No budget Mitogenomics: Assembling 14 new mitogenomes for the ant subfamily Pseudomyrmecinae from public data

Gabriel A Vieira Corresp., [1] , **Francisco Prosdocimi** Corresp. [1]

[1] Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

Corresponding Authors: Gabriel A Vieira, Francisco Prosdocimi
Email address: gabriel.vieira@bioqmed.ufrj.br, prosdocimi@bioqmed.ufrj.br

The advent of Next Generation Sequencing has reduced sequencing costs and increased genomic projects from a huge amount of organismal taxa, generating an unprecedented amount of genomic datasets publicly available. Often, only a tiny fraction of outstanding relevance of the genome data produced by researchers is used in their works. This fact allows the data generated to be recycled in further projects worldwide. The assembly of complete mitogenomes is frequently overlooked though it is useful to understand evolutionary relationships among taxa, especially those presenting poor mtDNA sampling at the level of genera and families. This is exactly the case for ants (Hymenoptera:Formicidae) and more specifically for the subfamily Pseudomyrmecinae, a group of arboreal ants with several cases of convergent coevolution without any complete mitochondrial sequence available. In this work, we assembled, annotated and performed comparative genomics analyses of 14 new complete mitochondria from Pseudomyrmecinae species relying solely on public datasets available from the Sequence Read Archive (SRA). We used all complete mitogenomes available for ants to study the gene order conservation and also to generate two phylogenetic trees using both (i) concatenated set of 13 mitochondrial genes and (ii) the whole mitochondrial sequences. Even though the tree topologies diverged subtly from each other (and from previous studies), our results confirm several known relationships and generate new evidences for sister clade classification inside Pseudomyrmecinae clade. We also performed a synteny analysis for Formcidae and identified possible sites in which nucleotidic insertions happened in mitogenomes of pseudomyrmecine ants. Using a data mining/bioinformatics approach, the current work increased the number of complete mitochondrial genomes available for ants from 15 to 29, demonstrating the unique potential of public databases for mitogenomics studies. The wide applications of mitogenomes in research and presence of mitochondrial data in different public dataset types makes the “no budget mitogenomics” approach ideal for comprehensive molecular studies, especially for subsampled taxa.

No budget Mitogenomics: 14 new mitogenomes for the ant subfamily Pseudomyrmecinae using public data

Gabriel Alves Vieira[1], Francisco Prosdocimi[1]

[1] Instituto de Bioquímica Médica Leopoldo de Meis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil;

Correspondig Authors:

Gabriel Vieira, Francisco Prosdocimi

Email address: gabriel.vieira@bioqmed.ufrj.br, prosdocimi@bioqmed.ufrj.br

Introduction

More than one decade after the advent of next-generation sequencing (NGS) (Margulies et al., 2005), it is now clear that this amazing and mature technology fostered an unprecedented increase in the generation of genomic data together with an important reduction in sequencing costs (Mardis, 2008; van Dijk et al., 2014; Goodwin, McPherson & McCombie, 2016). In order to gather and democratize the access to genomic data, the International Nucleotide Sequence Database Collaboration (INSDC, http://www.insdc.org/) has been established in 1987. This continuous effort comprises three international centers: (i) the National Center for Biotechnology Information (NCBI), (ii) the European Bioinformatics Institute (EBI) and (iii) the DNA Data Bank of Japan (DDBJ) (Karsch-Mizrachi, Takagi & Cochrane, 2017). As part of this outstanding initiative, the Sequence Read Archive (SRA) was created to host raw sequence reads and metadata generated by NGS projects (Kodama, Shumway & Leinonen, 2012). Making raw sequence data available is key for the experimental reproducibility, a pillar of scientific endeavor. SRA database has been recurrently used to support new research, such as: the evaluation of single nucleotide polymorphisms and deletions (Bordbari et al., 2017), the test of new bioinformatics software (Simpson et al., 2009; Langmead & Salzberg, 2012; Bolger, Lohse & Usadel, 2014), and also to evaluate the impacts of common procedures on data, such as trimming (Del Fabbro et al., 2013), among many other studies.

The availability of public data is continuously growing together with the potential uses of such databases to the scientific community. In a 2-year period (August-2015 to August-2017), 3000 trillion base pairs have been added to SRA, promoting a 233% growth of the repository (Karsch-Mizrachi, Takagi & Cochrane, 2017). However, potential uses of these data are far from being fully exploited, once public databases present resources that could be used to address diverse ranges of unexplored biological questions. Here we focus in searching for the presence of complete mitochondrial genomes in public genomics datasets.

Whole-Genome Sequencing (WGS) experiments and partial genome sequencing projects normally yield enough sequencing reads from mitochondria to allow the assembly of complete mitogenomes. These small organellar genomes can be often assembled in high coverage due to the high copy number of these organelles (Smith, 2015). Also, previous studies indicate that it is possible to recover complete and/or nearly complete mitochondrial sequences from RNA-Seq data (Tian & Smith, 2016; Rauch et al., 2017) and targeted sequencing strategies as exome

(Picardi & Pesole, 2012; Guo et al. 2013; Samuels et al. 2013) and UCE off-target data (Raposo do Amaral et al., 2015; Miller et al. 2016). The assembly of numerous complete mitogenomes and/or large mitochondrial contigs from the sequencing of pooled multi-species samples has also been performed successfully (Timmermans et al., 2015; Linard et al., 2018) under an approach named 'mito-metagenomics' (Tang et al., 2014) or 'mitochondrial metagenomics' (MMG) (Crampton-Platt et al., 2015).

Besides, some works have successfully used public data to assemble mitochondrial sequences (Diroma et al., 2014; Kayal et al., 2015; Linard et al., 2018). There remains a large number of species genomics data sampled in the SRA database without complete mitochondrial sequences available.

Due to their small sizes, high conservation and the absence of introns, mitogenomes are the most commonly sequenced chromosomes, especially for metazoans (Smith, 2015). Mitochondrial genomes are poorly sampled for many taxa and therefore our current knowledge about evolutionary biology of many clades could be improved with the use of public data. Being primarily maternally inherited and non-recombinant, such sequences are often used to study evolutionary biology, population genetics, phylogeography, systematics and conservation of various clades (Avise, 1994; Moritz, 1994), specially from subsampled taxa and non-model organisms.

An example of poor mitogenome taxon sampling occurs in ants (Hymenoptera: Formicidae). Despite being an ubiquitous, ecologically dominant and hyper diverse group (Wilson, 1990) with over 13,000 species described (Bolton, 2012), complete mitogenome records are available for mere 15 species in GenBank.

Although UCE sequencing data has been used to study ant phylogeny (Blaimer et al., 2015; Ward & Branstetter, 2017; Brasnstetter et al., 2017), there are limited attempts to recover mitochondrial sequences from these off-target data generated (Ströher et al., 2017) and use these information to further understand evolutionary relationships for the clade.

One particular ant group that suffers from poor mitogenome sampling is the ant subfamily Pseudomyrmecinae that contains 3 genera: (i) the New-World genus *Pseudomyrmex*, consisting of ~137 species, most of which can be classified in one of the ten morphological species groups described (Ward 1989, 1993, 1999, 2017); (ii) the Paleotropical *Tetraponera*,

with ~93 species; and (iii) the South American *Myrcidris*, that has only one species described, *Myrcidris epicharis* (Ward & Downie, 2005; Bolton, 2012; Ward, 2017).

There are two known groups of Pseudomyrmecinae: (i) a group composed of generalist arboreal species, and another group of (ii) ants specialized in plant colonization (Janzen, 1966; Ward, 1991). While ants from the Group 1 nest in dead sticks of various types of plants and are generally passive in relation to foreign objects, ants from the Group 2 are obligate inhabitants of hollow cavities in live tissues (domatia) of plants and are often aggressive towards other insects or plants. Also, ants from Group 2 provide protection from herbivory and competition to its host plant in a relationship commonly associated with coevolved mutualism (Janzen, 1966; Ward, 1991).

Previous works using morphological and molecular data (Ward, 1991; Ward & Downie, 2005) suggest that this kind of mutualism has evolved independently at least 12 times in the Pseudomyrmecinae subfamily. For instance, *Pseudomyrmex* ants evolved similar behaviors by convergence, despite coevolving with different plant hosts. Commonly associated to these ants are species from the Fabaceae genera *Vachellia* (formerly part of the *Acacia* genus), *Platymisicum* and *Tachigali;* and the Polygonaceae genera *Triplaris* and *Ruprechtia* (Ward & Downie, 2005; Chomicki, Ward & Renner, 2015).

Cases of convergent evolution are frequently characterized using phylogenetics approaches (Ward & Branstetter, 2017). Evolutionary analyses of mitochondrial genomes often allow a better understanding about the history of higher clades like genera, families and orders because of the slow pace of mitochondrial evolution (DeSalle et al., 2017). The estimated divergence time between *Pseudomyrmex* and *Tetraponera* is measured in ~95.8 MYA (Million Years Ago), according to Gómez-Acevedo et al. (2010), supporting a vicariance event during the separation of South America from Africa.

Several molecular studies have been described on *Pseudomyrmex*, some addressing co-evolutionary questions, such as the impact of mutualistic associations in the rate of genome evolution (Rubin & Moreau, 2016). Other works focused on phylogenetic relationships and biogeography to characterize ant-plant associations (Chomicki, Ward & Renner, 2015; Ward & Branstetter, 2017). However, complete mitogenomes analyses have never been performed for Pseudomyrmecinae due to absence of these data. In the current “no budtget mitogenomics” approach (defined here as the usage of public raw data to assemble large mitochondrial

sequences unavailable at public databases), we used publicly available genomic data generated elsewhere (S1 Table) to assemble and analyze the complete mitochondrial sequence for 12 *Pseudomyrmex* and 2 *Tetraponera* species from Pseudomyrmecinae subfamily. Thus, we present the first dozen of mitogenomes for this subfamily and performed evolutionary analyses of them together with all other available Formicidae mitogenomes, trying to better understand the sister clade relationships inside this highly diverse clade. The current manuscript presents new complete mitochondrial sequences for ant species that cover 5 out of 10 *Pseudomyrmex* species groups and it almost duplicates the number of mitochondrial genomes available for ants, increasing this number from 15 to 29.

Methods

Data acquisition

Fourteen Illumina paired-end datasets were downloaded from EMBL Nucleotide Archive (https://www.ebi.ac.uk/ena) in SRA file format (see Table S1). The datasets were converted to FASTQ using fastq-dump (with –readids and –split-files parameters) from SRAtoolkit.2.8.2.

Mitochondrial genome assembly and annotation

The complete datasets with different number of sequencing reads were used as input for *de novo* assembly using NOVOPlasty2.6.3 (Dierckxsens, Mardulyn & Smits, 2016), except by the dataset for *Tetraponera rufonigra*, that had to be trimmed with Trimmomatic v.0.36 (Bolger, Lohse & Usadel, 2014) to produce sequences with the same length, in bp. NOVOPlasty assemblies needs a seed sequence to start the assembly. Seeds were selected using COI (Cytochrome Oxidase I) sequences from the same species (when available) or using COI regions from closely-related species. Preliminary mitogenome assemblies by NOVOPlasty were used as reference to a second round of genome assembly using MIRA v.4.0.2 (Chevreux, Wetter & Suhai, 1999). MIRA has been used to map raw sequencing reads to the consensus mitochondrial sequence. When necessary, we also used MITObim v.1.9 (Hahn, Bachmann & Chevreux, 2013) to close small gaps and/or to obtain a more reliable mitogenome assembly.

Tablet software (Milne et al., 2012) was used to check read coverage and circularization of complete mitogenomes. Automatic annotation performed using MITOSWebServer (Bernt et al., 2013) was followed by manual curation using Artemis (Carver et al., 2012) and aided by BLAST (Altschul et al., 1997) to allow a rational decision on gene boundaries.

Phylogenomics analyses

Formicidae phylogenetics relationships were reconstructed using (i) the 14 complete mitogenomes produced by us together with (ii) all other 15 complete mitochondrial genomes currently available for the clade; and (iii) two mitogenomes of bees (Apidae family) used as outgroups. Two phylogenetics trees have been built using (i) the whole mitochondrial sequence; and (ii) the concatenated gene set of all 13 protein-coding genes. For the former, we manually edited the sequences to start at the COI gene when necessary and aligned the whole mitogenomes using ClustalW (Thompson, Gibson & Higgins, 2003). For the latter, we aligned

and concatenated the nucleotides for all protein-coding genes (PCGs) using the Phylomito package (https://github.com/igorrcosta/phylomito). Modeltest (Posada & Crandall, 1998) was run with the two datasets and identified the model GTR+G+I as the nucleotide substitution model that better explained sequence variation. Aligned sequences were used as input to a Maximum Likelihood (ML) analysis in MEGA7 (Kumar, Stecher & Tamura, 2016). Resampling was conducted by bootstrap using 1000 replicates. Blast Ring Image Generator (BRIG) software was run (Alikhan et al. 2011) to compare and visualize all mitogenomes of Pseudomyrmecinae produced here.

159 ## Results
160
161 *Mitogenome assembly and annotation of Pseudomyrmecinae*
162
163 The 14 genomic datasets used to assemble complete mitogenome sequences for
164 pseudomyrmecine ants were downloaded from SRA database (Table S1). Two different dataset
165 types were used: (i) Whole Genome Sequencing (WGS), that often contained a higher amount of
166 sequencing data, being calculated as 212.7 Gbp for 6 species; an average of 35.45 Gbp per
167 species (Rubin & Moreau, 2016); and (ii) UCE (Ultra Conserved Elements) experiments, on
168 which we have downloaded 5.94 Gbp for 8 species; an average of 742.5 Mbp per species
169 (Branstetter et al., 2017; Ward & Branstetter, 2017).
170 The complete dataset downloaded for each species was used as input for a *de novo*
171 sequence assembly using NOVOPlasty. After this first round of genome assembly, we used a
172 subset containing either 2 or 4 million sequencing reads as input for a second round of genome
173 assembly using MIRA software. This procedure was performed to both map the sequencing
174 reads into the preliminary assembly and improve the mitogenome quality. For some
175 mitogenomes MIRA could not produce the complete, circularized mitochondrion genome; and a
176 third round of assembly was needed. In that case, the largest contig generated by MIRA has been
177 used as backbone to finish the assembly using MITObim (Table 1). This pipeline was capable to
178 assemble the whole mitochondria of all Pseudomyrmecinae except for *T. aethiops*, on which we
179 have had to use the entire sequencing read dataset for MIRA and MITObim instead of filtering
180 the subset of reads on round 2. The 14 mitochondrial genomes built here were checked for
181 circularity and confirmed to present, as expected for metazoans, 13 protein-coding genes, 22
182 tRNAs, 2 rRNAs and a control region (Wolstenholme, 1992). The genome annotation for all
183 complete mitogenomes is presented (Table S2). All mitochondrial genomes produced here were
184 submitted to GenBank under the Third Party Annotation (TPA) database (Cochrane et al., 2006)
185 that provided accession numbers allowing sequence retrieval (Table 1).
186 The sequencing read coverage for mitogenomes ranged between 85x and 292x for
187 mitogenomes on which a subset of reads was used. For *T. aethiops,* the coverage was higher
188 once the entire dataset was used (712x). Assembly coverage was observed to be evenly

distributed, except in cases of AT-rich regions that presented low coverage, generally close to poly-T sequences.

Mitogenome size variation in the Pseudomyrmex genus

Pseudomyrmex mitogenomes have shown significant variation in size, ranging from 15704 to 18835 bp (Table 1). We observed 3 distinct mitogenome size ranges for the clade (Table 1). Mitogenome size in the genus varied from: (i) less than 16 kb in *P. gracilis* and *P. concolor*; (ii) between 17kb and 18kb in *P. pallidus* and *P. dendroicus*; and (iii) higher than 18kb in other species, that belong to *P. ferrugineus* group. A comparative genomics analysis using BRIG software identified four variable regions as putative insertion segments (Figure 1). After genome annotation, we identified these presumed insertions to be located between (i) *COII* and *trn-K*; (ii) *ATP8* and *ATP6*; (iii) *trn-N* and *trn-F*; and (iv) trn-W and *COI*.

Gene order arrangements in ant mitogenomes (Formicidae)

Regardless the limited sample of mitochondria analyzed for ants, in general, five slightly different synteny rearrangements (Figure 2) could be observed in Formicidae family (Duan, Peng & Qian, 2016). All Pseudomyrmecinae and Dolichoderinae mitogenomes analyzed showed a single conserved gene arrangement for all species that is also shared by most of Formicinae species. Formicinae and Myrmicinae clades present a modal synteny arrangement suggesting a possible ancestral gene arrangement for each group. One single species of Formicinae (*Camponotus atrox)* present inversions between *trn-M*, *I* and *Q* that differs from other mitogenomes from this subfamily, possibly representing a derived variation. Myrmicinae also present two other unique rearrangements restricted to a single species each, suggesting derived syntenies: (i) *P. punctatus* has an inversion between *trn-K* and *D*; and (ii) *W. auropunctata* presents both an inversion between *trn-V* and D-loop and a feature (*trnY*) on the opposite strand when compared to the others.

Phylogenetic analyses of Formicidae using mitogenome data

In order to assess the phylogeny of the group, two Maximum Likelihood trees were produced using slightly different input data: (i) the aligned and concatenated sequences for all 13 mitochondrial PCG's (Figure 3); and (ii) the complete mitochondrial genomes (Figure 4). We analyzed all ant species presenting complete mitogenomes available on Genbank (Gotzek, Clarke & Shoemaker, 2010; Hasegawa et al., 2011; Berman, Austin & Miller, 2014; Babbucci et al., 2014; Kim, Hong & Kim, 2015; Duan, Peng & Qian, 2016; Liu et al., 2016; Yang et al., 2016) and two Apidae bees as outgroups (Crozier & Crozier, 1993; Cha et al., 2007) (see accession numbers and references for all sequences on Table S3). The trees reconstructed from mitochondrial data corroborated most of the phylogenetic relationships known for ants, with several clades observed as monophyletic with high confidence (bootstrap = 100). Both trees showed similar results, though differences can be observed in several nodes regarding tree topology and/or statistical support. The major difference observed is that the gene-concatenation tree displayed all subfamilies as monophyletic, while Myrmicinae was recovered as paraphyletic in the tree based on complete mitogenomes.

Discussion

In this study, we used public data to assembly, annotate, compare and provide evolutionary analyses of 14 complete mitochondrial genome sequences from the ant subfamily Pseudomyrmecinae plus 15 other ant mitogenomes downloaded from GenBank.

Genome coverage and NuMTs

Even though pieces of the mitochondrion genome may be copied to the nucleus forming NuMTs (Nuclear Mitochondrion Sequences), the genome coverage obtained for the assemblies often presented uniform distributions, even for *Pseudomyrmex gracilis* on which NuMTs have been previously identified (Rubin & Moreau, 2016). The correct assembly of mitochondrial genomes were possible because the number mitochondrial reads is probably much higher than the number of reads coming from NuMTs.

The low coverage in segments with a pronounced AT-bias should be expected because AT-rich regions are known to have reduced amplification in Illumina library preparation protocols (Dohm et al., 2008; Aird et al., 2011; Oyola et al., 2012). It has been shown that ant mitogenomes have a remarkable AT bias in the control region that exceeds 90% (Berman, Austin & Miller, 2014; Liu et al., 2016). Also, this region has already been proved to be particularly difficult to sequence in hymenopterans (Castro & Dowton, 2005; Dowton et al., 2009; Rodovalho et al., 2014).

Comparative mitogenomics

Aside from the identification of 4 putative insertion sites that could explain the differences observed in mitogenome size (pointed by arrows in Figure 1), we also observed that all 7 mitogenomes included in *P. ferrugineus* group have approximately the same genome size in bp, suggesting that this group is monophyletic. On the other hand, there is a significant difference in mitogenome size between *P. concolor* (15906 bp) and *P. dendroicus* (17362 bp), both belonging to *P. viidus* species group. This corroborates previous works indicating that this species group is paraphyletic (Ward, 1989; Ward & Downie, 2005).

The multiple syntenies encountered within Myrmicinae and Formicinae are probably associated to the remarkable biodiversity observed for these two subfamilies: Myrmicinae is the largest ant subfamily in species richness, with over 6,600 species described, almost half of all biodiversity documented for ants; and Formicidae is the second most biodiverse, featuring over 3,100 species. Other Formicidae subfamilies in this study are not nearly as diverse: Dolichoderinae has ~713 species while Pseudomyrmecinae presents ~231 species documented (Bolton, 2012). Ancestral gene arrangement for Formicinae is identical to the one observed in Pseudomyrmecinae and Dolichoderinae, signaling that Formicinae is closely related to this group than to Myrmicinae.

A higher number of mitogenomes and broader taxon coverage will improve the assessment of correlation between mitochondrial gene order and subfamily biodiversity, allowing a better understanding of synteny evolution in ant mitochondria.

Phylogenomic relationships of Formicidae inferred using mitogenome data

Overall, in the phylogenomic trees generated for the whole Formicidae family, the phylogeny of the subfamily Pseudomyrmecinae was strongly recollected as monophyletic, and the phylogenetic positions of most clades were well resolved. The monophyly for the Pseudomyrmecinae subfamily and also for *Pseudomyrmex* and *Tetraponera* genera were recovered with 100% bootstrap support (BS) in both trees. The genus *Pseudomyrmex* presented few unsupported nodes, but *Tetraponera* was completely resolved on both trees (BS = 100). In both trees, both the monophyletic status of the *P. flavicornis* group and the paraphyletic status of the *P. viidus* group confirms (i) previous observations based exclusively on morphology (Ward, 1989), (ii) phylogenies using both morphological characters and few nuclear markers (Ward & Downie, 2005), and (iii) our own observations regarding mitogenome size. Although the morphological division in species groups has not been formalized or regulated under nomenclatures (Ward, 2017), the work using a hybrid morphological/molecular approach of Ward & Downie, 2005 shows that only two out of nine groups defined at the time were paraphyletic: *P. pallens* and *P. viidus* groups. This issue confirms the relevance of using morphological characters in determining relationships between clades, but also reinforces that molecular evidence can clarify and complement such studies, refining and improving the overall

support of the phylogenies reconstructed. Under this work, we generated complete mitochondrial sequences for ants representing 5 out of the 10 groups described for *Pseudomyrmex* species, covering at least half of *Pseudomyrmex* genetic diversity and adding a new source of molecular evidence for further studies on the clade.

Both trees suggest strongly that ant-plant mutualisms are paraphyletic in Pseudomyrmecinae (please check species labeled with asterisks in trees), also adding evidence to previous assumptions of generalist behavior as a basal trait in the *Pseudomyrmex* genus (Ward & Branstetter, 2017). This suggests that ant-plant coevolution developed later (and independently) several times in the clade. Mutualistic species are more common in the *P. ferrugineus* species group, strengthening the hypothesis of mutualism being a derived trait in Pseudomyrmecinae. In the *Pseudomyrmex* genus, the *P. ferrugineus* group features may present two independent lineages of mutualistic ants (considering that *P. feralis* is often considered to display generalist behavior; BS = 50), while other two independent mutualistic lineages can be observed by the phylogenetic placement of the species *P. concolor* and *P. dendroicus*. Considering the *Tetraponera* genus, *T. aethiops* and *T. rufonigra* are closely related species and only *T. aethiops* presents exclusive ant-plant mutualistic behavior. This shows that evolution of mutualistic traits in Pseudomyrmecinae may have occurred in short periods of time. So, considering the limited number of species sampled here, we were able to identify 5 out of the 12 times that mutualistic associations have been reported to appear in the clade (Ward, 1991; Ward & Downie, 2005). With a better taxonomic coverage, this number can be increased and new analyses performed, further improving our understanding about these coevolutionary events.

Well resolved relationships for several Pseudomyrmecinae species (such as *P. peperi*, *P. veneficus*, *P. particeps*, *P. gracilis*, *T. aethiops* and *T. rufonigra*) corroborate both the results of Ward & Downie (2005) and the ML tree generated using UCE data from Ward & Branstetter (2017). The sister group relationship between *P. dendroicus* and *P. elongatus* is also well supported (BS = 100 in complete mitochondria tree; and BS = 99 in gene-concatenation tree), in line with a recent work using concatenated WGS scaffolds as input for ML tree reconstruction (Rubin & Moreau, 2016).

However, subtle differences were observed between our results and the inferred UCE multiloci phylogenetic relationships (Ward & Branstetter, 2017). Using UCE data, *P. janzeni* was observed as sister group to *P. ferrugineus*. Here, the complete mitogenome tree recaptured

this same relationship with a bootstrap replicate value of 77. On the other hand, in the concatenated gene set, the sister group relationship observed between *P. janzeni* and *P. flavicornis* showed a lower support (BS = 47). Overall, this relationship seemed to be better recollected by the analysis of the complete mitochondrial sequence, also corroborating the UCE analyses.

Within Pseudomyrmecinae, we observed two species whose phylogenetic positions were not well resolved by the current mitochondrial analyses and, therefore, their relationship can be seen as inconclusive: (i) *P. feralis* in the whole mitogenome tree; and (ii) *P. pallidus* in the concatenated gene tree. Both positions feature a bootstrap replicate value of 50.

Both trees showed Dolichoderinae subfamily as monophyletic, even though this result was not recovered in all replicates. Dolichoderinae is a highly diverse subfamily and contains over 700 species, but it has been represented here by merely two species. Thus, we believe that a higher coverage of species will improve the robustness of the phylogenetic analyses.

Previous work with morphological characters and/or nuclear genes presents evidence of sister group relationship between Pseudomyrmecinae and Myrmeciinae (Ward & Downie, 2005; Brady et al., 2006). However, complete mitochondrial genomes are not available for the subfamily Myrmeciinae. Therefore, Dolichoderinae is expected to be the closest relative to Pseudomyrmecinae in our trees. In fact, both trees corroborate large-scale molecular phylogenies using few nuclear genes (Brady et al., 2006) and UCE data (Branstetter et al., 2017). Shared synteny between all Pseudomyrmecinae and Dolichodrinae sampled also supports the sister group relationship observed. Our results evidence Myrmicinae as sister taxa to a clade containing both Pseudomyrmecinae and Dolichoderinae, while Formicinae has been observed as a more basal group in the Formicidae family. This position for Formicinae is highly supported in the gene-concatenation tree but not in the tree using complete mitogenomes. This position is not supported by other works using nuclear data that evidence a sister group relationship between Myrmicinae and Formicinae (Brady et al., 2006; Branstetter et al., 2017).

The monophyly of the subfamily Formicinae and all its nodes show maximum support on both trees (BS = 100). These trees also confirm the monophyly for the genus *Formica* and show genera *Camponotus* and *Polyrhachis* as closely related to each other, as observed in the work of Blaimer and collaborators (2015) that used UCE loci for tree inference. The only issue in this subfamily concerns the unsupported phylogenetic placement of Formicinae in relation to other

subfamilies. Mitogenome data successfully delivered sound phylogenetic relationships even for *Camponotus atrox* that showed a unique synteny but have had its position well resolved in both trees, including in the complete mitochondrial tree, that may be prone to suffer from synteny changes. This issue confirms the robustness of mitochondrial sequences to infer ant phylogenies.

Overall, the most controversial results obtained here are related to the position of the subfamily Myrmicinae. For that clade, the gene concatenation tree was capable to indicate monophyly (BS = 74) but whole mitogenome data produced paraphyly. In the latter case, a clade consisting of myrmicine ants *Myrmica scabrinois*, *Pristomyrmex punctatus* and *Atta texana* appear as basal group to all other ants. On the other hand, both trees successfully recaptured the monophyly of the genus *Solenopsis* and the relationships between their species (*S. geminata* as sister group of the clade consisting of *S. invicta* and *S. geminata*) with 100% bootstrap support. The sister group relationship between *Solenopsis* spp. and *Vollenhovia emeryi* is also recovered. These results corroborate those obtained by the use of concatenated amino acid sequences of all mitochondrial PCGs for tree inference (Duan, Peng & Qian, 2016). However, our assessment of the position of *V. emeryi* was better supported (BS = 90 on gene-concatenation tree and BS = 99 on complete mitochondria tree) than that of this previous work (BS = 75). Considering that Duan and collaborators (2016) used a similar approach to ours (gene concatenation under a Maximum Likelihood method), we may conclude that these better results indicate that nucleotidic data presents more reliable information for these clades than amino acidic data.

The position of other myrmicine ants in our gene-concatenation tree are not well resolved, such as the placement of *Myrmica scabrinodis* (BS = 52), *Wasmannia auropunctata* (BS = 43) and *Pristomyrmex punctatus* (BS = 16). However, the position of these species in the amino acid tree of Duan, Peng & Qian (2016) is also inconclusive and differs from here by grouping *W. auropunctata* and *M. scabrinodis* together, a relationship supported only in 35% of the bootstrap replicates. This clade is placed as sister group to *Solenopsis* spp. and *V. emeryi* with a lower support (BS = 21) and *P. punctatus* assumes a more basal position in tree on 46% of the replicates. However, *Atta laevigata* appears at the base of all Myrmecinae with high 100% bootstrap support in the amino acid tree. As the mitogenome of *A. laevigata* available is not complete, it was not used as input for the nucleotide gene-concatenation performed here, as opposed to its congeneric *Atta texana. A. texana* also appears at the base of the Myrmicinae subfamily, but under a low resolution sister group relationship with *M. scabrinodis* (BS = 52).

This clade is sister to all other myrmecine ants (BS = 74). Lastly, the position of *Cardiocondyla obscurior* was also not well supported (BS = 43), but since it is a recently published mitogenome, it was absent in the work of Duan and collaborators.

In both works, mitogenome analyses were not fully capable of resolving important nodes of the myrmicine branch and several factors may be involved in these unsatisfactory results. It is necessary to highlight that Myrmicinae is the most biodiverse ant subfamily (Bolton, 2012) and it is known to feature several dubious monophyletic groups (Brady et al., 2006; Ward, 2011; Ward et al., 2015). This diversity is evidenced by the fact that, despite only nine mitogenomes are available for the group, three different mitochondrial gene arrangements can be observed, suggesting a high rate of mitochondrial evolution in this subfamily.

Also, there have been divergences in the Myrmicinae branch of previous molecular phylogenetic studies attempting to study the Formicidae family (Brady et al., 2006; Moreau et al., 2006). On the other hand, Ward et al. (2015) focuses on the subfamily by reconstructing a large-scale phylogeny using 11 nuclear markers from 251 species sampled across all 25 myrmicine tribes, most of them nonmonophyletic. By using such huge amounts of data covering a great part of Myrmicinae species diversity, they managed to propose a new classification of Myrmicinae consisting of exclusively monophyletic tribes, which also reduced the number of genera that are not monophyletic.

Thus, the hyperdiverse nature of this clade, associated to poor taxon sampling and a possible high rate of mitochondrial genome evolution may have contributed to produce inconclusive results in mitochondrial analyses. Also, even though some relationships were not elucidated by mitochondrial phylogenomics alone, the information provided by the mitogenome has been proven several times to be useful in the study of evolutionary relationships for several taxa, either confirming (Prosdocimi et al., 2012; Finstermeier et al., 2013) or refuting previous phylogenetic hypotheses (Kayal et al., 2015; Uliano-Silva et al., 2016). So, we still recommend the use of mitochondrial data, preferably alongside other markers (i.e., nuclear genes), to increase phylogenetic signal and recapture phylogenies. However, we also believe that mitochondrial data alone will yield better results if we address the shortage of mitogenomes available for this clade and improve mitochondrial taxon coverage. In that sense, results present here are extremely relevant to show that information already available in public databases should be used to obtain such sequences at no additional sequencing costs.

421

422 *No budget mitogenomics*

423

424 The results presented here confirm that both UCE and WGS data publicly available can
425 be used to assemble complete mitochondrial genomes with high coverage (Table 1), which can
426 be explained by the high copy number of mitochondrial genome reads compared to nuclear
427 genomes sequencing reads that may reach something between 0.25% to 0.5% of the total number
428 of bases generated (Prosdocimi et al., 2012), sometimes reaching percentages as high as 2% of
429 reads mapping to mtDNA (Ekblom, Smeds & Ellegren, 2014). We also confirm the potential of
430 UCE data as a low-cost alternative to sequence complete mitogenomes with high coverage as
431 described by Raposo do Amaral et al. (2015). Mitogenome data is used in various types of
432 analyses and mitochondrial sequences are encountered in several types of datasets, normally
433 providing enough information to assemble the entire mitochondrial sequence. This versatility and
434 ubiquity of mitogenome information should be used in favor of biodiversity studies, especially
435 considering the increasingly available public datasets for a great number of species.
436 The potential of these sequences in unveiling phylogenies must not be overlooked,
437 especially if we consider that there are different dataset types available for different species
438 (WGS, RNA-Seq, UCE enrichment, among others). These different resources makes it difficult
439 to achieve an integrated phylogenetic/phylogenomic analyses using the public data, that often
440 depends on sequence orthology to be performed (Kuzniar et al., 2008). Thus, the use of different
441 types of data to assemble the complete or nearly complete mitogenomes for species with publicly
442 available data presents a solution to this problem, with the mitochondrial genome acting as a
443 “normalizing sequence” that allows the comparison of different datasets. For instance, in this
444 work some species had only UCE data publicly available, while others presented standard WGS
445 datasets. Yet, by assembling, annotating and analyzing the complete mitogenome for these
446 species, we were able to broaden our scope and study all of them together. Thus, we suggest that
447 the use of mitogenomes obtained from public data has the potential to become an important
448 source of phylogenetic information. Besides, the study of mitochondrial sequences may be one
449 of the fastest routes towards a high-quality comprehensive species-level tree for hyperdiverse
450 taxa such as insects. Steps have been taken that way, as it can be seen on recent work by Linard
451 et al. (2018), where data mining from Genbank and assembly of metagenomic datasets provided

mitochondrial contigs (>3kbp) for almost 16,000 coleopteran species. This huge amount of data was used to generate the largest phylogenetic tree for the clade.

Studies that attempt to assemble complete mitogenomes using public data are yet scarce whereas the size and breadth of public databases is ever growing, along with its potential to answer phylogenetic questions. No budget mitogenomics represents an unprecedented opportunity to reconstruct and analyze large-scale phylogenies for various groups at different taxa levels, which in its turn may help other evolutionary and conservation biology studies and promote an overall increase on our knowledge about non-model species and their diversity.

Conclusion

Here we assembled and annotated the first 14 mitogenomes for the ant subfamily Pseudomyrmecinae using a pipeline that relies solely on public data from different sources and types, making profit of open-source bioinformatics software. These sequences were used to study synteny, comparative genomics and phylogenomic analyses providing valuable information regarding Pseudomyrmecinae phylogeny and evolution. Mitochondrial data on other ant clades, though limited, were useful in both synteny and phylogenomic analyses to broaden our scope and allow the study other ant groups. The mitochondrial sequences assembled cover a considerable portion of Pseudomyrmecinae biodiversity and will be useful for further evolutionary and conservational studies. This work practically doubles the number of complete ant mitogenomes available at no additional sequencing costs since mitogenome taxon coverage is still lacking for Formicidae and its improvement is absolutely necessary for better resolution and robustness of large scale phylogenies. Based on these results, we emphasize that the ever-increasing breadth of public databases, associated to the possibility of obtaining mitochondrial sequences from different types of sequencing data makes no budget mitogenomics the ideal approach for the study of species diversity and it is possibly the fastest route toward species-level phylogenetic trees.

References

Aird D., Ross MG., Chen W-S., Danielsson M., Fennell T., Russ C., Jaffe DB., Nusbaum C., Gnirke A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology 12:R18. DOI: 10.1186/gb-2011-12-2-r18.

Alikhan N-F., Petty NK., Ben Zakour NL., Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12. DOI: 10.1186/1471-2164-12-402.

Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25:3389–3402. DOI: 10.1093/nar/25.17.3389.

Avise JC. 1994. Molecular Markers, Natural History and Evolution. Boston, MA: Springer US. DOI: 10.1007/978-1-4615-2381-9.

Babbucci M., Basso A., Scupola A., Patarnello T., Negrisolo E. 2014. Is It an Ant or a Butterfly? Convergent Evolution in the Mitochondrial Gene Order of Hymenoptera and Lepidoptera. Genome Biology and Evolution 6:3326–3343. DOI: 10.1093/gbe/evu265.

Berman M., Austin CM., Miller AD. 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 41:1179–1187. DOI: 10.1007/s11033-013-2657-5.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M., Stadler PF. 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69:313–319. DOI: 10.1016/j.ympev.2012.08.023.

Blaimer BB., Brady SG., Schultz TR., Lloyd MW., Fisher BL., Ward PS. 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evolutionary Biology 15. DOI: 10.1186/s12862-015-0552-5

Bolger AM., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

Bolton B. 2012. AntCat. An online catalog of the ants of the world. Available at http://www.antcat.org/ (accessed 22 June 2018).

Bordbari MH., Penedo MCT., Aleman M., Valberg SJ., Mickelson J., Finno CJ. 2017. Deletion of 2.7 kb near HOXD3 in an Arabian horse with occipitoatlantoaxial malformation. Animal Genetics 48:287–294. DOI: 10.1111/age.12531.

Brady SG., Schultz TR., Fisher BL., Ward PS. 2006. Evaluating alternative hypotheses for the early evolution and diversification of ants. Proceedings of the National Academy of Sciences 103:18172–18177. DOI: 10.1073/pnas.0605858103.

Branstetter MG., Longino JT., Ward PS., Faircloth BC. 2017. Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods in Ecology and Evolution 8:768–776. DOI: 10.1111/2041-210X.12742.

Carver T., Harris SR., Berriman M., Parkhill J., McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28:464–469. DOI: 10.1093/bioinformatics/btr703.

Castro LR., Dowton M. 2005. The position of the Hymenoptera within the Holometabola as inferred from the mitochondrial genome of Perga condei (Hymenoptera: Symphyta: Pergidae). Molecular Phylogenetics and Evolution 34:469–479. DOI: 10.1016/j.ympev.2004.11.005.

Cha SY., Yoon HJ., Lee EM., Yoon MH., Hwang JS., Jin BR., Han YS., Kim I. 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, Bombus ignitus (Hymenoptera: Apidae). Gene 392:206–220. DOI: 10.1016/j.gene.2006.12.031.

Chevreux, B., Wetter, T. and Suhai, S., 1999, October. Genome sequence assembly using trace signals and additional sequence information. In German conference on bioinformatics (Vol. 99, No. 1, pp. 45-56).

Chomicki G., Ward PS., Renner SS. 2015. Macroevolutionary assembly of ant/plant symbioses: Pseudomyrmex ants and their ant-housing plants in the Neotropics. Proceedings of the Royal Society B: Biological Sciences 282:20152200. DOI: 10.1098/rspb.2015.2200.

Cochrane G., Bates K., Apweiler R., Tateno Y., Mashima J., Kosuge T., Mizrachi IK., Schafer S., Fetchko M. 2006. Evidence Standards in Experimental and Inferential INSDC Third Party Annotation Data. OMICS: A Journal of Integrative Biology 10:105–113. DOI: 10.1089/omi.2006.10.105.

Crozier RH., Crozier YC. 1993. The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization. Genetics 133:97–117.

Crampton-Platt A., Timmermans MJTN., Gimmel ML., Kutty SN., Cockerill TD., Vun Khen C., Vogler AP. 2015. Soup to Tree: The Phylogeny of Beetles Inferred by Mitochondrial Metagenomics of a Bornean Rainforest Sample. Molecular Biology and Evolution 32:2302–2316. DOI: 10.1093/molbev/msv111.

Del Fabbro C., Scalabrin S., Morgante M., Giorgi FM. 2013. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. PLoS ONE 8:e85024. DOI: 10.1371/journal.pone.0085024.

Dierckxsens N., Mardulyn P., Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research:gkw955. DOI: 10.1093/nar/gkw955.

Diroma MA., Calabrese C., Simone D., Santorsola M., Calabrese FM., Gasparre G., Attimonelli M. 2014. Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. BMC Genomics 15:S2. DOI: 10.1186/1471-2164-15-S3-S2.

Dohm JC., Lottaz C., Borodina T., Himmelbauer H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Research36:e105–e105. DOI: 10.1093/nar/gkn425.

Dowton M., Cameron SL., Dowavic JI., Austin AD., Whiting MF. 2009. Characterization of 67 Mitochondrial tRNA Gene Rearrangements in the Hymenoptera Suggests That Mitochondrial tRNA Gene Position Is Selectively Neutral. Molecular Biology and Evolution 26:1607–1617. DOI: 10.1093/molbev/msp072.

Duan X-Y., Peng X-Y., Qian Z-Q. 2016. The complete mitochondrial genomes of two globally invasive ants, the Argentine ant Linepithema humile and the little fire ant Wasmannia auropunctata. Conservation Genetics Resources 8:275–277. DOI: 10.1007/s12686-016-0555-6.

Ekblom R., Smeds L., Ellegren H. 2014. Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. BMC Genomics15:467. DOI: 10.1186/1471-2164-15-467.

Finstermeier K., Zinner D., Brameier M., Meyer M., Kreuz E., Hofreiter M., Roos C. 2013. A Mitogenomic Phylogeny of Living Primates. PLoS ONE 8:e69504. DOI: 10.1371/journal.pone.0069504.

Gómez-Acevedo S., Rico-Arce L., Delgado-Salinas A., Magallón S., Eguiarte LE. 2010. Neotropical mutualism between Acacia and Pseudomyrmex: Phylogeny and divergence times. Molecular Phylogenetics and Evolution 56:393–408. DOI: 10.1016/j.ympev.2010.03.018.

Goodwin S., McPherson JD., McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17:333–351. DOI: 10.1038/nrg.2016.49.

Gotzek D., Clarke J., Shoemaker D. 2010. Mitochondrial genome evolution in fire ants (Hymenoptera: Formicidae). BMC Evolutionary Biology 10:300. DOI: 10.1186/1471-2148-10-300.

Hahn C., Bachmann L., Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41:e129–e129. DOI: 10.1093/nar/gkt371.

Hasegawa, E., Kobayashi, K., Yagi, N. and Tsuji, K., 2011. Complete mitochondrial genomes of normal and cheater morphs in the parthenogenetic ant Pristomyrmex punctatus (Hymenoptera: Formicidae). Myrmecol News, 15(85), p.90.

Janzen DH. 1966. Coevolution of Mutualism Between Ants and Acacias in Central America. Evolution 20:249. DOI: 10.2307/2406628.

Karsch-Mizrachi, I., Takagi, T., Cochrane, G. and International Nucleotide Sequence Database Collaboration, 2017. The international nucleotide sequence database collaboration. Nucleic acids research, 46(D1), pp.D48-D51. DOI: 10.1093/nar/gkx1097.

Kayal E., Bentlage B., Cartwright P., Yanagihara AA., Lindsay DJ., Hopcroft RR., Collins AG. 2015. Phylogenetic analysis of higher-level relationships within Hydroidolina (Cnidaria: Hydrozoa) using mitochondrial genome data and insight into their mitochondrial transcription. PeerJ 3:e1403. DOI: 10.7717/peerj.1403.

Kim MJ., Hong EJ., Kim I. 2016. Complete mitochondrial genome of Camponotus atrox (Hymenoptera: Formicidae): a new tRNA arrangement in Hymenoptera. Genome 59:59–74. DOI: 10.1139/gen-2015-0080.

Kodama Y., Shumway M., Leinonen R., on behalf of the International Nucleotide Sequence Database Collaboration 2012. The sequence read archive: explosive growth of sequencing data. Nucleic Acids Research 40:D54–D56. DOI: 10.1093/nar/gkr854.

Kumar S., Stecher G., Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33:1870–1874. DOI: 10.1093/molbev/msw054.

Kuzniar A., van Ham RCHJ., Pongor S., Leunissen JAM. 2008. The quest for orthologs: finding the corresponding gene across genomes. Trends in Genetics24:539–551. DOI: 10.1016/j.tig.2008.08.009.

Langmead B., Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359. DOI: 10.1038/nmeth.1923.

Linard B., Crampton-Platt A., Moriniere J., Timmermans MJTN., Andújar C., Arribas P., Miller KE., Lipecki J., Favreau E., Hunter A., Gómez-Rodríguez C., Barton C., Nie R., Gillett CPDT., Breeschoten T., Bocak L., Vogler AP. 2018. The contribution of mitochondrial metagenomics to large-scale data mining and phylogenetic analysis of Coleoptera. Molecular Phylogenetics and Evolution 128:1–11. DOI: 10.1016/j.ympev.2018.07.008.

Liu N., Duan X-Y., Qian Z-Q., Wang X-Y., Li X-L., Ding M-Y. 2016. Characterization of the complete mitochondrial genome of the myrmicine ant Vollenhovia emeryi (Insecta: Hymenoptera: Formicidae). Conservation Genetics Resources 8:211–214. DOI: 10.1007/s12686-016-0535-x.

Margulies M., Egholm M., Altman WE., Attiya S., Bader JS., Bemben LA., Berka J., Braverman MS., Chen Y-J., Chen Z., Dewell SB., Du L., Fierro JM., Gomes XV., Godwin BC., He W., Helgesen S., Ho CH., Irzyk GP., Jando SC., Alenquer MLI., Jarvie TP., Jirage KB., Kim J-B., Knight JR., Lanza JR., Leamon JH., Lefkowitz SM., Lei M., Li J., Lohman KL., Lu H., Makhijani VB., McDade KE., McKenna MP., Myers EW., Nickerson E., Nobile JR., Plant R., Puc BP., Ronan MT., Roth GT., Sarkis GJ., Simons JF., Simpson JW., Srinivasan M., Tartaro KR., Tomasz A., Vogt KA., Volkmer GA., Wang SH., Wang Y., Weiner MP., Yu P., Begley RF., Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380. DOI: 10.1038/nature03959.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends in Genetics 24:133–141. DOI: 10.1016/j.tig.2007.12.007.

Miller MJ., Aguilar C., De León LF., Loaiza JR., McMillan WO. 2016. Complete mitochondrial genomes of the New World jacanas: Jacana spinosa and Jacana jacana. Mitochondrial DNA 27:764–765. DOI: 10.3109/19401736.2014.915530.

Milne I., Stephen G., Bayer M., Cock PJA., Pritchard L., Cardle L., Shaw PD., Marshall D. 2013. Using Tablet for visual exploration of second-generation sequencing data. Briefings in Bioinformatics 14:193–202. DOI: 10.1093/bib/bbs012.

Moritz C. 1994. Applications of mitochondrial DNA analysis in conservation: a critical review. Molecular Ecology 3:401–411. DOI: 10.1111/j.1365-294X.1994.tb00080.x.

Moreau CS. 2006. Phylogeny of the Ants: Diversification in the Age of Angiosperms. Science 312:101–104. DOI: 10.1126/science.1124891.

Oyola SO., Otto TD., Gu Y., Maslen G., Manske M., Campino S., Turner DJ., MacInnis B., Kwiatkowski DP., Swerdlow HP., Quail MA. 2012. Optimizing illumina next-generation sequencing library preparation for extremely at-biased genomes. BMC Genomics 13:1. DOI: 10.1186/1471-2164-13-1.

Picardi E., Pesole G. 2012. Mitochondrial genomes gleaned from human whole-exome sequencing. Nature Methods 9:523–524. DOI: 10.1038/nmeth.2029.

Posada D., Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818. DOI: 10.1093/bioinformatics/14.9.817.

Prosdocimi F., de Carvalho DC., de Almeida RN., Beheregaray LB. 2012. The complete mitochondrial genome of two recently derived species of the fish genus Nannoperca (Perciformes, Percichthyidae). Molecular Biology Reports 39:2767–2772. DOI: 10.1007/s11033-011-1034-5.

Raposo do Amaral F., Neves LG., Resende MFR., Mobili F., Miyaki CY., Pellegrino KCM., Biondo C. 2015. Ultraconserved Elements Sequencing as a Low-Cost Source of Complete Mitochondrial Genomes and Microsatellite Markers in Non-Model Amniotes. PLOS ONE 10:e0138446. DOI: 10.1371/journal.pone.0138446.

Rauch C., Christa G., de Vries J., Woehle C., Gould SB. 2017. Mitochondrial Genome Assemblies of Elysia timida and Elysia cornigera and the Response of Mitochondrion-Associated Metabolism during Starvation. Genome Biology and Evolution 9:1873–1879. DOI: 10.1093/gbe/evx129.

Rodovalho C de M., Lyra ML., Ferro M., Bacci M. 2014. The Mitochondrial Genome of the Leaf-Cutter Ant Atta laevigata: A Mitogenome with a Large Number of Intergenic Spacers. PLoS ONE 9:e97117. DOI: 10.1371/journal.pone.0097117.

Rubin BER., Moreau CS. 2016. Comparative genomics reveals convergent rates of evolution in ant–plant mutualisms. Nature Communications 7:12679. DOI: 10.1038/ncomms12679.

Simpson JT., Wong K., Jackman SD., Schein JE., Jones SJM., Birol I. 2009. ABySS: A parallel assembler for short read sequence data. Genome Research 19:1117–1123. DOI: 10.1101/gr.089532.108.

Smith DR. 2015. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? Briefings in Functional Genomics:elv027. DOI: 10.1093/bfgp/elv027.

Ströher PR., Zarza E., Tsai WLE., McCormack JE., Feitosa RM., Pie MR. 2017. The mitochondrial genome of Octostruma stenognatha and its phylogenetic implications. Insectes Sociaux 64:149–154. DOI: 10.1007/s00040-016-0525-8.

Tang M., Tan M., Meng G., Yang S., Su X., Liu S., Song W., Li Y., Wu Q., Zhang A., Zhou X. 2014. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward

678 biodiversity analysis using mito-metagenomics. Nucleic Acids Research42:e166–e166. DOI:
679 10.1093/nar/gku917.
680 Thompson JD., Gibson TJ., Higgins DG. 2003. Multiple Sequence Alignment Using
681 ClustalW and ClustalX. Current Protocols in Bioinformatics 00:2.3.1-2.3.22. DOI:
682 10.1002/0471250953.bi0203s00.
683 Uliano-Silva M., Americo JA., Costa I., Schomaker-Bastos A., de Freitas Rebelo M.,
684 Prosdocimi F. 2016. The complete mitochondrial genome of the golden mussel Limnoperna
685 fortunei and comparative mitogenomics of Mytilidae. Gene 577:202–208. DOI:
686 10.1016/j.gene.2015.11.043.
687 Timmermans MJTN., Viberg C., Martin G., Hopkins K., Vogler AP. 2016. Rapid
688 assembly of taxonomically validated mitochondrial genomes from historical insect
689 collections. Biological Journal of the Linnean Society 117:83–95. DOI: 10.1111/bij.12552.
690 van Dijk EL., Auger H., Jaszczyszyn Y., Thermes C. 2014. Ten years of next-generation
691 sequencing technology. Trends in Genetics 30:418–426. DOI: 10.1016/j.tig.2014.07.001.
692 Ward, P.S., 1989. Systematic studies on pseudomyrmecine ants: revision of the
693 Pseudomyrmex oculatus and P. subtilissimus species groups, with taxonomic comments on other
694 species. Quaestiones Entomologicae, 25(4), pp.393-468.
695 Ward, P.S., 1991. Phylogenetic analysis of pseudomyrmecine ants associated with
696 domatia-bearing plants. Ant-plant interactions. Oxford University Press, Oxford, pp.335-352.
697 Ward, P.S., 1993. Systematic studies on Pseudomyrmex acacia-ants (Hymenoptera:
698 Formicidae: Pseudomyrmecinae). Journal of Hymenoptera Research, 2, pp.117-168.
699 Ward P. 1999. Systematics, biogeography and host plant associations of the
700 Pseudomyrmex viduus group (Hymenoptera: Formicidae), Triplaris - and Tachigali -inhabiting
701 ants. Zoological Journal of the Linnean Society 126:451–540. DOI: 10.1006/zjls.1998.0158.
702 Ward, P.S., 2011. Integrating molecular phylogenetic results into ant taxonomy
703 (Hymenoptera: Formicidae). Myrmecological News, 15, pp.21-29.
704 Ward PS. 2017. A review of the Pseudomyrmex ferrugineus and Pseudomyrmex goeldii
705 species groups: acacia-ants and relatives (Hymenoptera: Formicidae). Zootaxa 4227:524. DOI:
706 10.11646/zootaxa.4227.4.3.
707 Ward PS., Brady SG., Fisher BL., Schultz TR. 2015. The evolution of myrmicine ants:
708 phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae): Phylogeny
709 and evolution of myrmicine ants. Systematic Entomology40:61–81. DOI: 10.1111/syen.12090.
710 Ward PS., Branstetter MG. 2017. The acacia ants revisited: convergent evolution and
711 biogeographic context in an iconic ant/plant mutualism. Proceedings of the Royal Society B:
712 Biological Sciences 284:20162569. DOI: 10.1098/rspb.2016.2569.
713 Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and
714 evolution. International Review of Cytology 141:173–216.
715 Yang S., Li X., Cai L-G., Qian Z-Q. 2015. Characterization of the complete
716 mitochondrial genome of Formica selysi (Insecta: Hymenoptera: Formicidae:
717 Formicinae). Mitochondrial DNA:1–3. DOI: 10.3109/19401736.2015.1018229.

Table 1(on next page)

Information about mitochondrial genome assemblies of Pseudomyrmecinae

1

Pseudomyrmecinae Species	Species group	*Mitogenome TPA accession number*	MITObim third assembly round needed	Mitogenome Coverage	Low coverage Region	Mitogenome Size (bp)	GC content(%)
P. concolor	*P. viidus*	*BK010475*	No	193.2x	No	15906	24.5
P. dendroicus	*P. viidus*	*BK010473*	Yes	123.9x	No	17362	19.4
P. elongatus	*P. oculatus*	*BK010474*	No	115.4x	No	17304	22.1
P. feralis	*P. ferrugineus*	*BK010379*	No	128.0x	No	18835	21.9
P. ferrugineus	*P. ferrugineus*	*BK010380*	Yes	87.0x	No	18480	22.8
P. flavicornis	*P. ferrugineus*	*BK010381*	Yes	152.7x	No	18498	23.2
P. gracilis	*P. gracilis*	*BK010472*	No	165.5x	13761-13928	15704	23.4
P. janzeni	*P. ferrugineus*	*BK010382*	No	125.8x	15848-15867	18380	23.4
P. pallidus	*P. pallidus*	*BK010383*	No	91.9x	No	17117	25.5
P. particeps	*P. ferrugineus*	*BK010384*	No	126.8x	15799-15820	18524	20.2
P. peperi	*P. ferrugineus*	*BK010385*	Yes	87.4x	16006-16023	18709	22.4
P. veneficus	*P. ferrugineus*	*BK010386*	No	155.4x	15889-15928	18410	20.6
T. aethiops	NE	*BK010476*	Yes	712.9x	13934-13982	15988	21.3
T. rufonigra	NE	*BK010387*	No	292.2x	13889-13982	15907	25.9

2

Figure 1

Comparative genomics analysis of all 14 Pseudomyrmecinae ants

BLAST comparison of all Pseudomyrmecinae mitochondrial genomes against a reference (*Pseudomyrmex janzeni*) generated by Blast Ring Image Generator (BRIG). Gaps in rings correspond to regions with less than 50% identity to the reference sequence. Most mitochondrial features are conserved within the clade, even though ATP8 and some tRNAs (trn-S, trn-E and trn-T) were observed to be less conserved. Four regions (identified by arrows) present nucleotide size variations between (i) *COII* and *trn-K*; (ii) *ATP8* and *ATP6*; (iii) *trn-N* and *trn-F* and; (iv) *trn-W* and *COI* .

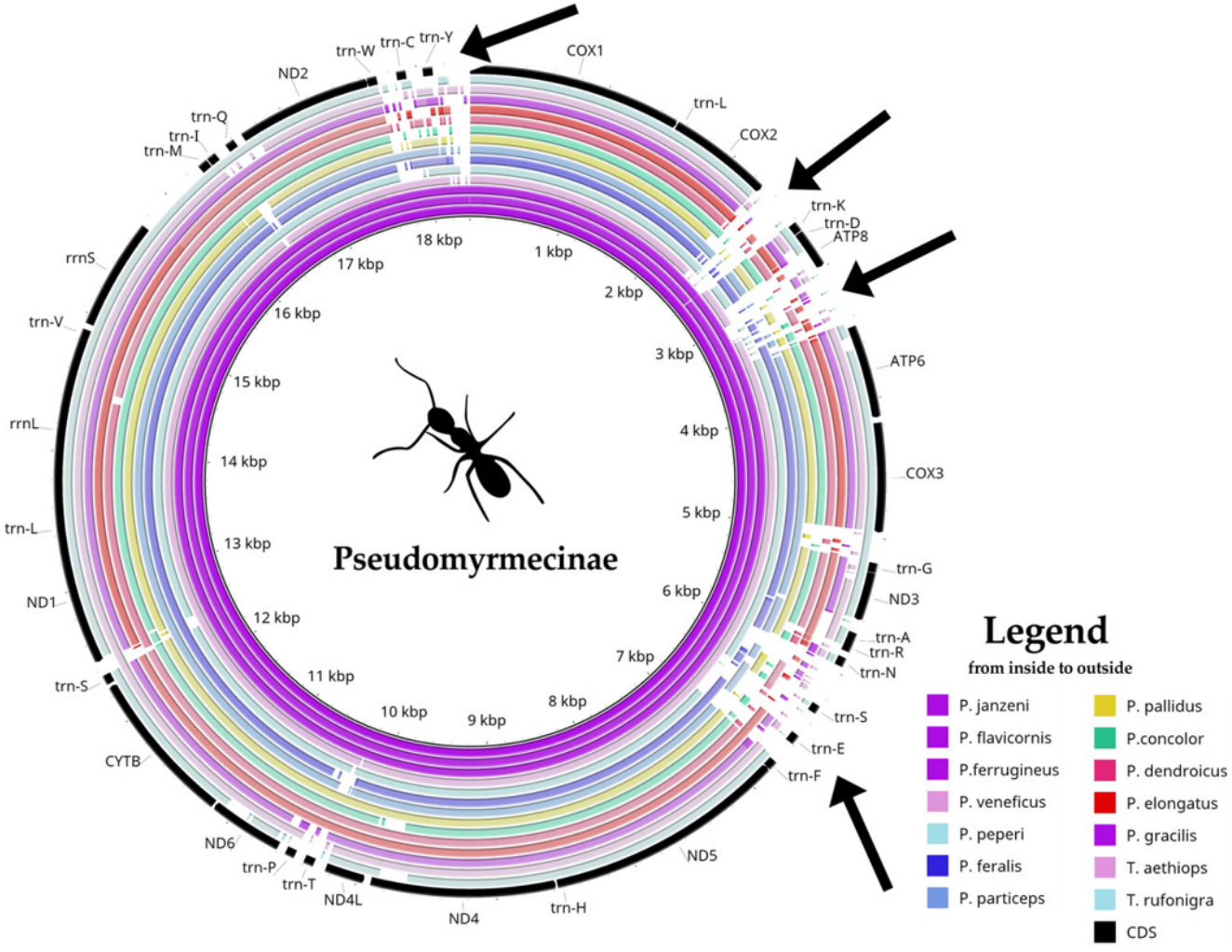

Figure 2

Synteny of all complete Formicidae mitogenomes available on Genbank

The gene arrangements inside the horizontal rectangle are present in most species analyzed and believed to be an ancestral feature for their clades, while the ones outside correspond to unique gene orders encountered in single species. Vertical rectangles and arrows indicate regions where synteny changes occurred and the asterisk (*) and arrow in the *trn-Y* of *W. auropunctata* indicates that it is the only feature in Formicinae mitochondria that changed its coding strand and transcription direction.

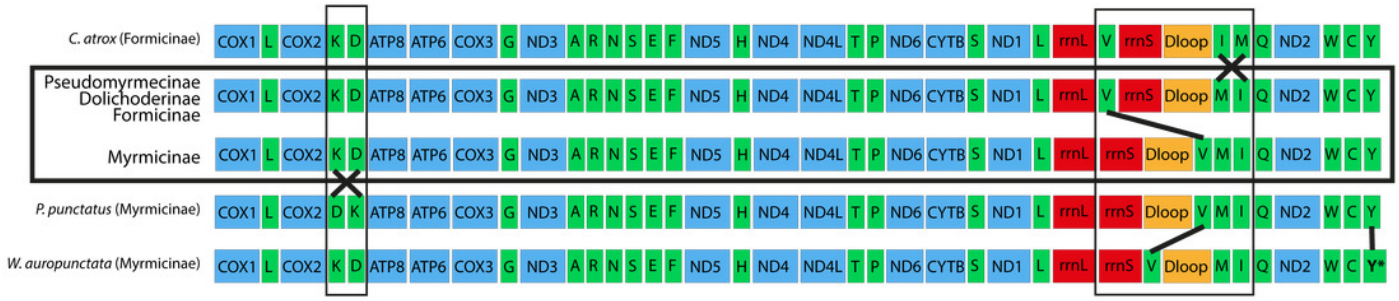

Figure 3

Gene-concatenation phylogenomic tree for all Formicidae complete mitogenomes available on Genbank

The tree was built using the aligned and concatenated nucleotidic sequences for all 13 protein-coding mitochondrial genes. Modeltest identified 'GTR+G+I' as the most adequate substitution model and phylogeny was reconstructed by Maximum Likelihood using MEGA7 software, with 1000 bootstrap replicates. Bees from the Apidae family were used as outgroup. *Pseudomyrmex* species groups are described and mutualistic pseudomyrmecines are evidenced by the presence of an asterisk "*".

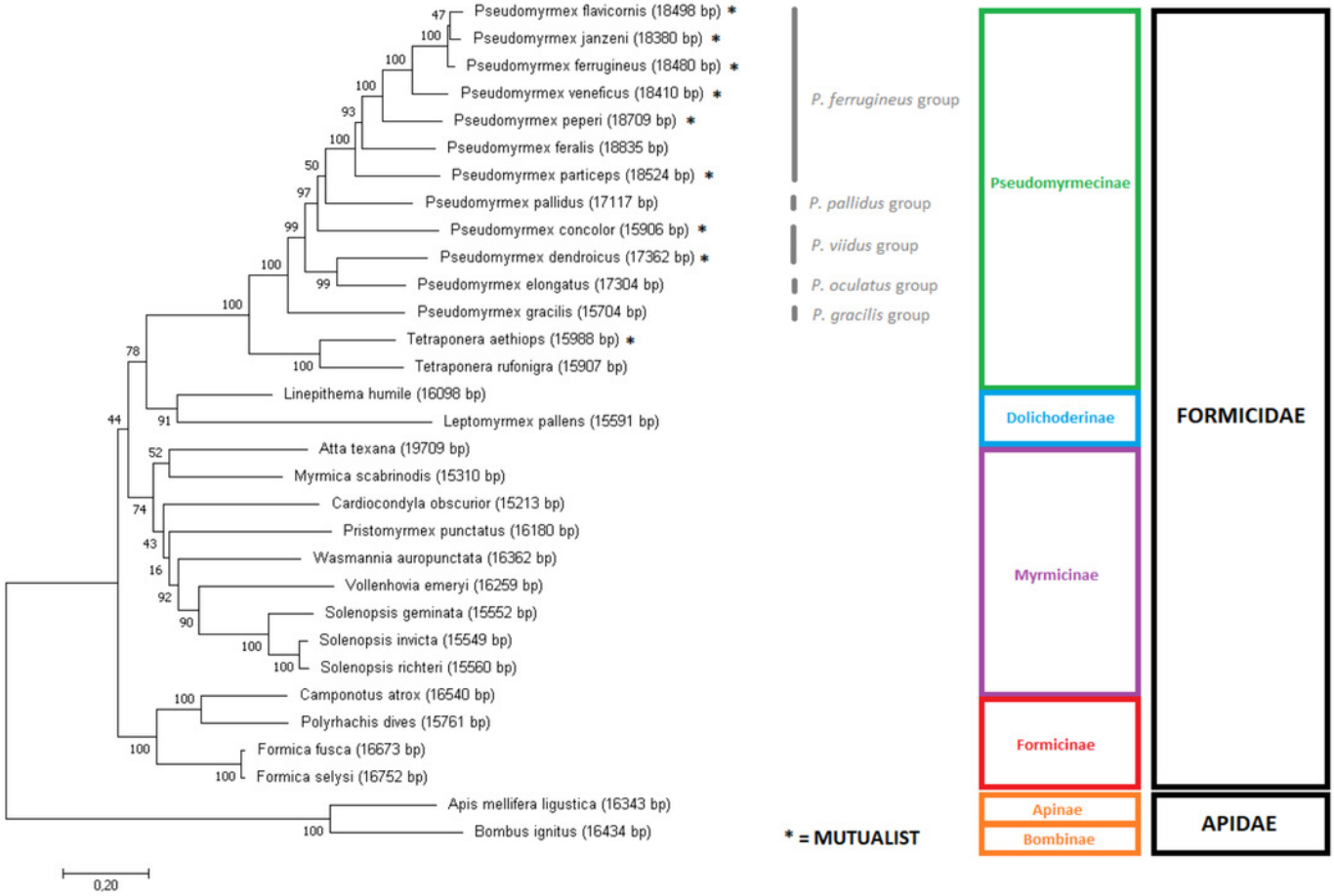

Figure 4

Phylogenetic tree using the complete mitochondrial sequence of all complete ant mitogenomes available on Genbank

'GTR+G+I' was chosen as substitution model as suggested by Modeltest. The tree was built with MEGA7 using Maximum Likelihood with 1000 bootstrap replicates. Mitogenomes from bees were used as outgroups. *Pseudomyrmex* species groups and mutualistic pseudomyrmecines are evidenced.