Characterisation of the complete mitochondrial genome and 13 microsatellite loci through next-generation sequencing for the New Caledonian spider-ant *Leptomyrmex pallens*

Maïa Berman[1,2,3,*], Chris M. Austin[2,4], Adam D. Miller[5]

[1] CSIRO Ecosystem Sciences, PMB 44 Winnellie, Northern Territory 0822, Australia

[2]Research Institute for the Environment and Livelihoods, Charles Darwin University, Northern Territory, Australia

[3]Université de Montpellier II, UMR AMAP, Montpellier, France

[4] Monash University Sunway Campus, Jalan Lagoon Selatan, Bandar Sunway, 46150 Petaling Jaya, Selangor, Malaysia

[5] Department of Zoology, The University of Melbourne, Parkville, Victoria, 3010 Australia

*Corresponding author: Maïa Berman, CSIRO Ecosystem Sciences, PMB 44 Winnellie, Northern Territory 0822, Australia, Phone: +61889448406, email: abeille2@gmail.com

Abstract

The complete mitochondrial genome and a set of polymorphic microsatellite markers were identified by 454 pyrosequencing (1/16th of a plate) for the New Caledonian rainforest spider-ant *Leptomyrmex pallens*. *De novo* genome assembly recovered the entire mitochondrial genome with mean coverage of 8.9-fold (range 1 to 27). The mitogenome consists of 15,591 base pairs including 13 protein-coding genes, 2 ribosomal subunit genes, 22 transfer RNAs, and a non-coding AT-rich region. The genome arrangement is typical of insect taxa and very similar to the only other published ant mitogenome from the *Solenopsis* genus, with the main differences consisting of translocations and inversions of tRNAs. A total of 13 polymorphic loci were also characterized using 41 individuals from a single population in the Aoupinié region, corresponding to workers from 21 nests and 16 foraging workers. We observed moderate genetic variation across most loci (mean number of alleles per locus = 4.50; mean expected heterozygosity = 0.53) with evidence of only two loci deviating significantly from Hardy–Weinberg equilibrium due to null alleles. Marker independence was confirmed with tests for linkage disequilibrium. Most loci cross amplified for three additional *Leptomyrmex* species. The annotation of the mitogenome and characterization of microsatellite markers will provide useful tools for assessing the colony structure, population genetic patterns, and dispersal strategy of *L. pallens* in the context of rainforest fragmentation in New Caledonia. Furthermore, this paper confirms a recent line of evidence that comprehensive mitochondrial data can be obtained relatively easily from small next-generation sequencing analyses. Greater synthesis of next-generation sequencing data will play a significant role in expanding the taxonomic representation of mitochondrial genome sequences.

Keywords: 454-pyrosequencing; mitogenome characterization; microsatellites; Formicidae species cross amplification; population genetics

Introduction

Ants are of great ecological importance and are commonly used as key indicator species for assessing the ecological impacts of environmental disturbance [1]. However the population genetic responses of ant communities to deforestation and fragmentation in tropical rainforests [2] have been seldom investigated (but see Bickel et al. [3] and Tawato et al. [4]). This is partly due to the limited knowledge of the biology, ecology and social behaviour of most tropical rainforest ant species. Social structure in particular is a strong determinant of relatedness and dispersal behaviour in ants [5], but such information is often lacking. Understanding this structure is critically important for conducting reliable population assessments, particularly at the genetic level. Genetic markers such as microsatellites allow colony structure to be easily inferred [6]. Today these markers can be developed quickly and at low cost for most species using next-generation sequencing (NGS) [7].

Recent studies have demonstrated that NGS can be used effectively to recover additional molecular markers including whole mitochondrial genome (mitogenome) sequences [8,9]. Investigation of genetic variation at mitochondrial loci can be particularly useful for assessing patterns of sex-biased dispersal, species phylogeography and demographic histories [10]. However, while several complete ant nuclear genome sequences have been described [11-17], only one study has successfully characterised complete mitogenomes in ants and these are limited to the genus *Solenopsis* [18].

In this study we develop a suite of molecular markers for the New Caledonian spider-ant *Leptomyrmex pallens*, a widespread, conspicuous tropical rainforest species [19]. Rainforests of New Caledonia have been severely fragmented and converted to species-poor savannas over the last 3500 yrs, mostly due to anthropogenic fires [20]. In order to determine the impacts of habitat fragmentation on the population genetics of *L. pallens*, we undertook a modest 454 pyrosequencing run, from which we designed a set of 13 polymorphic

microsatellite markers and recovered the full mitochondrial DNA sequence. Combined, these markers will provide a valuable resource for investigating population genetics and demographic history of *L. pallens* in response to fire induced habitat fragmentation in New Caledonia.

Material and methods

Next-generation sequencing

Approximately 10 µg of genomic DNA was extracted from a single whole, crushed *L. pallens* specimen, using a QIAGEN DNA Easy kit. DNA was subsequently processed by the Australian Genome Research Facility (Melbourne), where it was nebulized, ligated with 454 sequencing primers and tagged with a unique oligo sequence allowing sequences to be separated from pooled species DNA sequences, using post-run bioinformatic tools. The DNA sample was analyzed using high throughput DNA sequencing on $1/16^{th}$ of a 70 x 75 mm Pico Titer Plate using the Roche GS FLX (454) system [21].

Microsatellite isolation and characterisation

Unique sequence contigs possessing microsatellite motifs were identified using the open-source QDD version 2 [22]. Primer3 [23] was used to design optimal primer sets for each unique contig, where possible. A selection of contigs including di-, tri-, and tetra-nucleotide repeats was used for subsequent analysis. Loci were screened for polymorphism using template DNA from eight individuals, representing two sample locations from New Caledonia: the Aoupinié region in the Northern Province, and the Montagne des Sources region in the Southern Province. Loci were pooled into groups of four, labeled with unique fluorophores (FAM, NED, VIC, PET) and co-amplified by multiplex PCR using a QIAGEN multiplex kit and an Eppendorf Mastercycler *S* gradient PCR machine following the protocol

described by Blacket et al. [24]. Genotyping was subsequently performed using an Applied Biosystems 3730 capillary analyzer (http://www.agrf.org.au) and product lengths were scored manually and assessed for polymorphisms using GENEMAPPER version 4.0 (Applied Biosystems).

Polymorphic loci were selected, pooled into two groups for multiplexing, based on observed locus specific allele size ranges, and further characterized using DNA from 41 individuals sampled in 2010-12 from 21 nests and 16 locations in the Aoupinié region (c.a. 4900 ha). Unpublished analyses justify the spatial sampling design, as it represents a set of unrelated individuals within a panmictic population (M. Berman, unpub. data). Microsatellite profiles were again examined using GENEMAPPER version 4 and alleles scored manually. The EXCEL MICROSATELLITE TOOLKIT [25] was then used to estimate expected (H_E) and observed (H_O) heterozygosities and number of alleles (N_A), while conformation to Hardy-Weinberg equilibrium (HWE), inbreeding coefficient (F_{IS}) and linkage disequilibrium estimates between all pairs of loci were examined using the open-source GENEPOP on the web version 4 [26]. Where necessary, significance values were adjusted for multiple comparisons using Bonferroni corrections [27]. All loci were assessed using MICRO-CHECKER to check for null alleles and scoring errors [28]. The frequency of null alleles per locus was obtained using the 'Brookfield 1' formula, as no evidence of null homozygotes across loci was found [29]. Finally, cross-species amplifications were conducted on each of the two other New Caledonian Leptomyrmex species (*L. nigriceps*, 2 individuals; *L. geniculatus*, 2 individuals) and an Australian representative (*L. nigriventris*, 1 individual).

Mitochondrial assembly and annotation

Sequence reads in SFF format were edited by trimming 454 adaptor sequences, and converted to fasta and quality file format using the open-source GALAXY software [30].

Genomic sequence contigs were assembled using *de novo* default 454 parameters in the open-source Unix based software MIRA version 3.4.0.1 [31]. Coverage statistics were obtained with the software GENEIOUS version 5.6.6 [32]. Gene positions, codon usage, transcriptional orientations, transfer RNA (tRNA) cove-scores [33] and secondary structures were obtained with the open-source DOGMA software [34], using a 30% identity cutoff for protein coding genes, due to the paucity of related mitochondrial genomes available in the database. Annotations and reading frames were confirmed by visual inspection in GENEIOUS using the *Solenopsis invicta* mitochondrial sequence as reference [18]. The ribosomal subunits (rRNA) gene boundaries were estimated with alignments implemented in GENEIOUS using the *Solenopsis* genomes, with a high degree of conservatism at the beginning and end of the respective genes across taxa. GENEIOUS was used to generate the finalized annotated mitochondrial genome map (Genebank accession number KC160533). Gene arrangement, composition and transcriptional polarity was compared with the ant *Solenopsis invicta* (Hymenoptera: Vespoidea: Formicidae: Myrmicinae; GenBank accession number NC014672 [18]), the wasp *Abispa ephippium* (Hymenoptera: Vespoidea: Vespidae: Eumeninae; NC011520 [35]), the bee *Apis mellifera ligustica* (Hymenoptera: Apoidae: Apidae; NC001566 [36]), and the fruitfly *Drosophila melanogaster* (Diptera: Drosophilidae: Drosophilinae; NC001709 [37]).

Results and discussion

Next-generation sequencing output

A total of 49,014 reads with an average length of 527 bp (after trimming 454 adapter sequences), and covering up to 25.6 Mb of the *L. pallens* genome was obtained by NGS. These data would represent ~ 8.8% of the genome, based on previous ant genome size

estimates (~290 Mb, *Atta cephalotes* [15]). However, these figures are probably overestimated due to expected read redundancy.

Microsatellite isolation and characterization

A total of 3207 unique sequence contigs possessing microsatellite motifs were identified by QDD analysis, of which 2741 contigs were found to possess optimal priming sites. A total of 40 contigs were screened for polymorphism, with 29 containing di-nucleotide repeats, 8 containing tri-nucleotide repeats, and 3 containing tetra-nucleotide repeat motifs. The screening analysis found 21 loci to be polymorphic, 10 were monomorphic and 9 failed to amplify. A final subset of 13 polymorphic loci were selected for further investigation based on marker profiles and allele sizes.

All loci were characterized by low to moderate genetic variation, with an average of 4.5 alleles per locus (range = 2 – 12 alleles) and heterozygosity estimates ranging between 0.33 and 0.77 (mean = 0.53; Table 1). Linkage disequilibrium analyses confirmed marker independence as none of the 78 pairwise tests were significant after Bonferroni correction. All loci were found to conform with Hardy-Weinberg expectations, apart from LP14 and LP26, which showed an excess of homozygotes, as indicated by high F_{IS} values (Table 1). MICRO-CHECKER detected the presence of null alleles at these loci only. Cross-species amplifications are reported in Table 1. Some loci were found to be polymorphic in each species, however more comprehensive analyses at the population level are needed to provide reliable estimates of marker viability.

Mitochondrial genome of L. pallens

Approximately 0.7% of the total NGS reads (343 reads) were of mitochondrial origin, and *de novo* assembly of mtDNA sequence contigs revealed complete genome coverage with a mean

coverage of 8.9-fold (range 1-27). The absence of internal stop codons, ambiguous base calls, and evidence of heteroplasmy, suggests that our sequences are authentic mitochondrial targets rather than nuclear mitochondrial-like sequences (numts) which are common in ants [38]. The nucleotide composition of the α-strand is 5,758 adenine (36.9%), 3,431 cytosine (22%), 1,317 guanine (8.4%), and 5,082 thymine (32.6%). The A-T bias is a common feature of arthropods [39,36,18] but is less pronounced in Formicidae so far (Table 3).

Genome composition

The mitochondrial genome of *L. pallens* is a circular molecule 15,591 bp in length, with a typical metazoan gene composition including 13 protein-coding genes, 2 ribosomal subunits (rRNA), and 22 tRNAs, all of which had a cove-score > 20 (Figure 1; Table 2). The gene arrangement overall, including respective transcriptional polarities of genes, is very similar to that described in the *Solenopsis* genus and typical of insect taxa generally (Figure 2).

Protein-coding genes

The majority-strand (α) encodes 9 genes, while the minority-strand (β) encodes 4 genes (Table 2; Figure 1). Nucleotide overlap was observed between ATP8 and 6 (1 nucleotide; Table 2), a common feature in insect mitochondrial genomes [35,36]. We did not observe the common overlap between the ND4L and ND4 genes as observed in insects and metazoans generally [18,35,37], instead these genes were separated by 16 intergenic nucleotides. Non-overlapping ND4 genes have also found previously in *A. mellifera* [36]. The standard methionine (ATN) initiation codon was inferred for 12 of the 13 genes while the ND1 gene appears to use a valine (TTG) codon, as in the firefly [40] (Table 2). Open reading frames were terminated with the typical TAA and TAG codons for all genes (Table 2).

Ribosomal subunits, transfer RNAs, and non-coding regions

Both ribosomal subunit genes are encoded by the *β*-strand, with the s-rRNA (12S) separated from the l-rRNA (16S) by trnV $^{Val(uac)}$. The genomic position and transcriptional polarity of the rRNA genes is typical of insect species (Figure 2).

A total of 22 tRNAs corresponding with the standard metazoan gene set were identified on the basis of their respective anticodons and secondary structures (Table 2). Gene lengths are largely congruent with other insect species. All tRNAs could be folded into the canonical cloverleaf structure except for trnS$^{Ser\ (ucu)}$, which had four nucleotides in the unpaired loop but lacked the DHU arm. This feature is common among insects and metazoans generally [41]. The arrangement of the tRNAs differed from the other taxa at positions known to be variable [42,18]. At the COX2-ATP8 junction, *L. pallens* presents a large overlap of 66bp between trnK$^{Lys\ (cuu)}$ and trnD$^{Asp\ (guc)}$, with an inverted orientiation of trnK$^{Lys\ (cuu)}$ compared with other insect mitogenomes (Figure 2). At the s-rRNA-ND2 junction, which encompasses the highly variable AT-rich region, a translocation of the trnN$^{Asn\ (auu)}$ is evident in the *L. pallens* and *S.invicta* in respect to other insect genomes, and it is also inverted in *S. invicta* (Figure 2). Finally, the orientation of the trnG$^{Gly\ (ucc)}$ in *L. pallens* differs from *S. invicta* and *A. mellifera*, but is common in other insect taxa including *A. ephippium* or *D. melanogaster* [42].

Finally, a total of 837 noncoding nucleotides were identified, with 575 bp spread across 27 intergenic regions and a large contiguous 262 bp noncoding region (Table 2). The large noncoding region probably represents the putative AT-rich region based on its relative position between the tRNA$^{Asn(auu)}$ and tRNA$^{Met(cau)}$ [18], and sequence characteristics (A+T-rich and noncoding).

Conclusion

The 454 NGS platform has become a commonly used tool for the development of genetic markers for systematic research [43]. In this study, we successfully isolated 3,207 microsatellite-containing contigs for *L. pallens* from a total of 49,014 reads covering approximately 8.8% of the genome. From these contigs, 13 polymorphic microsatellite markers were successfully characterized. We also used NGS in combination with bioinformatic tools to assemble and annotate the complete mitochondrial DNA sequence of *L. pallens*, only the second ant genus to have its complete mitogenome characterized to date. These results are consistent with those of Miller et al. [8] in demonstrating that unexplored bulk data produced by NGS can be easily mined to recover full mitochondrial genomes from modest 454 analyses (in this case only $1/16^{th}$ of a 70 x 75 mm Pico Titer Plate). The 'molecular toolbox' presented here for *L. pallens* will be used to assess its population genetics, social structure and evolutionary history. Such studies will provide a valuable framework for quantifying the effects of rainforest fragmentation on biodiversity in New Caledonia.

Acknowledgements

This work has been partly funded by a competitive funding grant from Charles Darwin University to MB as part as her PhD thesis, and by the Agence Nationale de la Recherche BDIV-07-008, project 'Incendies Nouvelle Calédonie'. We are very grateful to the New Caledonian Gohapin tribe who has allowed us to access and sample on their land, and to Jasmin Packer, Quentin Auriac, Barbara Pianu and Viviane Degret for their help in the field. We thank Alan Andersen for providing the *Leptomyrmex nigriventris* specimen from the CSIRO-TERC collection.

226 **References**

227 1. Andersen AN, Majer J (2004) Ants show the way Down Under: invertebrates as
228 bioindicators in land management. Frontiers in ecology and the environment 2 (6):291-298.
229 doi:10.1890/1540-9295(2004)002[0292:astwdu]2.0.co;2
230 2. Laurance WF, Lovejoy TE, Vasconcelos HL, Bruna EM, Didham RK, Stouffer PC,
231 Gascon C, Bierregaard RO, Laurance SG, Sampaio E (2002) Ecosystem decay of Amazonian
232 forest fragments: a 22-year investigation. Conservation Biology 16:605-618
233 3. Bickel TO, Brühl CA, Gadau JR, Hölldobler B, Linsenmair KE (2006) Influence of habitat
234 fragmentation on the genetic variability in leaf litter ant populations in tropical rainforests of
235 Sabah, Borneo. Biodiversity and Conservation 15 (1):157-175. doi:10.1007/s10531-004-
236 4248-1
237 4. Tawato N, Harper NE, Mohamed M, Khen CV, Searle JB, Hill JK (2011) Impacts of forest
238 fragmentation on the genetic diversity and population structure of *Pachycondyla obscurans* in
239 Sabah, Malaysian Borneo. Asian Myrmecology 4:59-68
240 5. Pamilo P, Gertsch P, Thoren P, Seppa P (1997) Molecular population genetics of social
241 insects. Annual Review of Ecology and Systematics 28:1-25
242 6. Crozier RH, Oldroyd BP, Tay WT, Kaufmann BE, Johnson RN, Carew ME, Jennings KM
243 (1997) Molecular advances in understanding social insect population structure.
244 Electrophoresis 18 (9):1672-1675. doi:10.1002/elps.1150180934
245 7. Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011) Rise of the machines –
246 recommendations for ecologists when using next generation sequencing for microsatellite
247 development. Molecular Ecology Resources 11 (6):1093-1101. doi:10.1111/j.1755-
248 0998.2011.03037.x
249 8. Miller A, Good R, Coleman R, Lancaster M, Weeks A (2012) Microsatellite loci and the
250 complete mitochondrial DNA sequence characterized through next generation sequencing

and de novo genome assembly for the critically endangered orange-bellied parrot, Neophema chrysogaster. Molecular Biology Reports:1-8. doi:10.1007/s11033-012-1950-z

9. Prosdocimi F, Carvalho D, Almeida R, Beheregaray L (2012) The complete mitochondrial genome of two recently derived species of the fish genus *Nannoperca* (Perciformes, Percichthyidae). Molecular Biology Reports 39 (3):2767-2772. doi:10.1007/s11033-011-1034-5

10. Avise JC (2000) Phylogeography: the history and formation of species. President and Fellows of Harvard College, United States of America

11. Li J, Heinz KM (2000) Genome complexity and organization in the red imported fire ant *Solenopsis invicta* Buren. Genetics Research 75 (02):129-135. doi:doi:null

12. Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D, Dijkstra MB, Oettler J, Comtesse F, Shih C-J, Wu W-J, Yang C-C, Thomas J, Beaudoing E, Pradervand S, Flegel V, Cook ED, Fabbretti R, Stockinger H, Long L, Farmerie WG, Oakey J, Boomsma JJ, Pamilo P, Yi SV, Heinze J, Goodisman MAD, Farinelli L, Harshman K, Hulo N, Cerutti L, Xenarios I, Shoemaker D, Keller L (2011) The genome of the fire ant *Solenopsis invicta*. Proceedings of the National Academy of Sciences 108 (14):5679-5684. doi:10.1073/pnas.1009690108

13. Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Fave M-J, Fernandes V, Gadau J, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Helmkampf M, Holley J-A, Hu H, Viniegra ASI, Johnson BR, Johnson RM, Khila A, Kim JW, Laird J, Mathis KA, Moeller JA, Muñoz-Torres MC, Murphy MC, Nakamura R, Nigam S, Overson RP, Placek JE, Rajakumar R, Reese JT, Robertson HM, Smith CR, Suarez AV, Suen G, Suhr EL, Tao S, Torres CW, van Wilgenburg E, Viljakainen L, Walden KKO, Wild AL, Yandell M, Yorke JA, Tsutsui ND (2011) Draft genome of the globally widespread and

invasive Argentine ant (*Linepithema humile*). Proceedings of the National Academy of Sciences 108 (14):5673-5678. doi:10.1073/pnas.1008617108

14. Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Favé M-J, Fernandes V, Gibson JD, Graur D, Gronenberg W, Grubbs KJ, Hagen DE, Viniegra ASI, Johnson BR, Johnson RM, Khila A, Kim JW, Mathis KA, Munoz-Torres MC, Murphy MC, Mustard JA, Nakamura R, Niehuis O, Nigam S, Overson RP, Placek JE, Rajakumar R, Reese JT, Suen G, Tao S, Torres CW, Tsutsui ND, Viljakainen L, Wolschin F, Gadau J (2011) Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proceedings of the National Academy of Sciences 108 (14):5667-5672. doi:10.1073/pnas.1007901108

15. Suen G, Teiling C, Li L, Holt C, Abouheif E, Bornberg-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A, Denas O, Elhaik E, Favé M-J, Gadau J, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Harkins TT, Helmkampf M, Hu H, Johnson BR, Kim J, Marsh SE, Moeller JA, Muñoz-Torres MC, Murphy MC, Naughton MC, Nigam S, Overson R, Rajakumar R, Reese JT, Scott JJ, Smith CR, Tao S, Tsutsui ND, Viljakainen L, Wissler L, Yandell MD, Zimmer F, Taylor J, Slater SC, Clifton SW, Warren WC, Elsik CG, Smith CD, Weinstock GM, Gerardo NM, Currie CR (2011) The Genome Sequence of the Leaf-Cutter Ant *Atta cephalotes* Reveals Insights into Its Obligate Symbiotic Lifestyle. PLoS Genetics 7 (2). doi:10.1371/journal.pgen.1002007

16. Nygaard S, Zhang G, Schiøtt M, Li C, Wurm Y, Hu H, Zhou J, Ji L, Qiu F, Rasmussen M, Pan H, Hauser F, Krogh A, Grimmelikhuijzen CJP, Wang J, Boomsma JJ (2011) The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Res 21 (8):1339-1348. doi:10.1101/gr.121392.111

17. Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, Zhang P, Huang Z, Berger SL, Reinberg D, Wang J, Liebig J (2010) Genomic Comparison of

300 the Ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329 (5995):1068-1071.
301 doi:10.1126/science.1192428
302 18. Gotzek D, Clarke J, Shoemaker D (2010) Mitochondrial genome evolution in fire ants
303 (Hymenoptera: Formicidae). BMC Evolutionary Biology 10 (1):300
304 19. Lucky A, Ward PS (2010) Taxonomic revision of the ant genus *Leptomyrmex* Mayr
305 (Hymenoptera: Formicidae). Zootaxa 2688:1-67
306 20. Jaffré T, Bouchet P, Veillon J-M (1998) Threatened plants of New Caledonia: Is the
307 system of protected areas adequate? Biodiversity and Conservation 7:109-135
308 21. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J,
309 Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC,
310 He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-
311 B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H,
312 Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R,
313 Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR,
314 Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF,
315 Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors.
316 Nature 437 (7057):376-380.
317 doi:http://www.nature.com/nature/journal/v437/n7057/suppinfo/nature03959_S1.html
318 22. Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin J-F (2010) QDD:
319 a user-friendly program to select microsatellite markers and design primers from large
320 sequencing projects. Bioinformatics 26 (3):403-404. doi:10.1093/bioinformatics/btp670
321 23. Rozen S, Skaletsky H (1999) Primer3 on the WWW for General Users and for Biologist
322 Programmers. In: Misener S, Krawetz S (eds) Bioinformatics Methods and Protocols, vol 132.
323 Methods in Molecular Biology. Humana Press, pp 365-386. doi:10.1385/1-59259-192-2:365

24. Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. Molecular Ecology Resources 12 (3):456-463. doi:10.1111/j.1755-0998.2011.03104.x

25. Park S (2001) The Excel microsatellite toolkit. Website http://animalgenomics ucd ie/sdepark/ms-toolkit/[accessed August 2007]

26. Raymond M, Rousset F (1995) Genepop (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity 86:248-249

27. Rice WR (1989) Analyzing tables of statistical tests. Evolution 43 (1):223-225

28. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4 (3):535-538. doi:10.1111/j.1471-8286.2004.00684.x

29. Brookfield J (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 5 (3):453-455

30. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) Galaxy: A platform for interactive large-scale genome analysis. Genome Res 15 (10):1451-1455. doi:10.1101/gr.4086505

31. Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S (2004) Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. Genome Res 14 (6):1147-1159. doi:10.1101/gr.1917404

32. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2011) Geneious. 5.5.6 edn., Available from http://www.geneious.com/

348 33. Eddy SR, Durbin R (1994) RNA sequence analysis using covariance models. Nucleic
349 Acids Research 22 (11):2079-2088. doi:10.1093/nar/22.11.2079
350 34. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes
351 with DOGMA. Bioinformatics (Oxford, England) 20 (17):3252-3255.
352 doi:10.1093/bioinformatics/bth352
353 35. Cameron SL, Dowton M, Castro LR, Ruberu K, Whiting MF, Austin AD, Diement K,
354 Stevens J (2008) Mitochondrial genome organization and phylogeny of two vespid wasps.
355 Genome 51 (10):800-808. doi:10.1139/g08-066
356 36. Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis*
357 *mellifera*: complete sequence and genome organization. Genetics 133 (1):97-117
358 37. Lewis OL, Farr CL, Kaguni LS (1995) *Drosophila melanogaster* mitochondrial DNA:
359 completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular
360 Biology 4 (4):263-278. doi:10.1111/j.1365-2583.1995.tb00032.x
361 38. Martins J, Solomon SE, Mikheyev AS, Mueller UG, Ortiz A, Bacci M (2007) Nuclear
362 mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini).
363 Insect Molecular Biology 16 (6):777-784. doi:10.1111/j.1365-2583.2007.00771.x
364 39. Clary D, Wolstenholme D (1985) The mitochondrial DNA molecule of *Drosophila*
365 *yakuba*: Nucleotide sequence, gene organization, and genetic code. Journal of Molecular
366 Evolution 22 (3):252-271. doi:10.1007/bf02099755
367 40. Bae JS, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly,
368 *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis
369 with other insects. Molecular Phylogenetics and Evolution 32 (3):978-985.
370 doi:10.1016/j.ympev.2004.03.009
371 41. Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution.
372 International Review of Cytology 141:173-216

373 42. Dowton M, Cameron SL, Dowavic JI, Austin AD, Whiting MF (2009) Characterization
374 of 67 Mitochondrial tRNA Gene Rearrangements in the Hymenoptera Suggests That
375 Mitochondrial tRNA Gene Position Is Selectively Neutral. Molecular Biology and Evolution
376 26 (7):1607-1617. doi:10.1093/molbev/msp072
377 43. Mardis ER (2008) The impact of next-generation sequencing technology on genetics.
378 Trends in Genetics 24 (3):133-141. doi:10.1016/j.tig.2007.12.007
379
380
381
382
383
384
385
386
387
388

Table 1. Primer sequence and characteristics of the 13 microsatellite loci developed for *L. pallens*, and subsequently amplified in two multiplex (Mpex) reactions. FAM-, NED-, VIC- and PET- are fluorescent universal tags, and reverse primers were modified with a pigtail 5' sequence (see Blacket et al. [29]). Number of individuals (n), number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities, Hardy-Weinberg equilibrium *P*-values (HWE; bold values indicate significant deviations after Bonferroni correction) and inbreeding coefficient (F_{IS}) are indicated. Successful cross-species amplifications are indicated as follows: a = *L. geniculatus* (NC); b = *L. nigriceps* (NC) and c = *L. nigriventris* (AU)

Mpex	Locus	Primer sequence (5'-3')	Repeat motif	*n*	N_A	Size range (bp)	H_O/H_E	HWE	F_{IS}	Null alleles frequency	Other sp.	GenBank accession #
1	LP1	FAM-AACATTCGCAAACTCGTATTTAT Pigtail- TATAAGCGTCAATCGGCACA	$(AT)_9$	41	2	125-127	0.42/0.51	0.35	0.18	0.06	b, c	KC190501
	LP2	FAM-TTGGTCCCGCGAATTAAATA Pigtail- CCTCGCCAAGAAAATATTGC	$(CCG)_{10}$	41	4	146-161	0.44/0.45	0.27	0.03	-0.01	b, c	KC190502
	LP16	VIC-GGTCGGTAGTGCGGTATCTT Pigtail- CGTCATTGTGTTTCGTAAATTG	$(AG)_{10}$	41	3	136-140	0.46/0.44	1	-0.06	-0.02	b	KC190503
	LP23	NED-TCACTCGATTTCGTGACTGC Pigtail- ATCGCGCACAGGAATCTTAC	$(AG)_{10}$	41	4	163-175	0.54/0.46	0.69	-0.17	0	a, b, c	KC190504
	LP24	NED-GGAACAGGTGCTGAGAATCC Pigtail- TGGCTAGTCCATGATTGTGC	$(AT)_{14}$	41	12	189-217	0.68/0.76	0.52	0.11	0.04	b	KC190505
	LP25	NED-GAATCGAGCACGATCAACAA Pigtail- TACGCGTGCTCACTCAAGTC	$(ACG)_{18}$	41	3	112-118	0.54/0.62	0.24	0.14	0.05	b, c	KC190506
2	LP4	FAM-CTCGCGAGACTTCGATAAGG Pigtail- CCTGGAACGAGATCGACAAT	$(AGC)_9$	41	4	184-211	0.59/0.63	0.08	0.08	0.03	b, c	KC190507
	LP7	FAM-TGACTGCATATCTGTAAGGAATCTG Pigtail- TGACAAATAAACACGTACGCAA	$(AC)_{12}$	41	2	117-119	0.37/0.33	0.66	-0.10	-0.03	b	KC190508
	LP12	VIC-TGTTATCTAGCATTCATTCTATTGCTG Pigtail- TTCTCAATATAAATCAGTGAACGC	$(AT)_{10}$	41	6	126-137	0.56/0.57	0.50	0.02	0.03	b	KC190509
	LP14	VIC-TTGCATCTATCTATCACACTATCGC Pigtail- TGTTCGATCCTCGATGTATCTC	$(AC)_{11}$	41	9	139-160	0.50/0.77	**<0.01**	**0.35**	0.13	-	KC190510
	LP21	NED-ATGATGAACGAAACCCAAGC Pigtail- AGTTGTTTCAGAAGGTGCCG	$(AG)_9$	41	4	186-194	0.37/0.38	0.01	0.04	0.01	-	KC190511
	LP26	NED-AACGTCGAAATCCGATGAAT Pigtail- TGCTTGAGACAGATAGCCCA	$(AT)_9$	41	3	166-170	0.34/0.51	**<0.01**	**0.33**	0.11	a, b, c	KC190512
	LP37	PET-ACGAGACGAGAGGGACAGAA Pigtail- GGAGGACGTGGGTAATGTGA	$(AG)_{10}$	41	3	138-142	0.39/0.46	0.40	0.15	0.04	b, c	KC190513

Table 2. Mitochondrial gene profile of *Leptomyrmex pallens*. Parentheses around the feature position indicate a transcription on the β-strand

Position		Feature	tRNA	Codon	Codon
Start	End		cove score	start	stop
1	1530	COX1		ATG	TAA
1530	1596	trnL $^{Leu(uaa)}$	51.17		
1596	2309	COX2		ATT	TAA
(2394	2463)	trnK $^{Lys(cuu)}$			
2398	2464	trnD $^{Asp(guc)}$	45.96		
2464	2613	ATP8		ATT	TAA
2613	3272	ATP6		ATA	TAA
3281	4078	COX3		ATG	TAA
4102	4170	trnG $^{Gly(ucc)}$	52.57		
4167	4517	NAD3		ATA	TAA
4528	4594	trnA $^{Ala(ugc)}$	39.70		
4599	4666	trnR $^{Arg(ucg)}$	27.97		
4740	4798	trnS $^{Ser(ucu)}$	21.83		
4807	4878	trnE $^{Glu(uuc)}$	46.77		
(4877	4941)	trnF $^{Phe(gaa)}$	44.32		
(4949	6601)	NAD5		ATA	TAA
(6603	6675)	trnH $^{His(gug)}$	31.36		
(6707	8050)	NAD4		ATG	TAA
(8067	8351)	NAD4L		ATA	TAA
8354	8417	trnT $^{Thr(ugu)}$	49.48		
(8433	8504)	trnP $^{Pro(ugg)}$	44.52		
8543	9079	NAD6		ATG	TAA
9087	10205	CYTB		ATG	TAA
10221	10288	trnS $^{Ser(uga)}$	33.96		
(10298	11230)	NAD1		TTG	TAG
(11246	11313)	trnL $^{Leu(uag)}$	35.77		
(11326	12599)	l-rRNA			
(12633	12700)	trnV $^{Val(uac)}$	45.18		
(12675	13448)	s-rRNA			
13613	13681	trnN $^{Asn\ (auu)}$	23.89		
13681	13875	AT-rich region			
13944	14010	trnM $^{Met(cau)}$	49.66		
14011	14076	trnI $^{Ile(gau)}$	27.57		
(14083	14153)	trnQ $^{Gln(uug)}$	38.10		
14299	15294	NAD2		ATT	TAA
15304	15373	trnW $^{Trp(uca)}$	54.91		
(15426	15488)	trnC $^{Cys(gca)}$	39.71		
(15497	15562)	trnY $^{Tyr(gua)}$	43.00		

Table 3. Genomic composition of the mitochondrial genome of *L. pallens* and other insects.

Family	Subfamily	Genus	Species	Genbank accession #	*α* -strand		13 Protein-coding genes		2 rRNAs		22 tRNAs*		AT-rich region	
					Length (bp)	A + T (%)	Length (bp)	A +T (%)	Length (bp)	A +T (%)	Length (bp)	A +T (%)	Length (bp)	A + T (%)
Formicidae	Dolichoderinae	*Leptomyrmex*	*pallens*	KC160533	15,591	69.5	11,070	67.2	2,048	70.6	1,419	80.6	195	95.4
	Myrmicinae	*Solenopsis*	*invicta*	NC_014672	15,549	77.1	11,047	74.2	2,114	82.1	1,507	85.3	377	93.1
Vespoidae	Eumeninae	*Abispa*	*ephippium*	NC_011520	16,953	80.7	11,305	78.7	2,180	81.9	1,787	83.5	308	89.9
Apidae	Apinae	*Apis*	*mellifera*	NC_001566	16,343	84.9	11,067	83.3	2,157	83.5	1,437	87	827	96
Drosophilidae	Drosophilinae	*Drosophila*	*melanogaster*	NC_001709	19,517	82.2	11,179	77.2	2,111	81.9	1,263	77.8	4,601	95.6

*all taxa had 22 tRNA except *A. ephippium*, which has 26 tRNA

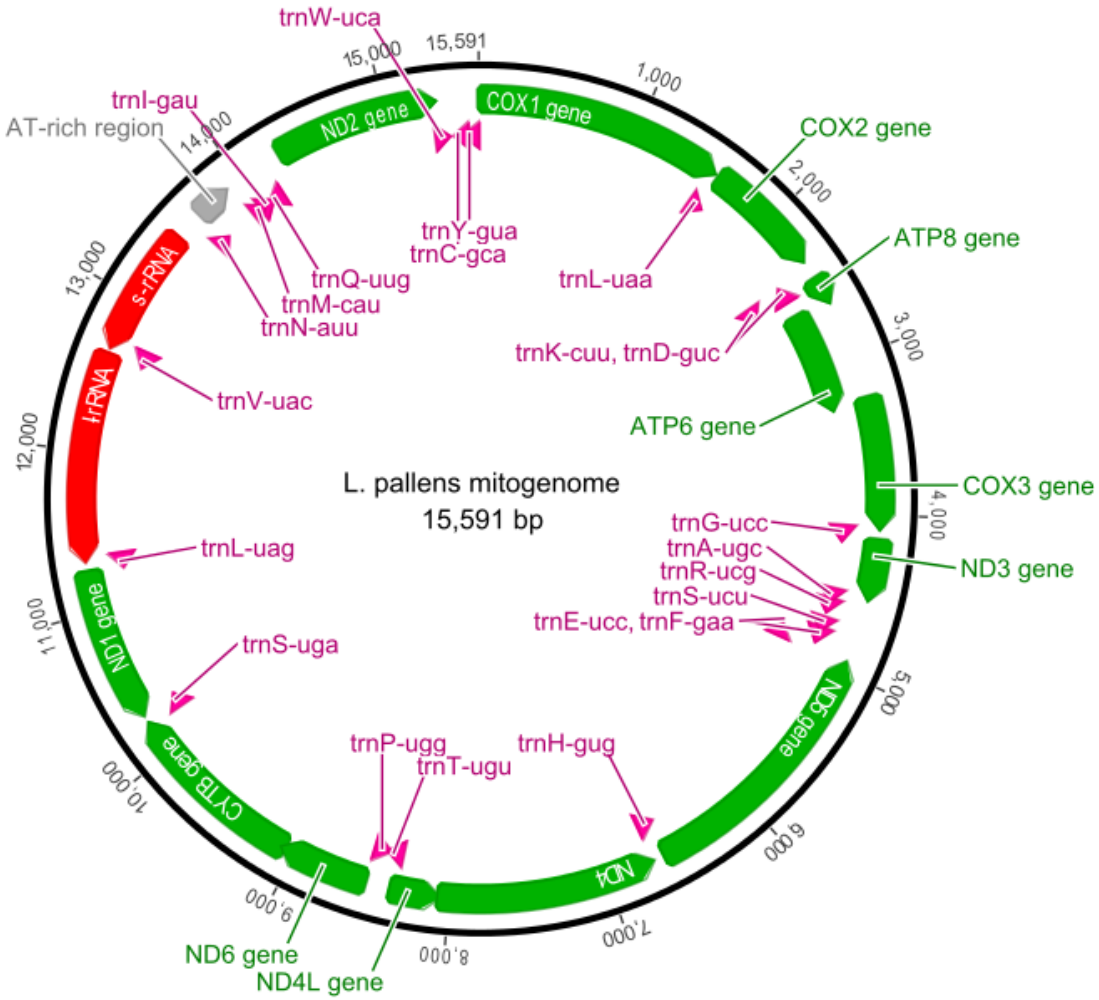

Fig. 1 Map of the *Leptomyrmex pallens* mitochondrial genome. COX1-3 indicates cytochrome c oxidase subunits 1–3; CYTB, cytochrome b; ATP6–8, ATPase subunits 6 and 8; ND1–6/4L, NADH dehydrogenase subunits 1–6/4L. Transfer RNA genes are designated by single-letter amino acid codes and corresponding anti-codon (Table 2). Arrow heads indicate direction of transcription

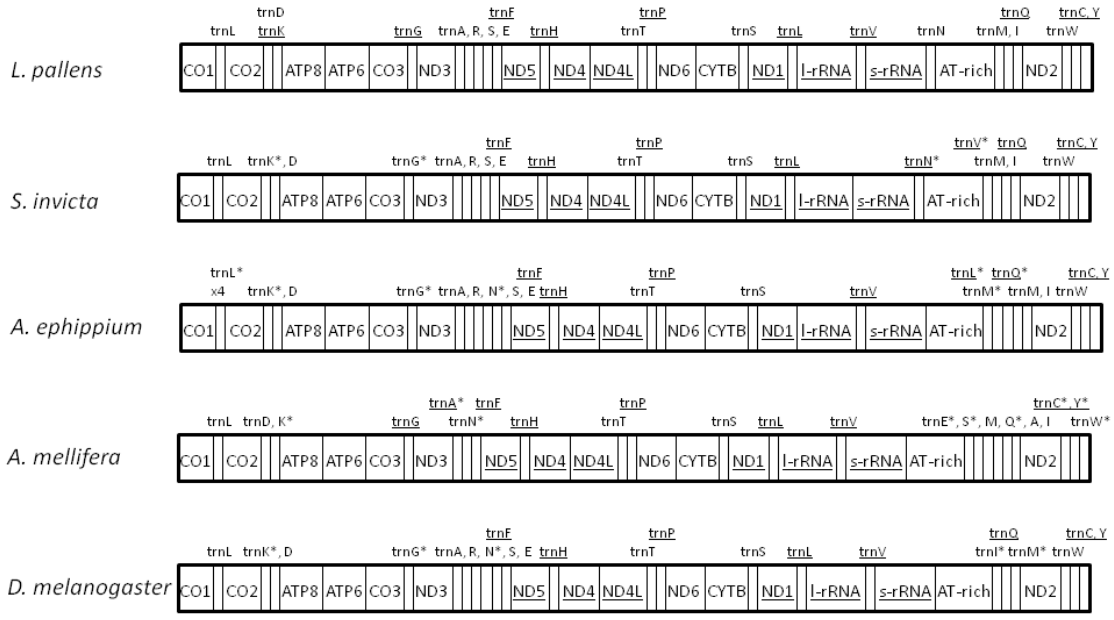

Fig. 2 Comparison of the mitogenomic organization of *L. pallens* and other insect taxa. Differences are highlighted by an asterisk. Genes are transcribed by the α-strand, except those underlined, which are transcribed on the β-strand

Author/s:
Berman, M;Austin, CM;Miller, AD

Title:
Characterisation of the complete mitochondrial genome and 13 microsatellite loci through next-generation sequencing for the New Caledonian spider-ant Leptomyrmex pallens

Date:
2014-03

Citation:
Berman, M., Austin, C. M. & Miller, A. D. (2014). Characterisation of the complete mitochondrial genome and 13 microsatellite loci through next-generation sequencing for the New Caledonian spider-ant Leptomyrmex pallens. MOLECULAR BIOLOGY REPORTS, 41 (3), pp.1179-1187. https://doi.org/10.1007/s11033-013-2657-5.

Persistent Link:
http://hdl.handle.net/11343/283221