peng Wang, Cen Guo, Lei Zhao, Xiao-Yan Wang, Ying Guo, Li Li, De-Zhu Li, and Zhen-Hua Guo. 2016. “Development of a Universal and Simplified DdRAD Library Preparation Approach for SNP Discovery and Genotyping in Angiosperm Plants.” Plant Methods 12 (1): 39. https://doi.org/10.1186/s13007-016-0139-1.
Yang, In Seok, and Sangwoo Kim. 2015. “Analysis of Whole Transcriptome Sequencing Data: Workflow and Software.” Genomics & Informatics 13 (4): 119–25. https://doi.org/10.5808/GI.2015.13.4.119.
Yang, Jian, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, et al. 2010. “Common SNPs Explain a Large Proportion of the Heritability for Human Height.” Nature Genetics 42 (7): 565–69. https://doi.org/10.1038/ng.608.
Yang, Jinliang, Haiying Jiang, Cheng-Ting Yeh, Jianming Yu, Jeffrey A. Jeddeloh, Dan Nettleton, and Patrick S. Schnable. 2015. “Extreme-Phenotype Genome-Wide Association Study (XP-GWAS): A Method for Identifying Trait-Associated Variants by Sequencing Pools of Individuals Selected from a Diversity Panel.” The Plant Journal 84 (3): 587–596. https://doi.org/10.1111/tpj.13029.
Yang, Qian, Wei Deng, Xuefeng Li, Qin Yu, Lianyang Bai, and Mingqi Zheng. 2016. “Target-Site and Non-Target-Site Based Resistance to the Herbicide Tribenuron-Methyl in Flixweed (Descurainia Sophia L.).” BMC Genomics 17 (October): 551–564. https://doi.org/10.1186/s12864-016-2915-8.
Yang, Wanneng, Hui Feng, Xuehai Zhang, Jian Zhang, John H. Doonan, William David Batchelor, Lizhong Xiong, and Jianbing Yan. 2020. “Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives.” Molecular Plant 13 (2): 187–214. https://doi.org/10.1016/j.molp.2020.01.008.
Yang, Yun Young, and Jae Geun Kim. 2016. “The Optimal Balance between Sexual and Asexual Reproduction in Variable Environments: A Systematic Review.” Journal of Ecology and Environment 40 (1): 12. https://doi.org/10.1186/s41610-016-0013-0.
Yasumizu, Yoshiaki, Saori Sakaue, Takahiro Konuma, Ken Suzuki, Koichi Matsuda, Yoshinori Murakami, Michiaki Kubo, Pier Francesco Palamara, Yoichiro Kamatani, and Yukinori Okada. 2020. “Genome-Wide Natural Selection Signatures Are Linked to Genetic Risk of Modern Phenotypes in the Japanese Population.” Molecular Biology and Evolution 37 (5): 1306–16. https://doi.org/10.1093/molbev/msaa005.
Yates, Katherine L, Phil J Bouchet, M Julian Caley, Kerrie Mengersen, Christophe F Randin, Stephen Parnell, Alan H Fielding, et al. 2018. “Outstanding Challenges in the Transferability of Ecological Models.” *Trends in Ecology & Evolution* 33 (10): 790–802. https://doi.org/10.1016/j.tree.2018.08.001.
Yoo, Andy B., Morris A. Jette, and Mark Grondona. 2003. “SLURM: Simple Linux Utility for

Resource Management." In Job Scheduling Strategies for Parallel Processing, edited by Dror Feitelson, Larry Rudolph, and Uwe Schwiegelshohn, 44–60. Berlin, Heidelberg: Springer Berlin Heidelberg.

You, Frank M., Jin Xiao, Pingchuan Li, Zhen Yao, Gaofeng Jia, Liqiang He, Santosh Kumar, et al. 2018. "Genome-Wide Association Study and Selection Signatures Detect Genomic Regions Associated with Seed Yield and Oil Quality in Flax." International Journal of Molecular Sciences 19 (8). https://doi.org/10.3390/ijms19082303.

Yuan, Joshua S., Patrick J. Tranel, and C. Neal Stewart. 2007. "Non-Target-Site Herbicide Resistance: A Family Business." Trends in Plant Science 12 (1): 6–13. https://doi.org/10.1016/j.tplants.2006.11.001.

Yuan, Zhaofeng, Qiang Cao, Ke Zhang, Syed Tahir Ata-Ul-Karim, Yongchao Tian, Yan Zhu, Weixing Cao, and Xiaojun Liu. 2016. "Optimal Leaf Positions for SPAD Meter Measurement in Rice." Frontiers in Plant Science 7. https://doi.org/10.3389/fpls.2016.00719.

Yu, Huihui, Weibo Xie, Jing Li, Fasong Zhou, and Qifa Zhang. 2014. "A Whole-Genome SNP Array (RICE6K) for Genomic Breeding in Rice." Plant Biotechnology Journal 12 (1): 28–37. https://doi.org/10.1111/pbi.12113.

Yu, Jianming, Gael Pressoir, William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, et al. 2006. "A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness." Nature Genetics 38 (2): 203–8. https://doi.org/10.1038/ng1702.

Yu, Jianming, James B. Holland, Michael D. McMullen, and Edward S. Buckler. 2008. "Genetic Design and Statistical Power of Nested Association Mapping in Maize." Genetics 178 (1): 539–551. https://doi.org/10.1534/genetics.107.074245.

Yu, Q., A. Collavo, M.-Q. Zheng, M. Owen, M. Sattin, and S. B. Powles. 2007. "Diversity of Acetyl-Coenzyme A Carboxylase Mutations in Resistant Lolium Populations: Evaluation Using Clethodim." Plant Physiology 145 (2): 547–558. https://doi.org/10.1104/pp.107.105262.

Yu, Q., H. Han, G. R. Cawthray, S. F. Wang, and S. B. Powles. 2013. "Enhanced Rates of Herbicide Metabolism in Low Herbicide-Dose Selected Resistant Lolium Rigidum." Plant, Cell and Environment 36 (4). https://doi.org/10.1111/pce.12017.

Yu, Qin, Heping Han, and Stephen B. Powles. 2008. "Mutations of the ALS Gene Endowing Resistance to ALS-Inhibiting Herbicides in Lolium Rigidum Populations." Pest Management Science 64 (12): 1229–36. https://doi.org/10.1002/ps.1624.

Yu, Qin, Ibrahim Abdallah, Heping Han, Mechelle Owen, and Stephen Powles. 2009. "Distinct Non-Target Site Mechanisms Endow Resistance to Glyphosate, ACCase and ALS-Inhibiting Herbicides in Multiple Herbicide-Resistant Lolium Rigidum." Planta 230 (4): 713–23. https://doi.org/10.1007/s00425-009-0981-8.

Yu, Q, M S Ahmad-Hamdani, H Han, M J Christoffers, and S B Powles. 2013. "Herbicide Resistance-Endowing ACCase Gene Mutations in Hexaploid Wild Oat (Avena Fatua): Insights into Resistance Evolution in a Hexaploid Species." Heredity 110 (3): 220–231. https://doi.org/10.1038/hdy.2012.69.

Zaitlen, Noah, Peter Kraft, Nick Patterson, Bogdan Pasaniuc, Gaurav Bhatia, Samuela Pollack, and Alkes L. Price. 2013. "Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits." PLoS Genetics 9 (5). https://doi.org/10.1371/journal.pgen.1003520.

Zappa Nardelli, Francesco, Julia Belyakova, Artem Pelenitsyn, Benjamin Chung, Jeff Bezanson, and Jan Vitek. 2018. "Julia Subtyping: A Rational Reconstruction." Proc. ACM Program. Lang. 2 (OOPSLA): 113:1–113:27. https://doi.org/10.1145/3276483.

Zeng, Zhao-Bang. 2005. "QTL Mapping and the Genetic Basis of Adaptation: Recent Developments." Genetica 123 (1): 25–37. https://doi.org/10.1007/s10709-004-2705-0.

Zhang, C J, S H Lim, J W Kim, G Nah, A Fischer, and D S Kim. 2016. "Leaf Chlorophyll Fluorescence Discriminates Herbicide Resistance in Echinochloa Species." Edited by Per Kudsk. *Weed Research* 56 (6): 424–433. https://doi.org/10.1111/wre.12226.

Zhang, Xiao Qi, and Stephen B. Powles. 2006. "The Molecular Bases for Resistance to Acetyl Co-Enzyme A Carboxylase (ACCase) Inhibiting Herbicides in Two Target-Based Resistant Biotypes of Annual Ryegrass (Lolium Rigidum)." Planta 223 (3): 550–557. https://doi.org/10.1007/s00425-005-0095-x.

Zhang, Xu-Sheng, and William G. Hill. 2005. "Genetic Variability under Mutation Selection Balance." Trends in Ecology & Evolution 20 (9): 468–70.

https://doi.org/10.1016/j.tree.2005.06.010.
Zhao, Can, Wenrong Xu, Xiaoling Song, Weimin Dai, Lei Dai, Zheng Zhang, and Sheng Qiang. 2018. “Early Flowering and Rapid Grain Filling Determine Early Maturity and Escape from Harvesting in Weedy Rice.” Pest Management Science 74 (2): 465–76. https://doi.org/10.1002/ps.4730.
Zhao, Chunjiang, Ying Zhang, Jianjun Du, Xinyu Guo, Weiliang Wen, Shenghao Gu, Jinglu Wang, and Jiangchuan Fan. 2019. “Crop Phenomics: Current Status and Perspectives.” Frontiers in Plant Science 10. https://doi.org/10.3389/fpls.2019.00714.
Zhao, Yun, Ke Wang, Wen-li Wang, Ting-ting Yin, Wei-qi Dong, and Chang-jie Xu. 2019. “A High-Throughput SNP Discovery Strategy for RNA-Seq Data.” BMC Genomics 20 (1): 160. https://doi.org/10.1186/s12864-019-5533-4.
Zhao, Yusheng, Jian Zeng, Rohan Fernando, and Jochen C. Reif. 2013. “Genomic Prediction of Hybrid Wheat Performance.” Crop Science 53 (3): 802–10. https://doi.org/10.2135/cropsci2012.08.0463.
Zhou, Xiang, and Matthew Stephens. 2012. “Genome-Wide Efficient Mixed-Model Analysis for Association Studies.” Nature Genetics 44 (7): 821–24. https://doi.org/10.1038/ng.2310.
Zhu, Yuan, Alan O. Bergland, Josefa González, and Dmitri A. Petrov. 2012. “Empirical Validation of Pooled Whole Genome Population Re-Sequencing in Drosophila Melanogaster.” Edited by Jeong-Sun Seo. PLoS ONE 7 (7): e41901. https://doi.org/10.1371/journal.pone.0041901.
Zimdahl, Robert L. 2007. Fundamentals of Weed Science. Vol. 1. https://doi.org/10.1016/0378-4290(95)90065-9.
Zingaretti, Laura M., Salvador Alejandro Gezan, Luis Felipe V. Ferrão, Luis F. Osorio, Amparo Monfort, Patricio R. Muñoz, Vance M. Whitaker, and Miguel Pérez-Enciso. 2020. “Exploring Deep Learning for Complex Trait Genomic Prediction in Polyploid Outcrossing Species.” Frontiers in Plant Science 11. https://doi.org/10.3389/fpls.2020.00025.
Zuk, Or, Eliana Hechter, Shamil R. Sunyaev, and Eric S. Lander. 2012. “The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability.” Proceedings of the National Academy of Sciences 109 (4): 1193–98. https://doi.org/10.1073/pnas.1119675109.

Appendices

8. Appendices

Chapter 2 Appendix

Table S2.1. List of parameters used in Quantinemo2 simulations.

Parameters	Value	Description
mating_system	0	random mating & hermaphrodite
patch_number	100	100 populations
generations	200	200 generations
patch_capacity	1000	carrying capacity for 1000 individuals per population
regulation_model_offspring	1	regulation of population size to carrying capacity via random culling of offspring
regulation_model_adults	1	regulation of population size to carrying capacity via random culling of adults
mating_nb_offspring_model	0	total number of offspring is set to carrying capacity
mean_fecundity	1	constant population size
dispersal_model	3	2-dimensional stepping-stone model
dispersal_lattice_range	1	dispersal range of 8 adjacent patches: vertical, horizontal and diagonal
dispersal_border_model	2	absorbing boundaries: migration beyond the border is lost
quanti_genome	1 – 7	7 chromosomes
quanti_all	2	biallelic loci
quanti_mutation_rate	0	no mutations
quanti_environmental_model	0	environmental variance constant
quanti_heritability	0.50	narrow-sense heritability set to 50%
quanti_selection_model	2	directional selection
quanti_dir_sel_min	0.0	minimum fitness
quanti_dir_sel_max	1.0	maximum fitness

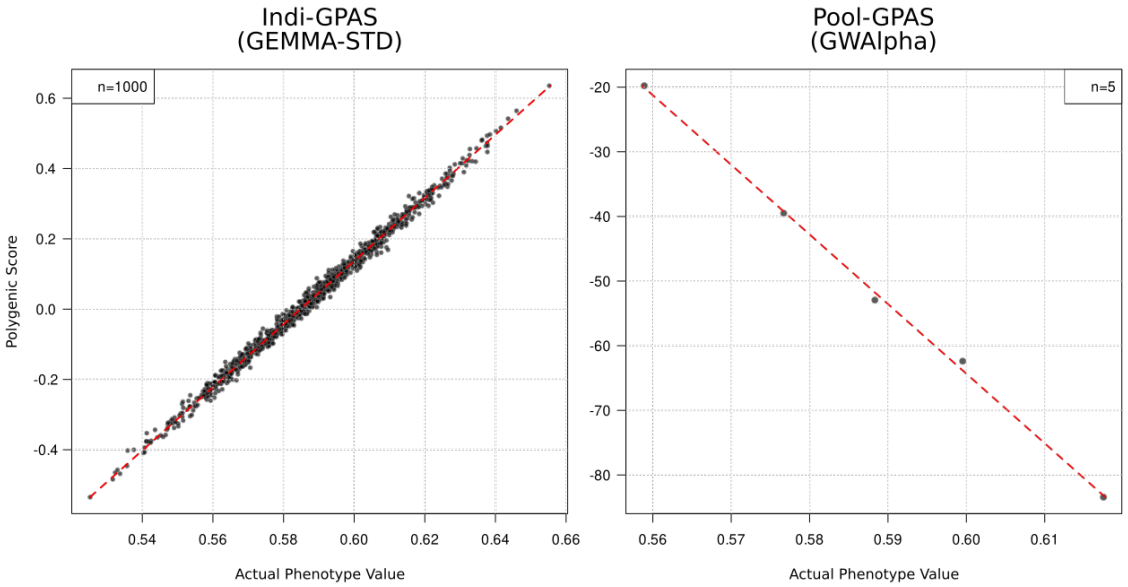


Figure S2.1. Representative plots showing the linear relationships between the actual phenotypic values and the polygenic scores computed from single-SNP estimates using Indi-GPAS and Pool-GPAS models.

Table S2.2. Mean comparison using Tukey's honest significant differences, and the linear (full-factorial) model-derived adjusted coefficients of the GPAS models.

Genotyping Scheme	Model	AUC (GWAS performance)		RMSE (1 - GP performance)	
		Mean	Coefficient	Mean	Coefficient
Individual	EMMAX (GRM)	0.8578 [b]	0.82	0.0573 [b]	0.0411
Individual	EMMAX (STD)	0.8658 [b]	0.80	0.0570 [b]	0.0414
Individual	GCTA (GRM)	0.8521 [d]	0.91	0.0568 [b]	0.0472
Individual	GCTA (STD)	0.8649 [bc]	0.88	0.0596 [a]	0.0106
Individual	GEMMA (GRM)	0.8723 [a]	0.89	0.0547 [c]	0.0523
Individual	GEMMA (STD)	0.8724 [a]	0.89	0.0545 [c]	0.0519
Pool	LMM (Fst Hivert)	0.7269 [f]	0.94	0.0568 [b]	0.0469
Pool	LMM (Fst Weir & Cockerham)	0.7269 [f]	0.94	0.0569 [b]	0.0468
Pool	GWAlpha	0.7936 [e]	1.17	0.0588 [a]	0.0591

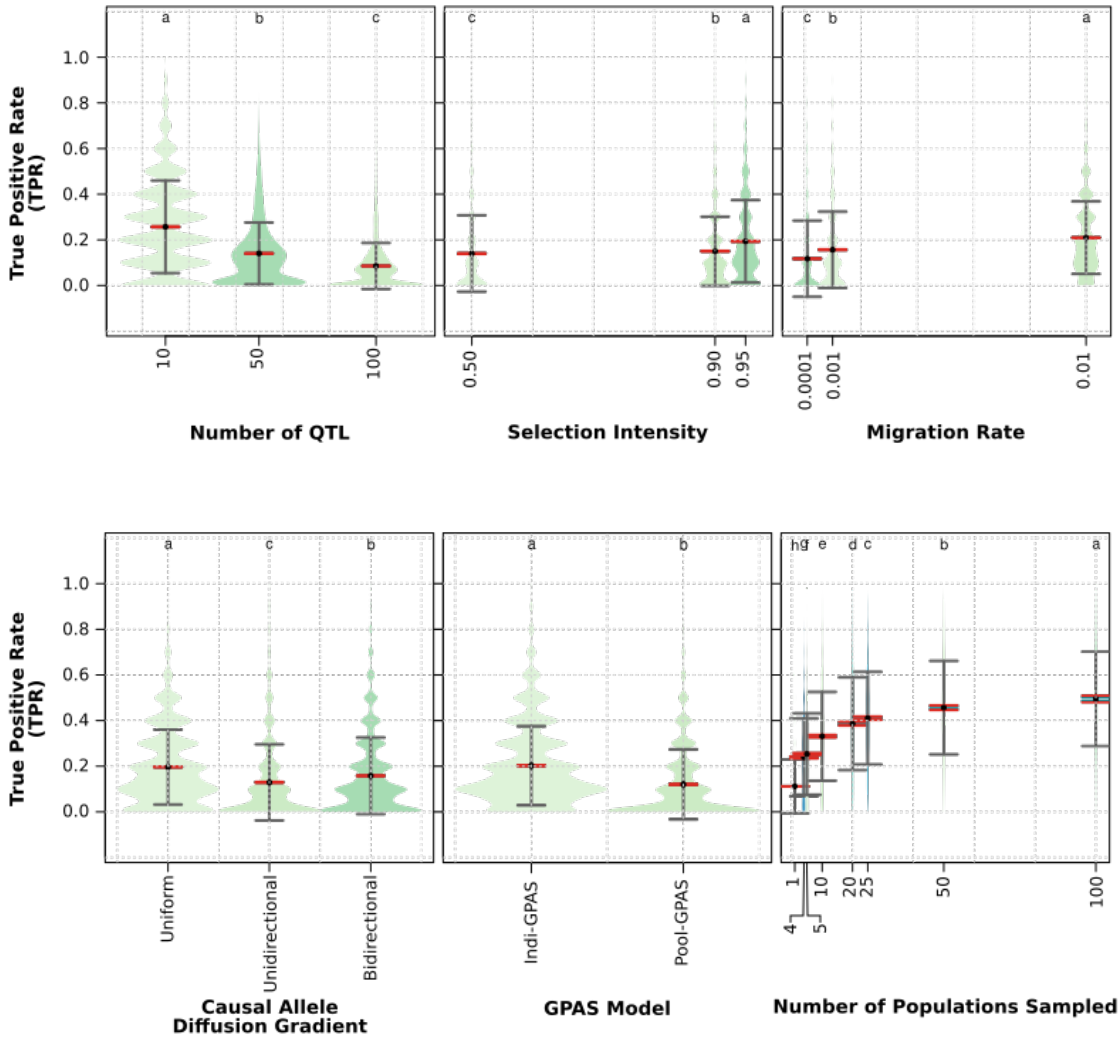


Figure S2.2. Violin plots of the true positive rate (TPR) for the landscape parameters, GPAS models and the number of populations sampled. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. levels with the same letter are not significantly different at $p<0.05$).

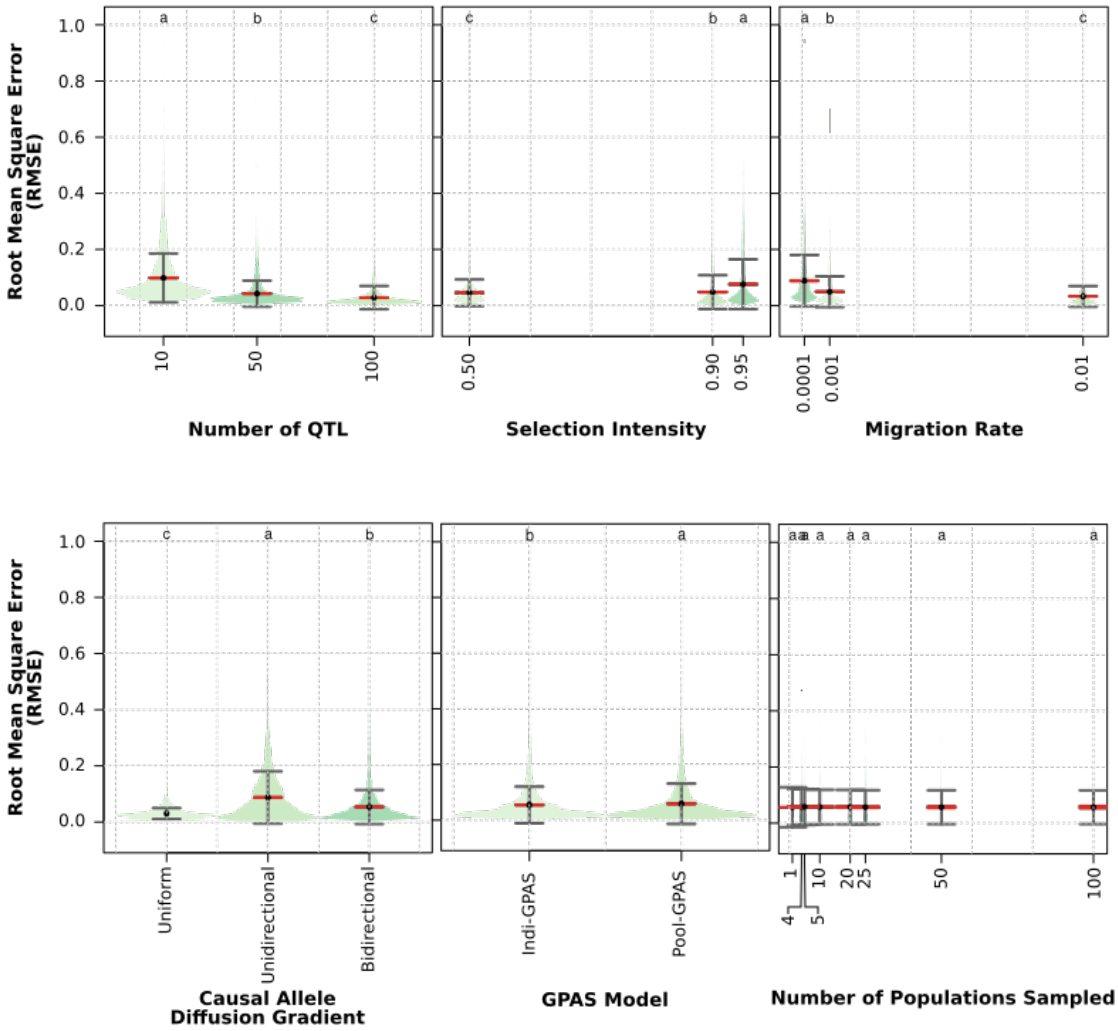


Figure S2.3. Violin plots of the root mean square error (RMSE) for the landscape parameters, GPAS models and the number of populations sampled. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. levels with the same letter are not significantly different at $p<0.05$).

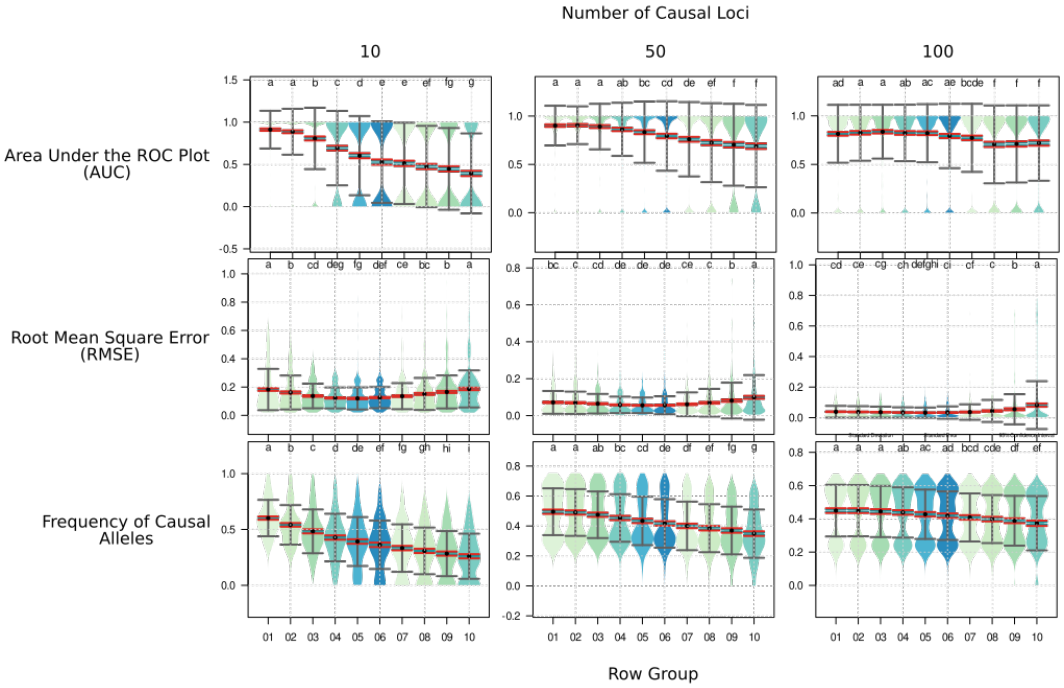


Figure S2.4. Violin plots of GPAS performances across the landscape under unidirectional causal allele diffusion gradient for 10, 50 and 100 simulated QTL. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

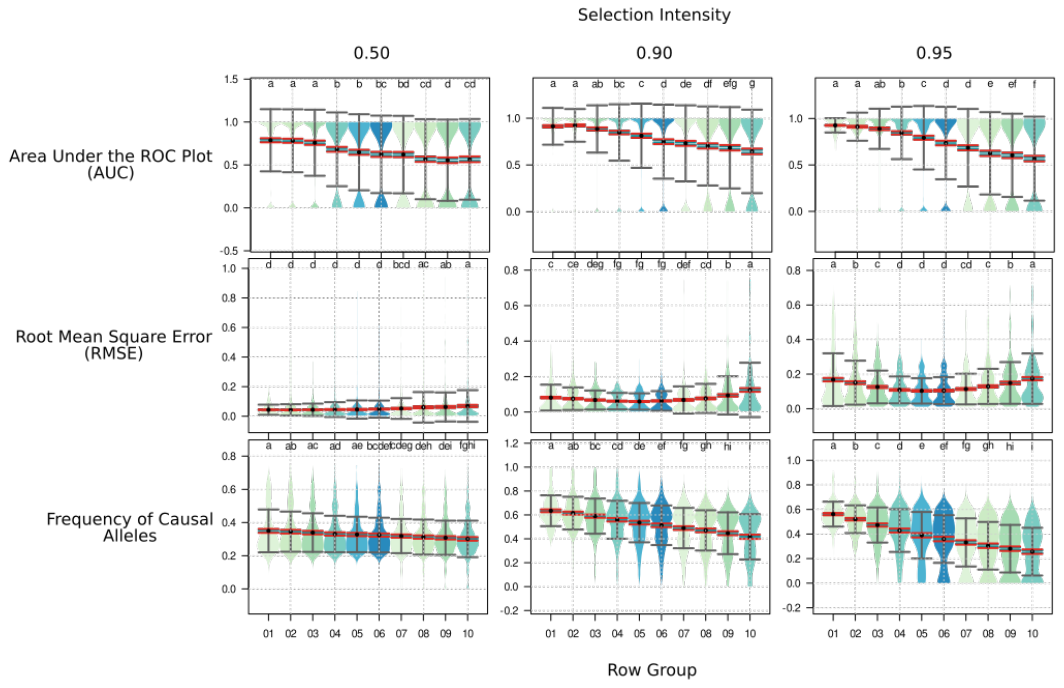


Figure S2.5. Violin plots of GPAS performances across the landscape under unidirectional causal allele diffusion gradient for each level of the simulated selection intensities. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

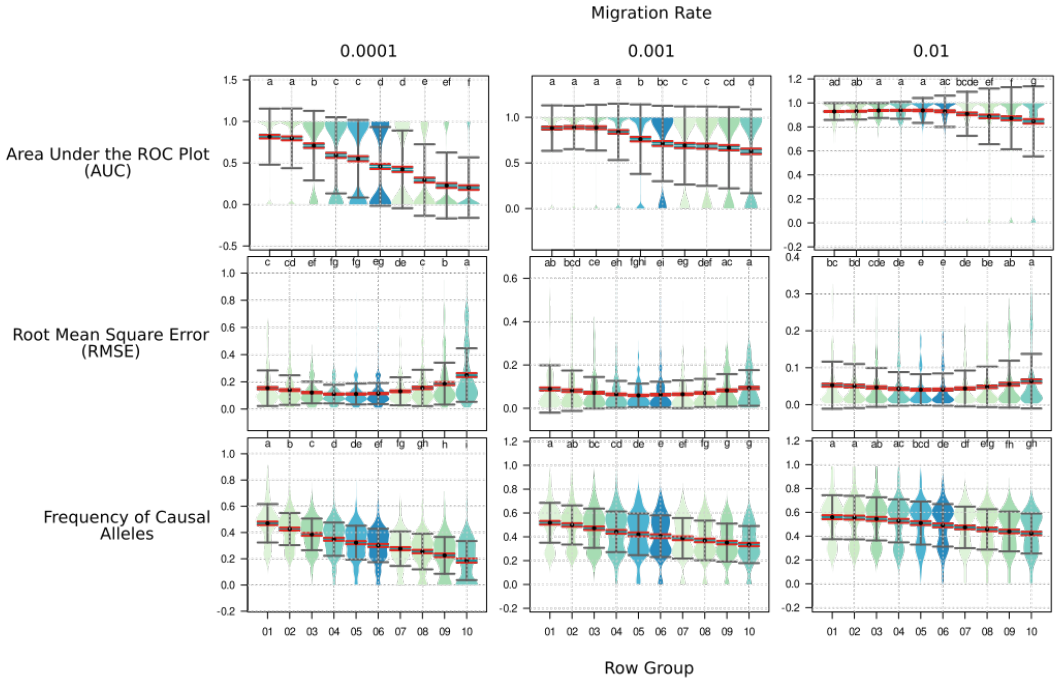


Figure S2.6. Violin plots of GPAS performances across the landscape under unidirectional causal allele diffusion gradient for each level of the simulated migration rates. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

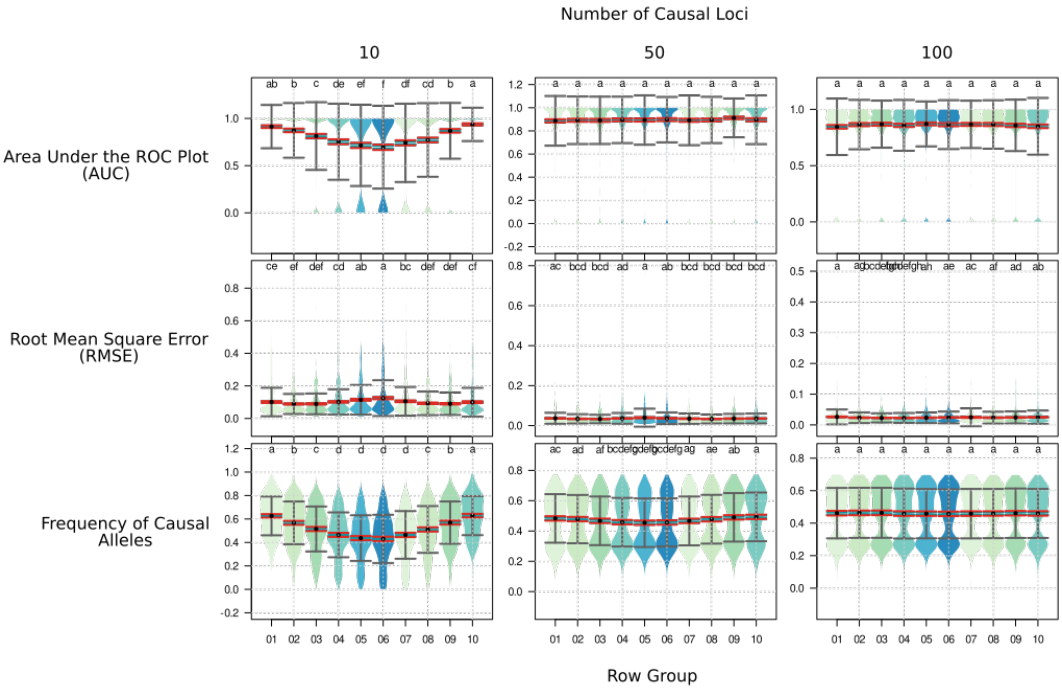


Figure S2.7. Violin plots of GPAS performances across the landscape under bidirectional causal allele diffusion gradient for 10, 50 and 100 simulated QTL. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

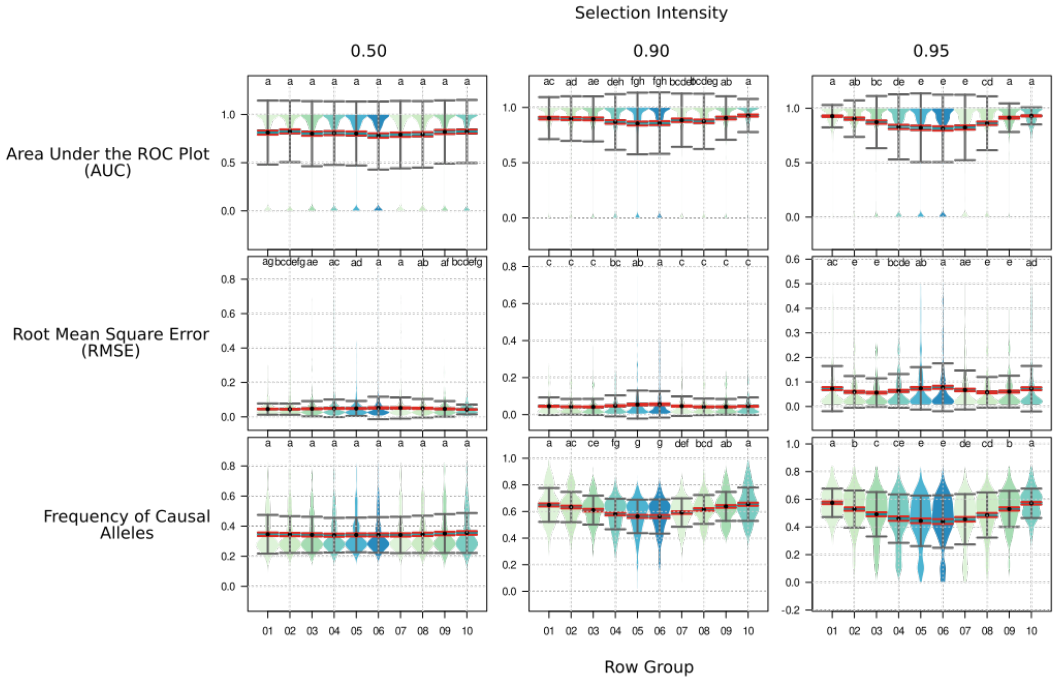


Figure S2.8. Violin plots of GPAS performances across the landscape under bidirectional causal allele diffusion gradient for each level of the simulated selection intensities. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

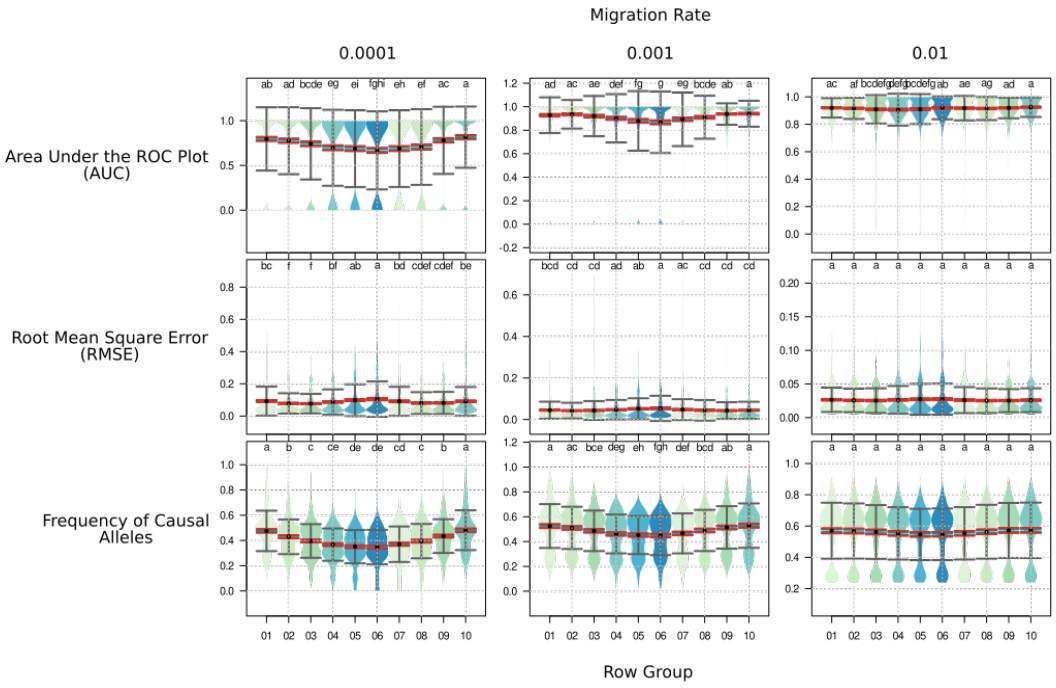


Figure S2.9. Violin plots of GPAS performances across the landscape under bidirectional causal allele diffusion gradient for each level of the simulated migration rates. Black dots represent the mean, black whiskers show ±1 standard deviation, red whiskers are the 95% confidence interval, and the letters on top of each plot are Tukey's honest significant difference (HSD)-based grouping (i.e. row groups with the same letter are not significantly different at $p<0.05$).

Chapter 4 Appendix

Table S4.1. Parameters used in bwa mem for the alignment of Illumina DNA sequencing reads.

Parameter	Value
Minimum seed length	19
Maximum band (gap) length	100
Off-diagonal drop-off (max. extension score difference to stop extension)	100
Trigger re-seeding (decrease for accuracy and increase for speed)	1.5
Maximum MEM (maximal exact match) otherwise discard	500
Minimum chain length fraction with longest overlapping chain otherwise drop	0.5
Minimum chain length	0
Matching score	1
Mismatch penalty	4
Gap open penalty	6
Gap extension penalty Illumina HiSeq X platform in 150bp paired-end reads mode	1
Clipping penalty	5
Penalty for unpaired read pair	17

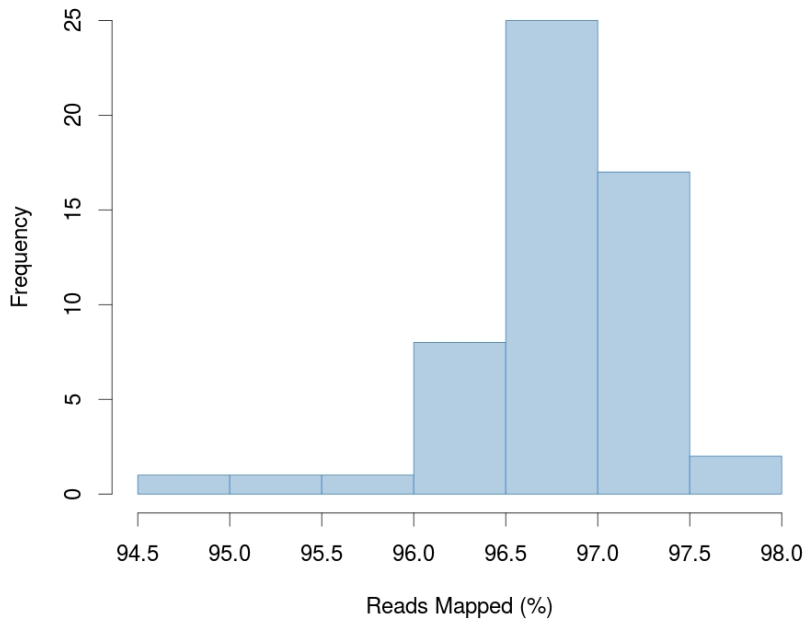


Figure S4.1. Distribution of the proportions of Illumina reads mapped to the *Lolium perenne* draft reference genome.

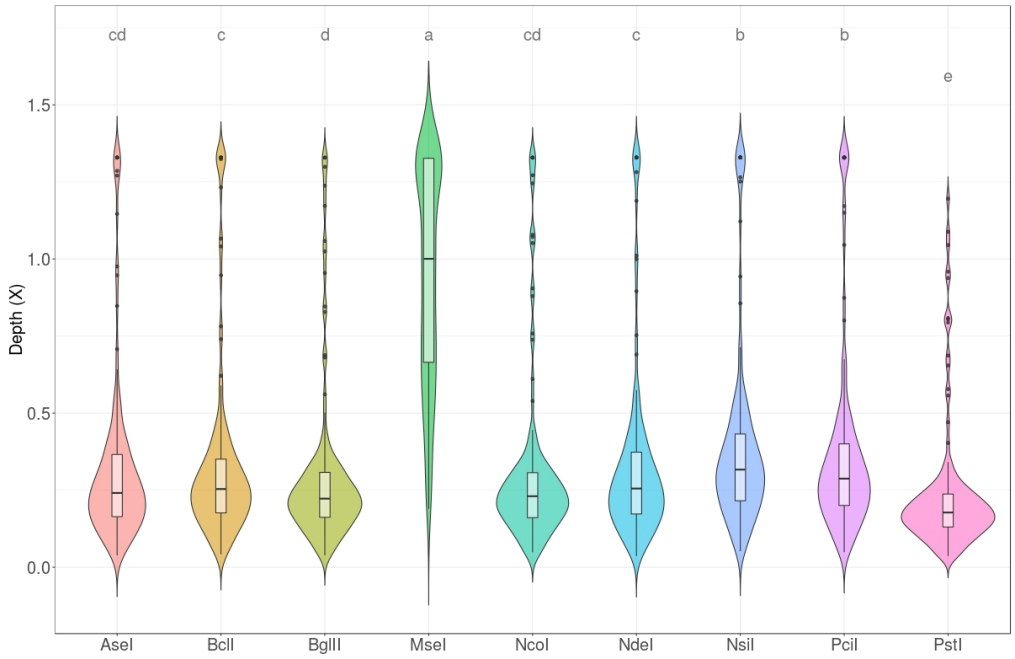


Figure S4.2. Simulated depth of RADseq across the *Lolium perenne* genome using pair-wise combinations of 8 restriction enzymes (letters refer to Tukey's HSD grouping at α=0.05).

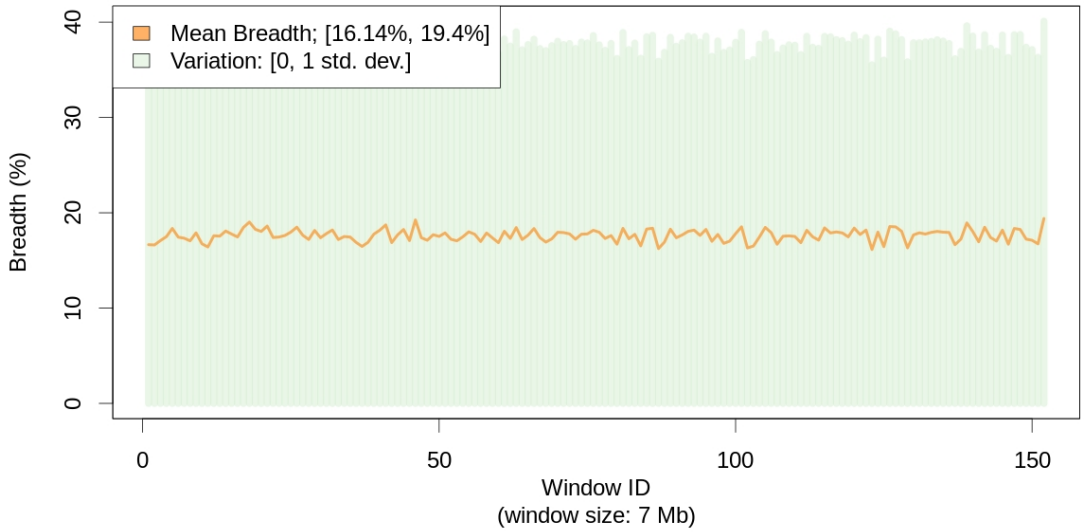


Figure S4.3. Breadth of sequencing coverage across the genome.

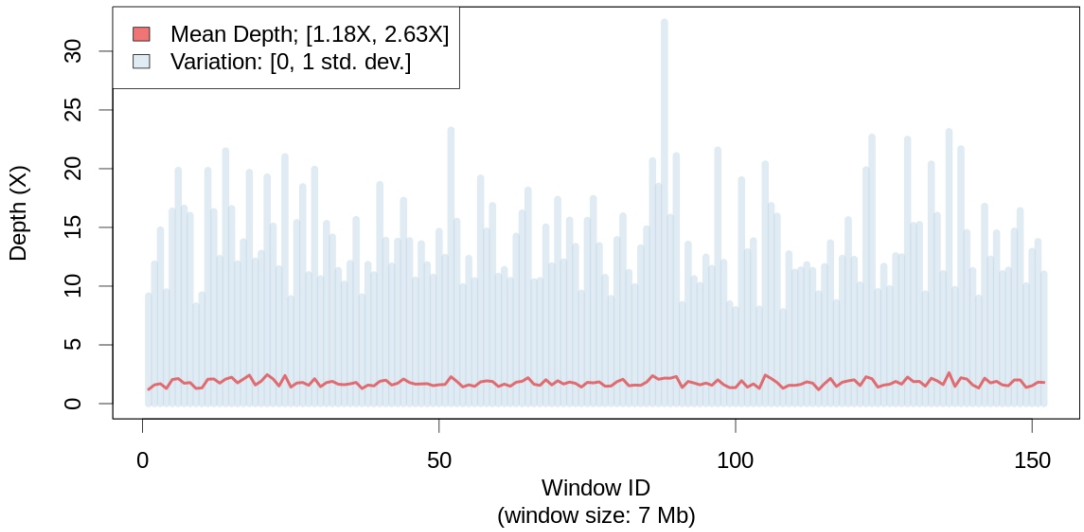


Figure S4.4. Depth of sequencing coverage across the genome.

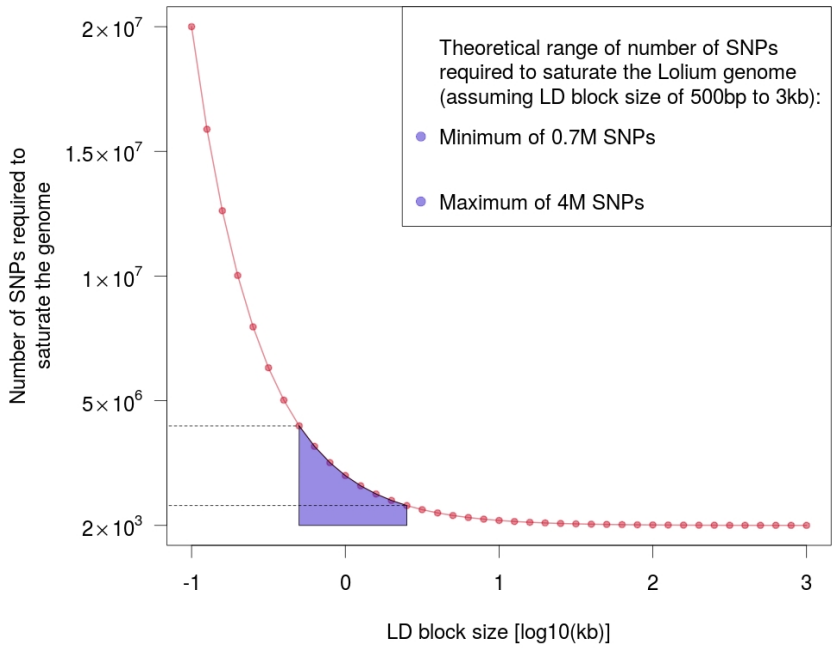


Figure S4.5. Theoretical minimum number of SNPs required to saturate the *Lolium rigidum* genome assuming a genome size of 2Gb as a function of linkage block size (purple area under the curve and dashed lines indicate the range of SNP counts required based on empirical *Lolium perenne* linkage block sizes ranging from 0.5-3kb.

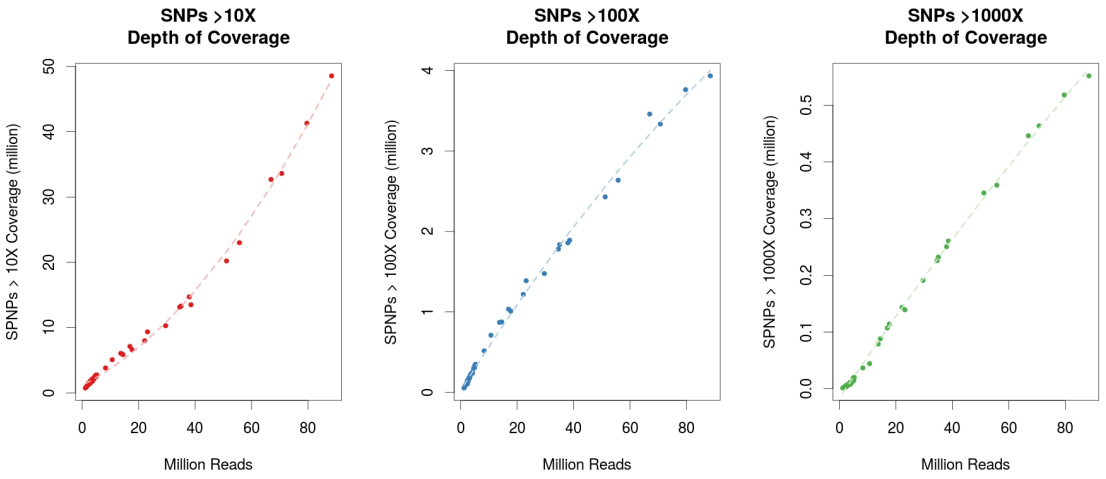


Figure S4.6. Number of SNPs with at least 10X, 100X and 1000X depths as functions of the number of sequencing reads.

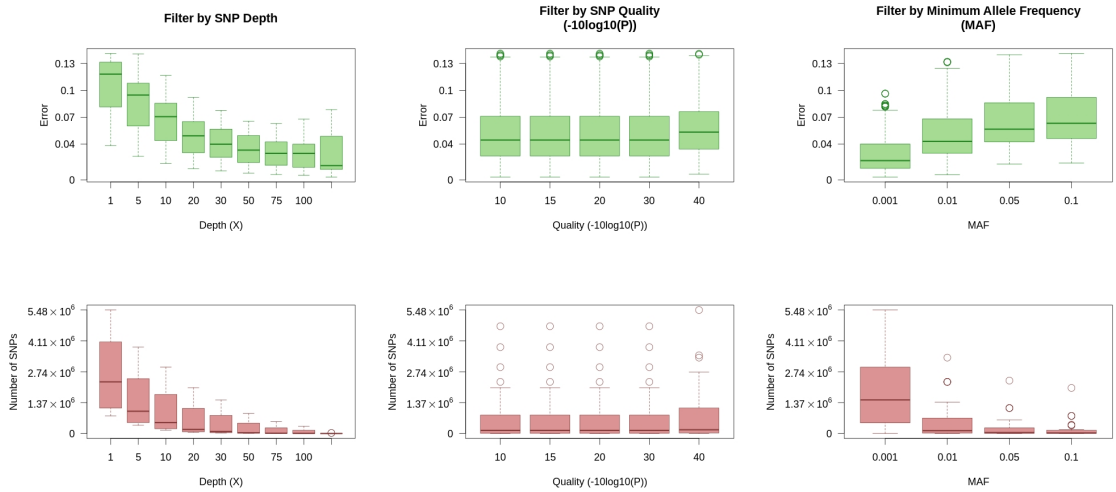


Figure S4.7. Responses of allele frequency accuracy and number of SNPs to varying stringency of the SNP filtering parameters: minimum sequencing depth, minimum base quality, and minimum allele frequency (MAF).

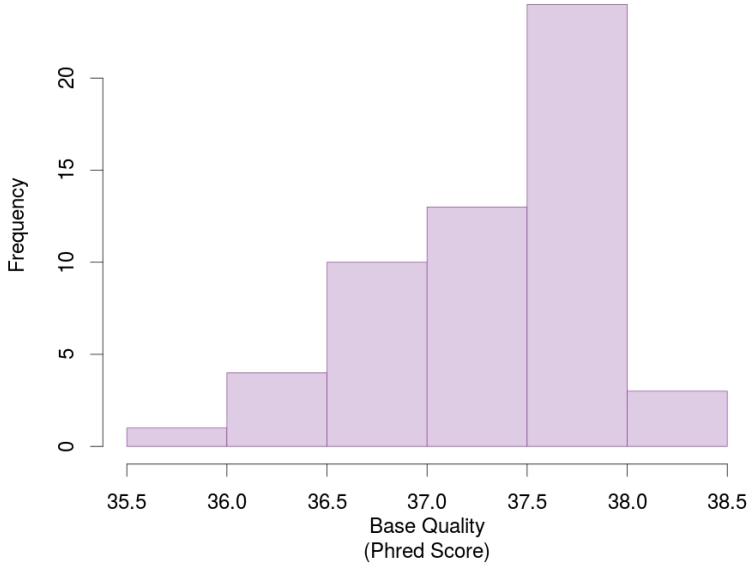


Figure S4.8. Distribution of base calling qualities of the Illumina sequencing data.

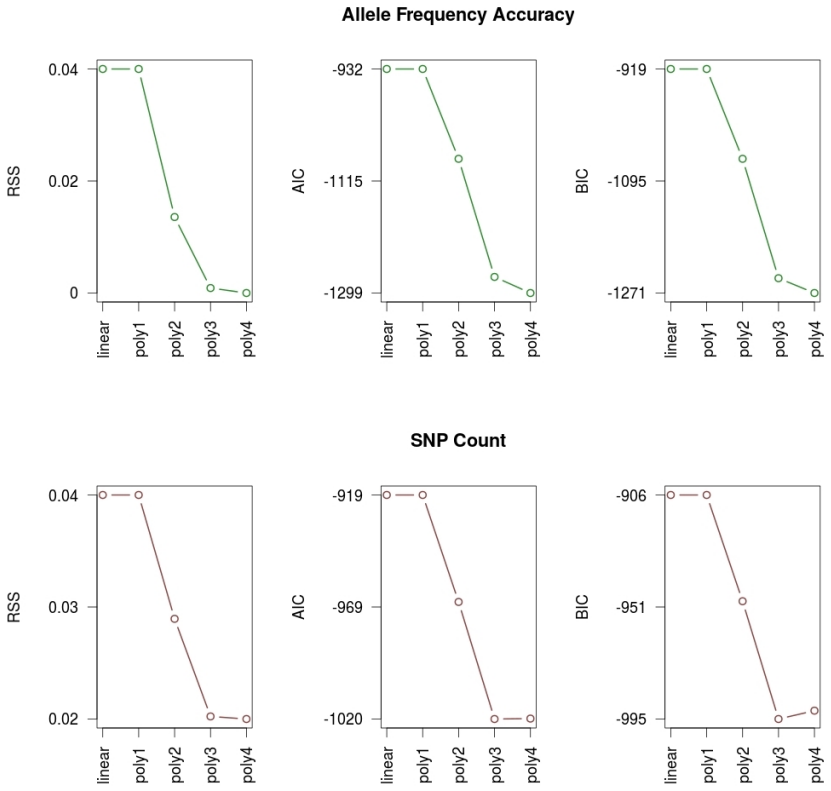


Figure S4.9. Model comparisons between linear and polynomial models of 1st, 2nd, 3rd and 4th degrees using F-test (RSS: residual sum of squares), Akaike information criterion (AIC) and Bayesian information criterion (BIC).

Chapter 5 Appendix

Table S5.1. List of neutral scaffold used in the estimation of likelihoods of various convergent evolutionary models.

Scaffold Name	Maximum -log(p-value)
scaffold_4013|ref0041948	0.30
scaffold_197|ref0030887	0.55
scaffold_37347|ref0031916	0.56
scaffold_44379|ref0025591	0.60
scaffold_1160|ref0004473	0.76
scaffold_21888|ref0019008	0.78
scaffold_359|ref0027898	0.80
scaffold_2421|ref0036714	0.81
scaffold_11246|ref0048313	0.81
scaffold_13444|ref0044543	0.82
scaffold_15405|ref0015202	0.83
scaffold_35844|ref0030074	0.85
scaffold_24983|ref0011970	0.91
scaffold_1078|ref0045090	0.91
scaffold_8396|ref0012761	0.95
scaffold_9549|ref0015617	0.97

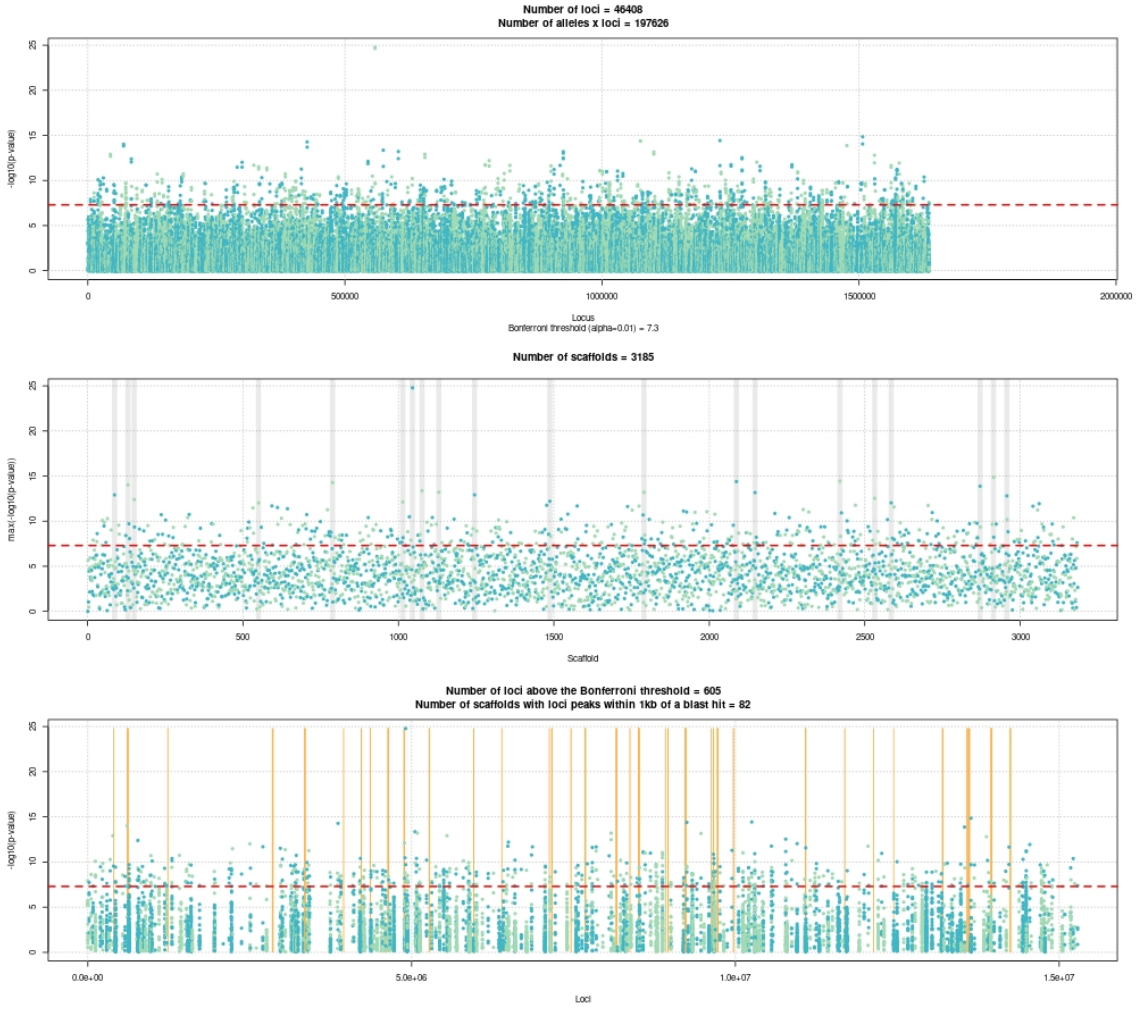


Figure S5.1. Manhattan plots for clethodim resistance using the merged *Lolium ridum* dataset and the linear mixed model with F_{ST} as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum $-\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

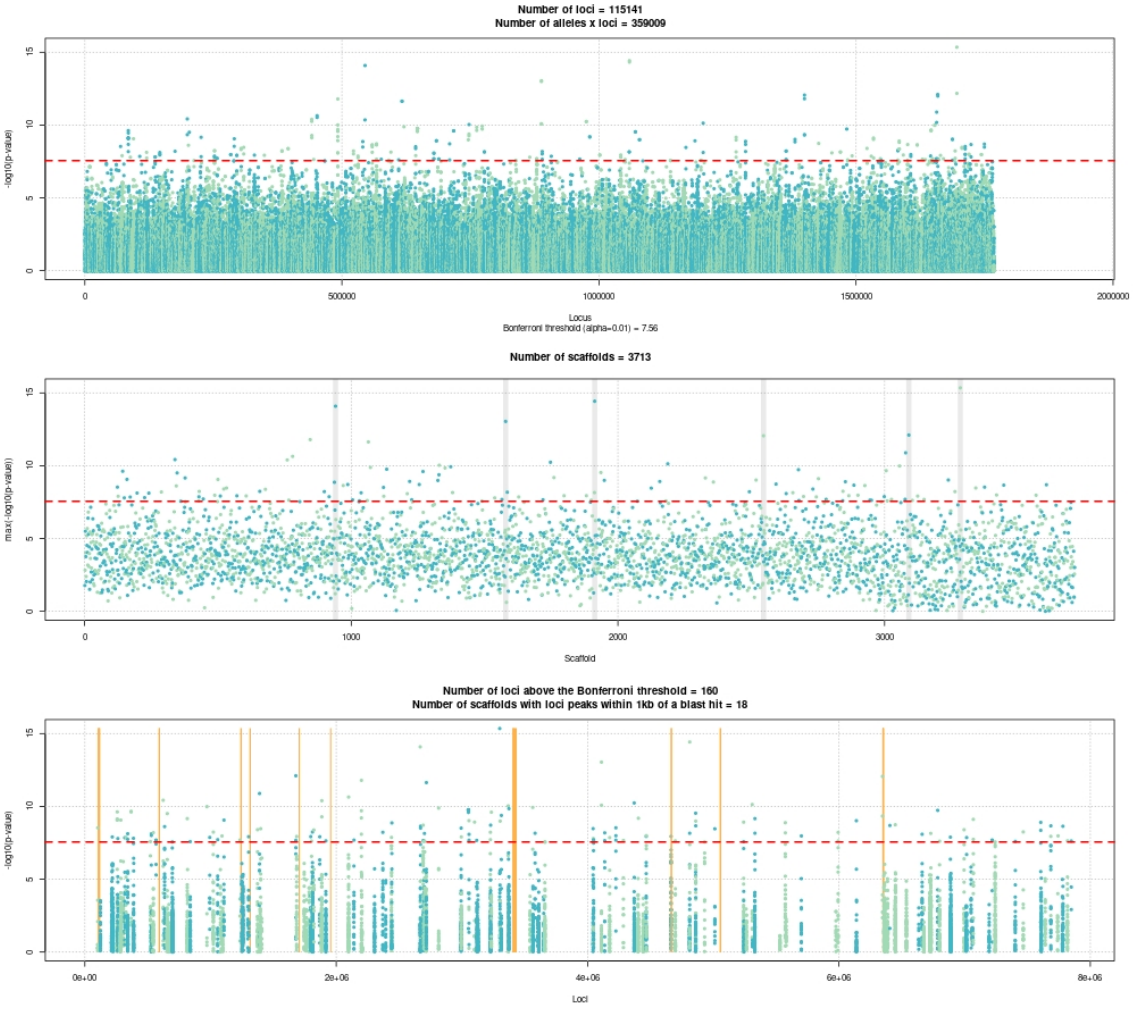


Figure S5.2. Manhattan plots for clethodim resistance using ACC13 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

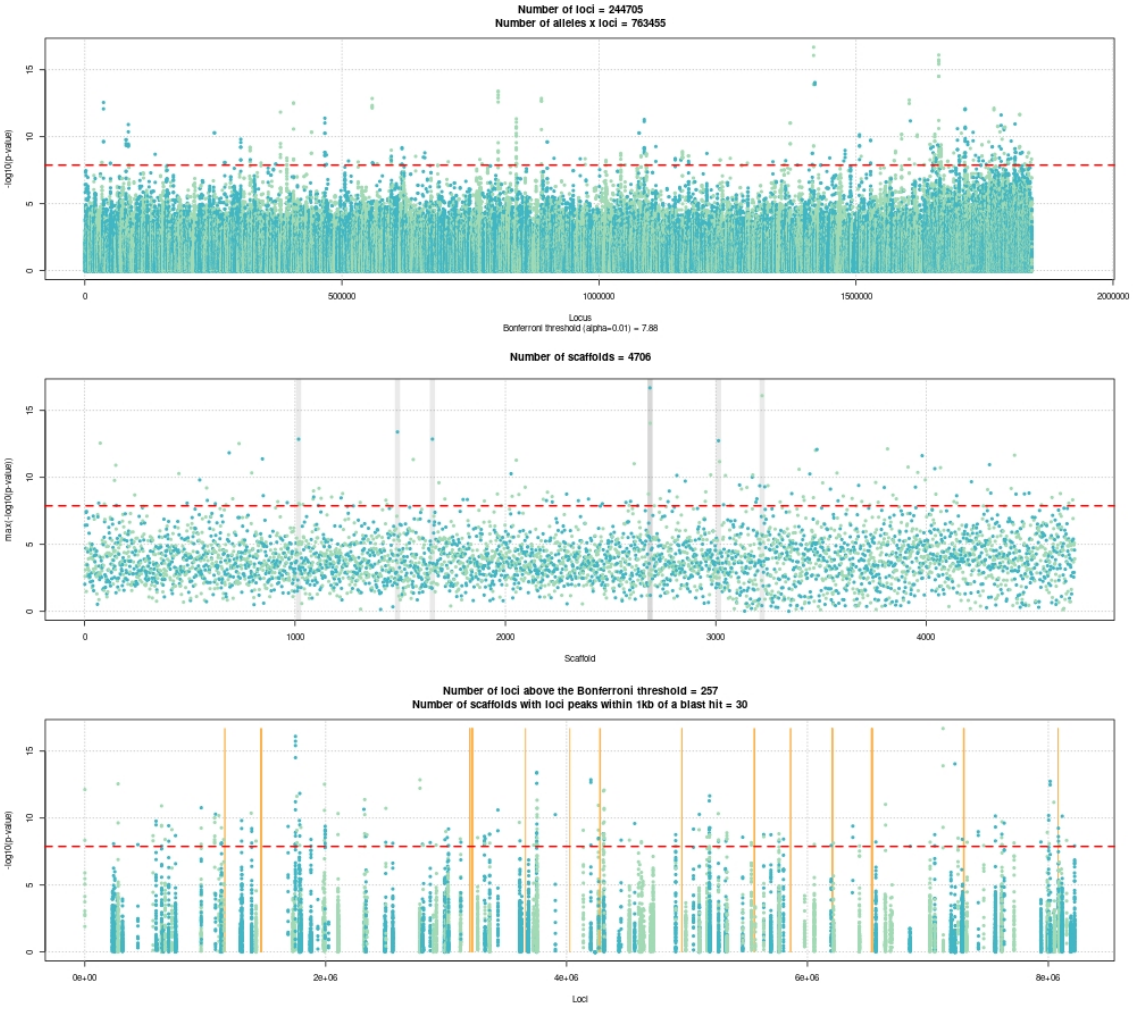


Figure S5.3. Manhattan plots for clethodim resistance using ACC49 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

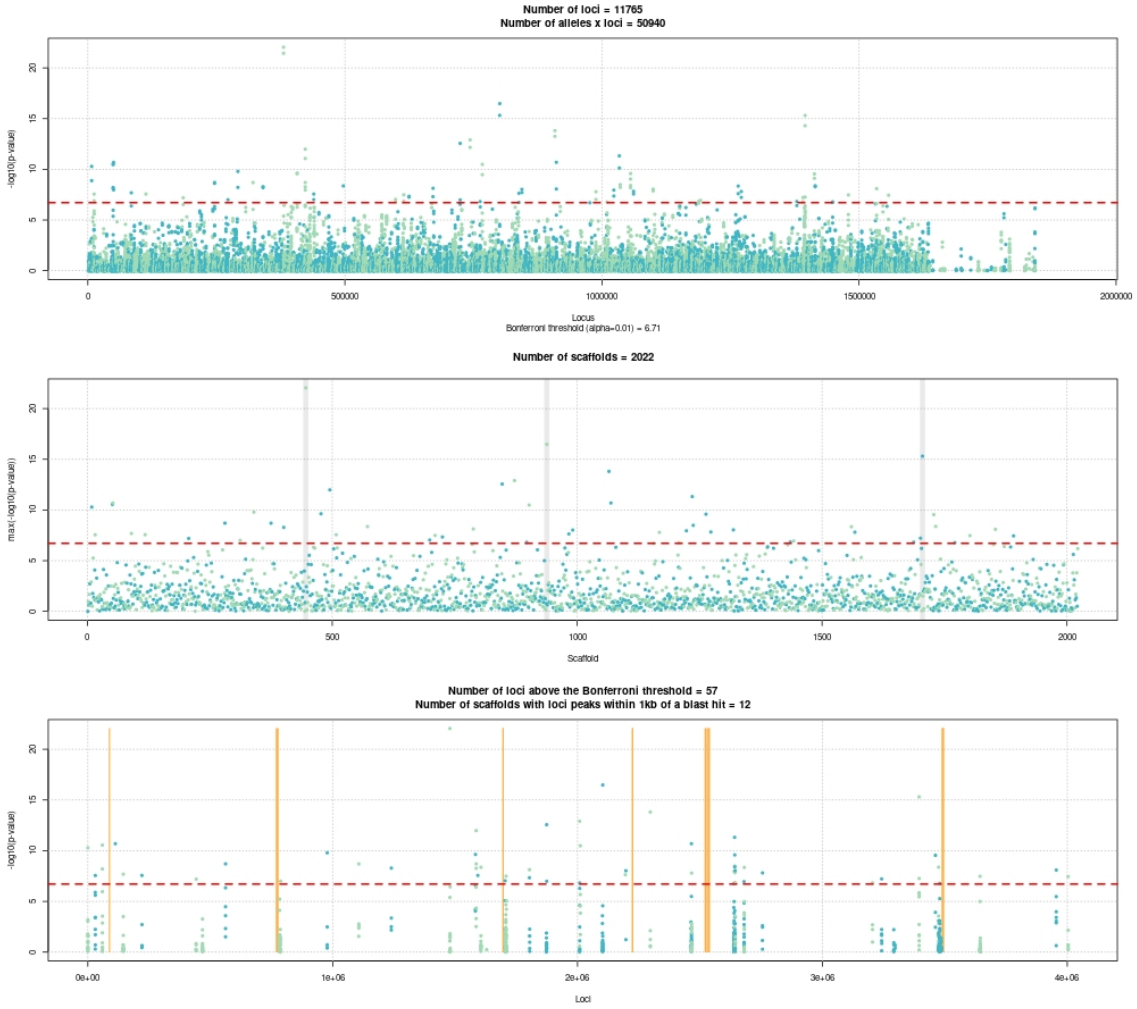


Figure S5.4. Manhattan plots for glyphosate resistance using the merged *Lolium ridum* dataset and the linear mixed model with F_{ST} as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum $-\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

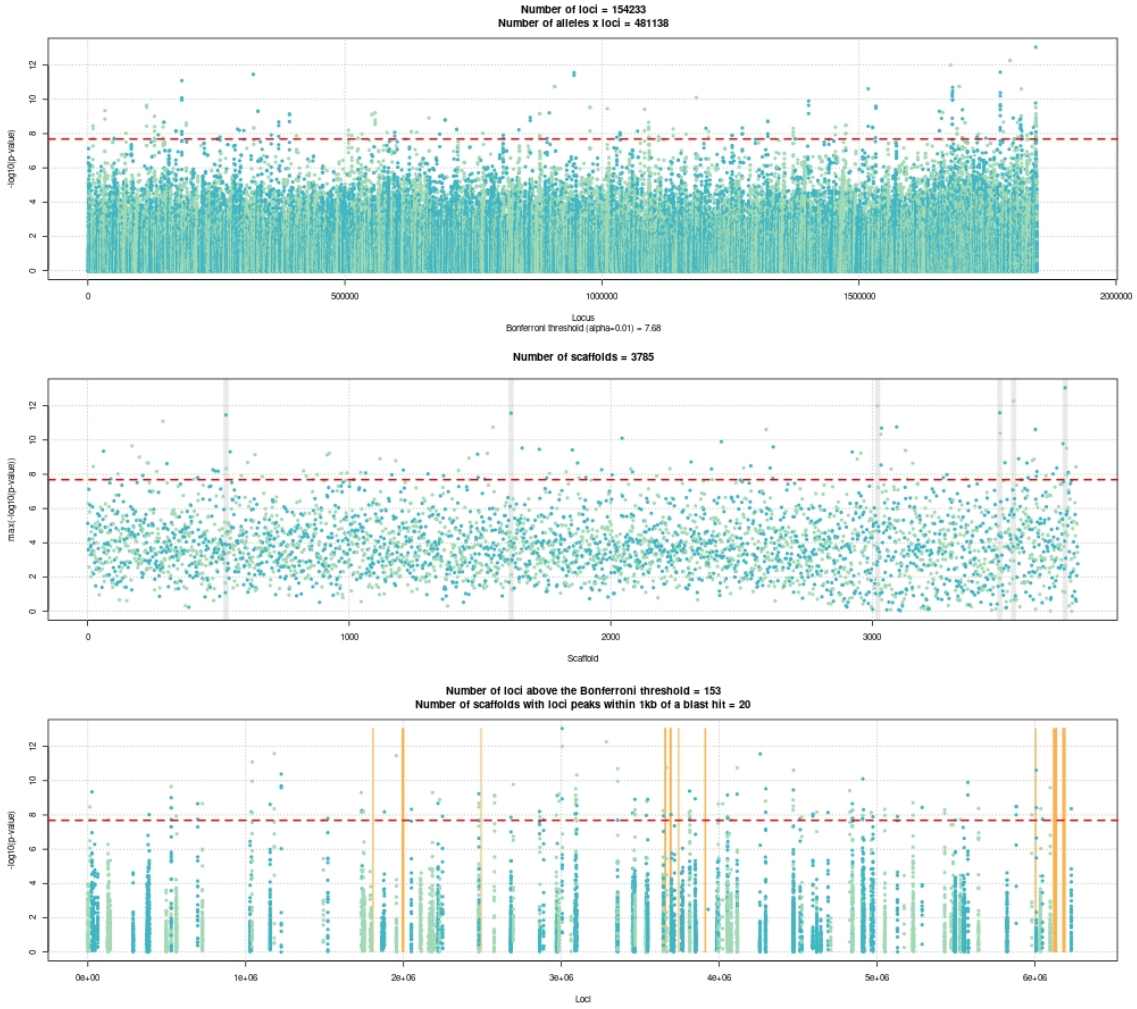


Figure S5.5. Manhattan plots for glyphosate resistance using ACC31 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

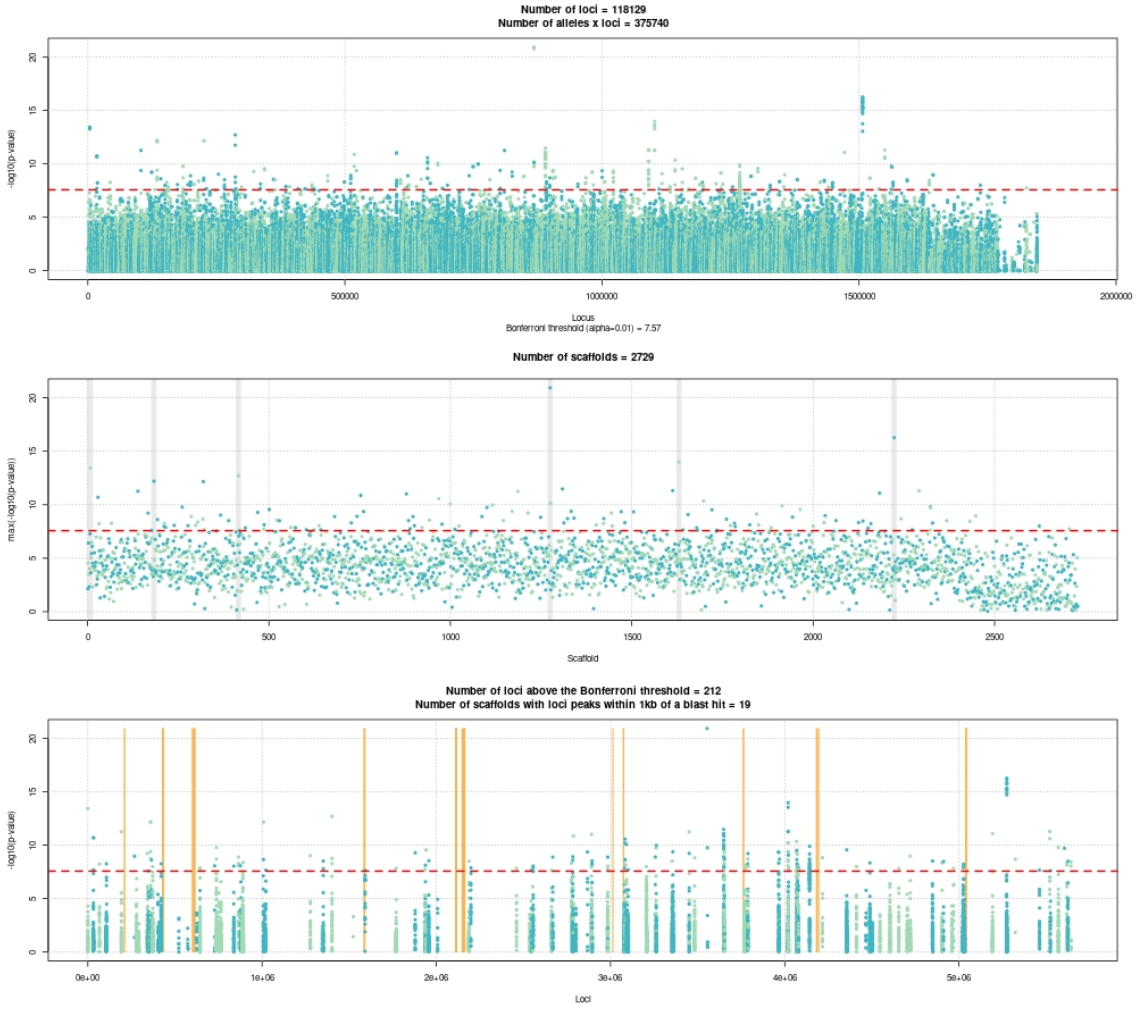


Figure S5.6. Manhattan plots for glyphosate resistance using ACC62 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

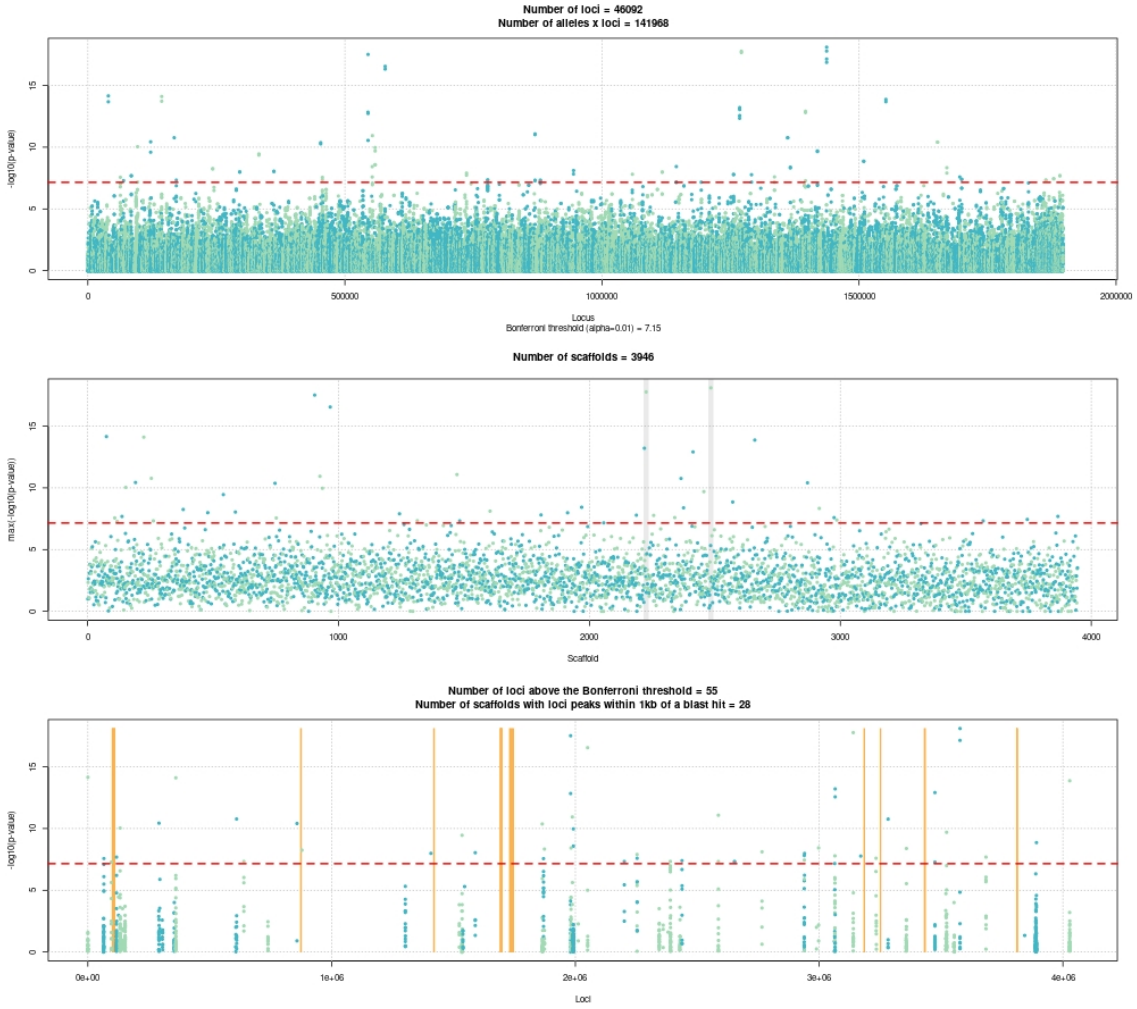


Figure S5.7. Manhattan plots for glyphosate resistance using Inverleigh *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

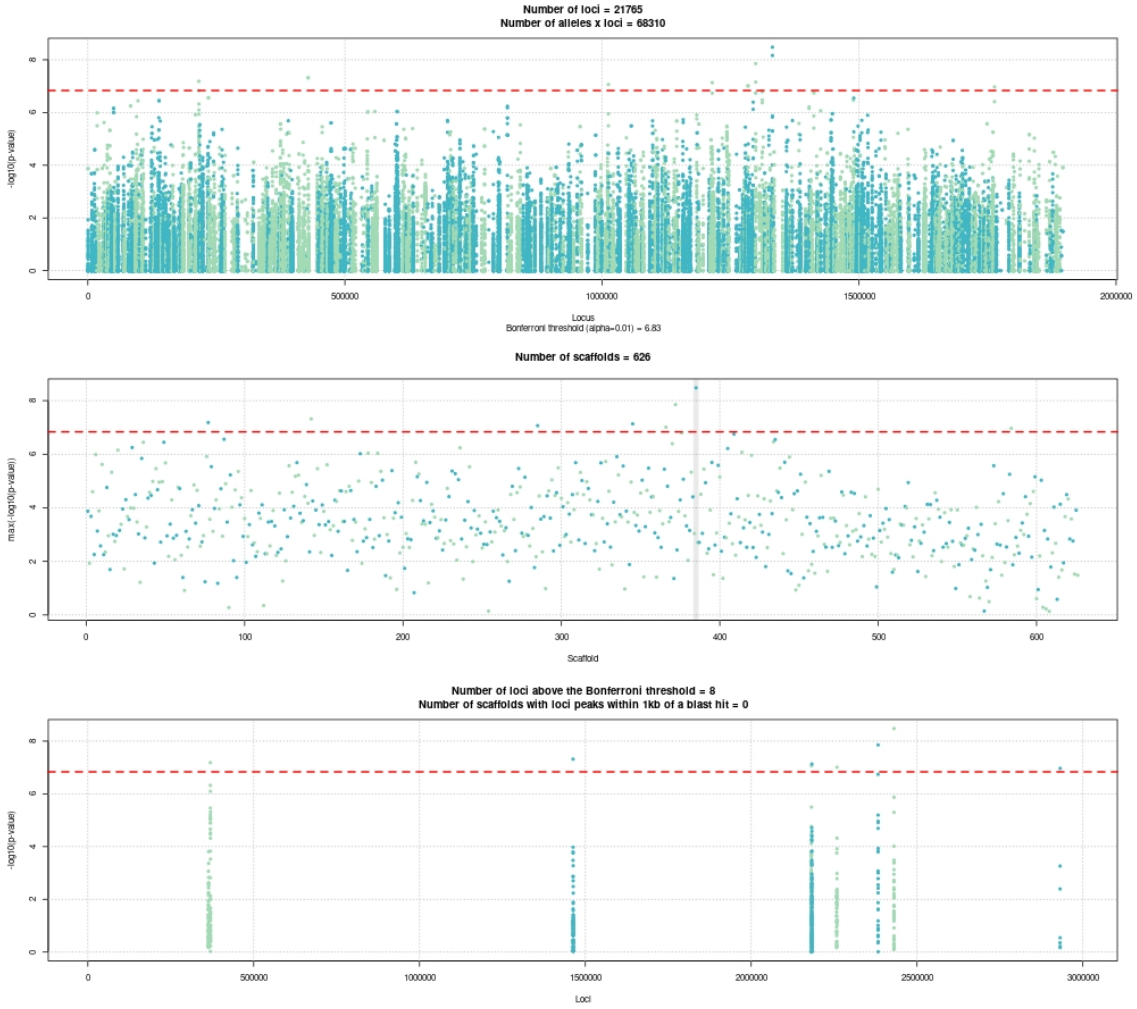


Figure S5.8. Manhattan plots for glyphosate resistance using Urana *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

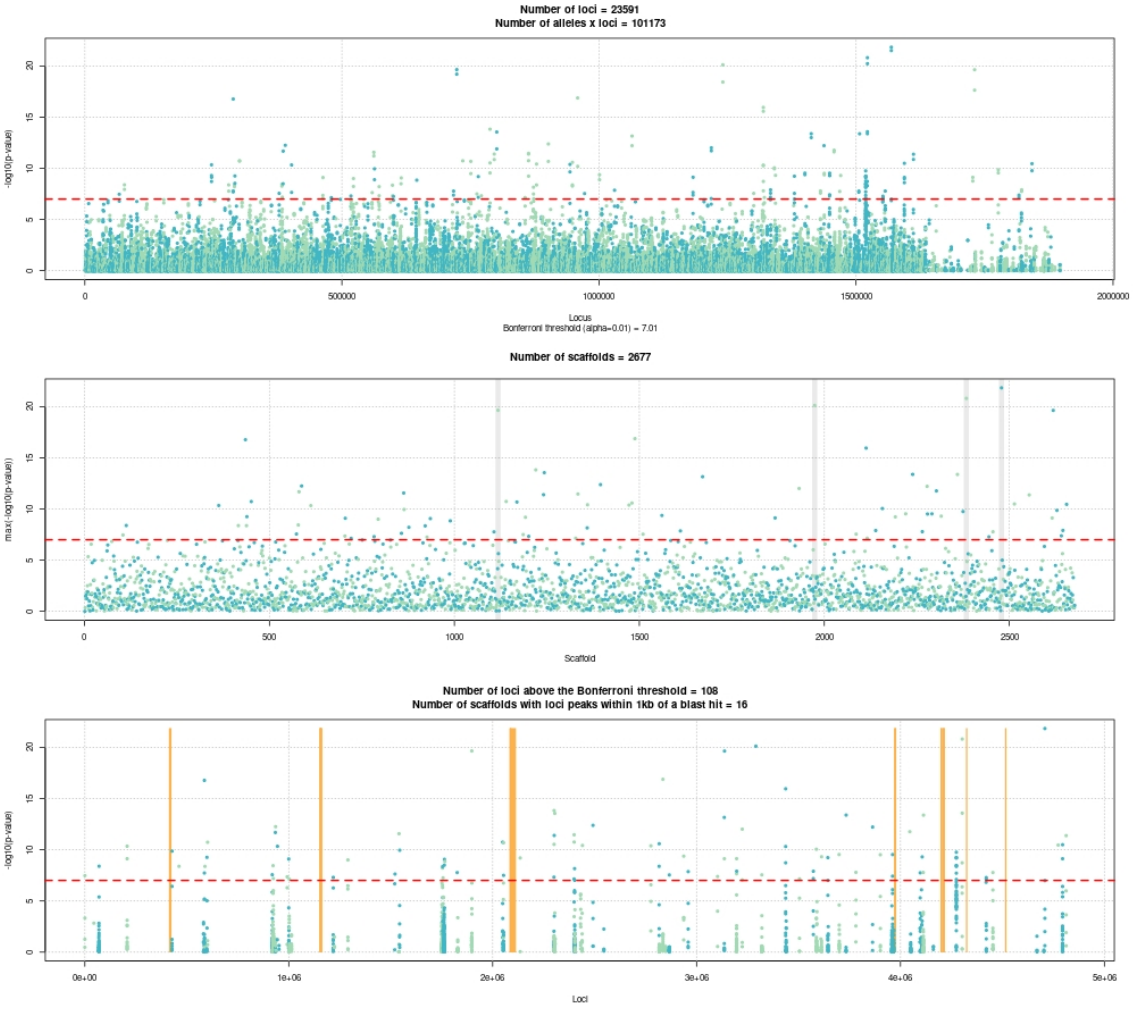


Figure S5.9. Manhattan plots for sulfometuron resistance using the merged *Lolium ridum* dataset and the linear mixed model with F_{ST} as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum $-\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

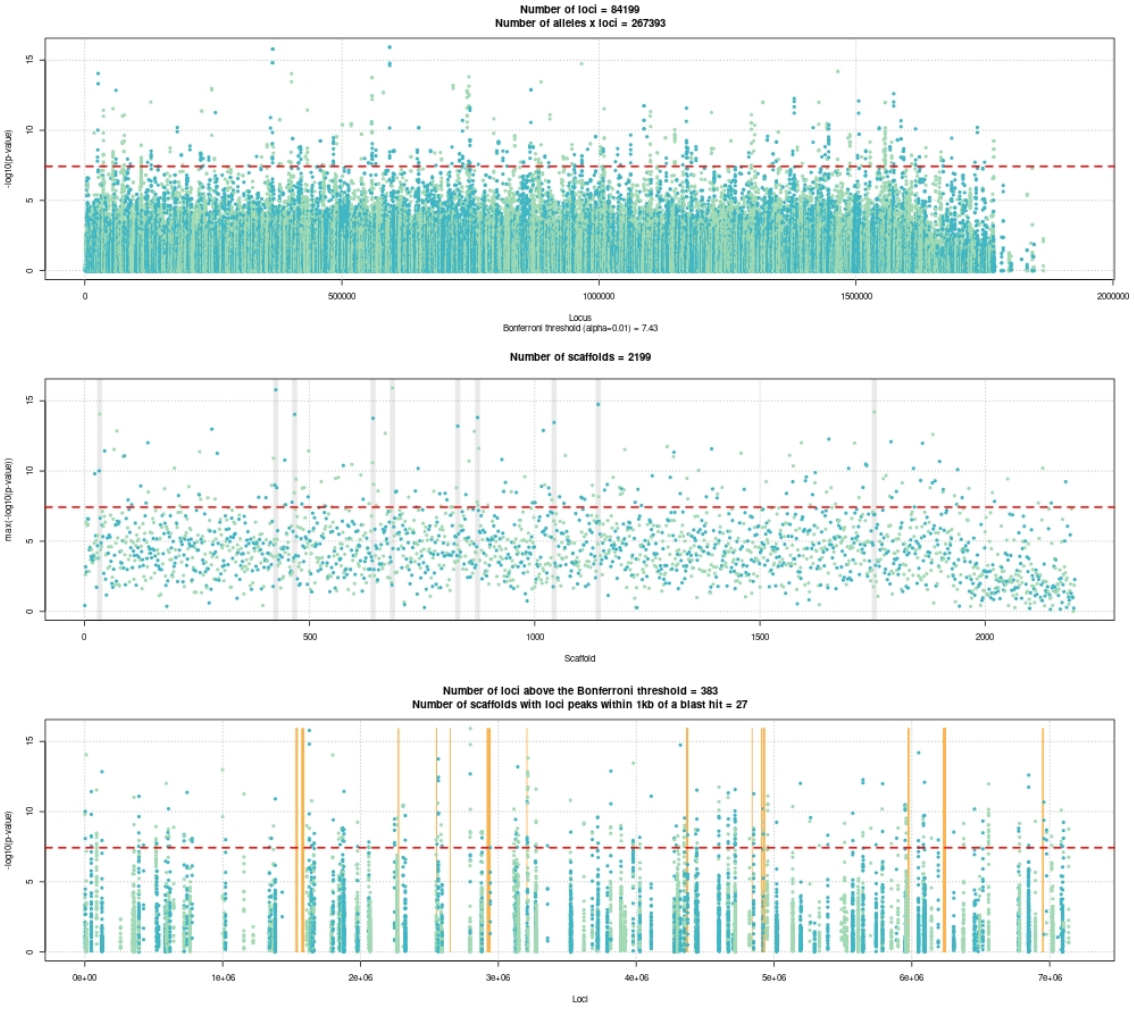


Figure S5.10. Manhattan plots for sulfometuron resistance using ACC11 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

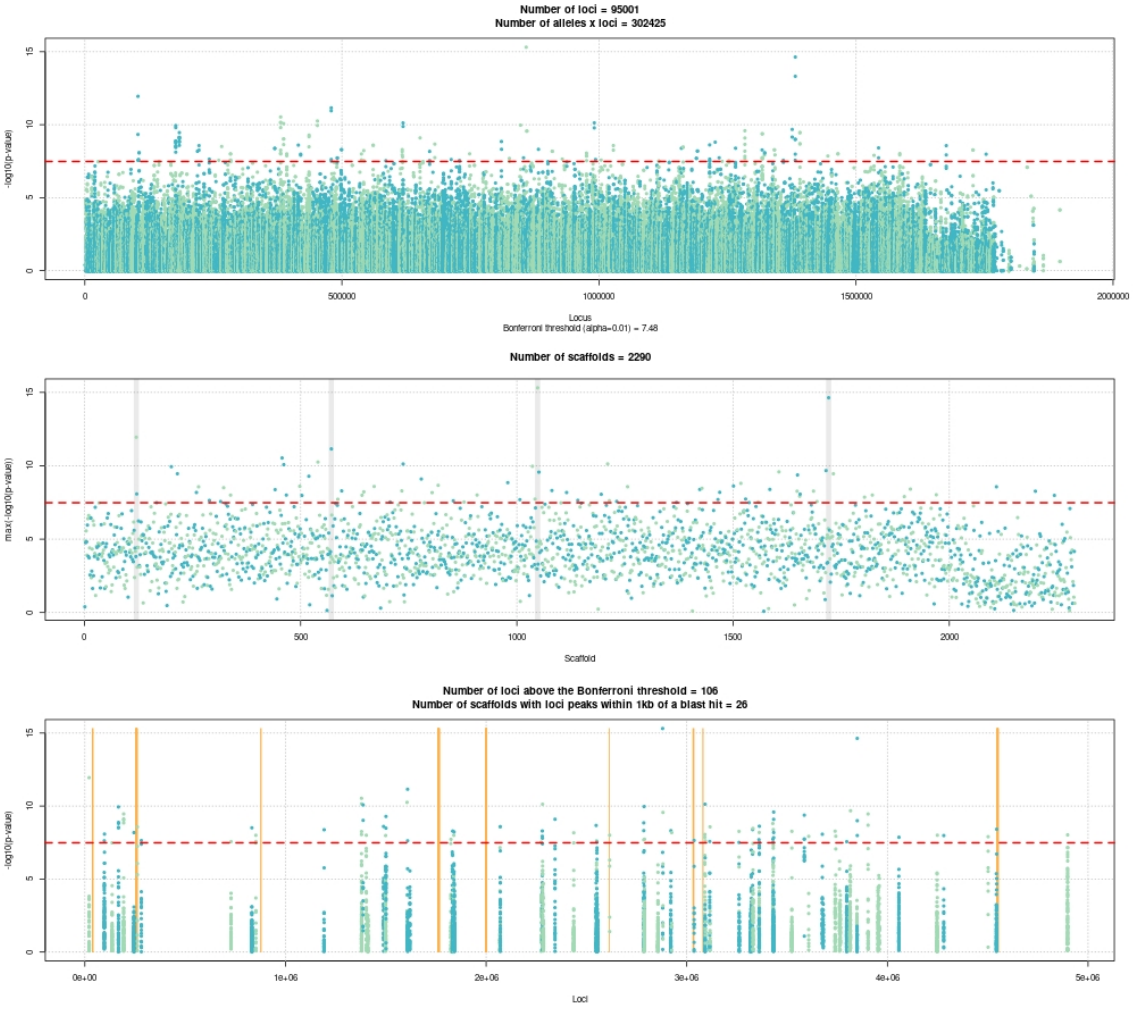


Figure S5.11. Manhattan plots for sulfometuron resistance using ACC59 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

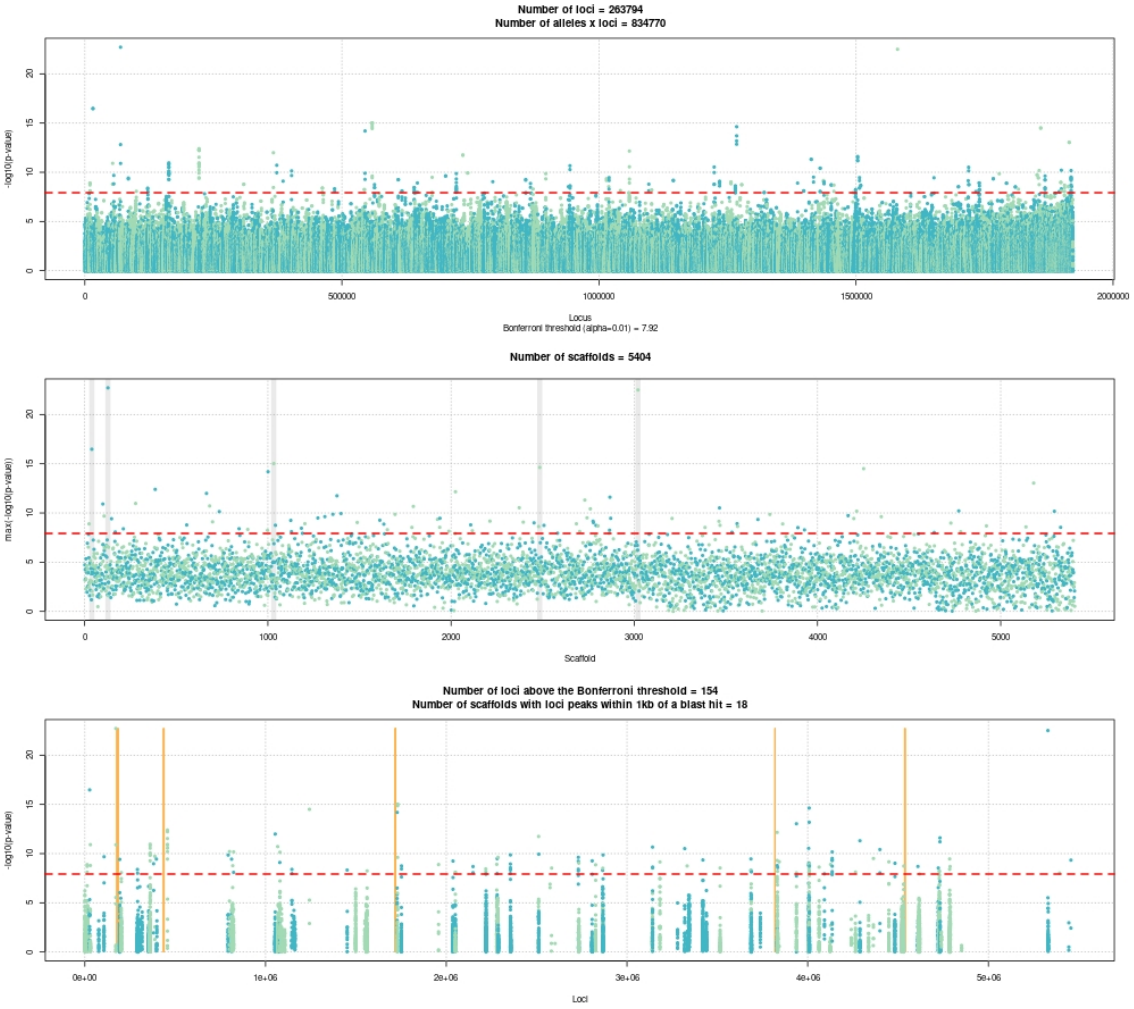


Figure S5.12. Manhattan plots for sulfometuron resistance using Inverleigh *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

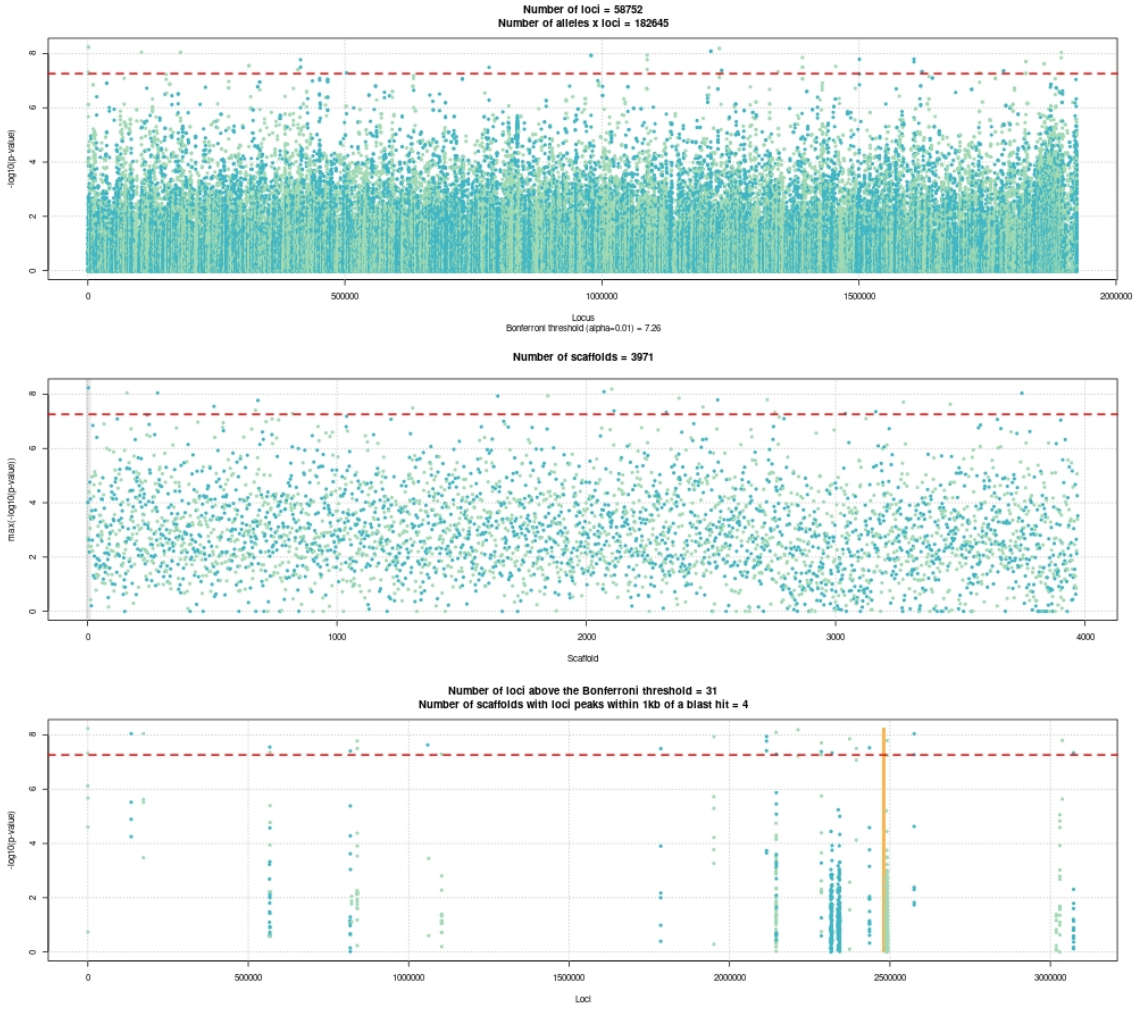


Figure S5.13. Manhattan plots for sulfometuron resistance using Urana *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

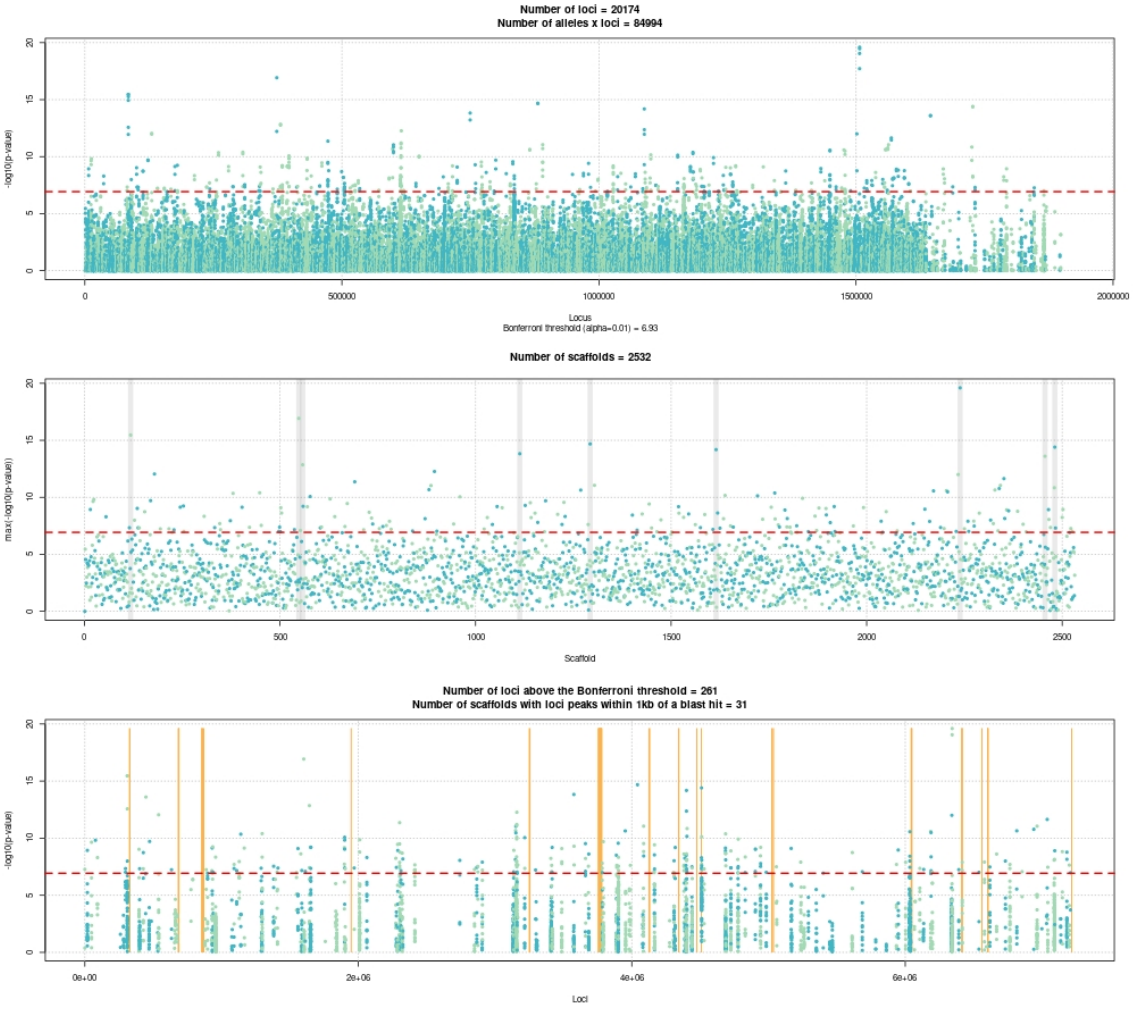


Figure S5.14. Manhattan plots for terbuthylazine resistance using the merged *Lolium ridum* dataset and the linear mixed model with F_{ST} as random covariate (Top: plot across the entire genome; Middle: plot per scaffold using maximum $-\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

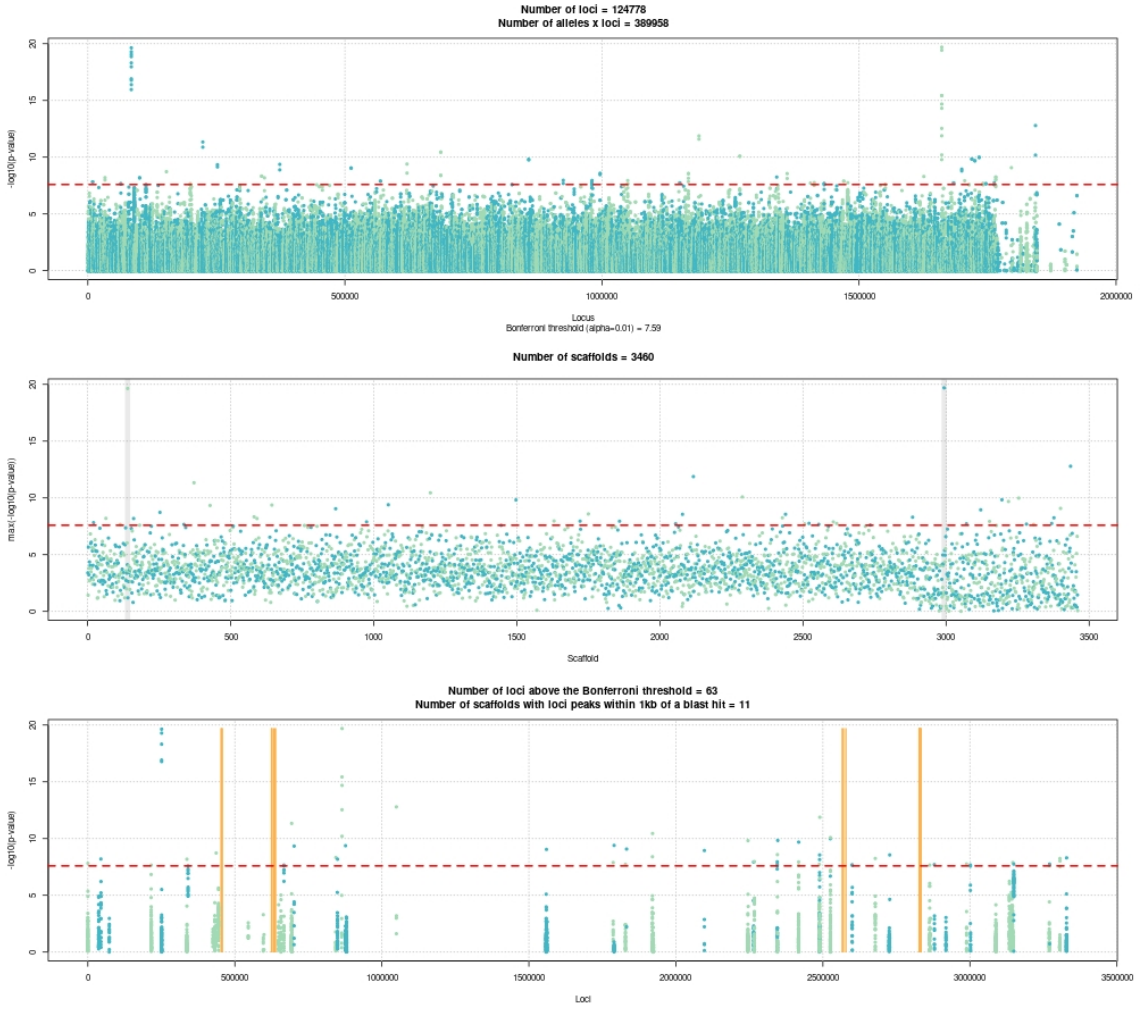


Figure S5.15. Manhattan plots for terbuthylazine resistance using ACC09 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

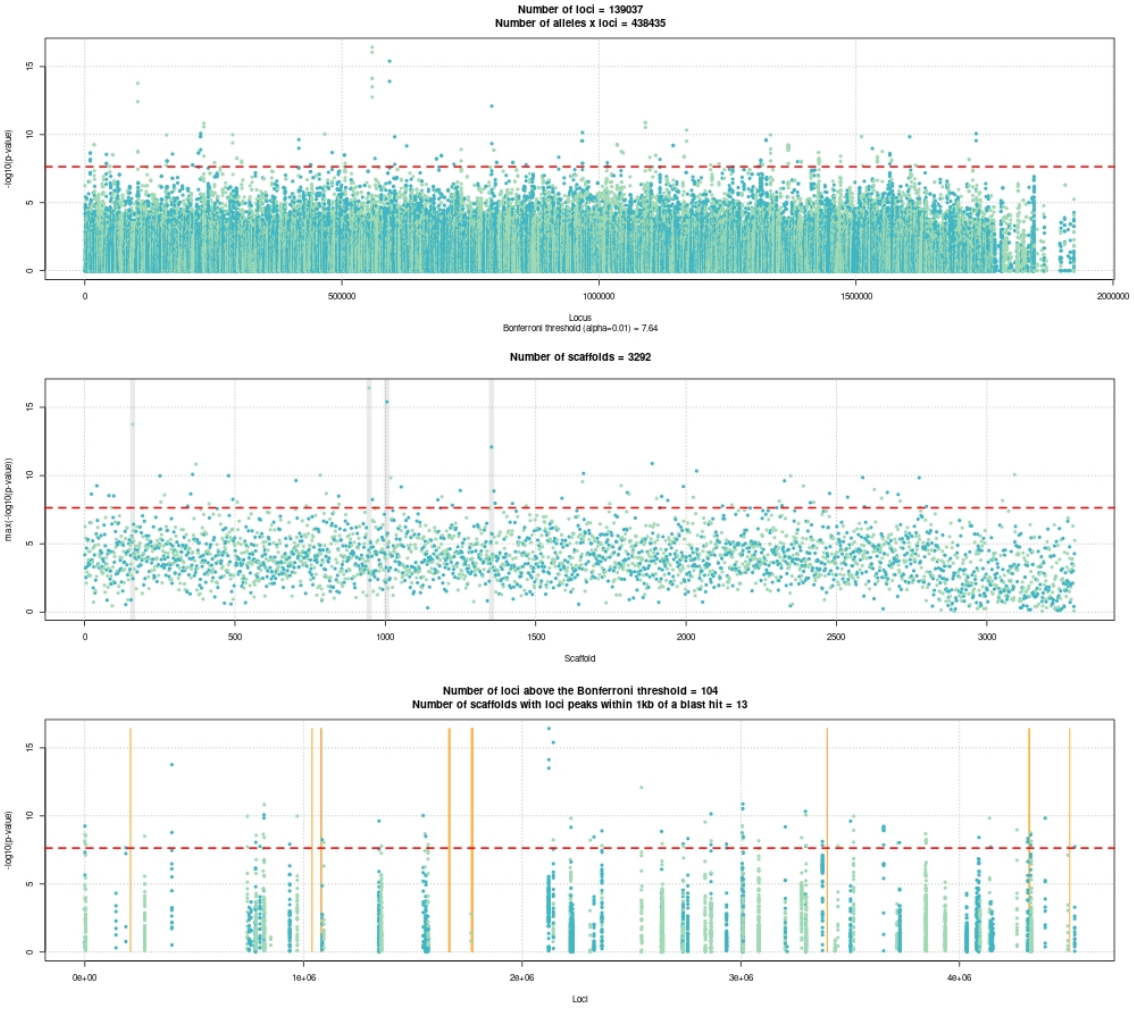


Figure S5.16. Manhattan plots for terbuthylazine resistance using ACC54 *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

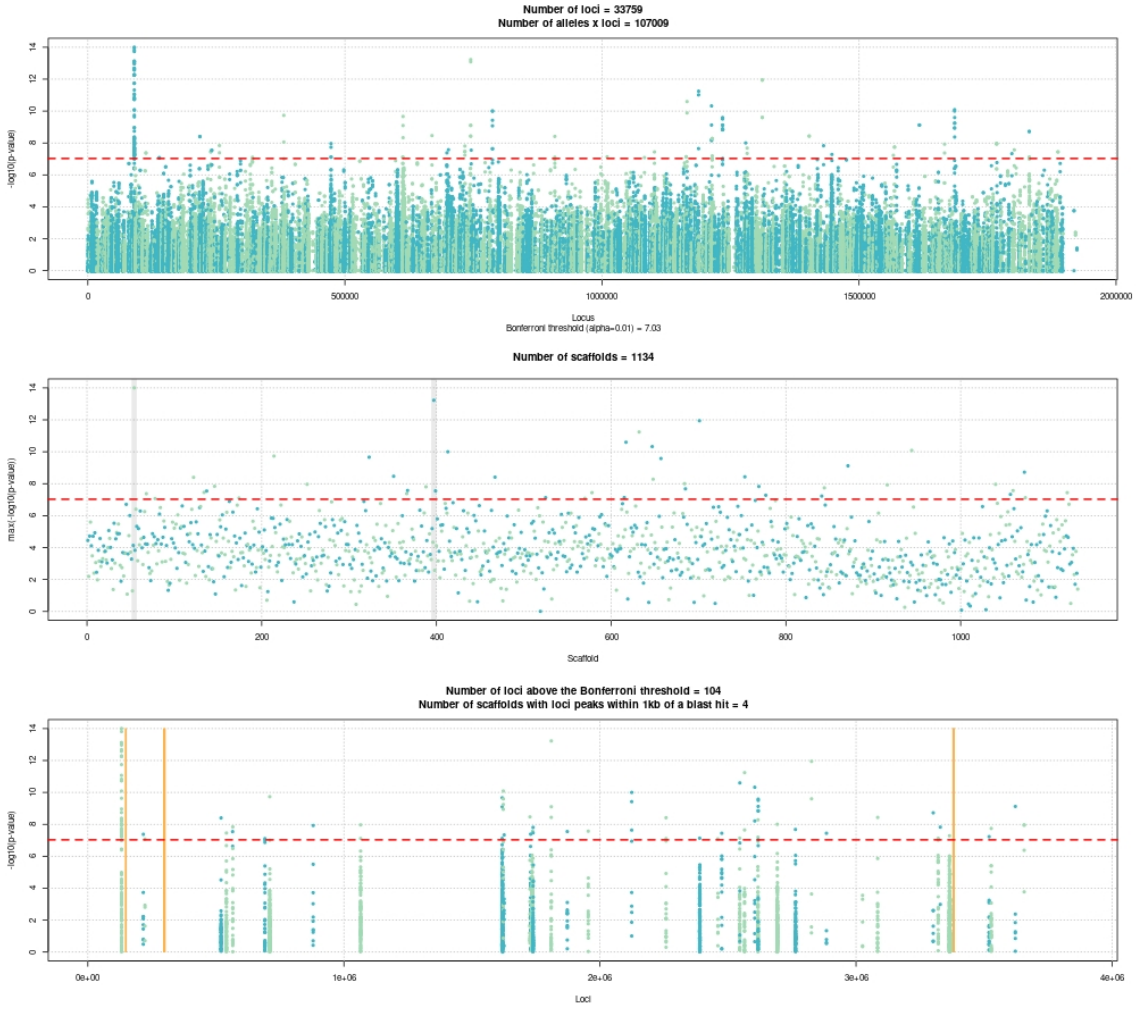


Figure S5.17. Manhattan plots for terbuthylazine resistance using Inverleigh *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

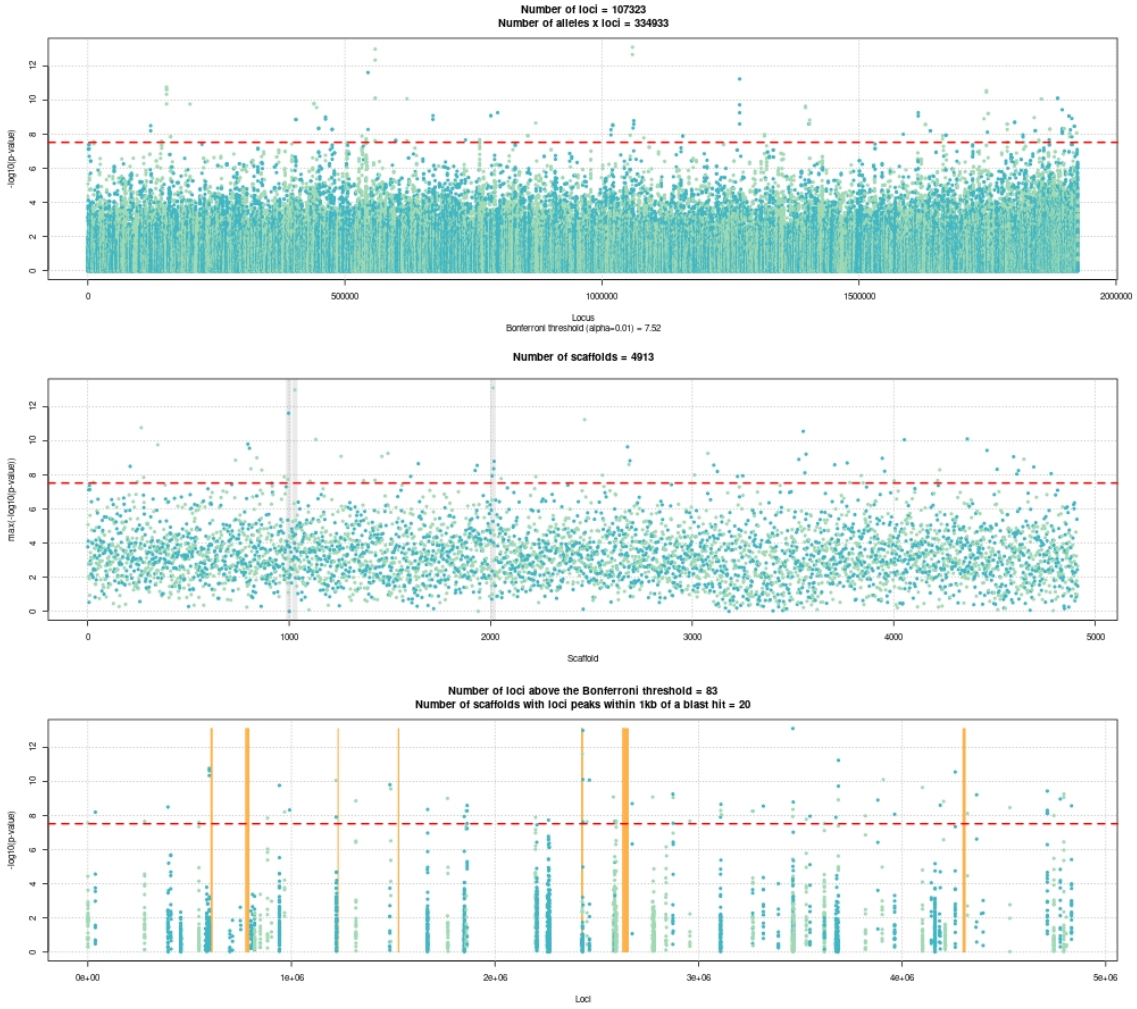


Figure S5.18. Manhattan plots for terbuthylazine resistance using Urana *Lolium rigidum* population and univariate GWAlpha model (Top: plot across the entire genome; Middle: plot per scaffold using maximum -$\log_{10}$(p-value) per scaffold; Bottom: plot of scaffolds with at least one peak; Red horizontal dash line: Bonferroni threshold at 1% alpha; Gray thick vertical lines: top 5% of the peaks; Yellow vertical lines: location of predicted gene sequences with blast hits).

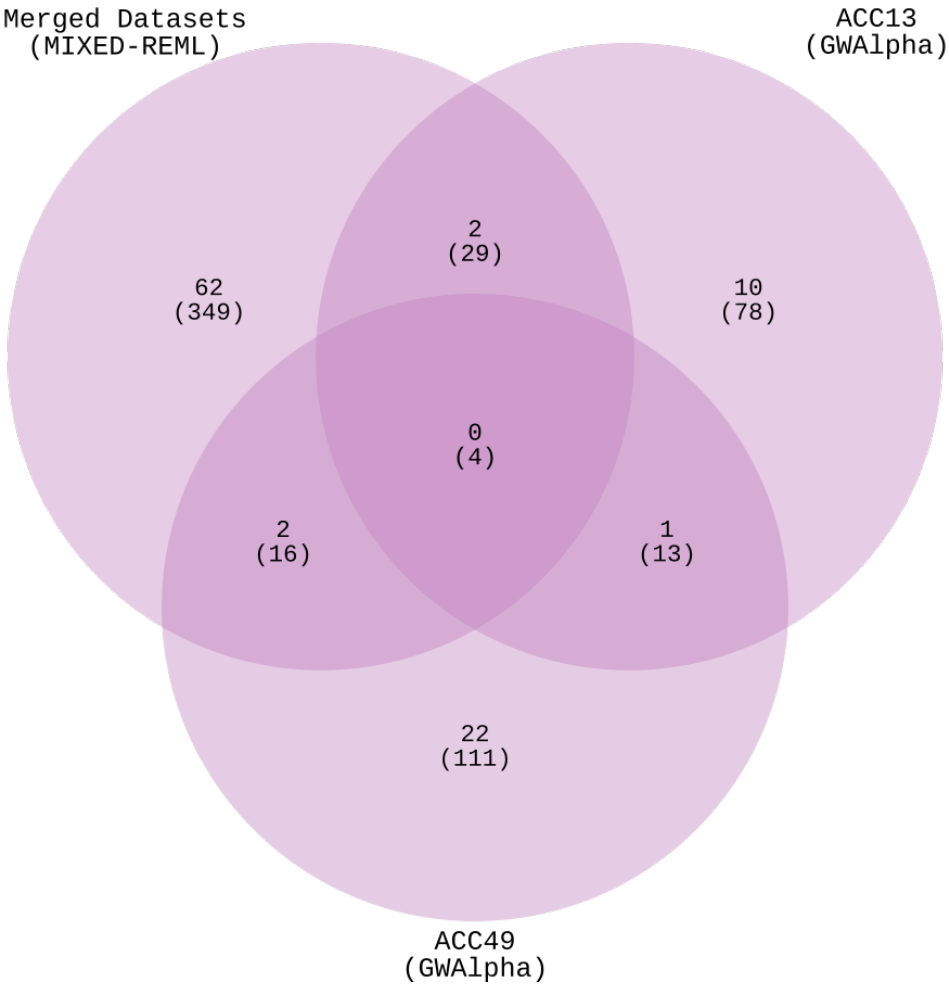


Figure S5.19. Venn diagram of candidate clethodim resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.

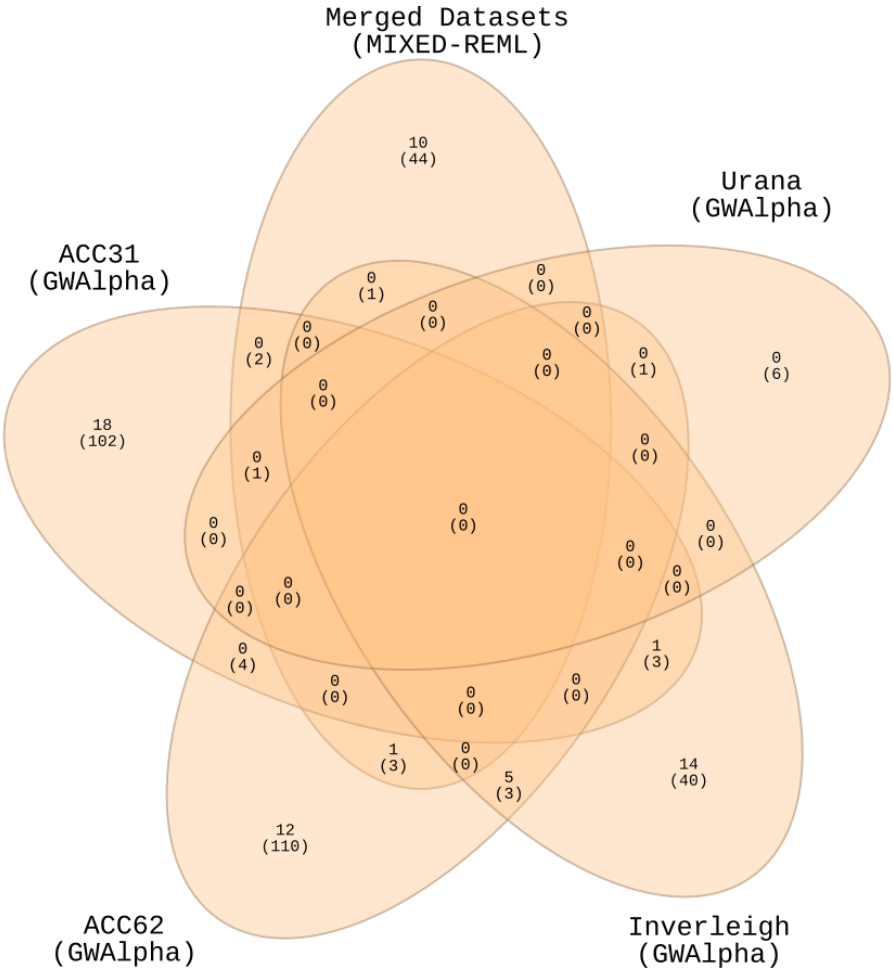


Figure S5.20. Venn diagram of candidate glyphosate resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.

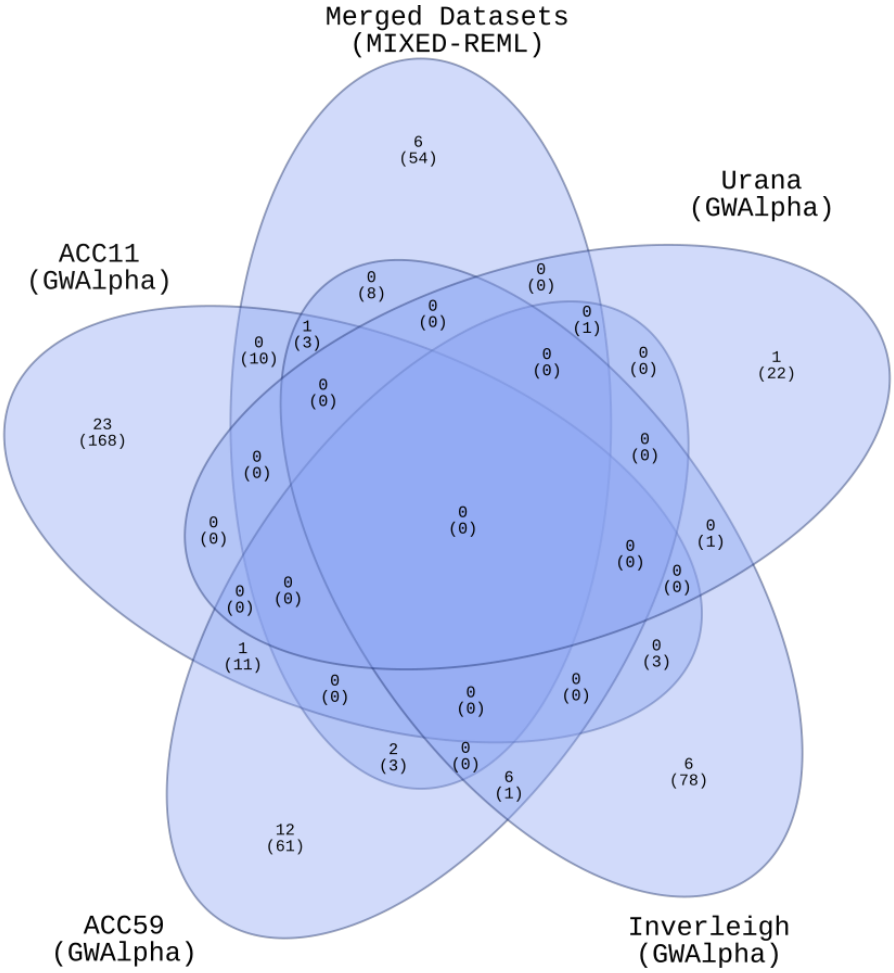


Figure S5.21. Venn diagram of candidate sulfometuron resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.

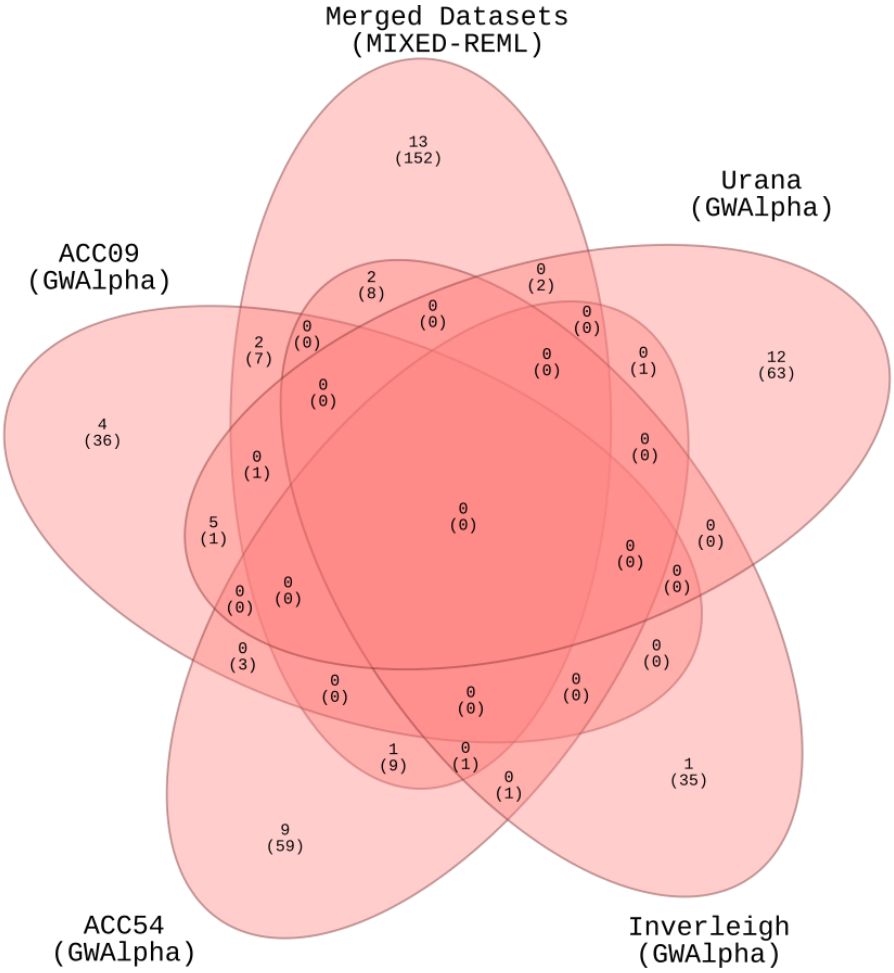


Figure S5.22. Venn diagram of candidate terbuthylazine resistance genes and scaffolds containing the candidate loci across Pool-GWAS experiments. The number of genes are outside parenthesis and the number of scaffolds are inside parenthesis.

Table S5.2. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for clethodim resistance using scaffold_3626|ref0003919 and the following populations: Susceptibles: ACC01,ACC05,ACC06 and Resistants: [ACC27,ACC34] & ACC36.

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	0.60	35642	0.6	NA	NA	NA
MODEL2	0.63	35642	0.6	NA	NA	$1x10^{-5}$
MODEL3	1.60	35642	0.6	0.01	0	NA
MODEL4	0.56	35642	0.6	NA	NA	$1x10^{-5}$
MODEL5	1.15	35642	0.6	$7.55x10^{-7}$	25	NA

Table S5.3. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for clethodim resistance using scaffold_3626|ref0003919 and the following populations: Susceptibles: ACC06,ACC38,ACC50 and Resistants: [ACC08,ACC09],ACC11.

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	0.02	35642	0.4	NA	NA	NA
MODEL2	0.11	35642	0.6	NA	NA	0.001
MODEL3	0.49	35642	0.6	0.1	0	NA
MODEL4	0.10	35642	0.6	NA	NA	$1x10^{-5}$
MODEL5	0.35	35642	0.6	0.01	0	NA

Table S5.4. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for glyphosate resistance using scaffold_198|ref0030880 and the following populations: Susceptibles: ACC01,ACC02,ACC15 and Resistants: ACC18,[ACC47,ACC50].

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	$1.75x10^{-11}$	7918	0.6	NA	NA	NA
MODEL2	$5.37x10^{-7}$	7918	0.6	NA	NA	1
MODEL3	0.01	7918	0.6	0.1	0	NA
MODEL4	$1.75x10^{-11}$	7918	0.6	NA	NA	$1x10^{-5}$
MODEL5	$9.25x10^{-4}$	7918	0.3	0.1	0	NA

Table S5.5. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for glyphosate resistance using scaffold_198|ref0030880 and the following populations: Susceptibles: ACC38,ACC39,ACC55 and Resistants: [ACC04,ACC09],ACC24.

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	0	7918	0.0001	NA	NA	NA
MODEL2	0	7918	0.0001	NA	NA	$1x10^{-5}$
MODEL3	$7.42x10^{-15}$	7918	0.2	0.001	25	NA
MODEL4	0	7918	0.0001	NA	NA	$1x10^{-5}$
MODEL5	$7.11x10^{-16}$	7918	0.2	0.001	50	NA

Table S5.6. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for sulfometuron resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC01,ACC19,ACC21 and Resistants: ACC14,[ACC51,ACC53].

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	8.45	22922	0.6	NA	NA	NA
MODEL2	11.98	22922	0.6	NA	NA	1
MODEL3	59.62	22922	0.5	0.1	0	NA
MODEL4	8.54	22922	0.6	NA	NA	$1x10^{-5}$
MODEL5	37.00	22922	0.6	0.1	0	NA

Table S5.7. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for sulfometuron resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC03,ACC20,ACC57 and Resistants: [ACC11,ACC12],ACC53.

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	0.40	22366	0.60	NA	NA	NA
MODEL2	0.51	22366	0.60	NA	NA	$1x10^{-5}$
MODEL3	1.68	22366	0.40	0.1	0	NA
MODEL4	0.37	22366	0.40	NA	NA	$1x10^{-5}$
MODEL5	1.02	22366	0.25	0.1	0	NA

Table S5.8. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for terbuthylazine resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC07,ACC21,ACC26 and Resistants: [ACC04,ACC06],ACC55.

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	15.94	22442	0.6	NA	NA	NA
MODEL2	20.08	22442	0.6	NA	NA	1
MODEL3	80.27	22442	0.4	0.1	0	NA
MODEL4	18.00	22442	0.6	NA	NA	1
MODEL5	24.69	22442	0.6	$9.58x10^{-7}$	0	NA

Table S5.9. Maximum likelihood estimates of the site of selection, selection coefficient, initial frequency of favourable allele, age of this standing favourable allele, and migration rate for terbuthylazine resistance using scaffold_6912|ref0027973 and the following populations: Susceptibles: ACC07,ACC21,ACC26 and Resistants: ACC09,[ACC60,ACC62].

Model	Mean Relative Composite - logLikelihood	Site (bp)	Selection Coefficient	Initial frequency of resistance allele	Age of resistance allele	Migration rate
MODEL1	23.20	22442	0.6	NA	NA	NA
MODEL2	28.47	22442	0.6	NA	NA	1
MODEL3	76.57	22442	0.6	0.1	0	NA
MODEL4	23.20	22442	0.6	NA	NA	$1x10^{-5}$
MODEL5	67.59	22442	0.6	0.1	0	NA