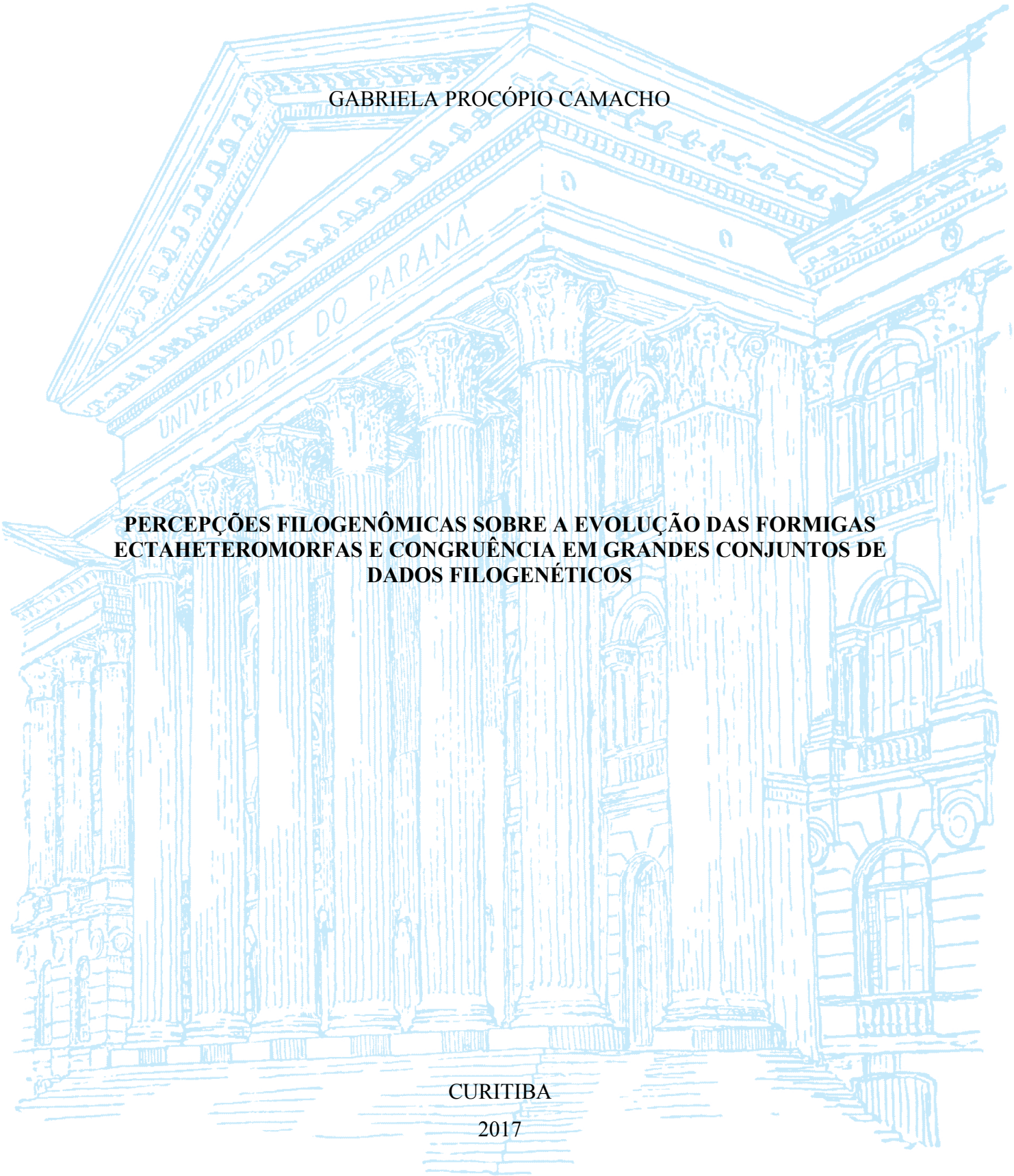
UNIVERSIDADE FEDERAL DO PARANÁ

GABRIELA PROCÓPIO CAMACHO



PERCEPÇÕES FILOGENÔMICAS SOBRE A EVOLUÇÃO DAS FORMIGAS ECTAHETEROMORFAS E CONGRUÊNCIA EM GRANDES CONJUNTOS DE DADOS FILOGENÉTICOS

CURITIBA

2017

GABRIELA PROCÓPIO CAMACHO

PERCEPÇÕES FILOGENÔMICAS SOBRE A EVOLUÇÃO DAS FORMIGAS ECTAHETEROMORFAS E CONGRUÊNCIA EM GRANDES CONJUNTOS DE DADOS FILOGENÉTICOS

Tese apresentada como requisito parcial à obtenção do grau de Doutora em Ciências Biológicas (Entomologia), no Curso de Pós-Graduação em Entomologia, Setor de Ciências Biológicas, Departamento de Zoologia, da Universidade Federal do Paraná.

Orientador: Prof. Dr. Rodrigo dos S. Machado Feitosa

Coorientador: Prof. Dr. Márcio Roberto Pie

CURITIBA

2017

MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS BIOLÓGICAS
(ENTOMOLOGIA)

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS BIOLÓGICAS (ENTOMOLOGIA) da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **GABRIELA PROCOPIO CAMACHO** intitulada: **"Percepções filogenômicas sobre a evolução das formigas Ectaheteromorfas e congruência em grandes conjuntos de dados filogenéticos"**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua Aprovação no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 22 de Dezembro de 2017.

RODRIGO DOS SANTOS MACHADO FEITOSA
Presidente da Banca Examinadora

JOHN EDWIN LATTKE BRAVO
Avaliador Externo

PAULO LABIAK EVANGELISTA
Avaliador Externo

ALINE CRISTINA MARTINS
Avaliador Externo

EDUARDO ANDRADE BOTELHO DE ALMEIDA
Avaliador Externo

Universidade Federal do Paraná
Sistema de Bibliotecas

Camacho, Gabriela Procópio
Percepções filogenômicas sobre a evolução das formigas ectaheteromorfas e congruência em grandes conjuntos de dados filogenéticos. / Gabriela Procópio Camacho. – Curitiba, 2017.
98 f.: il. ; 30cm.

Orientador: Rodrigo dos S. Machado Feitosa
Co-orientador: Márcio Roberto Pie

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Entomologia.

1. Formiga 2. Genômica 3. Filogenia I. Título II. Feitosa, Rodrigo dos Santos Machado III. Pie, Márcio Roberto IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Entomologia.

CDD (20. ed.) 595.796

Aos meus pais, Dorinha e Kennedy, que
me trouxeram de mãos dadas até aqui.

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, Dr. Rodrigo Feitosa pela confiança e pelas oportunidades incríveis que me deu de crescer e prosperar. Muito além de um orientador, ele é um amigo, um compasso moral e ético, um exemplo de como a simpatia e seriedade podem nos levar longe, de humildade e de generosidade. Devo a ele não só este trabalho, mas toda uma carreira que espero começar aqui.

Agradeço também ao meu coorientador, Dr. Márcio Pie, que sempre esteve disponível e acessível em todos os momentos, com suas ideias incríveis e excelentes discussões.

Mais que tudo, agradeço à minha família que, apesar de muitas vezes sofrerem com as ausências que essa carreira exige, sempre entenderam e apoiaram a minha paixão e foram muito mais que companheiros, foram pilares na construção de cada um dos meus sonhos. Meus pais, Dorinha e Kennedy, que sempre me fizeram acreditar que eu poderia ser o que eu quisesse ser, são os responsáveis pela minha coragem de enfrentar os obstáculos, de perseguir os meus sonhos e por conquistar os meus próprios caminhos. A eles eu devo cada palavra escrita aqui, cada vitória (e cada vez que eu me levantei depois de uma derrota) e toda a minha vida. Meu irmão, que sempre foi um exemplo de força e verdade, sempre esteve presente quando eu precisei dele, mesmo quando eu fui menos presente do que ele merecia. A ele eu devo uma amizade da vida toda e levo comigo uma admiração sem igual por quem ele é. Seu amor esteve comigo em cada passo do caminho e estará comigo sempre. Meus avós, João Francisco e Ném e Eliana e José são os exemplos de força, seriedade, compromisso, carinho e perseverança que sempre me fizeram ter certeza de que, aonde quer que eu vá, eu nunca estarei sozinha. Aos meus padrinhos, Dalbert, Marta e Márcia, a quem o amor e a devoção eu nunca serei capaz de agradecer o suficiente, agradeço por tudo o suporte e carinho durante essa caminhada. Aos meus tios, Guida, Cidinha, Mércia, Francisco, Madra, Juarez, Zilda, Hebert, Vanelli, Virgínia, Alírio, Beth, Hamilton, Paulo, Lindamar, Eduardo, Celi, Sônia, Nacip, meus primos João Vinícius, Júnior, Andrey, Andrezza, Érica, Yara, Flávia, Arthur, Matheus, Bernardo, Luciana, Leonardo, Ana Carolina, Glênio, Lizandra e Camila e a toda a nossa família, muito obrigada pela compreensão, pelo apoio e pelo carinho ao longo dos anos. A todos eu gostaria de dizer que, apesar das ausências, das distâncias ou das divergências, vocês sempre

foram e sempre serão o meu porto seguro, meu exemplo, minha motivação, meu orgulho e meu maior amor.

Agradeço mais do que profundamente aos colegas do Laboratório de Biologia e Sistemática de Formigas da UFPR, nosso FeitosaLab, pelos momentos maravilhosos que passamos juntos, pelo aprendizado na ciência e na vida, pelas conversas engrandecedoras e pelo carinho e amizade de sempre. Aline, Paloma, Thiago, Weslly, Sebastian, Natália, Juliana, Alexandre, Mila, Tainara, Isabela, Jaqueline, Yohan, Yasmin, Ana Carolina, foi um prazer trabalhar com vocês e será sempre um prazer poder chamar vocês de amigos.

Agradeço também aos colegas do PieLab, Patrícia, Deborah, André, Ricardo e Raquel, que sempre dividiram comigo todo seu conhecimento e foram insubstituíveis durante a construção deste trabalho. Agradeço ao Dr. John Lattke, por dividir comigo seu amplo conhecimento sobre a taxonomia dos grupos, e ao Mayron, por ser um ótimo amigo e companheiro de trabalho. Aos colegas do Programa de Pós-Graduação em Entomologia da UFPR pela companhia e bons momentos. Ao docentes e funcionários do Programa de Pós-Graduação em Entomologia da UFPR pela importante participação nesta etapa da minha formação. Ao Dr. Jack Longino e Dr. Michael Branstetter, pela disponibilização de amostras, de estrutura laboratorial e ajuda em laboratório para aquisição dos UCE's. Ao Dr. Ted Schultz e aos membros do Smithsonian AntLab, Ana, Bonnie, Fred, Dietrich, Mike, Eugenia, Jeffrey, pelo apoio, incentivo e amizade. Ao Dr. Scott Powell e aos membros do Powell lab, Jignasha, Flávio e Carol, por me acolherem e me presentearem com uma amizade sem valor e excelentes momentos.

Tenho muito a agradecer também aos queridos amigos Maíra, Dalana, Zhé, Renata, Heraldo, Bernando, Babi, João, Larissa e Lívia pela amizade, suporte, apoio e conselhos não só nesta jornada, mas em todas as outras.

Por fim, agradeço ao povo brasileiro e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela bolsa concedida durante estes dois anos e ao Programa de Pós Graduação em Entomologia pela estrutura necessária para realização do trabalho.

"A vida não é fácil para nenhum de nós. Mas, e daí? Nós devemos ter perseverança e, acima de tudo, confiança em nós mesmos. Temos que acreditar que temos talento para algo e que este algo deve ser alcançado. Não há o que temer na vida, apenas o que entender. "

Marie Curie

RESUMO

Percepções filogenômicas sobre a evolução das formigas ectaheteromorfas e congruência em grandes conjuntos de dados filogenéticos

Entre os diferentes métodos que utilizam dados genômicos para o estudo da história evolutiva dos grupos recentes, os elementos ultraconservados (UCE's) provaram ser marcadores ideais para inferências em diferentes escalas de tempo. A taxonomia e a filogenia interna de Formicidae têm sido exaustivamente estudadas e têm se estabilizado significativamente nas últimas décadas devido ao estudo ativo da sistemática de formigas em escala macroevolutiva. As subfamílias Ectatomminae e Heteroponerinae, referidas atualmente como ectaheteromorfas, no entanto, nunca tiveram sua história evolutiva extensivamente estudada através de dados moleculares. Dessa forma, no presente estudo, usamos os elementos ultraconservados para reconstruir a filogenia das ectaheteromorfas, com o intuito de permitir a compreensão das relações internas do grupo, bem como elucidar os padrões biogeográficos dos gêneros, além de auxiliar na delimitação das tribos e gêneros internos às subfamílias. Para isso, amostramos um conjunto de 1629 UCE's através do enriquecimento de sequências alvo e sequenciamento multiplexado para 145 espécies que compõem as subfamílias Ectatomminae e Heteroponerinae. Obtivemos uma filogenia altamente resolvida e fortemente suportada para o grupo, recuperando ambas as subfamílias como monofiléticas. Os gêneros *Acanthoponera*, *Rhytidoponera*, *Typhlomyrmex* e *Ectatomma* também aparecem como monofiléticos em nossas análises. *Gnamptogenys* é recuperado como parafilético em relação a *Typhlomyrmex*, enquanto *Heteroponera* é parafilético em relação a *Acanthoponera*. Todos os gêneros parecem ter uma origem Neotropical, com exceção de *Rhytidoponera*, que tem origem australiana. As subfamílias se originaram no Cretáceo, apresentando sua maior diversificação durante o Mioceno e o Eoceno. Neste trabalho, propomos a sinonímia de Heteroponerinae sob Ectatomminae, com o intuito de obter uma classificação estável que, além de manter o grupo monofilético, seja suportada por sinapomorfias morfológicas que possibilitem seu fácil reconhecimento. Ao combinar métodos de inferência filogenética com datação e análises biogeográficas, fornecemos uma estrutura filogenética para o conhecimento da história evolutiva destas formigas. Estes resultados também fornecem as bases para uma classificação revisada para as subfamílias.

Adicionalmente, apesar dos ótimos resultados apresentados por estudos filogenômicos, uma suposição tácita de muitos é que os grandes problemas em filogenética molecular estariam sendo resolvidos através de "força bruta", isto é, com a implementação massiva de dados em análises moleculares. No entanto, o fato de que diferentes conjuntos de dados filogenômicos podem levar a resultados altamente suportados, porém diferentes, sugere que a inconsistência entre genes em conjuntos de dados reais pode se provar um grave problema. Para testar a robustez e confiabilidade de resultados obtidos através de grandes conjuntos de dados filogenéticos, nós analisamos a variação topológica em árvores de genes para fornecer uma exploração detalhada da variação e a inconsistência presentes em conjuntos de dados filogenômicos. Para tal, utilizamos os dados de Johnson et al. (2013) e sua hipótese para a posição de formigas em Aculeata como estudo de caso. Nós encontramos um alto grau de incongruência entre o sinal filogenético das árvores de genes, mas este sinal não é enviesado em relação a nenhuma topologia em particular. Além disso, encontramos evidências para um suporte emergente na recuperação de alguns dos nós, enquanto a relação entre formigas, vespas esfeciformes e abelhas não apresenta um forte suporte emergente, de modo que todos os genes do conjunto de dados são necessários para que essa relação seja recuperada. Nossos resultados sugerem que, embora centenas de loci nem sempre sejam essenciais para uma inferência topológica precisa, o uso de grandes conjuntos de dados genômicos para a inferência filogenética pode fornecer resultados mais robustos. Como o número de loci é grande, é possível recuperar genes suficientes que produzam árvores genéticas confiáveis, e a probabilidade de gerar árvores de genes com sinal filogenético tendencioso é menor. No entanto, essa robustez não é concedida apenas pelo grande volume de dados, mas principalmente pela presença de um suporte emergente, em um cenário em que os dados adicionais ampliam desproporcionalmente o sinal filogenético que suporta os nós finais.

Palavras-chave: Filogenômica, Elementos Ultra-Conservados, Formicidae, Aculeata, Sistemática filogenética, Taxonomia, Ectatomminae, Heteroponerinae, Formigas

ABSTRACT

Phylogenomic insights into the evolution of ectaheteromorph ants and congruence in big-data phylogenetics

Among the different methods that use genomic data to study the evolutionary history of recent groups, the ultraconserved elements (UCEs) have proven to be ideal markers for such inferences at different time scales. The taxonomy and internal phylogeny of Formicidae have been extensively studied and have stabilized significantly in the last decades due to the active study of ants systematic on a macroevolutionary scale. The subfamilies Ectatomminae and Heteroponerinae, currently referred to as ectaheteromorphs, however, have never had their evolutionary history extensively studied through molecular data. Thus, in the present study, we used the ultraconserved elements to reconstruct the phylogeny of ectaheteromorphs, in order to allow the understanding of the internal relations of the group, as well as to elucidate the biogeographic patterns of the genera, and assist in the delimitation of the tribes and genera within subfamilies. For this, we sampled a set of 1629 UCE's using target enrichment and multiplexing sequencing of 145 species that make up the subfamilies Ectatomminae and Heteroponerinae. We recovered a highly resolved and strongly supported phylogeny for the group, recovering both subfamilies as monophyletic. The genera *Acanthoponera*, *Rhytidoponera*, *Typhlomyrmex* and *Ectatomma* also appear as monophyletic in our analyzes. *Gnamptogenys* is recovered as paraphyletic in relation to *Typhlomyrmex*, while *Heteroponera* is paraphyletic in relation to *Acanthoponera*. All genera appear to have a Neotropical origin, with the exception of *Rhytidoponera*, which is has an Australasian origin. The subfamilies had their origin during the Cretaceous period, showing their greater diversification during the Miocene and the Eocene. In this work, we propose the synonymy of Heteroponerinae under Ectatomminae, with the intention of obtaining a stable classification that, besides maintaining the group monophyletic, be supported by morphological synapomorphies that allow its easy recognition. By combining phylogenetic inference methods with dating and biogeographic analysis, we provide a phylogenetic framework for the knowledge of the evolutionary history of these ants. These results also provide the basis for a revised classification of the subfamily.

In addition, despite the excellent results recovered by phylogenetic studies, a tacit assumption by many is that the major problems in molecular phylogenetics would be solved by "brute force", that is, by the simple addition of large amounts of molecular data. However, the fact that different phylogenetic data sets may lead to highly supported but different results suggests that the inconsistency between genes in real data sets may prove to be a serious problem. To test the robustness and reliability of results obtained through large phylogenetic datasets, we explored the topological variation in gene trees to provide a detailed exploration of the variation and inconsistency present in phylogenetic data sets, using data from Johnson et al. (2013) and their hypothesis for the position of ants in Aculeata as a case study. We found a high degree of incongruity between the phylogenetic signal of the gene trees, but this signal is not bias in relation to any particular topology. In addition, we found evidence for an emerging support for some of the nodes, while the relationship between ants, spheciform wasps and bees does not appear to have a strong emergent support, requiring all genes from the data set to be recovered. Our results suggest that although hundreds of loci are not always essential for accurate topological inference, the use of large sets of genomic data for phylogenetic inference may provide more robust results. With a large number of loci, it is possible to retrieve enough genes that produce reliable gene trees, and the probability of generating gene trees with a biased phylogenetic signal is lower. However, this robustness is not only due to the large volume of data, but mainly due to the presence of an emerging support, where the additional data disproportionately amplifies the phylogenetic signal that supports the final nodes.

Key-words: Phylogenomics, Ultra-Conserved Elements, Formicidae, Aculeata, Phylogenetic Systematics, Taxonomy, Ectatomminae, Heteroponerinae, Ants.

SUMÁRIO

Introdução geral **14**
Capítulo 1: A genomic perspective for Ectaheteromorph ants with the use of ultra-conserved elements **17**
Introduction 17
Methods 19
Taxon sampling *19*
Molecular data collection *20*
Processing and alignment of UCE data *21*
Phylogenetic inference *22*
Dating *23*
State reconstruction analysis *23*
Results 24
UCE capture statistics *24*
Phylogenetic results *24*
Taxonomic results *27*
Taxonomic account *28*
Dating and State reconstruction results *29*
Discussion 32
Phylogenomics resolves the relationships among Ectatomminae genera *32*
Taxonomy of Ectatomminae revisited *33*
Ectatomminae biogeography *37*
Conclusions 38
References 40
Capítulo 2: Exploring Gene Tree Incongruence at the Root of the Aculeata Tree (Hymenoptera) **47**
Introduction 47
Material and methods 49
Data set *49*
Gene tree and species tree estimation *50*
Evaluation of individual genes support *50*
Evaluation of emergent support *51*
Results 53
Species tree estimation *53*
Evaluation of individual genes support *53*
Evaluation of emergent support *56*
Discussion 58
Conclusions 63
References 65
Referências Bibliográficas **70**
Anexo I **81**
Supplementary material S1. Summary of UCE capture statistics 81
Anexo II **93**
Supplementary material S2. Taxon distribution matrix used in the biogeographic analyses. 93
Anexo III **96**
Supplementary material S3. Phylogenetic trees from analyses not illustrated in the main text. A) RAxML bootstrap tree for 90p-matrix data set; B) RAxML bootstrap tree for 95p-matrix; C) Species tree with local posterior probability values for the 1603 UCE loci in the 90p-pruned data set. 96

Introdução geral

O advento do sequenciamento de terceira geração permitiu o acúmulo de dados genéticos a uma taxa sem precedentes, oferecendo oportunidades interessantes em uma variedade de disciplinas, incluindo a genética populacional, filogenética e filogenética molecular (Carstens and Lemmon 2012; Puritz et al. 2012; McCormack et al. 2013; Ruane et al. 2015; Romiguier et al. 2016; Welch et al. 2016). A análise de extensos conjuntos de dados multi-locus possibilitou a resolução de problemas complexos em filogenética zoológica, como as relações basais entre linhagens de pássaros (McCormack et al. 2013), a filogenia de mamíferos placentários (Mccormack et al. 2012), ou a posição filogenética das tartarugas nos tetrápodes (Crawford et al. 2012). Para formigas, dentre as 17 subfamílias reconhecidas, onze já foram sistematicamente estudadas de forma ampla com base em dados moleculares, representando 94% de toda a diversidade conhecida para a família. No entanto, a maioria destes estudos limita-se a analisar um pequeno número de genes mitocondriais e nucleares que amostram uma pequena porção do genoma. A filogenômica, por outro lado, utiliza o maior número de genes possível, ou mesmo o genoma completo de um indivíduo, para fornecer inferências filogenéticas e, atualmente, tem se mostrado o método mais eficaz para recuperar a história evolutiva de um grupo particular (Johnson et al. 2013; Blaimer et al. 2015; Branstetter et al. 2017a). Esta abordagem aumenta o número de caracteres que podem ser usados em análises filogenéticas em centenas ou milhares de vezes, reduzindo problemas de amostragem baixa para inferência filogenética (Delsuc et al. 2005).

Entre os diferentes métodos que utilizam dados genômicos, os elementos ultraconservados (doravante denominados UCE's) provaram ser marcadores ideais para o estudo de relações evolutivas em diferentes escalas de tempo (Faircloth et al. 2015). O enriquecimento dos loci UCE tem sido utilizado para investigar questões envolvendo divergências filogenéticas mais antigas para vários grupos de vertebrados (Crawford et al. 2012; Faircloth et al. 2013; McCormack et al. 2013), insetos (Faircloth et al. 2015) e, mais recentemente, formigas (Blaimer et al. 2015; Blaimer et al. 2016a, Branstetter et al. 2017a). A técnica também é útil para a compreensão das relações ao nível da população, incluindo divergências recentes (Smith et al. 2013). O emprego de UCE's para inferências filogenéticas tem diversos atributos positivos, pois se mostra efetivo mesmo com o uso de espécimes subotimamente preservados e DNA degradado (Blaimer et al. 2016b), além do baixo custo para sua aquisição (Branstetter et al. 2017a). A taxonomia e a filogenia interna de Formicidae têm sido exaustivamente estudadas e têm se estabilizado significativamente nas últimas décadas devido ao estudo ativo da sistemática de formigas em escala macroevolutiva (Baroni Urbani et al. 1992; Bolton 1995; Brady et al. 2006; Moreau et al. 2006; Ward 2014).

Muitas das descobertas provenientes destes estudos mostraram uma grande congruência entre as hipóteses morfológicas e moleculares existentes, porém, nem sempre essa congruência se mostra infalível. As subfamílias Ectatomminae e Heteroponerinae, representantes da linhagem das ectaheteromorfas, por exemplo, apesar de apresentarem morfologia e comportamento semelhante aos das linhagens basais de formigas (ditas poneroides) (Bolton 2003), aparecem de fato como parte do clado mais apical entre as formigas, o clado formicoide (Brady et al. 2006; Moreau et al. 2006; Ouellette et al. 2006), de acordo com os dados moleculares.

Apesar dos ótimos resultados apresentados por tais estudos filogenômicos, e em parte devido às incongruências entre hipóteses geradas através da morfologia e dados moleculares, uma suposição tácita de muitos é que os grandes problemas em filogenética molecular estariam sendo resolvidos através de "força bruta", isto é, com a implementação massiva de dados em análises moleculares. Por exemplo, estudos usando apenas poucos loci frequentemente apresentam alguns nós com suporte relativamente baixo, que são seguidos pela sugestão do autor de que mais dados são necessários para resolver esse problema específico. No entanto, o fato de que diferentes conjuntos de dados filogenômicos podem levar a resultados altamente suportados, mas diferentes, sugere que a inconsistência entre genes em conjuntos de dados reais pode se provar um grave problema. Por exemplo, a posição das formigas entre os himenópteros aculeados tem sido alvo de diferentes estudos filogenômicos, que sugerem uma relação de grupo-irmão entre Formicidae e Apoidea (abelhas e vespas esfeciformes) (Johnson et al. 2013; Branstetter et al. 2017b; Peters et al. 2017). No entanto, embora as relações entre formigas, abelhas e vespas pareçam relativamente bem resolvidas agora, outros estudos apresentaram resultados conflitantes no passado (Brothers 1990; Heraty et al. 2008; Pilgrim et al. 2011; Faircloth et al. 2014) e a causa e o grau dessas fontes conflitantes de sinal filogenético são atualmente desconhecidos. Dado que os conjuntos de dados são bastante extensos, não se pode simplesmente atribuir essa inconsistência a um ruído nas análises.

No presente estudo, usamos os elementos ultraconservados para reconstruir a filogenia das ectaheteromorfas, com o intuito de permitir a compreensão das relações internas do grupo, bem como elucidar os padrões biogeográficos dos gêneros, além de auxiliar na delimitação das tribos e gêneros internos às subfamílias. Para isso, nós amostramos um conjunto de 1629 UCE’s através do enriquecimento de sequências alvo e sequenciamento multiplexado para 145 espécies que compõem as subfamílias Ectatomminae e Heteroponerinae. Ao combinar métodos de inferência filogenética com datação e análises biogeográficas, fornecemos uma estrutura filogenética para o conhecimento da história evolutiva destas formigas. Estes resultados também fornecem as bases

para uma classificação revisada para as subfamílias. Além disso, analisamos a variação topológica em árvores de genes para fornecer uma exploração detalhada da variação e a inconsistência presentes em conjuntos de dados filogenômicos. Para tal, utilizamos os dados de Johnson et al. (2013) e sua hipótese para a posição de formigas em Aculeata como estudo de caso. Dessa forma, podemos entender de que maneira os dados filogenômicos recuperam dadas relações, visando elucidar a robustez e confiabilidade dos resultados obtidos através de grandes conjuntos de dados filogenéticos.

Capítulo 1: A genomic perspective for Ectaheteromorph ants with the use of ultra-conserved elements.

Introduction

Ants are a globally diverse and dominant lineage of eusocial aculeate and represent one of the great success stories of evolution, being the richest and most ecologically dominant group among all social insects (Hölldobler and Wilson 2008). The taxonomy and internal phylogeny of Formicidae have been extensively studied and significantly stabilized in the last decades due to the active study of ant systematics on a macroevolutionary scale (Baroni Urbani et al. 1992; Bolton 1995; Brady et al. 2006; Moreau et al. 2006; Ward 2014). Many of the findings from these studies have shown a great congruence between existing morphological and molecular hypotheses, such as the monophyly of the subfamily Proceratiinae and the recognition of the subfamily Paraponerinae as a distinct lineage among poneroid ants (Ouellette et al. 2006). Other findings, however, highlight the need for an increase in knowledge about the ancestral morphology and biology of ants. This is the case of the Ectatomminae and Heteroponerinae subfamilies that, although presenting morphology and behavior similar to those of the basal ant lineages (Bolton 2003), appear in fact as part of the most apical clade among the ants, the formicoid clade (Brady et al. 2006; Moreau et al. 2006; Ouellette et al. 2006).

In addition to the position of Ectatomminae and Heteroponerinae in relation to the other ant subfamilies, several studies based on molecular data have found strong evidence for the positioning of Heteroponerinae as the sister group to Ectatomminae (Brady et al. 2006; Moreau and Ouellette et al. 2006) and the use of the term ectaheteromorphs to refer to this grouping became common in the myrmecological literature. Feitosa (2011), based on a phylogenetic analysis using morphological data, presents ten unequivocal synapomorphies for the ectaheteromorph group, providing morphological support for this clade. The two subfamilies also share a disjoint distribution pattern, occurring in the Neotropical and Neartic regions, as well as in the Australian and Indo-Australian regions (Janicki et al. 2016). Currently, the subfamily Ectatomminae is divided in two tribes: Ectatommini, composed by the genera *Ectatomma* Fr. Smith, exclusive of the Neotropical Region, *Rhytidoponera* Mayr, occurring only in the Australian Region, and *Gnamptogenys* Roger, present in the Neotropical, Neartic, Indo-Australian and Australian regions; and Typhlomyrmecini, composed only by the genus and *Typhlomyrmex* Mayr, strictly Neotropical. Heteroponerinae has one single tribe, Heteroponerini, which harbors the genera *Acanthoponera* Mayr, exclusively Neotropical and *Heteroponera* Mayr, which presents a disjoint distribution in

106 the Neotropical and Australian regions, with *Aulacopone* Arnoldi, occurring only in the Palearctic
107 Region, as *incertae sedis* in the subfamily.

108 The ectaheteromorphs harbors 297 ant species (Bolton 2017), which can live and forage
109 both in the soil and vegetation, occurring in most tropical and subtropical regions of the world,
110 with an appreciable number of species in hot temperate environments (Camacho and Feitosa 2015;
111 Feitosa 2015). The species can nest underground, in rotting trunks, litter or even in trees, with
112 colony sizes ranging from a few dozens to a few hundred workers. These workers vary
113 morphologically, from large size ants, with robust body and well-developed compound eyes, to
114 tiny and totally blind workers; they also range from very short to very long appendages; and with
115 the cuticle varying from coarsely sculpted to polished and shiny, with discrete or highly
116 conspicuous coloration (Camacho and Feitosa 2015; Feitosa 2015).

117 The taxonomic limits of ectaheteromorphs have presented considerable stability since their
118 origin and, as currently circumscribed, the group has never received a treatment focused on the
119 relationships between the genera that it comprises (but see Lattke (1994) and Keller (2011)).
120 Individually, the Heteroponerinae genera were taxonomically revised and a complete and
121 comprehensive phylogeny based on morphological data was presented to the subfamily by Feitosa
122 (2011). Several studies have presented morphological phylogenies and taxonomic revisions for the
123 genera that compose Ectatomminae (Ward 1980; Ward 1984; Lattke 1995; Lattke 2004; Nettel-
124 Hernanz et al. 2015) and the relationships between them were systematically approached by Keller
125 (2011) using morphological data. However, due to the high demand for time and effort required to
126 explore morphological traits in ants, these studies addressed the relationships among species in a
127 limited biogeographic context, without covering the entire distribution known to the genera, and
128 relationships between genera were based on only a few species.

129 The use of molecular biology for phylogenetic inference has allowed significant advances
130 in understanding the evolutionary history and ecological success of ants. Among the 17 Formicidae
131 subfamilies, ten were systematically studied extensively through molecular data, accounting for
132 94% of the family diversity (Rabeling et al. (2008) (Martialinae); Ward and Fisher (2016)
133 (Amblyoponinae); Ward et al. (2015) (Agroecomyrmecinae and Myrmicinae); Schmidt (2013)
134 (Ponerinae); Brady et al. (2014) (Dorylinae); Ward et al. (2010) (Aneuretinae and Dolichoderinae);
135 Ward and Brady (2003) (Myrmeciinae); Blaimer et al. (2015) (Formicinae). However, most of
136 these studies are limited to analyzing a relatively low number of mitochondrial and nuclear genes,
137 which sample a small portion of the genome (except for Blaimer et al. 2015). Phylogenomics, on

138 the other hand, uses as many genes as possible, or even the complete genome of an individual, to
139 provide phylogenetic inferences, and has now proved to be the most effective method for retrieving
140 the evolutionary history of a particular group (Johnson et al. 2013; Blaimer et al. 2015; Branstetter
141 et al. 2017). This approach increases the number of characters that can be used in phylogenetic
142 analyzes hundreds or thousands of times, reducing low sampling problems for phylogenetic
143 inference (Delsuc et al. 2005).

144 Among the different methods that uses genomic data, the ultraconserved elements
145 (hereinafter UCE's) have proven to be ideal markers for the study of evolutionary relationships at
146 different time scales (Faircloth et al. 2015). Enrichment of UCE loci has been used to investigate
147 issues involving several older phylogenetic divergences for various vertebrates (Crawford et al.
148 2012; Faircloth et al. 2013; McCormack et al. 2013), insects (Faircloth et al. 2015), and more
149 recently, ants (Blaimer et al. 2015; Blaimer et al. 2016; Branstetter et al. 2017). The technique is
150 also useful for understanding relationships at the population level, including recent divergences
151 (Smith et al. 2013). The use of UCE's for phylogenetic inference has several positive attributes,
152 since it is effective even with the use of sub optimally preserved specimens and degraded DNA
153 (Blaimer et al. 2016), and has a comparative low cost for acquisition when compared to other
154 sequencing methods (Branstetter et al. 2017).

155 Here, the use of ultraconservative elements to reconstruct the phylogeny of the
156 ectaheteromorphs allows the understanding of the internal relations of the subfamily, as well as
157 elucidates the biogeographic patterns of the genera. To do so, we assembled a data set of 1629
158 UCE loci by means of target enrichment and multiplexed sequencing for 145 ectaheteromorph
159 taxa. We then use the combination of phylogenetic inference with divergence dating and
160 biogeographic analysis to provide a framework of the evolutionary history of these ants. These
161 results also provide the basis for a revised higher classification of the subfamily.

162 **Methods**

163 Taxon sampling

164 Our dataset comprises 450 individuals belonging to 145 species of ectaheteromorph ants,
165 representing six of the seven currently valid ectaheteromorph genera (for the species included, see
166 Supplementary material 1). We maximized the breadth of sampling by including at least one
167 representative from each biogeographic region in which a genus occurs (except for *Aulacopone*, a
168 monotypic genus known only by its holotype collected in the 1920's, which was gold coated and

169 submitted to scanning electron microscopy long ago) and aiming to sample across morphologically
170 disparate groups within genera. In addition, we included 16 taxa as "outgroup", representing six
171 subfamilies (Myrmicinae, Dorylinae, Pseudomyrmecinae, Formicinae, Myrmeciinae and
172 Dolichoderinae) belonging to the formicoid clade of ants (*sensu* Brady et al. (2006)). Trees were
173 rooted using Dorylinae, the subfamily most distantly related to the ectaheteromorphs. The total
174 sampling comprises 83 species of *Gnamptogenys* (322 terminals, 60% of the described species),
175 14 species of *Heteroponera* (37 terminals, 50% of the described species), four species of
176 *Acanthoponera* (10 terminals, 100% of the described species), three species of *Typhlomyrmex* (6
177 terminals, 42% of described species), 29 species of *Rhytidoponera* (29 terminals, 28% of the
178 described species) and 12 species of *Ectatomma* (30 terminals, 80% of the described species). All
179 specimens included in this study were collected in accordance with local regulations and all
180 necessary permits were obtained. Voucher specimens have been deposited at the Entomological
181 Collection *Padre Jesus Santiago Moure* of the Federal University of Paraná (DZUP), Brazil.

182 Molecular data collection

183 DNA was extracted destructively or non-destructively from adult workers using DNeasy
184 Blood and TissueKit (Qiagen, Valencia, CA, USA). We quantified DNA for each sample using a
185 Qubit fluorometer (High sensitivity kit, Life Technologies, Inc.) and sheared 8.8– 271 ng (92 ng
186 mean) DNA to a target size of approximately 600 bp by sonication (Qsonica). The sheared DNA
187 was used as input for a modified genomic DNA library preparation protocol (Kapa Hyper Prep
188 Library Kit, Kapa Biosystems) that incorporated “with-bead” cleanup steps (Fisher et al. 2011)
189 and a generic SPRI substitute (Roland and Reich, 2012), (“speedbeads” hereafter), as described by
190 (Faircloth et al. 2015). We used iTruSeq-style adapters during adapter ligation (Faircloth and
191 Gleen 2012), and PCR amplified 50% of the resulting library volume. After rehydrating and
192 purifying reactions, we combined groups of ten libraries at equimolar ratios into enrichment pools
193 having final concentrations of 153–178 ng/μL.

194 We enriched each pool using a set of 9,898 custom-designed probes (MYcroarray, Inc.)
195 targeting 2524 UCE loci specific for ants (Branstetter et al. 2017). We followed library enrichment
196 procedures for the MYcroarray MYBaits kit (Blumenstiel et al. 2010), except we used a 0.1X
197 concentration of the standard MYBaits concentration, and added 0.7 μL of 500 μM custom
198 blocking oligos designed against our custom sequence tags. We ran the hybridization reaction for
199 24 h at 65 °C, subsequently bound all pools to streptavidin beads (MyOne C1; Life Technologies),
200 and washed bound libraries according to a standard target enrichment protocol (Faircloth et al.

2012). We used the with-bead approach for PCR recovery of enriched libraries as described in Faircloth et al. (2012). We combined 15 μL of streptavidin bead-bound, enriched library with 25 μL HiFi Ready Mix (Kapa Biosystems), 5 μL of Illumina TruSeq primer mix (5 μM each) and 5 μL of ddH2O. We purified resulting reactions using 1.0X speedbeads, and we rehydrated the enriched pools in 22 μL EB. We quantified 2 μL of each enriched pool using a Qubit fluorometer (broad range kit).

Enrichment was verified by amplifying seven UCE loci targeted by the probe set. We set up a relative qPCR by amplifying two replicates of 1 ng of enriched DNA from each library at all seven loci and comparing those results to two replicates of 1 ng unenriched DNA for each library at all seven loci. We performed qPCR using a SYBR® FAST qPCR kit (Kapa Biosystems) on aViiATM 7 (Life Technologies). Following data collection, we computed the average of the replicate crossing point (Cp) values for each library at each amplicon, and we computed fold-enrichment values, assuming an efficiency of 1.78 and using the formula 1.78 abs (enriched Cp− unenriched Cp). We then created serial dilutions of each pool (1:200,000; 1:800,000; 1:1.000,000; 1:10.000,000) and performed qPCR library quantification, assuming an average library fragment length of 600 bp. Based on the size-adjusted concentrations estimated by qPCR, we pooled libraries at equimolar concentrations. Enriched sequences were sequenced on four lanes, one of them being a partial lane (included samples from other projects) of a 150-bp paired-end Illumina HiSeq 2500 run (High Throughput Genomics Lab at GNomEx Core Facilities, University of Utah). All the UCE laboratory work was conducted in and with support of the Dr. John Longino laboratory facilities of the University of Utah.

Processing and alignment of UCE data

The sequencing facilities demultiplexed and converted raw data from BCl to FASTQ format using either BASESPACE or BCL2FASTQ (available at http:// support. illumina. com/ downloads/ bcl2fastq_ conversion_ software_ 184. html). We trimmed the demultiplexed FASTQ data output for adapter contamination and low-quality bases using Illumiprocessor (Faircloth 2013), which is a wrapper program around TRIMMOMATIC (Bolger et al. 2014). All further data processing described in the following relied on scripts within the PHYLUCE v1.5. package. We computed summary statistics on the data using the get_fastq_stats.py script, and assembled the cleaned reads using the assemblo_trinity.py wrapper around the program Trinity (version trinityrnaseq_r20140717) (Grabher et al. 2011). Average sequencing coverage across assembled contigs was calculated using get_trinity_coverage.py. To identify assembled contigs representing

enriched UCE loci from each species, species-specific contig assemblies were aligned to a FASTA file of all enrichment baits using match_contigs_to_probes.py (min_coverage = 50, min_identity = 80), and sequence coverage statistics (avg, min, max) for contigs containing UCE loci were calculated using get_trinity_coverage_for_uce_loci.py. Subsequently, we used get_match_counts.py to query the relational database containing matched probes created in the previous step, in order to generate a list of UCE loci shared across all taxa. This list of UCE loci was then used in the get_fastas_from_match_counts.py script to create FASTA files for each UCE locus, which contain sequence data for taxa present at that particular locus. We aligned all data in all these FASTA files using MAFFT (Katoh et al. 2009) through seqcap_align_2.py (min-length=20, no-trim). Following alignment, we further trimmed our alignment using a wrapper script (get_gblocks_trimmed_alignment_from_untrimmed.py) for Gblocks (Castresana 2000) using the following settings: b1=0.5, b2=0.5, b3=12, b4=7.

We initially selected the following subsets of UCE alignments using get_only_loci_with_min_taxa.py: 1) 75% complete (75p-matrix) (containing alignment data from at least 337 of the 450 taxa), 2) 90% complete (90p-matrix) (>405 of 450 taxa), 3) 95% complete (95p-matrix) (>427 of 450 taxa). We also selected a reduced 161-taxa set (one representative of each of the 145 species and the 16 outgroups), including the subsets of UCE alignments: 1) 75p-pruned (>120 of 161 taxa), 2) 90p-pruned (>144 of 161 taxa) and 3) 95p-pruned (>152 of 161 taxa). We added missing data designators to each file with add_missing_data_designators.py, and generated alignment statistics across all alignments using get_align_summary_data.py. Finally, we concatenated individual alignments of UCE loci for each subset into one nexus alignment file with format_nexus_files_for_raxml.py for subsequent phylogenetic analyses.

Phylogenetic inference

We performed maximum likelihood best tree and bootstrap searches (N = 100) in RAxML v8.2.7 (Stamatakis 2014), initially on a 75%, 90% and 95% complete UCE matrix (see above). For each of the concatenated nucleotide UCE data sets we used two partitioning schemes: 1) partitioning by UCE locus and 2) an unpartitioned model. We also calculated the maximum likelihood best tree and bootstrap searches for a reduced 161 taxa dataset, containing only one individual per species, in order to access topological differences that could be influenced by taxon sampling. For subsequent analyses, however, we elected to proceed with the 90p-matrix. We also estimated gene trees for the 1483 UCE loci in the 90p-matrix by performing RAxML analyses (best tree and bootstrap) under GTR+4Γ model with 200 rapid bootstrap replicates, on individual

loci. Species tree analyses with local posterior probability support values were then performed in ASTRAL-III (Sayyari and Mirarab 2016) using the 1483 gene trees as input. Subsequently, we used individual locus characteristics to assess whether loci with different properties produce different phylogenies. Using a custom R v3.4.0 script (R Core Team 2017), we computed basic statistics, such as alignment length, proportion of parsimony informative sites, Robison-Foulds distance from the resulting species tree and GC content (proportion by loci) using the packages ips 0.0-7 and ape. We also computed average bootstrap support, used here as a proxy for phylogenetic signal, from which we filtered the data to a subset of 100 loci with the highest average bootstrap score across all nodes in the gene tree (100-best), to use in dating analyses. Additionally, we computed the site variability for each locus, used here as a proxy for site informativeness, and calculated the number of alignment sites that were informative, variable but not informative (autapomorphies) or constants. All the above phylogenetic analyses were performed on the Smithsonian Institution high performance cluster (SI/HPC), or on an iMac computer.

Dating

We inferred divergence dates within ectaheteromorphs from a reduced 161-taxa set (one representative of each of the 145 species and the 16 outgroups), since we were interested in the species level processes that describe species diversification. Because divergence time estimation is computationally very expensive, the divergence dating analysis were limited to the 100 best loci (as described above), using a Maximum Likelihood inference approach with the program MEGA7 (Kumar et al. 2016), run in a MacBook computer. Analyses were unpartitioned and a timetree was inferred using the RelTime method (Tamura et al. 2012) and the General Time Reversible model (Nei and Kumar 2000). The estimated log likelihood value was -1148388.11 and there was a total of 67234 positions in the final dataset. The timetree was computed using two calibration constraints. We employed a tree prior (the ML tree for the 100best data set) and defined secondary calibration priors for the Heteroponerinae and Ectatomminae clades, that were constrained as monophyletic, using the dates inferred by Brady et al. (2006) (30.9 to 47.1 ma and 50.9 to 61.1 ma, respectively), using a normal distribution. The root position and backbone relationships of the tree were sampled from the tree prior.

State reconstruction analysis

To evaluate the biogeographic history of the genera, seven biogeographic areas were considered: Neotropical, Nearctic, Palearctic, Afrotropical, Oriental, Indo-Australian and Australasian, using Wallace's line to divide the latter two areas, (following Ward et al. 2014) (see

297 Supplementary material S2). We used the resulting ML tree from our RAxML analysis on the 90%
298 complete UCE data set to trace the ancestral ranges with a marginal probability reconstruction with
299 model "Markov k-state 1 parameter model" (Mk1). The likelihood reconstruction method was
300 used, which finds the ancestral states that maximize the probability the observed states would
301 evolve under a stochastic model of evolution (Schluter et al. 1997; Pagel 1999). The Mk1 model
302 ("Markov k-state 1 parameter model") is a k-state generalization of the Jukes-Cantor model, and
303 corresponds to Lewis's (2001) Mk model, where the single parameter is the rate of change,
304 considering any particular change as equally probable.

305 **Results**

306 UCE capture statistics

307 An average of 2154.4 contigs with a mean length of 787.4 bp was assembled by Trinity
308 after adapter- and quality- trimming, with an average coverage of 17.4X. From all of the assembled
309 contigs, we recovered an average of 2,520 UCE loci per sample with a mean length of 536.07 bp.
310 The average coverage per captured UCE locus was 92.3X. Following alignment of individual UCE
311 loci, we filtered these data for loci captured for ≥75 % of taxa (75p-matrix), retaining 2206 loci,
312 loci captured for ≥90 % of taxa (90p-matrix), retaining 1483 loci and for loci captured for ≥95 %
313 of taxa (95p-matrix), retaining 375 loci. We further selected a data set of 100 loci with the best
314 average bootstrap support for subsequent dating analyses (100best-matrix), because this
315 represented a manageable size for dating analysis (whereas analysis of the full 2520 loci was not
316 feasible). Concatenation of UCE loci generated matrices of 1,227,825 bp (75p-matrix), 883,071
317 bp (90p-matrix), 220,331 bp (95p-matrix) and 12,895 bp (20best-matrix) (Supplementary material
318 1).

319 Phylogenetic results

320 The ML best tree and bootstrap searches for the 90p-pruned matrix are shown in Figure 1.
321 We recovered a highly resolved phylogeny for the ectaheteromorphs with most nodes displaying
322 100% bootstrap support. When the data set is pruned in order to contain only one representative
323 of each species (161 taxa), the same topology is recovered for all the data sets (75p, 90p, 95p and
324 100best) (Fig.1). The 75p-matrix and the 90p-matrix recovers the same topology, with very similar

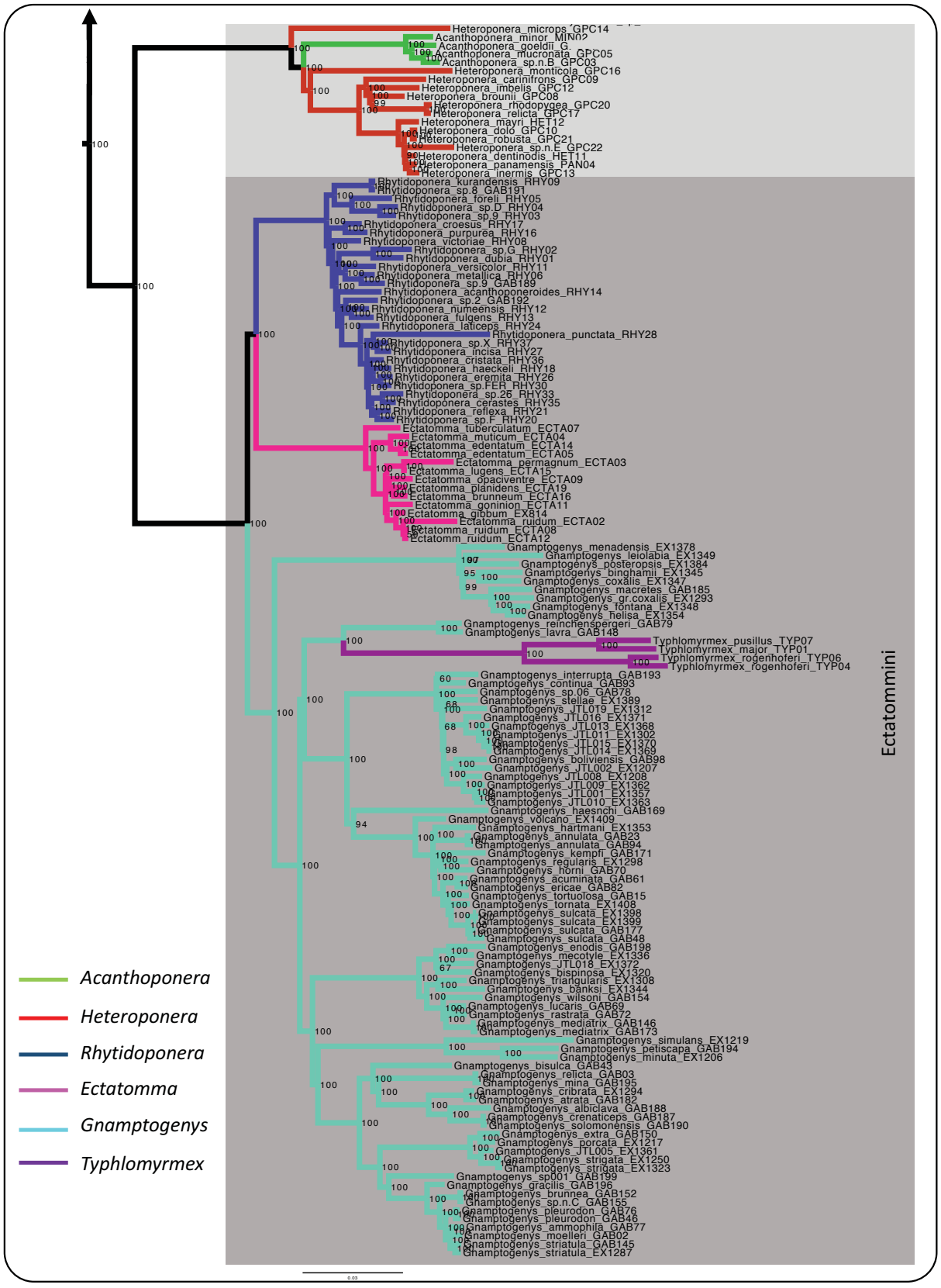


Figure 1. Phylogeny of the subfamily Ectatomminae estimated by the phylogenomic UCE 90% complete pruned data set (161 taxa). Figure is based on RAxML best tree searches, with RAxML bootstrap values mapped on the respective nodes. The bootstrap searches included 100 replicates. The six larger ectatommine lineages are indicated. See also Additional file 3.

bootstrap support values (Supplementary material S3A). Only a few nodes were recovered with a bootstrap score (BS) lower than BS = 98, mainly involving intraspecific relationships within South and Central American populations. The relationship between *Heteroponera monticola* Kempf & Brown, a South American species, and the clade formed by the Australasian species of *Heteroponera* also presents a lower support (BS=86). The 95p-matrix gives similar results in general (Supplementary material S3B), with the relationships among the heteroponerines mainly unchanged (except for the position of *Heteroponera* sp. E), and within all of the genera (with the exception of the position of *G. petiscapa* Lattke and some species of *G. striatula* group). Moreover, the relationship between the genera is not well resolved with this matrix. *Rhytidoponera* is recovered as sister to *Gnamptogenys* with low support, contrasting with the results of the 90p-matrix and the 75p-matrix. In contrast, with only 161 taxa, either the 75p-pruned (2221 loci), the 90p-pruned (1629 loci) and the 95p-pruned (610 loci) always revels the sister group relationship between *Rhytidoponera* and *Ectatomma*, with BS = 100 for the two first and BS = 60 for the latter. When the number of informative sites in each locus in the different data sets is computed, it is possible to see that the number of largely informative sites is bigger in the 90p-matrix than in the 95p-matrix, due simply to the larger number of loci in this data set (Fig. 2A, 2B). Similarly, the same pattern of locus informativeness is recovered for the pruned data sets, but in this case resulting in the same topology (Fig. 2C, 2D). Partitioning of the data matrices produced no changes in the topology and very low variation in the support values. The species tree estimated by ASTRAL-III from the 1603 gene trees of the 90p-pruned is very similar to the tree estimated in the concatenated analysis, with most nodes showing good local posterior probability (LPP) support values (Supplementary material S3C). The interspecific nodes with low support in the concatenated analysis also receive lowered support in the species tree analysis, although BS and LPP have different theoretical approaches to the calculation of support and cannot be easily compared. For the species tree estimated from the 373 gene trees of the 95p-matrix, the same sister group relationship between *Rhytidoponera* and *Gnamptogenys* in the concatenated analysis is recovered, with a similar LPP support value as the one found for the 90p-matrix relationship between *Rhytidoponera* and *Gnamptogenys* (95p: LPP = 41.31; 90p: LPP = 43.68).

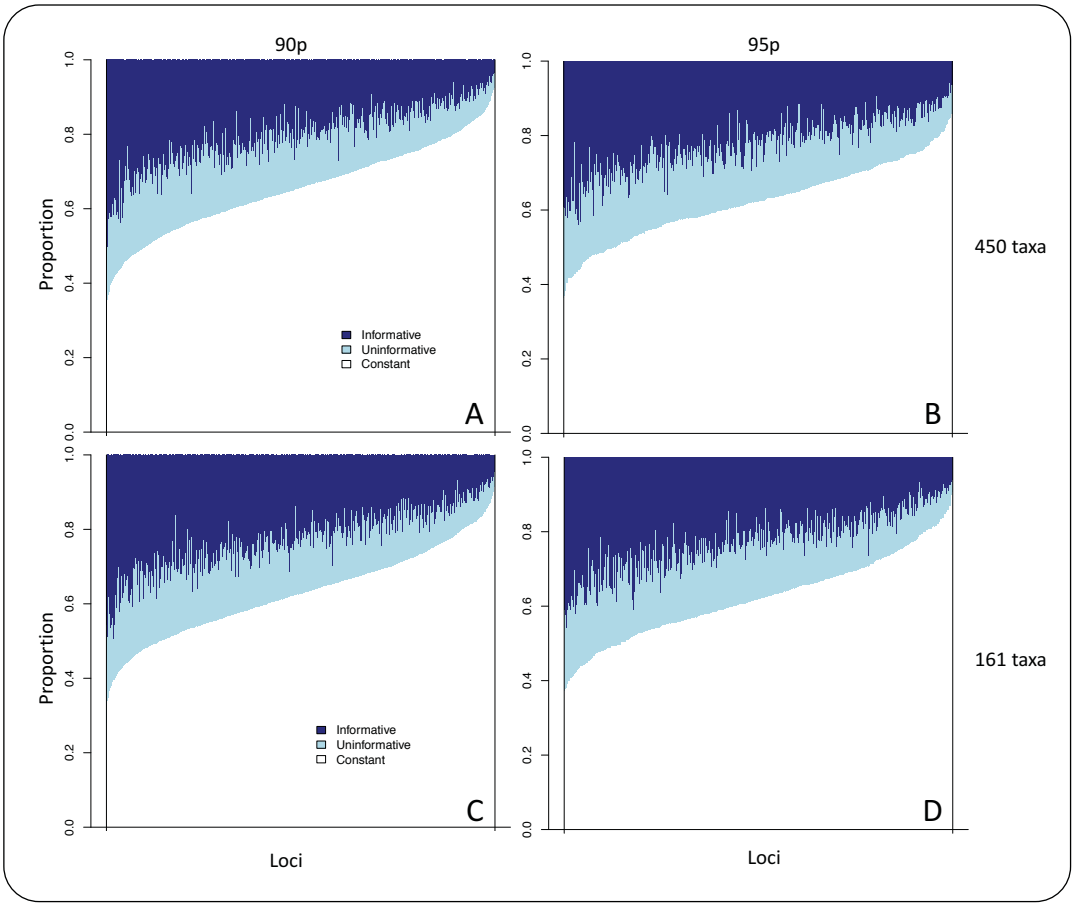


Figure 2. Phylogenetic informativeness per locus in the different matrices, showing the proportion of the loci that are variable and informative, variable but uninformative (autapomorphies) and constants. **a** Phylogenetic Informativeness (PI) as estimated with the ips package in R for the two UCE data sets (161 and 450 taxa) and two levels of completeness (90% and 95%).

Taxonomic results

In order to have a phylogenetic classification for the subfamily, with the tribes and subfamilies being monophyletic and mutually exclusive (Ward 2011), we propose some higher-level taxonomic changes. New and revived combinations implicitly include the junior synonyms of the species names listed below. Author and year of publication for all genus and species names can be found in AntCat (http://antcat.org/).

Clade 1: Ectatomminae (redefined)

The ectaheteromorphs, as currently defined, encompass two different subfamilies. The reciprocally monophyletic subfamilies Ectatomminae and Heteroponerinae together form a strongly supported clade (BS = 100; LPP = 72.88). We also find strong support for the monophyly of both subfamilies (Heteroponerinae: BS = 100, LPP = 98.66; Ectatomminae: BS = 100, LPP = 95.27). However, the morphological evidence strongly suggests that the division of these two

groups may not be necessary, and all the genera belonging to them could be combined in Ectatomminae, that represents the oldest available name and it takes priority as the name for all ectaheteromorphs. This establishes the following new synonymy and combination: Ectatomminae (Ectatommini, Heteroponerini **n. comb.**) = Heteroponerinae **n. syn.** Hence, we refer to the ectaheteromorphs as Ectatomminae throughout the remainder of this paper.

Clade 2: Heteroponerini (redefined)

This clade is recovered with high bootstrap support (BS = 100; LPP = 98.66) in all the analysis, and includes the genera *Heteroponera* and *Acanthoponera*. *Acanthoponera* is recovered as monophyletic in all our analysis (BS = 100; LPP = 97.93), with all the species represented in the analysis also appearing to be monophyletic. *Heteroponera* is recovered as paraphyletic with respect to *Acanthoponera* with a single species, *Heteroponera microps* Borgmeier, being clearly separated from the other species of *Heteroponera* with full support (BS = 100; LPP = 99.83), and presenting a sister group relationship to all other Heteroponerini genera. The remaining *Heteroponera* species are recovered as a monophyletic group, with moderate support (BS = 86; LPP = 37.08), suggesting that further investigation on the phylogenetic relationships within the genus is necessary. Hence, the tribe Heteroponerini **n. comb.** is redefined to contain the genus *Heteroponera* and *Acanthoponera*, with the genus *Aulacopone,* not included in this analysis, as *incertae sedis* in Ectatomminae.

Clade 3: Ectatommini (redefined)

The current definition of Ectatommini comprises four extant genera, all of them included in our analysis, forming a clade recovered with very good support (BS = 100; LPP = 94.6). We recovered two major clades within the Ectatommini with full support (BS = 100; LPP = 94.6): one formed by the reciprocally monophyletic genera *Rhytidoponera* (BS = 100; LPP = 96.47) and *Ectatomma* (BS = 100; LPP = 98.97); and other were *Gnamptogenys* is recovered as paraphyletic with respect to *Typhlomyrmex*, with high support (BS = 100; LPP = 78.92). Nevertheless, it is clear that the position of *Typhlomyrmex* in a separate tribe is not congruent with the relationships among the species, hence the necessity of the new synonymy: Ectatommini = Typhlomyrmecini **n. syn.**

Taxonomic account

The tribal classification of Ectatomminae is here modified to achieve consistency with our molecular phylogenetic results. We maintain the existing classification as far as possible, while striving to ensure that all recognized tribes are monophyletic. Genera known only from fossils are

signified with a dagger; most of these are unplaced to tribe and are treated as *incertae sedis* within the subfamily. For tribe diagnosis, see Bolton (2003). A formal taxonomic treatment with the possible generic changes discussed above will appear elsewhere (Camacho et al. in prep; Feitosa et al. in prep).

Ectatomminae Emery

= Heteroponerinae Bolton **new synonym**

Diagnosis: Presenting the characters of "poneromorph" subfamilies described by Bolton (2003: p.40). Clypeus broadly inserted between frontal lobes; anterior clypeal margin with a narrow lamellar apron[1]. Pronotum with the humeral corners angled, forming a distinct delimitation between the anterior and lateral margins[2]. Antero-ventral angle of pronotum triangular[3]. Anterior tarsus without arolium[2]. Petiole pedunculated[4]. Petiolar node as wide or wider than long[3]. Subpetiolar process very well developed, occupying more than one third of the ventral portion of the petiolar sternite[3]. Abdominal segment IV presclerites separated from the rest of segment by a constriction or slightly thickening[5]. Fourth abdominal tergite arched and larger than the sternite, giving the segment a curved appereance[4].

[1]Bolton (2003); [2]Lattke (2004); [3]Feitosa (2011); [4]Keller (2011);[5]Lattke (1994).

Comments: The characters presented here as diagnostic for the subfamily were recovered as synapomorphic for Ectatomminae by Feitosa (2011), after a comprehensive morphological phylogeny of Heteroponerini.

Tribe **Ectatommini** Emery

= Typhlomyrmecini Emery

Genera: *Ectatomma, Gnamptogenys, Rhytidoponera,* and *Typhlomyrmex*.

Tribe **Heteroponerini** Bolton

Genera: *Acanthoponera, Heteroponera*.

***Incertae sedis*:** *Aulacopone*, †*Canapone*, †*Electroponera*, †*Pseudectatomma*.

Dating and State reconstruction results

The time-calibrated phylogeny as estimated from the 100best data set is shown on Figure 3 and the ancestral ranges estimated by the ancestral state reconstructions is shown on Figure 4.

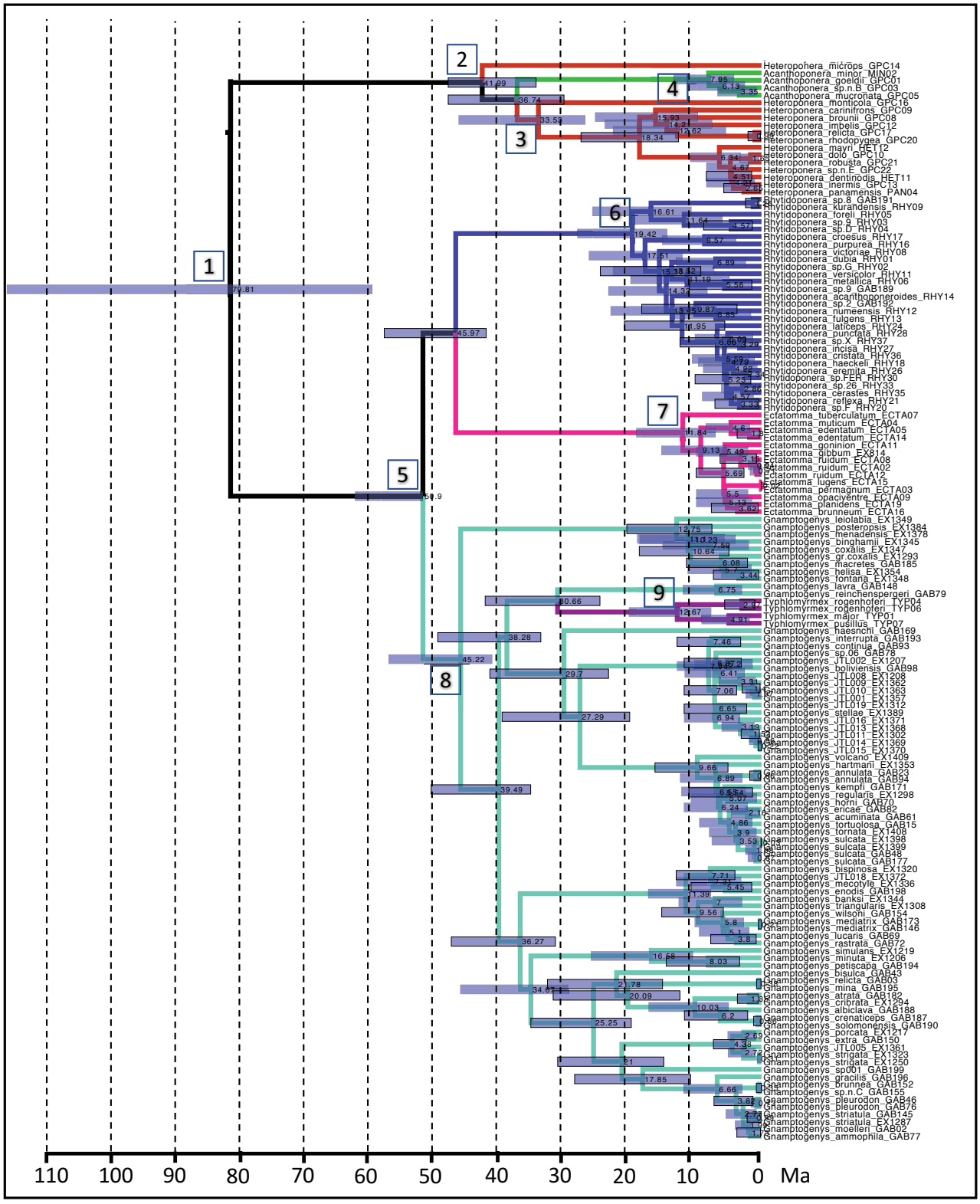
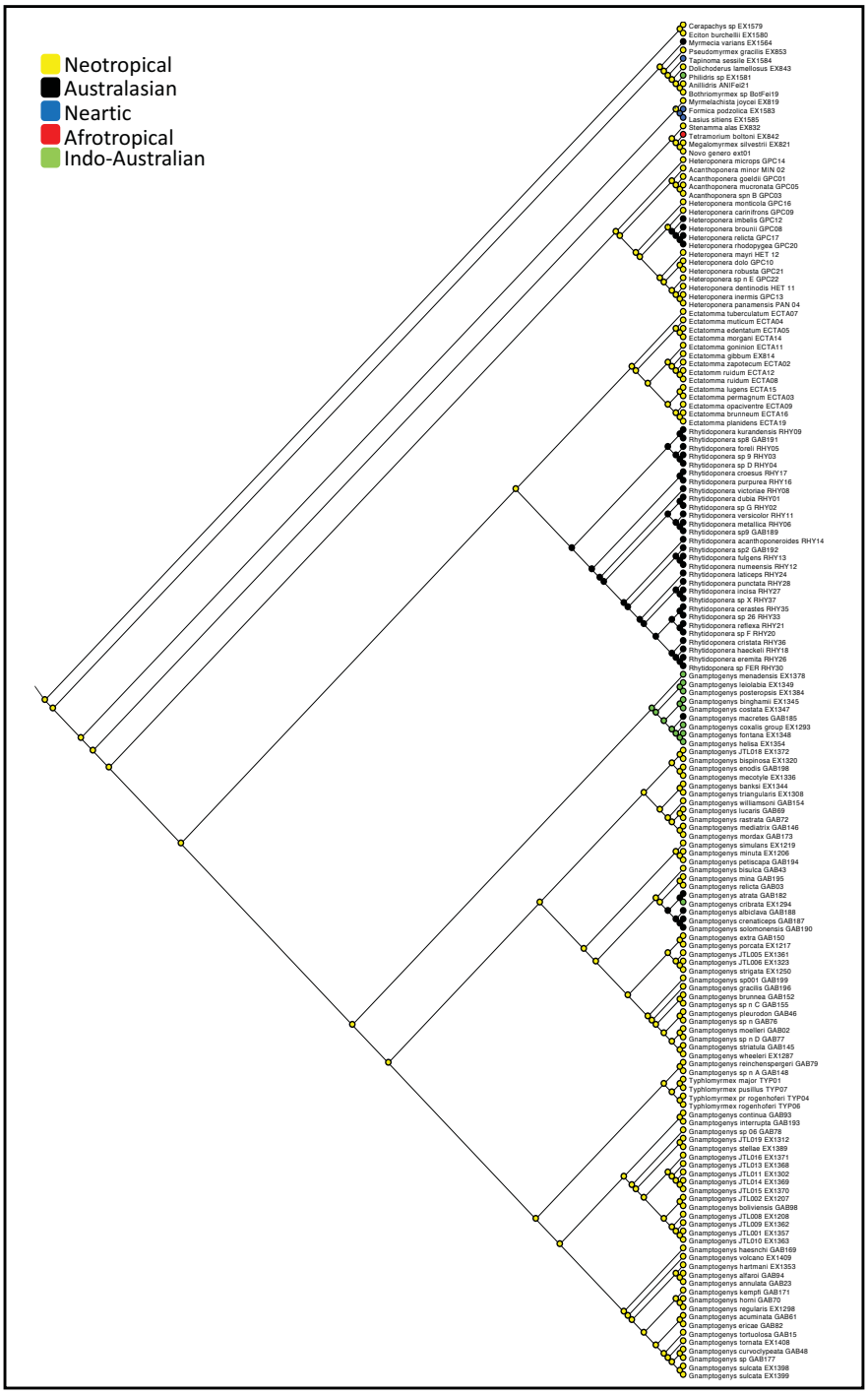


Figure 3. Time-calibrated phylogeny for the subfamily Ectatomminae. Timetree inferred using the RelTime method and the General Time Reversible mode, conducted in MEGA7. The timetree was computed using 2 calibration constraints on crown Ectatommini and Heteroponerini. The estimated log likelihood value is -1148388.11. Blue bars show the 95 % highest posterior density range for each node.

Support values, median crown group ages, select highest posterior density intervals (95 % HPD), and ancestral ranges are summarized in Table 1. According to our ancestral reconstructions, the species that compose the subfamily Ectatomminae share a Neotropical ancestor and crown-group Ectatomminae are estimated to have evolved in the late Cretaceous or early Paleogene, between 58.5–113.4 Ma. The Heteroponerini tribe is estimated to have a Neotropical origin, as well as Ectatommini and the two Ectatomminae tribes appears to have diversified throughout the late Eocene and early Paleocene, between 33.8–61.1 Ma. Our analyses suggest that the *Heteroponera* evolved in the early Oligocene (26.4-45.5 Ma) on the Neotropical region, with one lineage later reaching the Australasian region (Proportional Likelihood = 0.97). A Neotropical origin is further suggested for the common ancestor of *Ectatomma* and *Rhytidoponera* in the early Eocene (41.3-56.7 Ma), with the latter diversifying in the Australasian region throughout the early Miocene (14.1-37.6 Ma) and the former diversifying in the Neotropical region in the late Miocene (6.8-18.8 Ma). A similar pattern is found for the common ancestor of all *Gnamptogenys* (including *Typhlomyrmex*), which has its origin in the Neotropical region in the early Eocene (40.4-56 Ma), with an early diversification in the Indo-Australian region throughout the Miocene (7.4-20-2 Ma), with a later single invasion in the Australasian region. Furthermore, *Gnamptogenys* diversified mostly in the Neotropical region, with one lineage spreading to the Australasian region throughout the late Miocene (4.9-16.9 Ma), later reaching also the Indo-Australian region (Proportional Likelihood = 0.98). All the reconstructions computed were extremely well supported, with values always bigger than 99% proportional likelihood values for all the ingroup ancestors (except for the ones cited above). Notably, the oldest genera within the Ectatomminae are *Gnamptogenys* (40.4-56 Ma) and *Heteroponera* (26.4-45.6 Ma), while *Acanthoponera* (4.1-13.2 Ma) and *Ectatomma* (6.8-18.8 Ma) are recovered as the youngest lineages.

Table 1. Summary of crown group divergence ages and estimated ancestral ranges.

		ML		Time tree		Ancestral Range	
	Node	**BS support**	**LPP support**	**Mean Age**	**95% HPD**	**Area**	**Prob.**
Ectatomminae	1	100	72.9	79.8	58.5 - 113.4	Neotropical	0.99
Heteroponerini	2	100	98.7	42.0	33.9 - 47.1	Neotropical	0.99
Heteroponera	3	100	39.0	33.5	26.5 - 45.3	Neotropical	0.99
Acanthoponera	4	100	98.4	8.0	4.1 - 13.2	Neotropical	0.99
Ectatommini	5	100	95.3	50.9	50.9 - 61.1	Neotropical	0.99
Rhytidoponera	6	100	97.2	19.4	14.2 - 27.7	Australasian	0.99
Ectatomma	7	100	98.9	11.8	6.9 - 18.9	Neotropical	0.99
Gnamptogenys	8	100	77.8	45.2	40.8 - 56.1	Neotropical	0.99
Typhlomyrmex	9	100	98.5	12.7	7.6 - 19.9	Neotropical	0.99



463

464 **Figure 4.** Ancestral range estimates for the subfamily Ectatomminae. Ancestral ranges estimated by Mesquite are
465 mapped on MRCA nodes for each tribe and genus (regardless of the level of support).

466

467 **Discussion**

468 Phylogenomics resolves the relationships among Ectatomminae genera

469 Both our concatenated and species tree analyses recovered a fully resolved and mostly
470 highly supported phylogeny for Ectatomminae, estimating identical topologies at genus and

species level (Figs. 1 and 3). Uncertainty is confined only to one node in the phylogeny, for which the 95p-matrix showed different results. These results are congruent with previous works that suggests that a greater number of loci to work with is beneficial (Borowiec et al. 2015; Branstetter et al. 2017), although it remains unclear how many loci are necessary to resolve phylogenetic relationships. The 95p-matrix retained 375 loci, as opposed to the 1493 and 2206 retained by the 90p-matrix and 75p-matrix, respectively, suggesting that the comparatively low number of loci may not be sufficient to recover the sister-group relationship between *Rhytidoponera* and *Ectatomma*. However, it has long been recognized that simply increasing the amount of data can exacerbate systematic bias in phylogenetic estimation (Phillips et al. 2004; Philippe et al. 2011; Borowiec et al. 2015) and, in order to improve phylogenomic inference, the quality of the data is key (Borowiec et al. 2015). However, our analysis showed that there are no major differences in the quality between the data sets, all of them showing similar patterns of distribution of the informativeness of the data (Figure 2). Alternatively, previous works have been showing that the taxonomic balance within a data set has a large impact on phylogenetic results (Branstetter et al. 2017), emphasizing the importance of both a broad taxonomic sampling and taxonomic evenness across samples. In our complete data set, with 450 taxa, there's clearly an imbalance between the sampling of different genera, with *Gnamptogenys* representing 71% of the samples. By reducing the sampling to one representative of each species, the greater evenness among the taxa appears to have a significant effect on the topology, even though the smaller data set (95p-pruned, 610 loci) still presents a low support value for the relationship between *Rhytidoponera* and *Ectatomma*. This imbalance doesn't appear to influence most of the other clades in the phylogeny, suggesting to us that a greater sampling of *Rhytidoponera* species may be necessary to understand the evolution of those groups. In fact, although we have samples covering from 40 to 80% of each genus diversity, for *Rhytidoponera* our sampling represents only 28% of the species currently described, suggesting that a broader sampling of those species seems to be crucial to eliminate any doubts regarding its position among the Ectatomminae. However, even though these incongruences may be recovered, the fact that a more even sampling helps recover the same topology across all data sets give us confidence to assert that the relationships among the genera in Ectatomminae are well supported and represent the first results broadly exploring the evolution of the subfamily.

Taxonomy of Ectatomminae revisited

Based on our UCE phylogeny, we propose a few taxonomic changes for the subfamily that aim to improve ant systematics while simultaneously keeping names fairly stable. At the subfamily

level, our decision to synonymize Heteroponerinae under Ectatomminae does not affect the monophyly of these groups, since both are monophyletic as currently circumscribed, and their sister group relationship has been broadly discussed. Historically, the close relationship between both groups have been strongly suggested and supported by morphological (Brown 1958; Bolton 2003; Ward 2007; Keller 2011) and molecular data (Brady et al. 2006; Moreau et al. 2006; Moreau and Bell 2013; Branstetter et al. 2017). However, their morphology can be misleading, especially when defining the diagnostic characters for the groups separately. When describing Heteroponerinae, Bolton (2003) states that there is no unequivocal apomorphy for the subfamily, suggesting a number of characters that could have this status. Feitosa (2011) investigated the phylogenetic history of Heteroponerinae using morphological data, testing those characters suggested by Bolton (2003), as well as several others, and also could not recover any apomorphy for the group. However, in his work, Feitosa (2011) suggests at least ten unequivocal synapomorphies for the clade comprising both Ectatomminae and Heteroponerinae. For this reason, we propose to reclassify all ectaheteromorph ants as members of a single subfamily, ensuring the monophyly criterion that already applies to all other ant subfamilies but, most importantly, providing a clear and forthright diagnosis for the subfamily that will help ensure a greater understanding of the group morphology.

Regarding the changes at tribal level, our aim is to keep the classification stable. In this sense, the new combination of the tribe Heteroponerini and the synonymy of Typhlomyrmecini are made to ensure the correct placement of the former, and the monophyly of Ectatommini, in the case of the latter. At a generic level within the tribe Heteroponerini, the paraphyly of *Heteroponera* is a striking result, still unpredicted by morphology, with *H. microps* appearing as a separately diverging lineage. This result is congruent with previous hypothesis by Borgmeier (1957) and Feitosa (2011) which suggested that the diagnostic characters for this species are highly divergent from the morphological patterns for *Heteroponera*, but its placement as a separate genus is supported here for the first time. Similarly, the position of *H. monticola,* recovered as sister to all the other *Heteroponera* species, as well as the recovery of two separate clades, the first comprising *H. carinifrons* (from Chile) as sister to the Australasian species and the second comprising the remaining Neotropical species, are also brand new evolutionary hypothesis for the genus, with great impact in its biogeographical history. Unfortunately, the genus *Aulacopone* was not included in our analysis, due to the unavailability of specimens and difficulties of collecting in its type locality. The genus is monotypic and was collected only twice on the 1920's, with the only known specimen currently metalized, making it impossible to recover DNA information from the pinned

specimen. The distribution of this genus is singular within the Ectatomminae, being the only group to occur in the Paleartic region. *Aulacopone* is said to present several morphological similarities with the other heteroponerine (Brown 1958; Taylor 1980; Lattke 1994; Bolton 2003; Feitosa 2011), but its position among the Ectatomminae is still not well defined, due to the impossibility to examine important characters in the previous phylogenetic study (Feitosa 2011). These results regarding the relationships between the species in Heteroponerini are new and can shed new light to the study of their morphology. We believe that, in order to assure the stability of the classification and a high-level contribution for ant systematics, the assessment of the relationships between the species should combine both molecular and morphological approaches, in order to better understand the evolution of this group. For this reason, we are not proposing formal nomenclatural changes at this point, since a comprehensive phylogenetic and taxonomic assessment of the tribe is intended for later publication (Feitosa et al. *in prep*.).

The four genera that comprises Ectatommini are shown to form a well-supported clade, a result that is congruent with the morphological hypotheses for the group (Bolton 2003; Ward 2007; Keller 2011), although these works considered the four genera to be monophyletic. Regarding the molecular phylogenies published so far, only one or a few specimens of each genus were included, limiting their conclusions regarding the relationships among them (Brady et al. 2006; Moreau et al. 2006; Moreau and Bell 2013; Branstetter et al. 2017). In this sense, this is the first molecular study that aimed to investigate the genus level relationships in Ectatomminae. A fairly novel result, the sister group relationship between *Ectatomma* and *Rhytidoponera,* is congruent with previous morphological hypothesis by Keller (2000; 2011) and suggested by other broad Formicidae molecular phylogenies that do not focused on this groups (Brady et al. 2006; Moreau et al. 2006; Moreau and Bell 2013), but was never further explored. Brown (1958) noticed some similarities between the two genera, as the similar wing venations, the male genitalia and the lack of a metacoxal spine (present in most *Gnamptogenys*). Also, Brown (1958) brings attention to the similarities between *Ectatomma* workers and the largest species of *Rhytidoponera*. Our results are the first to include several specimens for those genera and to shed light on the evolution of these two groups, that should be further and comprehensively explored.

Perhaps the most striking novel result in our study is the strong support for the paraphyly of *Gnamptogenys* in relation to *Typhlomyrmex*. These are outstanding results that were never predicted by any morphological or molecular studies before. Historically, the position of *Typhlomyrmex* as related to the Ectatommini was first brought by Emery (1911), but Brown (1964)

later placed the genus in its own tribe, Typhlomyrmecini, considering them as similar to Amblyoponini. Lattke (1994) suggests that the genus similarities with Ectatommini should be further explored and Bolton (2003) finally considers Typhlomyrmecini as a member of Ectatomminae. However, the distinct and cryptic morphology of these ants appears to be misleading, suggesting a strong morphological convergence between *Typhlomyrmex* and other small-sized *Gnamptogenys* species.

Our results also reveal for the first time a phylogenetic structure for *Gnamptogenys* by identifying a series of independent lineages (Figure 1) as well as establishing some relationships among these lineages. Lattke (1995; 2004) provided revisions for both the Neotropical and Asian species of the genus based on morphological data, proposing five species groups for the neotropics, also recognizing the *minuta* group (Brandão and Lattke 1990), and five species groups for Asia. Some of these groups are partly recovered in our study, while others appear to be paraphyletic. A clade formed by the species that belong to the *coxalis* and *laevior* groups (sensu Lattke 2004) appears as sister to all other *Gnamptogenys* species (including *Typhlomyrmex*), strongly supported (BS = 100; LPP = 99.23). This finding suggests that the Indo-Australian *Gnamptogenys* species are a result of a separate evolutionary lineage in the group, not strictly related to the Australasian lineages. The basal position of this group is congruent with Lattke (2004) previous hypothesis, predicted by its morphological features and morphological phylogenetic analysis. We recover, with strong support (BS = 100; LPP = 74.39), the monophyletic relationship between the genus *Typhlomyrmex* and two small *Gnamptogenys* species (*G. reichenspergeri* (Sanstchi) and *G. lavra* Lattke). This clade is a very surprising result in our study, since these *Gnamptogenys* species were previously assigned to the *striatula* group of species (Lattke 1995) due to their remarkable morphological similarities with other small size species in the genus (i.e. *G. mina* (Brown), *G. haytiana* (Wheeler and Mann) and *G. relicta* (Mann)).

Another lineage recovered with high support (BS = 100; LPP = 97.31) is formed by large species such as *G. interrupta* (Mayr), *G. continua* (Mayr), *G. stellae* Lattke and others. Similarly, the largest species in the genus are grouped together (BS = 100; LPP = 91.56), recovering the *concinna* group, despite previous concerns about the heterogeneity of the species that it comprises (Lattke 1995). A group formed by species both from the *sulcata* and from the *mordax* groups (sensu Lattke 1995) forms a clade (BS = 100; LPP = 95.62), indicating that a morphological revision of these species is essential to understand the evolution of this lineage. A lineage comprised mainly by species specialized in praying myriapods and diplopods is recovered with

strong support (BS = 100; LPP = 97.37), formed by species such as *G. rastrata* (Mayr), *G. mediatrix* Brown, *G. triangularis* (Mayr), among others. The very distinctive *G. minuta* (Emery), *G. petiscapa* Lattke and *G. simulans* (Emery) forms a clade also recovered with high support (BS = 100; LPP = 98.01). And finally, we recovered a monophyletic clade (BS = 100; LPP = 89.33) comprising the Australasian *Gnamptogenys* species (along with *G. atrata* Lattke, and Indo-Australian species), as well as the species previously assigned by Lattke (1995) as the *striatula* group (except for *G. reichenspergeri* and *G. lavra*), suggesting a very interesting biogeographic history for this lineage. Thus, our molecular phylogenetic results provide strong justification for reclassifying the genus *Gnamptogenys* into a series of coherent monophyletic groups, providing the basis for future taxonomic work on these *Gnamptogenys* clades. However, we are not proposing formal nomenclatural changes, nor the synonymy of *Typhlomyrmex* under *Gnamptogenys*, since we believe that a comprehensive morphological assessment of this lineages should precede any classification changes in order to maintain the stability and feasibility of any decisions we may make. The nomenclatural changes, as well as a morphological discussion of the genus lineages is intended for later publication (Camacho et al. *in prep*.).

Ectatomminae biogeography

Our dating analyses places ectatommine evolution deep into the Cretaceous, also raising the possibility of a later evolution in the early Paleocene (58.5–113.4 Ma). The median crown-group age estimates are the same as the oldest known stem-group ectatommine fossil, *Canapone* Dlussky (~79 Ma). The identity of *Canapone* as part of Ectatomminae, however, is extremely controversial, limiting the inferences one may make regarding its importance for the subfamily. Regarding the crown-group age estimates of Ectatommini (50.9-61.1 Ma) and Heteroponerini (33.8–61.1 Ma), our results shows relatively earlier divergences than previous molecular dating estimates for the groups (Brady et al. 2006; Moreau et al. 2006; Moreau and Bell 2013). Divergence dating analyses can be sensitive with regard to incorrectly placed fossil calibrations (Brady 2011), and an imbalance of ingroup vs. outgroup sampling and a lack of calibrations in the outgroup part of the phylogeny may be driving our age estimates. Conversely, our estimates may represent a considerable improvement to previous studies for the very reason that our sampling of ectatommine lineages is more comprehensive. The origin of the Ectatomminae was placed in the Neotropics by Moreau and Bell (2013). Our inference of a Neotropical origin for the Ectatommini, the oldest tribe in the subfamily, and for the sister lineage Heteroponerini, agrees with this hypothesis. Heteroponerini appeared to have had a history of evolution mainly in the Neotropical

632 region, except for one dispersal to the Australasia in the early Miocene. For the Ectatommini, our
633 analyses reconstructed an ancestral dispersal from the Neotropical to the Indo-Australian in the
634 middle Miocene, and to the Australasian region in the early Miocene (Fig. 4), where some groups
635 then appears to have undergone the majority of its diversification. Moreau and Bell (2013) have
636 suggested that the Neotropics functioned as a cradle for ant diversification; our biogeographic
637 results appear to be fully consistent with this hypothesis. We have indications of a Ectatomminae
638 origin in the Neotropics, and the diversification of these ants appears to be strongly associated with
639 this particular region, besides its broad diversification in other regions.

640 Nevertheless, our results are consistent with a scenario in which Ectatomminae provides
641 an example of vicariance between South America and Australasia in Southern Hemisphere
642 biogeography. The inferred mean stem age of Ectatomminae, which is also the date of the split
643 between the two tribes (79.2 Ma; 95% CI, 58.5–113.4 Ma; Fig. 4), is on average older than the age
644 estimates of the tectonic splits between Australasia and South America/Antarctica (35.5–52 Ma),
645 and between South America and Antarctica (36 Ma) (Scotese et al. 1988; Veevers et al. 1991;
646 Woodburne and Case 1996). The divergence between *Ectatomma* (Neotropics) and *Rhytidoponera*
647 (Australasia) and between the basal Indo-Australian *Gnamptogenys* and the other Neotropical
648 lineages also occurred around this period. These results suggest that these groups diverged in
649 Gondwana, implying that the current distribution of Ectatomminae is mainly explained by a long-
650 distance dispersal event, as found for some ant taxa (e.g. army ants, Brady 2003; Brady et al. 2014).
651 The later diversifications in the Australasian region (i.e. some *Heteroponera* and *Gnamptogenys*
652 lineages) probably occurred trough dispersion events, that could be traced back to the Miocene.

653 **Conclusions**

654 Our phylogenomic UCE data robustly resolved the relationships between major lineages
655 within Ectatomminae for the first time. We show that, despite the reciprocal monophyly between
656 the ectaheteromorph subfamilies are recovered and strongly supported, the morphological
657 framework for these groups do not justify the classification into two subfamilies, hence the
658 synonym of Heteroponerinae under Ectatomminae can provide a more functional classification for
659 the group. For this, we provide a revised diagnosis for the Ectatomminae, listing ten unequivocal
660 synapomorphies for the group, based on previous morphological studies. The genera
661 *Acanthoponera*, *Rhytidoponera* and *Ectatomma* are all recovered as monophyletic, with their
662 divergence dates and ancestral range presented for the first time. However, two novel results
663 surprisingly show the paraphyly of two ecologically important and frequently collected genera,

Gnamptogenys and *Heteroponera*. We show that the large genus *Gnamptogenys* is not monophyletic in relation to *Typhlomyrmex*, constituting several distinct lineages, many of which appear to have diversified in a time period similar to the other genus in the subfamily, and with an early divergence to the Indo-Australian region through Gondwana. Morphology appear to be misleading in the case of *Gnamptogenys*, where taxa are presently grouped together in the genus or its species groups not due to unique common descent, but rather because collectively they retain a similar, generalized morphology. In contrast, *Typhlomyrmex*, which evolved within the broad assemblage of *Gnamptogenys*-like forms, acquired distinctively derived morphological features that were also misleading when considered as an evolutionary trait. Thus, our finding of *Gnamptogenys* paraphyly is not merely a simple error of taxonomy, but a case where morphology was not elucidative of the group's evolutionary history. The genus *Heteroponera* is also paraphyletic, appearing to have a Neotropical origin with a later diversification trough dispersion to the Australasian region. In this case, morphology of *H. microps* previously indicated that some lineages within the genus may present striking divergences that could indicate a separate lineage, a result recovered by our molecular data.

In order to assure that classification of Ectatomminae ants reflects monophyletic groups, the synonymy of some genera or the description of others may proof necessary. Nevertheless, we strongly believe that the molecular phylogenetic data should be combined with the study of morphological characteristics that could be diagnostic for the newly described genus or the new combinations, so the final classification can be functional and fully assessed by any researcher trough the observation of specimens in the laboratory or in the field. In this sense, our phylogenetic results are valuable for providing a framework for the study of the subfamily morphological evolution, as well as for encouraging new systematic and taxonomic work on the group. Here, we demonstrate that UCE data provide a robust phylogenomic source for the Ectatomminae ants. When viewed from the perspective of morphological evolution, it is now evident that anatomic traits represent a strong asset for the classification of these groups. We believe that the phylogenetic framework and the new classification proposed here will shed light on the study of Ectatomminae taxonomy and systematics, as well as for the morphological evolution of the groups that it comprises.

693 **References**

694 Baroni-Urbani C., Bolton B., Ward P.S. (1992). The internal phylogeny of ants (Hymenoptera:
695 Formicidae). Systematic Entomology, 17:301–29.

696 Blaimer B.B., Brady S.G., Schultz T.R., Lloyd M.W., Fisher B.L., Ward P.S. (2015).
697 Phylogenomic methods outperform traditional multi-locus approaches in resolving deep
698 evolutionary history: a case study of formicine ants. BMC Evolutionary Biology, 15:271.

699 Blaimer B.B., LaPolla J.S., Branstetter M.G., Lloyd M.W., Brady S.G. (2016a). Phylogenomics,
700 biogeography and diversification of obligate mealybug-tending ants in the genus
701 Acropyga. Molecular Phylogenetics and Evolution, 102: 20–9.

702 Blaimer B.B., Lloyd M.W., Guillory W.X., Brady S.G. (2016b). Sequence capture and
703 phylogenetic utility of genomic ultraconserved elements obtained from pinned insect
704 specimens. PLoS One, 11:1–20.

705 Blumenstiel B., Cibulskis K., Fisher S., DeFelice M., Barry A., Fennell T., Abreu J., Minie B.,
706 Costello M., Young G. (2010). Targeted exon sequencing by in-solution hybrid selection.
707 Current Protocols in Human Genetics, 66:18.4:18.4.1–18.4.24.

708 Bolger A.M., Lohse M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina
709 Sequence Data. Bioinformatics, 30: 2114–2120.

710 Bolton B. (1995). A new general catalogue of the ants of the world.. p. 504.

711 Bolton B. (2003). Synopsis and classification of Formicidae. Memoirs of American Entomological
712 Institute. 369 p.

713 Bolton, B. (2017). An online catalog of the ants of the world. Available from http://antcat.org.
714 (accessed 12/09/2017).

715 Borgmeier T. (1957). Myrmecologische Studien. Anais da Academia Brasileira de Ciencias, 29:
716 103-128.

717 Borowiec M.L., Lee E.K., Chiu J.C., Plachetzki D.C. (2015). Extracting phylogenetic signal and
718 accounting for bias in whole-genome data sets supports the Ctenophora as sister to
719 remaining Metazoa. BMC Genomics, 16:987.

720 Bossert S., Murray E.A., Blaimer B.B., Danforth B.N. (2017). The impact of GC bias on
721 phylogenetic accuracy using targeted enrichment phylogenomic data. Molecular
722 Phylogenetics and Evolution, 111:149–57.

723 Brady S.G. (2003). Evolution of the army ant syndrome: The origin and long-term evolutionary
724 stasis of a complex of behavioral and reproductive adaptations. Proceedings of the National
725 Academy of Sciences, 100:6575–9.

726 Brady S.G., Schultz T.R., Fisher B.L., Ward P.S. (2006). Evaluating alternative hypotheses for the
727 early evolution and diversification of ants. Proceedings of the National Academy of
728 Sciences, 103:18172–7.

729 Brady S.G. (2011). Effects of fossil calibration uncertainty on divergence dating in ants and bees.
730 American Entomologist, 57:56–7.

731 Brady S., Fisher B., Schultz T., Ward P. (2014). The rise of army ants and their relatives:
732 diversification of specialized predatory doryline ants. BMC Evolutionary Biology, 14:93.

733 Brandão C.R.F., Lattke J.E. (1990). Description of a new Ecuadorean *Gnamptogenys* species
734 (Hymenoptera: Formicidae), with a discussion on the status of the *Alfaria* group. Journal
735 of the New York Entomological Society, 98:489-494.

736 Branstetter M.G., Longino J.T., Ward P.S., Faircloth B.C. (2017). Enriching the ant tree of life:
737 enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera.
738 Methods in Ecology and Evolution, 8: 768–776.

739 Camacho, G.P., Feitosa, R.M. (2015). Estado da arte sobre a Filogenia, Taxonomia e Biologia de
740 Ectatomminae. In: As formigas Poneromorfas do Brasil. – pp. 23-32.

741 Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in
742 phylogenetic analysis. Molecular Biology and Evolution, 17: 540-552.

743 Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., Glenn T.C. (2012).
744 More than 1000 ultraconserved elements provide evidence that turtles are the sister group
745 of archosaurs. Biology Letters, 8: 783 LP-786.

746 Delsuc F., Brinkmann H., Philippe H. (2005). Phylogenomics and the reconstruction of the tree of
747 life. Nat Rev Genet, 6: 361-375.

748 Emery C. (1911). Hymenoptera, Fam. Formicidae, Subfam. Ponerinae. In: Wytsman, P. Genera
749 Insectorum. Fasc. 118. 124 pp.

750 Faircloth B.C., Glenn T.C. (2012). Not all sequence tags are created equal: designing and
751 validating sequence identification tags robust to indels. PLoS ONE, 7: e42543.

752 Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C.
753 (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple
754 evolutionary timescales. Systematic Biology, 61.

755 Faircloth, B.C. (2013). Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality
756 trimming. Available at: http://dx.doi.org/10.6079/J9ILL.

757 Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. (2013). A Phylogenomic Perspective on the
758 Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved
759 Elements (UCEs). PLOS ONE 8: e65923.

760 Faircloth B.C., Branstetter M.G., White N.D., Brady S.G. (2015). Target enrichment of
761 ultraconserved elements from arthropods provides a genomic perspective on relationships
762 among Hymenoptera. Molecular Ecology Resources, 15:489–501.

763 Feitosa R.M. (2011). Revisão taxonômica e análise filogenética de Heteroponerinae
764 (Hymenoptera, Formicidae). Universidade de São Paulo.

765 Feitosa, R.M. (2015). Estado da arte sobre a Filogenia, Taxonomia e Biologia de Heteroponerinae.
766 In: As formigas Poneromorfas do Brasil. – pp. 33-43.

767 Fisher S., Barry A., Abreu J., Minie B., Nolan J., Delorey T.M., Young G., Fennell T.J., Allen A.,
768 Ambrogio L. (2011). A scalable, fully automated process for construction of sequence-
769 ready human exome targeted capture libraries. Genome Biology, 12: R1.

770 Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L.
771 et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference
772 genome. Nature Biotechnology 29, 644–652.

773 Hölldobler B., Wilson E.O. (2008). The Superorganism: The Beauty, Elegance, and Strangeness
774 of Insect Societies. Norton, 320 p.

775 Janicki J., Narula N., Ziegler M., Guenard, B., Economo E.P. (2016). Visualizing and interacting
776 with large-volume biodiversity data using client-server web-mapping applications: The
777 design and implementation of antmaps.org. Ecological Informatics, 32: 185-193.

778 Johnson B.R., Borowiec M.L., Chiu J.C., Lee E.K., Atallah J., Ward P.S. (2013). Phylogenomics
779 resolves evolutionary relationships among ants, bees, and wasps. Current Biology, 23:
780 2058-2062.

781 Katoh K., Asimenos G., Toh H. (2009). Multiple alignment of DNA sequences with MAFFT. In:
782 Bioinformatics for DNA sequence analysis. Springer; 2009: 39–64.

783 Keller R.A. (2000). Cladistics of the tribe Ectatommini (Hymenoptera: Formicidae): a reappraisal.
784 Insect Systematics & Evolution, 31: 59-69.

785 Keller R.A. (2011). A Phylogenetic Analysis of Ant Morphology (Hymenoptera: Formicidae) with
786 Special Reference to the Poneromorph Subfamilies. Bulletin of the American Museum of
787 Natural History, 355:1–90.

788 Kumar S., Stecher G., Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis
789 Version 7.0 for Bigger Datasets, Molecular Biology and Evolution, 33: 1870–1874.

790 Lattke J.E. (1994). Phylogenetic relationships and classification of Ectatommine ants
791 (Hymenoptera, Formicidae). Entomologica Scandinavica, 25: 105-119.

792 Lattke J. (1995). Revision of the Ant Genus *Gnamptogenys* in the New world (Hymenoptera:
793 Formicidae). Journal of Hymenoptera Research, 4:137–93.

794 Lattke J.E. (2004). A taxonomic revision and phylogenetic analysis of the ant genus
795 *Gnamptogenys* Roger in Southeast Asia and Australasia (Hymenoptera: Formicidae:
796 Ponerinae). University of California Publications, Entomology, 122: 1-266.

797 Lewis P.O. (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological
798 Character Data. Systematic Biology, 50: 913–925.

799 McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T.
800 (2013). A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment
801 and High-Throughput Sequencing. PLoS ONE, 8: e54848.

802 Moreau C.S., Bell C.D., Vila R., Archibald S.B., Pierce N.E. (2006). Phylogeny of the Ants:

803 Diversification in the Age of Angiosperms. Science, 312:101–4.

804 Moreau C.S., Bell C.D. (2013). Testing the museum versus cradle tropical biological diversity
805 hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the
806 ants. Evolution; 67:2240–57.

807 Nei M., Kumar S. (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New
808 York.

809 Nettel-Hernanz A., Lachaud J., Fresneau D., López-Muñoz R.A., Poteaux C. (2015).
810 Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant
811 genus Ectatomma Smith, 1958 (Formicidae, Ectatomminae). Organisms Diversity &
812 Evolution, 15: 543-553.

813 Ouellette G.D., Fisher B.L., Girman D.J. (2006). Molecular systematics of basal subfamilies of
814 ants using 28S rRNA (Hymenoptera: Formicidae). Molecular Phylogenetics and
815 Evolution, 40: 359-369.

816 Pagel M. (1999). Inferring the historical patterns of biological evolution. Nature, 401:877.

817 Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G, et al.
818 (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough.
819 PLoS Biology, 9(3): e1000602.

820 Phillips M.J., Delsuc F., Penny D. (2004). Genome-scale phylogeny and the detection of
821 systematic biases. Molecular Biology and Evolution, 21:1455–8.

822 R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for
823 Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

824 Rabeling C., Brown J.M., Verhaagh M. (2008). Newly discovered sister lineage sheds light on
825 early ant evolution. Proceedings of the National Academy of Sciences, 105:14913–7.

826 Sayyari E., Mirarab S. (2016). Fast coalescent-based computation of local branch support from
827 quartet frequencies. Molecular Biology and Evolution, 33:msw079.

828 Schluter D., Price T., Mooers A.O., Ludwig D. (1997). Likelihood of ancestor states in adaptive
829 radiation. Evolution, 51: 1699-1711.

830 Schmidt C.A. (2013). Molecular phylogenetics and taxonomic revision of ponerine ants

(Hymenoptera: Formicidae: Ponerinae). Zootaxa, 3647:201–50.

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. (2013). Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. Systematic Biology, 63: 83-95.

Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30:1312–3.

Taylor R.W. (1980). Notes on the Russian endemic ant genus Aulacopone Arnoldi (Hymenoptera: Formicidae). Psyche, 86: 353-361.

Tamura K., Dudley J., Nei M., Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Veevers J.J., Powell C.M., Roots S.R. (1991). Review of seafloor spreading around Australia. I. synthesis of the patterns of spreading. Australian Journal of Earth Sciences, 38:373–89.

Ward P.S. (1980). A systematic revision of the *Rhytidoponera impressa* group (Hymenoptera: Formicidae) in Australia and New Guinea. Australian Journal of Zoology, 28: 341-365.

Ward P.S. (1984). A revision of the ant genus *Rhytidoponera* (Hymenoptera: Formicidae) in New Caledonia. Australian Journal of Zoology, 32: 131-175.

Ward P.S. (2007). Phylogeny, classification, and species-level taxonomy of ants (Hymenoptera: Formicidae). Zootaxa, 563:549–63.

Ward P.S., Brady S.G. (2003). Phylogeny and biogeography of the ant subfamily Prionomyrmecinae (Hymenoptera: Formicidae). Invertebrate Systematics, 17:361–86.

Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. (2010). Phylogeny and biogeography of dolichoderine ants: Effects of data partitioning and relict taxa on historical inference. Systematic Biology, 59:342–62.

Ward P.S. (2014). The Phylogeny and Evolution of Ants. Annual Review of Ecology, Evolution, and Systematics, 45: 23-43.

Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. (2015). The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). Systematic Entomology, 40:61–81.

Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. (2015). The evolution of myrmicine ants: Phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). Systematic Entomology, 40:61–81.

Ward P.S., Fisher B.L. (2016). Tales of dracula ants: The evolutionary history of the ant subfamily Amblyoponinae (Hymenoptera: Formicidae). Systematic Entomology, 41:683–93.

Ward P.S., Blaimer B.B., Fisher B.L. (2016). A revised phylogenetic classification of the ant subfamily Formicinae (Hymenoptera: Formicidae), with resurrection of the genera *Colobopsis* and *Dinomyrmex*. Zootaxa, 4072:343–57.

Woodburne M.O., Case J. (1996). Dispersal, vicariance, and the Late Cretaceous to early tertiary land mammal biogeography from South America to Australia. Journal of Mammal Evolution, 3:121–61.

Capítulo 2: Exploring Gene Tree Incongruence at the Root of the Aculeata Tree (Hymenoptera)

Introduction

The advent of third-generation sequencing has allowed for the accumulation of genetic data at an unprecedented rate, providing exciting opportunities in a variety of disciplines, including population genetics, phylogeography, and molecular phylogenetics (Carstens et al. 2012; Puritz et al. 2012; McCormack et al. 2013; Ruane et al. 2015; Romiguier et al. 2016; Welch et al. 2016). One of the hopes of analyzing extensive multi-locus datasets is the possibility of solving many vexing problems in phylogenetics, such as the higher relationships among bird lineages (McCormack et al. 2013), the phylogeny of placental mammals (Mccormack et al. 2012), or the phylogenetic position of turtles in tetrapods (Crawford et al. 2012). These problems are stimulating particularly due to the potential incongruence between species trees and their underlying gene trees (Maddison 1997). In fact, the assumption of a single evolutionary history for all loci that underlies the hitherto common practice of concatenation might cause the resulting inference to be statistically inconsistent, particularly near relatively short internodes (Degnan and Rosenberg 2009; Edwards et al. 2007). As a consequence, there has been a progressively stronger reliance on methods based on the multispecies coalescent model (Takahata et al. 1995; Rannala and Yang 2003), which generally assumes that discrepancies between gene trees and the species tree are exclusively due to deep coalescence (i.e. all loci are not under the influence of selection, the population is panmictic, and there is no recombination within each gene but free recombination between genes).

A tacit assumption of many phylogenomic efforts is that one could solve long-standing problems in molecular phylogenetics by "brute force". For instance, studies using only a handful of loci frequently show some nodes with relatively low support, which are followed by the author's suggestion that more data are necessary to resolve that particular issue. However, as more extensive datasets are increasingly common, many species tree methods based on the multispecies coalescent model such as BEST (Liu 2008) and *BEAST (Heled and Drummond 2010) are becoming increasingly impractical. This gives way to the use of a two-step process of first inferring gene trees for each locus using a robust method such as maximum likelihood and then using the obtained gene trees as input for species tree inference, as in STAR (Liu et al. 2009), STEAC (Liu

et al. 2009), GLASS (Mossel and Roch 2010), MP-EST (Liu et al. 2010), NJst (Liu and Yu 2011), and ASTRAL-II (Mirarab et al 2015). Those methods are statistically consistent to the multispecies coalescent model if gene trees are known without error (Liu et al. 2009; Liu and Yu 2011; Allman et al. 2013). Yet, the extent to which this assumption is violated in real datasets is poorly known. However, the fact that different phylogenomic datasets can lead to highly supported, but mutually inconsistent results is not only alarming, but suggest that inconsistency among gene trees in real datasets could be severe. For instance, Johnson et al. (2013) used transcriptome data on 19 species from all superfamilies of aculeate Hymenoptera (wasps, ants and bees) and found that ants and Apoidea (bees and spheciform wasps) were sister groups. This result is surprising given previous hypotheses that, although contradicting each other, never presented similar results (see Fig. 1a, b and c). Alternatively, Faircloth et al. (2014) used another dataset based on ultraconserved elements (UCEs, Faircloth et al. 2012) to infer the relationships among 44 taxa from six aculeate superfamilies, which placed ants at the base of the aculeate tree, as sister to the remaining aculeate lineages. More recently, Branstetter et al. (2017) showed that taxon sampling may be the cause for inconsistency among those two previous phylogenomic studies, since Faircloth et al. (2014) data set was biased towards ants and missed representatives of the Chrysidoidea, which proved to be an important outgroup. However, although the relationships between ants, bees and wasps seem relatively well resolved now (Johnson et al. 2013; Branstetter et al. 2017; Peters et al. 2017), the cause and degree of those conflicting sources of phylogenetical signal is currently unknown. Given that both datasets are fairly extensive, one cannot simply ascribe this inconsistency as noise. In the present study, we explore topological variation in gene trees to provide a detailed exploration of gene tree space in the data set of Johnson et al. (2013), to access the variation and inconsistency present in the data and how the final species tree recovers the position of ants as sister to apoid wasps and bees.

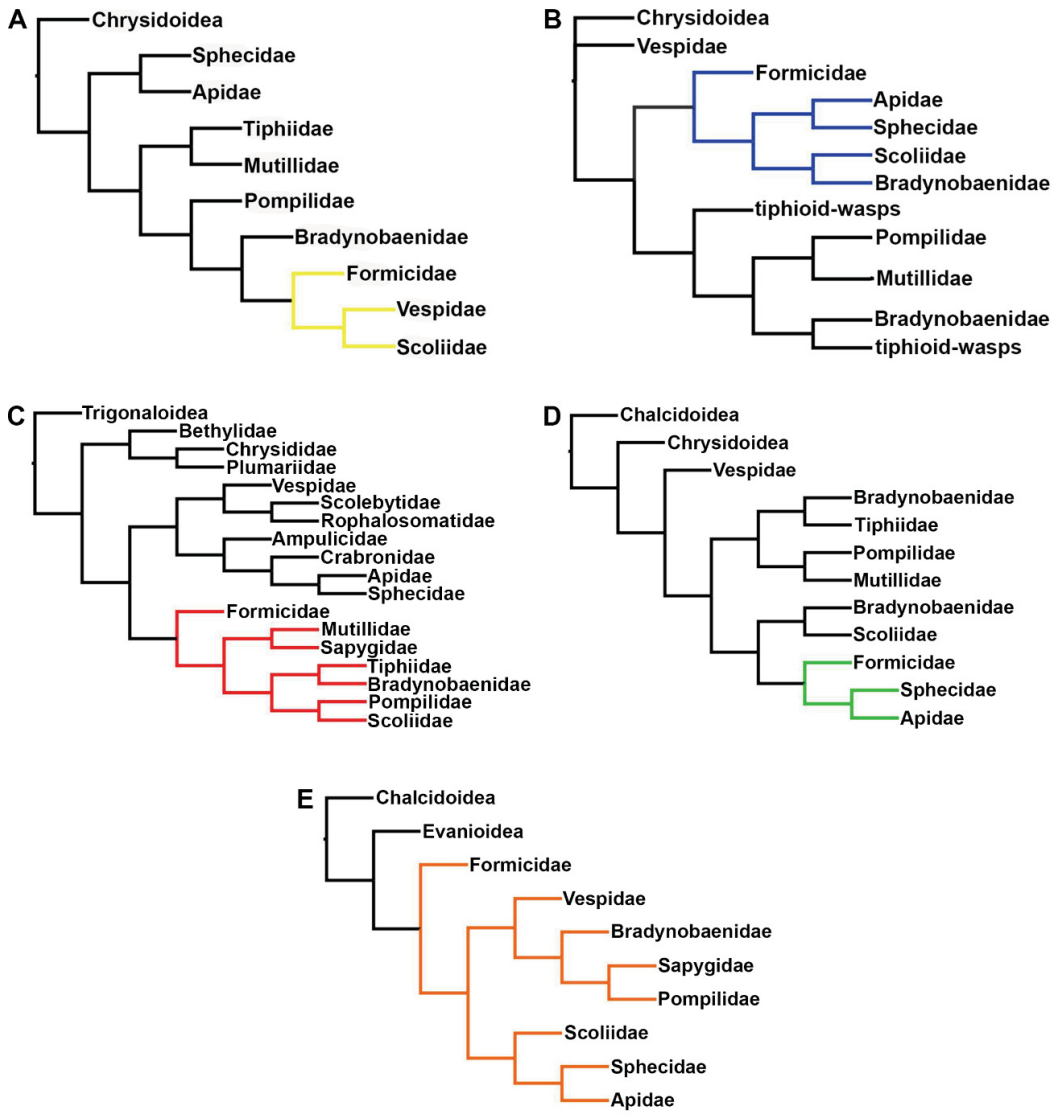


Figure 1. Previous hypotheses for the phylogenetic relationships of Formicidae among Aculeata. A) Brothers (1999), phylogenetic analysis of 92 morphological characters; B) Pilgrim et al. (2008), molecular analysis of four nuclear genes combined with the morphological matrix of Brothers (1999); C) Heraty et al. (2011), molecular analysis of three nuclear genes and one mitochondrial gene; D) Johnson et al. (2013), molecular analysis of 308 transcriptome loci; E) Faircloth et al. (2014), molecular analysis of 638 Ultra-Conserved Elements (UCEs) loci. The colored branches highlight the position of Formicidae among Aculeata, showing the clades formed by ants and their closest relatives.

Material and methods

Data set

We focused on the data generated by Johnson et al. (2013) as a test-case to assess incongruence among gene trees and their effects on phylogenetic inference. Although Johnson et al. (2013) analyzed four different datasets with varying levels of missing data, we chose to focus

938 only on their most complete matrix, with 308 genes x 19 taxa (175,404 sites, of which 73.42% are
939 coded as amino acids, 11.60% are gaps, and 14.98% are missing) to avoid conflating the effects of
940 missing data and gene tree incongruence. The matrix was obtained directly from Dryad
941 (http://doi.org/10.5061/dryad.jt440).

942 Gene tree and species tree estimation

943 Each gene tree was inferred by maximum likelihood (ML) with RAxML v8.2.x (Stamatakis
944 2014) using the Supercomputing Center of Ohio State University (OSU), USA. We used the same
945 model of evolution for each gene chosen in the original study, determined by comparing 36
946 possible protein models (see Johnson et al. 2013 for details). Branch support on each gene tree was
947 determined by 1000 bootstrap replicates in RAxML, using stabilization of the majority-rule
948 consensus tree as stopping criterion. The consensus gene trees were obtained from 100 randomly
949 selected bootstrap replicates for each of the 308 genes, to standardize the number of replicates for
950 every gene tree, since some of the genes reached convergence (i.e. bootstrapping was halted by
951 RAxML) before the completion of 1000 replicates. We used 100 replicates to generate the
952 consensus because this is the number usually referred as representing sufficient replicates for
953 phylogenomic data sets (see Johnson et al. 2013; Branstetter et al. 2017; Blaimer et al. 2016). Trees
954 were rooted using *Nasonia vitripennis* (Pteromalidae) as the outgroup.

955 We estimated species trees based on two different summary statistics approaches. First, we
956 used methods based on the average time of gene coalescence events (consensus method), as
957 implemented in STAR (Liu et al. 2009) and ASTRAL-II v4.7.12 (Mirarab et al. 2015). Second,
958 we used MP-EST (Liu et al. 2010) to generate a maximum pseudo-likelihood estimate under the
959 multispecies coalescent model (gene tree based coalescent method). As input for the species tree
960 analysis, we used all individual gene trees and accompanying bootstrap trees as input (100
961 replicates). Bootstrap support values where calculated based on multilocus bootstraping method
962 by Seo (2008) for all the species tree analyses.

963 Evaluation of individual genes support

964 To evaluate the informativeness of each of the loci, we used five metrics that could
965 potentially indicate locus informativeness for gene tree inference, namely the mean bootstrap
966 support across all nodes of the gene tree, locus length (number of base pairs), number of bootstrap
967 replicates needed for a stable consensus, Robinson-Foulds (RF) distance between individual gene
968 trees and the reference species tree based on all loci (Robinson and Foulds 1981), and the

969 proportion of parsimony informative sites. To calculate the percentage of parsimony informative
970 sites by gene, we used the package IPS 0.0-7 (Heibl 2014) in R 3.1.3 (R Core Team 2017). The
971 level of correlation between the metrics was assessed by a linear model using the package VEGAN
972 2.4-2. We also assessed support for the 18 nodes in the reference topology and support for each
973 one of the alternative hypothesis for the position of ants among Aculeata, by evaluating the fraction
974 of the 100 bootstrap pseudo-replicates computed for each gene tree that recovered each one of
975 those nodes and hypothesis. Results were summarized as heatmaps using the packages STATS
976 v3.4.0 and RCOLORBREWER v1.1-2 (Neuwirth 2007) on R v3.1.3.
977 The topological variation among all gene trees was assessed by their pairwise RF distances
978 using the package PHANGORN 2.0.3 (Schliep 2011). The RF distance is a metric that determines
979 the number of bipartitions that differ between a pair of trees to indicate the amount of topological
980 discordance between them. A matrix of RF distances among gene trees was subject to
981 multidimensional scaling (MDS, see Hillis et al. 2005). We focused on the first two ordination
982 axes and colored the obtained results according to their support for each of five alternative sister-
983 group relationships for ants (Fig. 1), namely Brothers (1999), Pilgrim et al. (2008), Heraty et al.
984 (2011), Johnson et al. (2013), and Faircloth et al (2014).
985 To better visualize the similarities between the gene trees topologies, we constructed a
986 METATREE v2 (Nye 2008), which cluster similar tree topologies together, so that conflicting
987 evolutionary histories within a set of trees are apparent as separate clades on the meta-tree (Nye
988 2008). METATREE builds a "tree-of-trees" that shows the relationships between alternative
989 phylogenies. Internal nodes correspond to the 50% majority-rule consensus of "sister" trees and
990 branch lengths computed as Robinson-Foulds distances. This program takes as input gene trees to
991 construct a tree that clusters together similar topologies, allowing the user to infer if genes
992 supporting any given topology tend to cluster together, indicating possible signal conflict with the
993 remaining dataset. For our METATREE analyses, we used as input the 308 gene trees generated by
994 RAxML analysis.
995 Evaluation of emergent support
996 The Random Addition Concatenation Analysis (RADICAL) is a method that evaluates the
997 effects of partitioning and combining genes in genome-level analyses (Narechania et al. 2012).
998 The method was developed to generate trees along a set of replicated random concatenation steps,
999 which range from one gene to all genes in the dataset. RADICAL catalogs tree heterogeneity while

allowing the visualization of emergent support through concatenation, i.e., how the addition of sets of genes influences tree topology and support (Narechania et al. 2012). Moreover, RADICAL monitors the dynamics of concatenation by calculating support statistics for the topologies generated at each step and comparing them to the whole library of trees (Narechania et al. 2012). Since our aim is to assess incongruence among individual gene trees and the resulting species trees, we developed a custom pipeline using the same logic, but based on coalescent analysis.

The analysis starts with 100 randomly chosen pairs of genes, each one giving rise to "analysis paths". A "path" is constructed by incrementally adding genes to the dataset. In RADICAL, this is done by concatenating genes randomly selected (with no replacement) and analyzing the new alignment using RAxML. Our approach estimates a new gene tree at each step of the path using MP-EST. In order to decrease execution time, steps were constructed using batches of 18 genes (approximately the square root of 308, so each path is also made up by 18 steps). Thus, the first step of the path is the species tree corresponding to 2 (randomly) chosen genes, the second is a species tree inferred from 20 genes (two genes from the first step plus 18 additional genes, also randomly sampled with no replacement), the third is inferred from 38 genes, etc. The analysis ends when all 308 genes were added to each replicate. Thus, the last tree in each path is the "total-evidence" species tree. With this data set in hands, we employed the same approach used by the authors of RADICAL to evaluate the phylogenetic support for each node of that tree. This was accomplished by calculating, at each step, the percentage of replicate species trees that contains a given node present in the total-evidence tree. If phylogenetic signal for any given node is strong in the majority of genes, it will be present in most species trees already at the first steps. If a node shows up in all (100%) species trees after a certain step, it has become "fixed" in the analysis. The strength of the phylogenetic signal supporting that node is thus inversely proportional to the number of steps required until fixation. In order words, the data provide much stronger evidence for a node that goes to fixation after, for instance, two steps in a path (when species tree were computed from random combinations of 20 genes) than after 10 steps, when much larger randomized datasets (164 genes) were used.

We also calculated the normalized Consensus Fork Index (CFI) (Colless 1980) by counting the number of identical nodes between the total-evidence species trees and each replicate species tree along the 100 paths, divided by the maximum number of nodes possible (18). Normalized CFIs vary between 0 and 1, where 0 indicates total-evidence and replicate species trees with no

nodes in common while 1 means identical trees. The distribution of normalized CFIs at each step were represented by density kernels, using the VIOPLOT v0.2 package (Hintze and Nelson 1998) in R v3.1.3. In order to test how the different alternative topological hypotheses are recovered throughout the paths, we used the function resolveAllNodes in PHYTOOLS v0.6-20 (Revell 2012), implemented in R v3.1.3.

Results

Species tree estimation

The species tree analyses produced fully resolved trees of the aculeate Hymenoptera, always supporting the hypothesis of ants as sister group to Apoidea, with robust support for all nodes with all three methods, as seen in Johnson et al. (2013). However, the tree topologies and support values varied among the different methods of estimation. STAR and ASTRAL-II analysis presented the same topology as Johnson et al. (2013), with slightly different bootstrap support values (Fig. 2B). Alternatively, the MP-EST analysis presented a very similar topology, but with the bootstrap support for the relationship between Formicidae and Apoidea much lower than both other methods (Fig. 2A). Also, the MP-EST analysis recovered a sister group relationship between Vespoidea and the thipioid-pompilioid wasps with low support, a result that is congruent with the findings of Branstetter et al. (2017), but not found in the STAR analysis by Johnson et al. (2013) or in the concatenated analysis by Peters et al. (2017). We choose the MP-EST species tree as our reference tree for the other analysis.

Evaluation of individual genes support

For the present data set, we noticed that many of the 308 loci are short (ranging from 141 to 1907 nucleotide sites, with a mean of 570 sites per locus) and the number of parsimony informative sites within each locus is low (ranging from 2 to 115, with a mean of 28 per locus). This renders many loci potentially lacking in phylogenetic information for gene tree estimation. However, the mean value of bootstrap support for each gene tree is not low, ranging from 21 to 79, with a grand mean of 48, which shows that despite the small number of parsimony informative sites per gene, the gene trees reconstructed seem reasonably well supported. Likewise, we noticed that bootstrapping was not halted until 1000 bootstrap replicates for most gene trees, suggesting that those genes lack strong phylogenetic signal supporting any particular topology. Additionally, the nodes that presented maximum bootstrap values on the final species tree also present high

support in each individual gene. This can be seen in the heatmap (Fig. 3A), where the support values per gene tree at each node of the resulting species tree are graphically represented.

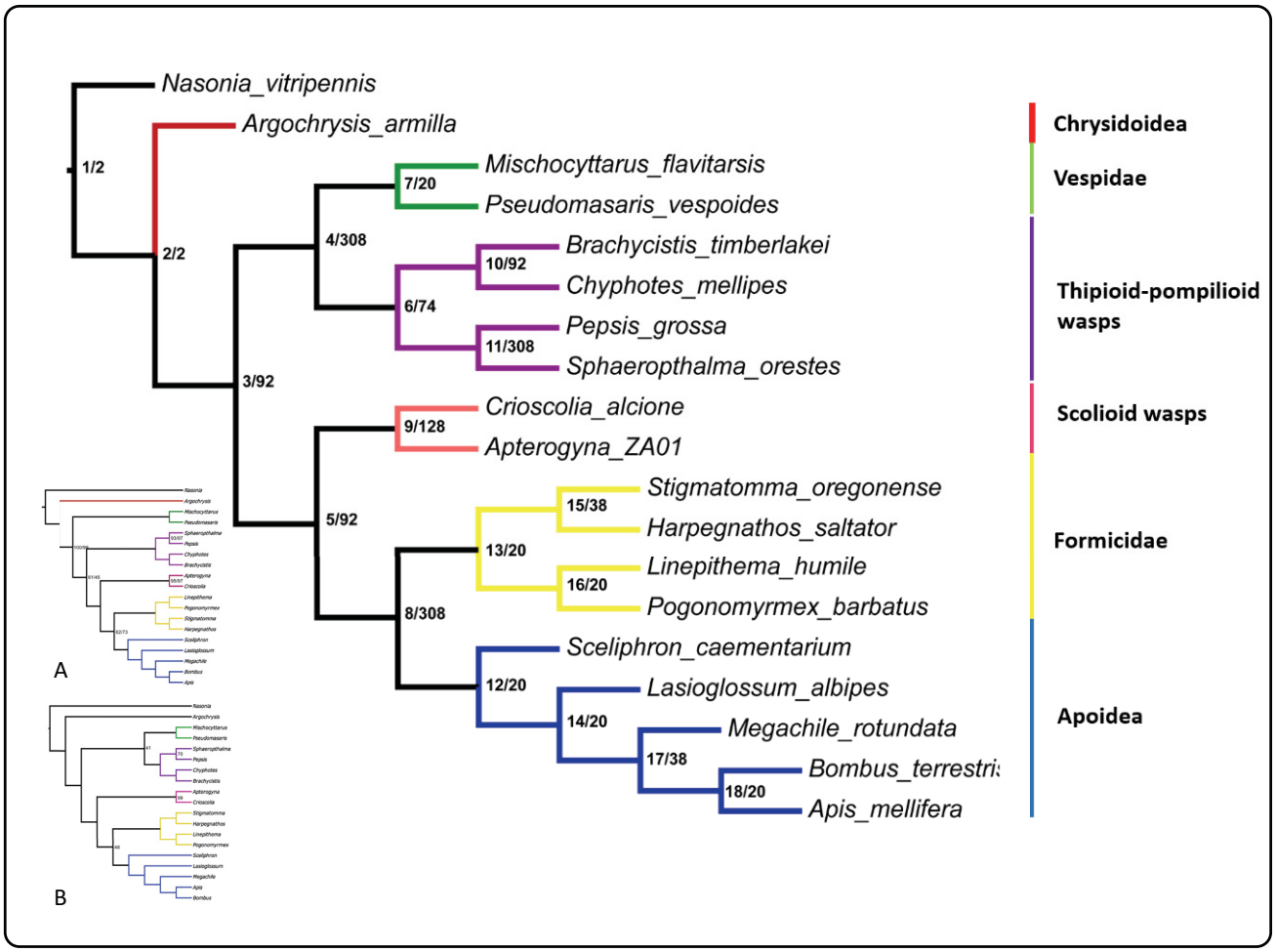


Figure 2. Aculeate Hymenoptera species tree resulting from a MP-Est analysis of 308 individual gene trees. Numbers are: node number/number of genes necessary for the clade to be present in 100 species tree replicates of the emergent support analysis. In detail, species trees generated with A) STAR / ASTRAL-II; B) MP-EST. The bootstrap support values not shown are equal to 100.

Regarding the topological variations, we can notice that very few gene trees recovers the alternative topologies in a very low frequency (Fig. 3B; Brothers 1999, Pilgrim et al. 2008). A larger number of gene trees recovers the relationship between ants and Apoidea (Fig. 3B, Johnson et al. 2013) at higher frequencies, while the hypothesis of Heraty et al. (2008) and Faircloth et al. (2014) are never recovered with this data set. The constraint analysis using the RF distance between gene trees showed that only 39 gene trees (12% of all the 308) were consistent with the final species tree hypothesis that ants and Apoidea are sister groups (Fig. 4, in purple). Additionally, we notice that these gene tree topologies that recover this hypothesis are not closely

related, that is, the data is not bias for this particular relationship. The 39 gene trees that are congruent with the hypothesis of a sister group relationship between ants and apoids are not clustered in clades in the MetaTree, but dispersed throughout the entire tree space (purple branches, Fig. 5). The leaf branches on our meta-tree are long, showing that the gene trees are quite dissimilar, and the internal vertices on the meta-tree correspondingly consist of highly unresolved trees (Fig. 5). We can notice that the meta-tree presents long edges radiating from a few central vertices, and this is typical for sets of trees with a high degree of conflict (Nye 2008).

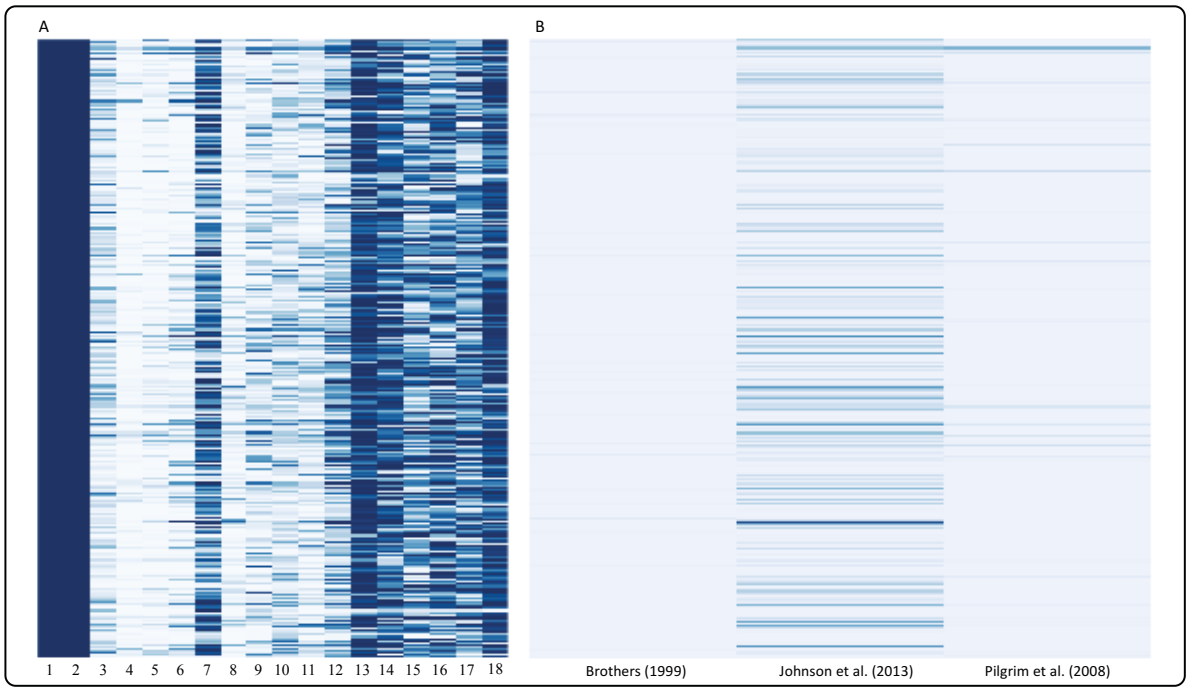


Figure 3. Heatmaps depicting the percentage of times that each node (A) of the reference topology (Fig. 2) or each alternative topology (B) previously published (Fig. 1) was found in 100 randomly selected bootstrap pseudo-replicates of the 308 gene trees. The two most basal nodes (1 and 2) were recovered in all pseudo-replicates of all genes. The topologies not depicted here (Heraty et al. 2011; Faircloth et al. 2014) presented a total frequency of 0. Each line along the columns corresponds to one gene and the number of bootstrap replicates recovering the node is color coded, the lightest colors being the lowest values.

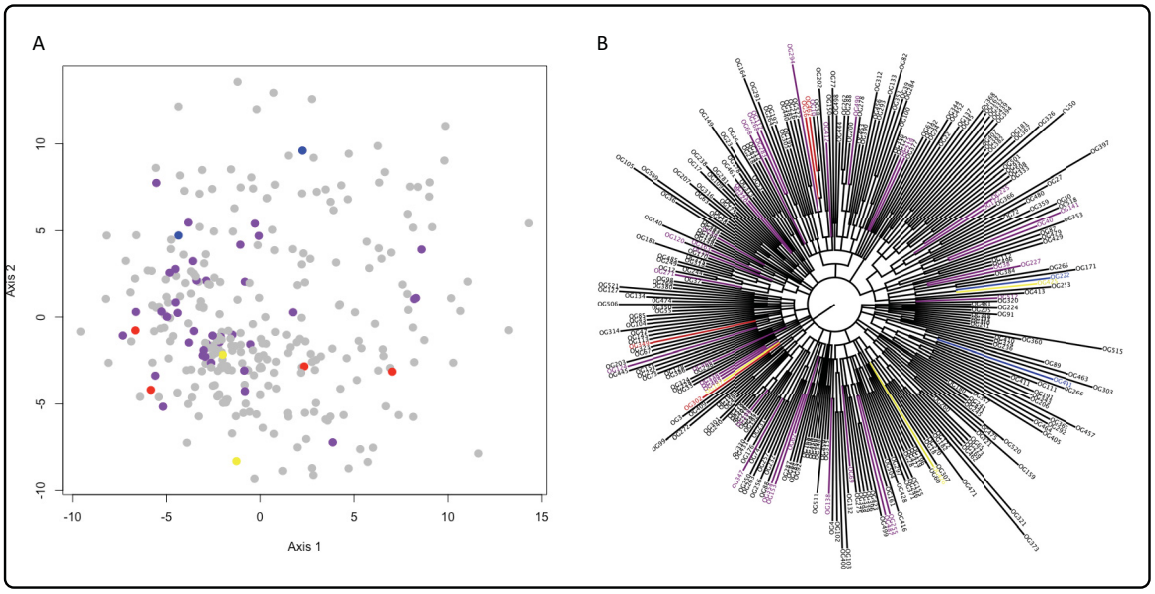


Figure 4. A) Non-metric multidimensional scaling (MDS) showing the Robinson-Foulds distance between the 308 gene trees. The colored points represent the alternative hypotheses of Brothers 1999 (yellow), Pilgrim et al. 2008 (blue), Heraty et al. 2011 (red) and the original hypotheses of Johnson et al. 2013 (purple). In grey are the genes there are not consistent with none of the previous hypotheses. B) MetaTree showing the relationships between all 308 gene trees generated in RAxML. Each leaf branch corresponds to a gene tree and interior nodes corresponds to consensus topologies between the gene trees in that clade. The differences in length between sister branches infers dissimilarities between the topologies, calculated as Robinson-Foulds distances from the consensus topologies of those trees. The colored branches represent the topologies that are consistent with the hypotheses for the position of ants among aculeates of Brothers 1999 (yellow), Pilgrim et al. 2008 (blue), Heraty et al. 2011 (red) and Johnson et al. 2013 (purple). In black are the genes tree topologies there are not consistent with the previous hypotheses.

Evaluation of emergent support

In the present data set, for each node in the resulting species tree (Fig. 2), the number of genes necessary to fix a given node varied and could be as low as two genes (for basal nodes). In fact, several of the nodes on the species tree achieved rapid fixation during the addition process (Fig. 5A), despite the substantial amount of incongruence among individual genes. Node 8, which corresponds to the relationship between Formicidae and Apoidea, exhibits a pattern that is more irregular than that of other nodes. The node exhibits a less stable trajectory when compared to the amount of emergent support present at other nodes. For instance, node 12 (Apoidea) appears in all (100) replicate species trees computed from randomized samples equal or larger than 20 genes, i.e. after 2 steps (Fig. 5A). Fixation of nodes 3 (Aculeata) and 10 (tiphioid complex), for example, were observed after 6 steps (92 genes), while node 9 (Scolioidea) required 8 steps (128 genes) (Fig. 5A). In contrast, after 17 steps (290 genes), node 8 (Ants and Apoidea) is only present in fewer than 80% of the replicate species trees (Fig. 5A). While the fixation paths for other nodes

show a clear asymptotic trend, increasing monotonically with the number of analyzed genes, node 8's path is much "wobblier" i.e., the percentage of replicate species trees containing this node decreased frequently among successive steps. Nevertheless, to recover the sister group relationship between ants and apoids (node 8), it is necessary to add all the 308 genes to the species tree analysis.

Averaged across all nodes, 50% of all the emergent support on the tree occurs by the time 20 genes have been added to the coalescence analysis, a data set size that comprises only 6.5% of the total gene space (Fig. 5B). The CFI values shows clear convergence towards the reference topology as more genes are added to the analysis (see node labels on Fig.2). Nevertheless, the kernels show that normalized CFI increases steadily with the number of genes. From step 7 (110 genes) onward, all replicate species tree recovers 15 of the 18 nodes (CFI 0.83) present in the reference topology. Note that, after step 8, median CFI (white dot) is 0.94, meaning that half of the replicate species trees differs by 1 or 0 nodes from the reference topology; the other half differs by 2 or 3 nodes. Median CFI becomes 1 at step 17.

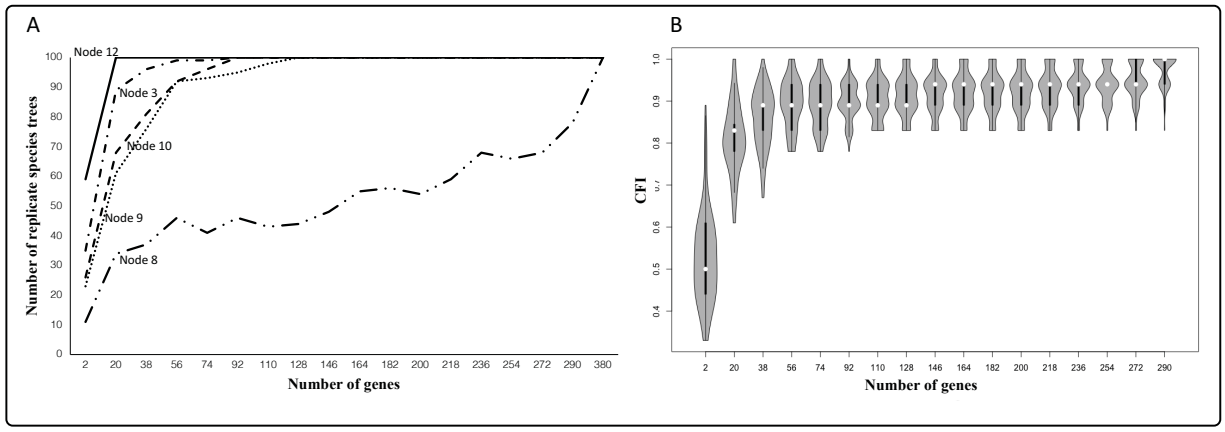


Figure 5. A) Example of fixation paths for selected nodes with different levels of support. Nodes are numbered according to Fig. 2. B) "Violin plot" of CFI values by the gene addition steps. The width of each kernel corresponds to the number of replicate species-trees, white dots within the kernels represent median CFI values, thick and thin lines are inter-quartile and min-max ranges, respectively. Normalized CFIs vary between 0 and 1, where 0 indicates trees with no nodes in common with the reference topology (Fig. 2) and 1 where all nodes are identical.

All nodes, except for 4 (Vespidae + Tiphioid-Pompilioid wasps), 8 (Formicidae + Apoidea) and 11 (Pompilidae + Mutillidae) are fixed after step 8 (128 genes). When collapsing these nodes and further resolving the polytomies generated, 27 distinct topologies are obtained. However, those topologies are not equally represented when adding more genes to the analysis. Figure 6A shows each of the 27 topologies and their frequencies among the 100 replicates at each step. With 128

genes, three topologies have equal representation among the trees, the red line being the same as the topology obtained with the 308 genes (Fig. 6A). The other two are even more frequent then the red one in several steps, but loose representation as more genes are added to the path.

When focusing on the position of Formicidae, only one of the alternative hypotheses (Fig. 1B) is recovered among the replicates with this data set. The sister group relationship between ants and a clade formed by Scolioidea + Apoidea is recovered in the same frequency as the Johnson et al. (2013) hypothesis of ants + Apoidea up to step 2 (20 genes). After that, with the addition of more genes, the former loses representation, as the later appears to be more and more frequent (Fig. 6B). A similar pattern is true for other hypothesis, not supported by any other work, for the sister group relationship between Apoidea and a clade formed by Formicidae + Scolioidea (Fig. 6B).

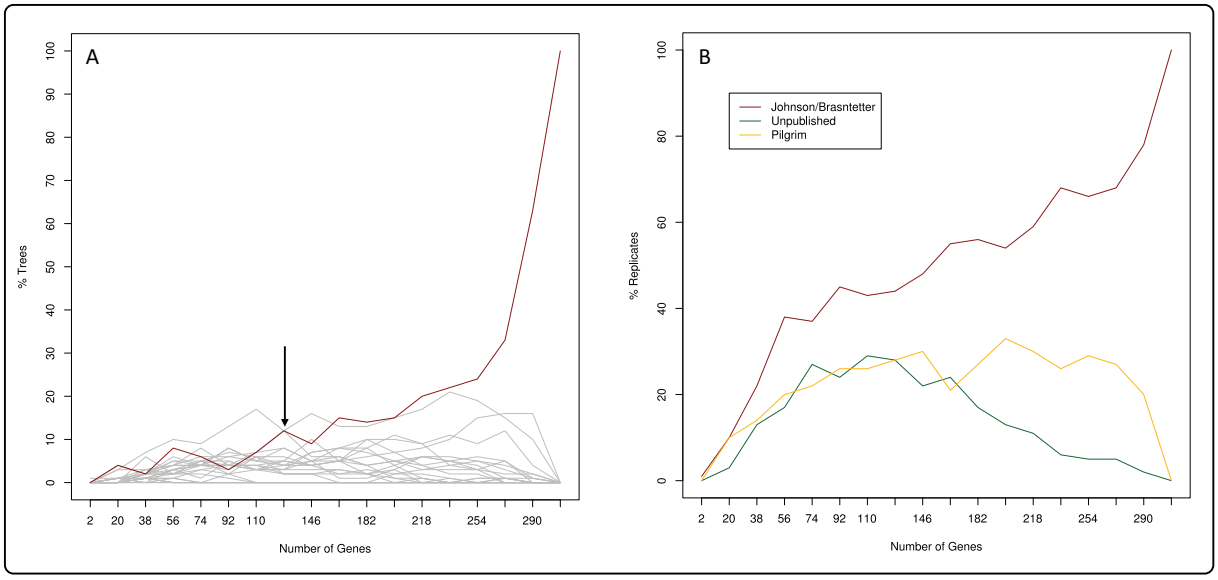


Figure 6. A) Graphical demonstration of the tree frequency of the 27 possible topologies (after the fixation of 15 out of 18 nodes in the main topology) in the addition path. The setae show step 8, where the frequency of three of those topologies is the same. B) Graphical demonstration of the frequency among replicates for the three-main hypothesis recovered by this data set.

Discussion

Here, using the dataset of Johnson et al. (2013) with a different summary statistics method (MP-EST), we find an almost identical result as the original authors, with the only difference being the sister group relationship between Vespidae and Tiphioid-Pompilioid wasps. This relationship

was further recovered by other study using a different phylogenomic data approach (target enrichment of Ultra-Conserved Elements) using Maximum likelihood, Bayesian Inference and Species tree analysis (Astral-II) (Branstetter et al. 2017). A different study using transcriptomes recovers the same topology as Johnson et al. (2013) does using STAR, through a concatenated analysis approach (ML) (Peters et al. 2017). These differences in performance by the three methods (STAR/ASTRAL-II and MP-EST) using the same data are remarkable, but fairly common. The main reason for these divergent results may be due to different algorithmic techniques, which can result in greater or lesser robustness to missing data (Springer and Gatesy 2015) and gene tree estimation error (Bayzid and Warnow 2012). Hence, the choice of coalescent-based method matters (Mirarab et al. 2015). These divergent results from different data types and analytical approaches shows that, although the large number of loci may be important for an accurate topological inference, principles such as obtaining sufficient taxon sampling (Leebens-Mack et al., 2005; Philippe et al., 2011), avoiding biases caused by alignment methods (Wong et al. 2008), incorporating secondary signal in concatenation analyses (Gatesy and Springer 2014), and performing rigorous tree searches (Simmons and Goloboff 2014) are extremely important for addressing systematic problems. However, there is a high congruence in the latest results regarding the position of ants among the Aculeata, with all studies recovering a sister group relationship among Formicidae and Apoidea. This suggests that, despite the previous morphological or single genes hypothesis for the position of ants, phylogenomic data strongly supports the novel results.

Regarding the quality of the loci used in the analysis, a little explored area that bears on the application of gene-tree-based coalescent methods to phylogenomic data is gene informativeness. Xi et al. (2015) demonstrated that genes with minimal phylogenetic information can produce unreliable gene trees (i.e., high error in gene tree estimation), which may in turn reduce the accuracy of species tree estimation using gene-tree-based coalescent methods. However, none of the parameters used here to measure gene informativeness (mean bootstrap support across all nodes of the gene tree, locus length, number of bootstrap replicates needed for a stable consensus, Robinson-Foulds (RF) distance between individual gene trees and the reference species tree based on all loci, and the proportion of parsimony informative sites) shows statistical correlation to each other, suggesting that there is not a single parameter that may be influencing the gene tree informativeness or the construction of reliable gene trees. In this sense, it is clear that finding which genes are better for gene tree estimation and a posterior species tree analysis can be a very difficult

1194 task, but sampling a large number of genes can alleviate problems of low phylogenetic
1195 informativeness. Such lack of phylogenetic information is likely to be especially problematic for
1196 regions of the species tree where internal branches are very short (Townsend 2007). Mirarab et al.
1197 (2015) demonstrated that gene tree estimation error is high for short genes (which can be less
1198 informative than longer genes), especially when species tree branch lengths are short.
1199 Nevertheless, under these circumstances, gene trees estimated from alignments with minimal
1200 phylogenetic information may reduce the accuracy of gene-tree-based coalescent methods (or any
1201 coalescent method for that matter), in the same way that weak or uninformative concatenated
1202 alignments will reduce the accuracy of concatenation methods.
1203 When looking for the mean support of individual nodes in the final species tree in each
1204 gene tree, is apparent that the inconsistency among gene trees in this dataset is uneven. While some
1205 nodes are strongly supported in most of the gene trees (i.e. node 13, family Formicidae), some
1206 nodes present a very low support individually among the gene trees (Fig. 3A). This is the case for
1207 the relationship between Formicidae and Apoidea (node 8), indicating that most of the genes used
1208 in the analyses do not support this hypothesis. Those results indicate that, when looking at
1209 individual genes, the relationship may not be recovered and the large number of gene trees that do
1210 not recover the same relationship for Formicidae and Apoidea as the final species tree suggests
1211 that inconsistency among gene trees in this dataset is severe. However, the hypothesis of
1212 relationships between ants and other aculeate species proposed by Brothers (1999) Pilgrim et al.
1213 (2008), Heraty et al. (2011) and Faircloth et al. (2014) were even more underrepresented, with
1214 only a few gene trees recovering some of these results, while none would recover the last two (Fig.
1215 3B). Additionally, despite of the high level of discordance among gene trees, the ones that do agree
1216 with the relationship between ants and Apoidea were not closely clustered with the final species
1217 tree (Fig. 4A) or with each other (Fig. 4B). These results show that the data set is cohesive and,
1218 although many differences can be found among the estimated gene trees, data is not biased toward
1219 a specific topology. Gene tree bias is one major problem in the construction of species trees based
1220 in the coalescent model, but this can be alleviated by sampling more genes and this applies even
1221 when these genes are minimally informative (Xi et al. 2015). Furthermore, the fact that very few
1222 gene trees recovers this clade shows that relying on a single locus or a few loci as a proxy for
1223 species trees could be a risk practice (Ruane et al. 2015). Our results suggest that the addition of a
1224 large number of genes can be important because, while doing so, it is possible to recover enough

genes that produce reliable gene trees, and the probability of generating gene trees with biased phylogenetic signal is lower. Therefore, if a reduced set of loci are discordant it is expected that numerous additional markers are required to generate a credible species tree. This is the case for this data set that, although highly incongruent, does not present any bias towards any topology, making it possible to recover a species tree topology that is both well resolved and strongly supported.

However, the individual genes support by itself does not shed light on questions of gene choice or the optimal amount of loci necessary in order to recover reliable phylogenetic results. The high level of incongruence among data suggests that there is an emerging support effect occurring, because once more genes are added to the analysis, some clades are favored above others, appearing in the final species tree with a stronger support than in any of the individual genes. The degree of emergent support is largely independent of the level of support for a node among the gene trees (Narechania et al. 2012). This is the case of nodes 3 (Aculeata) and 8 (ants, bees and spheciform wasps), for example, both of which have visually similar low individual gene support in this data set (Fig. 3A). However, the node recovering the monophyly of Aculeata (node 3) presents a much stronger emergent support than node 8, becoming fixated in the final species tree after the addition of 92 genes, as opposed to the 308 genes necessary to fixate node 8 (Fig. 2; 5A). In fact, when looking for the number of genes necessary to recover different clades in the analysis (Fig. 2; 5A), we notice that the support does not increases gradually. By the time 128 genes are added, 15 out of 18 nodes in the final species tree are recovered in all the replicates, despite the combination of genes used (Fig. 2), one of those nodes being the root position of *Nasonia vitripennis*.

It is important to emphasize that this fixation is not due only to the addition of a large number (128) of good quality genes, since those results refers to 100 pseudo-random combinations of 128 genes, resampled from the 308 total loci. We believe that these results are more robust than traditional bootstrapping, because we resample 40% of our data matrix, with a cut-off value of 100, while bootstrapping resamples between 85% and 90% of the data matrix, with a cut-off value of 70%. This rapid fixation reflects the presence of emergent support, a situation in which the accumulation of nodal support is more rapid than would be predicted based on the levels of support on individual gene trees (Gatesy et al. 1999; Gatesy and Baker 2005). In these cases, congruent phylogenetic signal is amplified as genes are combined during the stepwise addition, whereas

1256 divergent patterns of homoplasy specific to single genes or a small set of genes cancel each other
1257 out (Narechania et al. 2012).
1258 Hereupon, 14 out of 18 of the nodes in the final species tree are resolved using only 40%
1259 of the available data. Still, the remaining 60% of the loci are necessary to clarify the position of
1260 nodes 4 (Vespidae + Tiphioid-Pompilioid wasps), 8 (Formicidae + Apoidea) and 11 (Pompilidae
1261 + Mutillidae). The CFI values (Fig. 5B) shows that, after 128 genes, half of the topologies
1262 recovered differs from the final species tree in 1 or 0 nodes, but the frequency of occurrence of
1263 those different topologies is highly variable (Fig. 6A). At the step where we have 128 genes, three
1264 topologies appear with equal representation, while the others appear in different and lower
1265 frequencies. Up to the addition of 236 genes, two topologies seems to be recovered at similar
1266 frequencies, but after this point only the topology recovered by the final species tree (in red, Fig.
1267 6A) keeps gaining representation. The remaining topologies appears as only "noise", that cannot
1268 be dismissed as negligible throughout the analysis due to their combined frequency being quite
1269 large (Fig. 6A). These results show that there is not only gene tree incongruence among the data,
1270 but also phylogenetic conflict in this data set, since the phylogenetic signal appears to support
1271 several topologies but favors very few trees, emerging a threshold of an optimal number of genes
1272 after which one of them becomes dominant.
1273 In this sense, when looking for the favored topologies in order to understand the
1274 relationships recovered by them, we focused on the sister group relationship between ants, bees
1275 and spheciform wasps (i.e. the final species tree recovered by MP-EST). This node reaches a
1276 frequency of 100% only after the addition of all of the 308 loci, suggesting that the average gene
1277 has more support for relationships that conflict with this than for the node itself. This can be
1278 evidence that there is a low amount of emergent support at this node (Narechania et al. 2012).
1279 Nevertheless, as genes are added to the analysis, this conflicting support diminishes, recovering
1280 the final topology.
1281 In a similar sense, the support for alternative relationships (Fig. 1) for ants among the
1282 Aculeate appears to present this same pattern in the gene trees, being recovered by a few loci (Fig.
1283 4A and 4B) in very low frequencies (Fig. 3B). However, in those cases, the conflict between the
1284 average gene support was so high that, with the addition of more data, these relationships are easily
1285 degraded. In fact, only Pilgrim et al. (2008) and an unpublished hypothesis appears to have a
1286 relatively higher frequency after the addition of 128 genes, but still with in a much lower frequency

than the final species tree, being completely degraded after the addition of all 308 loci to the analysis (Fig. 6A). The dynamics of emergent support demonstrate that the addition of a large number of loci to the analysis is not simply a ‘‘brute force’’ method that produces a definitive topology as a result of overwhelming data set size, evidencing that the recovery of specific nodes is not simply a function of combining additional characters, but reflects a disproportionate amplification of phylogenetic signal with the increase of the amount of data (Narechania et al. 2012).

Conclusions

The position of ants among the Aculeata appear to be mostly resolve and robustly supported by many phylogenomic studies to this date (Johnson et al. 2013; Branstetter et al. 2017; Peters et al. 2017). The present scenario provides a framework for investigating the evolution of important behaviors and characteristics in Hymenoptera, such as nesting, feeding and social behavior (Johnson et al. 2013; Branstetter et al. 2017), as well as the genomic signatures of changes in these characteristics (Johnson et al. 2013). However, questions regarding the evolution of morphological characters in those groups still raise questions regarding their phenotypical evolution, with many hypotheses being considered as alternatives (Brothers 1990; Heraty et al. 2008; Pilgrim et al. 2011). The main criticism of those inferences comes from the fact that the addition of a large number of loci may be simply favoring a biased result by brute force. Our results suggest that while hundreds of loci are not always essential for accurate topological inference, the use of large genomic data sets for phylogenetic inference may provide more robust results. Because the number of loci is large, it is possible to recover enough genes that produce reliable gene trees, and the probability of generating gene trees with biased phylogenetic signal is lower. However, this robustness is not granted only by the sheer volume of the data, but mainly by the presence of an emergent support, where the additional data disproportionately amplifies the phylogenetical signal supporting the final nodes.

We have shown here that for most of the nodes in Johnson et al. (2013) data, there is a rapid convergence on well-accepted relationships, such as the monophyly of aculeates or ants. Indeed, not combining a large number of data may obscure the general agreement between gene tree topologies, by the fact that individual support for a given topology may not be as strong when conflicting signal is added. More importantly, the additional data may also increase the efficiency

1318 of a given gene's phylogenetic signal through the accumulation of hidden support (Gatesy et al.
1319 1999; Gatesy and Baker 2005). Those inferences can be made by the use of coalescence-based
1320 phylogenetic inference, as shown in our statistical tests, but also by the use of concatenation
1321 methods, such as demonstrated by Narechania et al. (2012) and recovered by Peters et al. (2017).
1322 However, the fact that the target relationship of ants as sister to the apoid wasps and bees is only
1323 recovered by the addition of the full data set suggests that internal conflicts persist throughout the
1324 data. Although several studies, with even larger data sets, are also recovering this same result,
1325 adding to the robustness of this relationship, phylogenomic studies can broadly benefit from an
1326 exploration of the dynamics of these data sets apart from the traditional measures of support, in
1327 order to recognize the sources of conflict and the robustness of given results.

References

Allman, E. S., Degnan, J. H., & Rhodes, J. A. (2013). Species tree inference by the STAR method and its generalizations. Journal of Computational Biology, 20(1), 50-61.

Bayzid M.S., Warnow T. (2012). Estimating optimal species trees from incomplete gene trees under deep coalescence. Journal of Computational Biology, 19:591–605.

Blaimer B.B., Lloyd M.W., Guillory W.X., Brady S.G. (2016). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS One, 11:1–20.

Branstetter M.G., Danforth B.N., Pitts J.P., Faircloth B.C., Ward P.S., Buffington M.L., et al. (2017). Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. Current Biology, 27:1019–25.

Brothers D.J. (1999). Phylogeny and evolution of wasps, ants and bees (Hymenoptera, Chrysidoidea, Vespoidea and Apoidea). Zoologica Scripta, 28: 233–250.

Carstens B., Lemmon A.R., Lemmon E.M. (2012). The promises and pitfalls of next-generation sequencing data in phylogeography. Systematic biology, 61: 713-715.

Colless D.H. (1980). Congruence between morphometric and alloyzme data for Menidia species: a reappraisal. Systematic Zoology, 29:288–299.

Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., Glenn T.C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters, 8: 783 LP-786.

Degnan J. H., Rosenberg N. A. (2006). Discordance of species trees with their most likely gene trees. PLoS genetics, 2(5): e68.

Edwards S.V., Liu, L., Pearl, D.K. (2007). High-resolution species trees without concatenation. Proceedings of the National Academy of Sciences, 104(14): 5936-5941.

Faircloth BC, McCormack JE, Crawford NG et al. (2012) Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology, 61, 717–726.

1355 Faircloth B.C., Branstetter M.G., White N.D., Brady S.G. (2014). Target enrichment of
1356 ultraconserved elements from arthropods provides a genomic perspective on relationships
1357 among Hymenoptera. Molecular Ecology Resources 15: 489-501.

1358 Gatesy J., O'Grady P., Baker R.H. (1999). Corroboration among data sets in simultaneous
1359 analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa.
1360 Cladistics, 15: 271–313.

1361 Gatesy J., Baker R.H. (2005). Hidden likelihood support in genomic data: can forty-five wrongs
1362 make a right? Systematic Biology, 54:483–492.

1363 Gatesy J., Springer M.S. (2014). Phylogenetic analysis at deep timescales: unreliable gene trees,
1364 bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular
1365 Phylogenetics and Evolution, 80:231-66.

1366 Heibl C. 2014. ips: Interfaces to Phylogenetic Software in R. Available from CRAN at
1367 https://CRAN.R-project.org/package=ips.

1368 Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data.
1369 Molecular biology and evolution, 27: 570-580.

1370 Heraty J., Ronquist F., Carpenter J.M., Hawks D., Schulmeister S., Dowling A.P., et al. (2011).
1371 Evolution of the hymenopteran megaradiation. Molecular Phylogenetics and Evolution,
1372 60:73–88.

1373 Hillis, D. M., Heath, T. A., & John, K. S. (2005). Analysis and visualization of tree space.
1374 Systematic Biology, 54(3), 471-482.

1375 Hintze J.L., Nelson R.D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. The
1376 American Statistician, 52:181–4.

1377 Johnson B.R., Borowiec M.L., Chiu J.C., Lee E.K., Atallah J., Ward P.S. (2013). Phylogenomics
1378 resolves evolutionary relationships among ants, bees, and wasps. Current Biology, 23:
1379 2058-2062.

1380 Leebens-Mack J., Raubeson L.A., Cui L., Kuehl J.V., Fourcase M.H., Chumley T.W., Boore J.L.,
1381 Jansen R.K., dePamphilis C.W. (2005). Identifying the basal angiosperm node in

chloroplast genome phylogenies: sampling one's way out of the Felsenstein Zone. Molecular Biology and Evolution, 22: 1948–1963.

Liu, L. (2008). BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics, 24: 2542-2543.

Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using coalescence times among sequences. Systematic Biology, 58: 468-477.

Liu L., Yu L., Edwards S.V. (2010). A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology, 10:302.

Liu L., Yu L. (2011). Estimating species trees from unrooted gene trees. Systematic Biology, 60: 661-667.

Maddison W.P. (1997). Gene trees in species trees. Systematic biology, 46: 523-536.

McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. (2012) Ultraconserved Elements Are Novel Phylogenomic Markers that Resolve Placental Mammal Phylogeny when Combined with Species Tree Analysis. Genome Research, 22: 746–754.

McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. (2013). A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment and High-Throughput Sequencing. PLoS ONE, 8: e54848.

Mirarab, Siavash, Tandy Warnow. "ASTRAL-II: Coalescent-Based Species Tree Estimation with Many Hundreds of Taxa and Thousands of Genes.". Bioinformatics, 31: i44–i52.

Mossel E., Roch S. (2010). Incomplete lineage sorting: consistent phylogeny estimation from multiple loci. Computational Biology and Bioinformatics, 7: 166-171.

Narechania A., Baker R.H., Sit R., Kolokotronis S.O., DeSalle R., Planet P.J. (2012). Random Addition Concatenation Analysis: A novel approach to the exploration of phylogenomic signal reveals strong agreement between core and shell genomic partitions in the cyanobacteria. Genome Biology and Evolution, 4:30–43.

Neuwirth E. (2007). RColorBrewer: ColorBrewer palettes. R package version 1.0-2.

1409 Nye T.M.W. (2008). Trees of trees: an approach to comparing multiple alternative phylogenies.
1410 Systematic Biology, 57:785–94.

1411 Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., et al. (2017).
1412 Evolutionary History of the Hymenoptera. Current Biology, 27:1013–8.

1413 Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G, et al.
1414 (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough.
1415 PLoS Biology, 9(3): e1000602.

1416 Pilgrim E.M., Von Dohlen C.D., Pitts J.P. (2008). Molecular phylogenetics of Vespoidea indicate
1417 paraphyly of the superfamily and novel relationships of its component families and
1418 subfamilies. Zoologica Scripta, 37:539–60.

1419 Puritz J.B., Addison J.A., Toonen R.J. (2012). Next-generation phylogeography: a targeted
1420 approach for multilocus sequencing of non-model organisms. PLoS One, 7: e34241.

1421 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for
1422 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

1423 Rannala B., Yang Z. (2003). Bayes estimation of species divergence times and ancestral population
1424 sizes using DNA sequences from multiple loci. Genetics, 164: 1645-1656.

1425 Revell L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other
1426 things). Methods in Ecology and Evolution, 3: 217–223.

1427 Robinson D.F., Foulds L.R. (1981). Comparison of phylogenetic trees. Mathematical Biosciences,
1428 53:131–147.

1429 Romiguier J., Cameron S.A., Woodard S.H., Fischman B.J., Keller L., Praz C.J.
1430 (2016). Phylogenomics controlling for base compositional bias reveals a single origin of
1431 eusociality in corbiculate bees. Molecular Biology and Evolution, 33: 670–678.

1432 Ruane S., Raxworthy C.J., Lemmon A.R., Lemmon E.M., Burbrink F.T. (2015). Comparing
1433 species tree estimation with large anchored phylogenomic and small Sanger-sequenced
1434 molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC
1435 Evolutionary Biology, 15:221.

1436 Schliep K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics, 27: 592-593.

1437 Seo T.K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data.
1438 Molecular Biology and Evolution, 25:960–71.

1439 Simmons M.P., Goloboff P.A. (2014). Dubious resolution and support from published sparse
1440 supermatrices: The importance of thorough tree searches. Molecular Phylogenetics and
1441 Evolution, 78:334–48.

1442 Springer M.S., Gatesy J. (2015). The gene tree delusion. Molecular Phylogenetics and Evolution,
1443 94:1–33.

1444 Takahata N., Satta Y., Klein J. (1995). Divergence time and population size in the lineage leading
1445 to modern humans. Theoretical Population Biology, p. 198–221.

1446 Townsend J.P. (2007). Profiling Phylogenetic Informativeness. Systematic Biology, 56: 222–231.

1447

1448 Welch A.J., Collins K., Ratan A., Drautz-Moses D.I., Schuster S.C., Lindqvist C. (2016). The
1449 quest to resolve recent radiations: Plastid phylogenomics of extinct and endangered
1450 Hawaiian endemic mints (Lamiaceae). Molecular Phylogenetics and Evolution, 99:16–33.

1451 Wong K.M., Suchard M.A., Huelsenbeck J.P. (2008). Alignment uncertainty and genomic
1452 analysis. Science, 319: 473–476.

1453 Xi Z., Liu L., Davis C.C. (2015). Genes with minimal phylogenetic information are problematic
1454 for coalescent analyses when gene tree estimation is biased. Molecular Phylogenetics and
1455 Evolution, 92:63–71.

1456
1457
1458

Referências Bibliográficas

Allman, E. S., Degnan, J. H., & Rhodes, J. A. (2013). Species tree inference by the STAR method and its generalizations. Journal of Computational Biology, 20(1), 50-61.

Baroni-Urbani C., Bolton B., Ward P.S. (1992). The internal phylogeny of ants (Hymenoptera: Formicidae). Systematic Entomology, 17:301–29.

Bayzid M.S., Warnow T. (2012). Estimating optimal species trees from incomplete gene trees under deep coalescence. Journal of Computational Biology, 19:591–605.

Blaimer B.B., Brady S.G., Schultz T.R., Lloyd M.W., Fisher B.L., Ward P.S. (2015). Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evolutionary Biology, 15:271.

Blaimer B.B., LaPolla J.S., Branstetter M.G., Lloyd M.W., Brady S.G. (2016a). Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus Acropyga. Molecular Phylogenetics and Evolution, 102: 20–9.

Blaimer B.B., Lloyd M.W., Guillory W.X., Brady S.G. (2016b). Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLoS One, 11:1–20.

Blumenstiel B., Cibulskis K., Fisher S., DeFelice M., Barry A., Fennell T., Abreu J., Minie B., Costello M., Young G. (2010). Targeted exon sequencing by in-solution hybrid selection. Current Protocols in Human Genetics, 66:18.4:18.4.1–18.4.24.

Bolger A.M., Lohse M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, 30: 2114–2120.

Bolton B. (1995). A new general catalogue of the ants of the world.. p. 504.

Bolton B. (2003). Synopsis and classification of Formicidae. Memoirs of American Entomological Institute. 369 p.

Bolton, B. (2017). An online catalog of the ants of the world. Available from http://antcat.org. (accessed 12/09/2017).

Borgmeier T. (1957). Myrmecologische Studien. Anais da Academia Brasileira de Ciencias, 29: 103-128.

1487 Borowiec M.L., Lee E.K., Chiu J.C., Plachetzki D.C. (2015). Extracting phylogenetic signal and
1488 accounting for bias in whole-genome data sets supports the Ctenophora as sister to
1489 remaining Metazoa. BMC Genomics, 16:987.

1490 Bossert S., Murray E.A., Blaimer B.B., Danforth B.N. (2017). The impact of GC bias on
1491 phylogenetic accuracy using targeted enrichment phylogenomic data. Molecular
1492 Phylogenetics and Evolution, 111:149–57.

1493 Brady S.G. (2003). Evolution of the army ant syndrome: The origin and long-term evolutionary
1494 stasis of a complex of behavioral and reproductive adaptations. Proceedings of the National
1495 Academy of Sciences, 100:6575–9.

1496 Brady S.G., Schultz T.R., Fisher B.L., Ward P.S. (2006). Evaluating alternative hypotheses for the
1497 early evolution and diversification of ants. Proceedings of the National Academy of
1498 Sciences, 103:18172–7.

1499 Brady S.G. (2011). Effects of fossil calibration uncertainty on divergence dating in ants and bees.
1500 American Entomologist, 57:56–7.

1501 Brady S., Fisher B., Schultz T., Ward P. (2014). The rise of army ants and their relatives:
1502 diversification of specialized predatory doryline ants. BMC Evolutionary Biology, 14:93.

1503 Brandão C.R.F., Lattke J.E. (1990). Description of a new Ecuadorean *Gnamptogenys* species
1504 (Hymenoptera: Formicidae), with a discussion on the status of the *Alfaria* group. Journal
1505 of the New York Entomological Society, 98:489-494.

1506 Branstetter M.G., Longino J.T., Ward P.S., Faircloth B.C. (2017). Enriching the ant tree of life:
1507 enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera.
1508 Methods in Ecology and Evolution, 8: 768–776.

1509 Brothers D.J. (1999). Phylogeny and evolution of wasps, ants and bees (Hymenoptera,
1510 Chrysidoidea, Vespoidea and Apoidea). Zoologica Scripta, 28: 233–250.

1511 Camacho, G.P., Feitosa, R.M. (2015). Estado da arte sobre a Filogenia, Taxonomia e Biologia de
1512 Ectatomminae. In: As formigas Poneromorfas do Brasil. – pp. 23-32.

1513 Carstens B., Lemmon A.R., Lemmon E.M. (2012). The promises and pitfalls of next-generation
1514 sequencing data in phylogeography. Systematic biology, 61: 713-715.

1515 Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in
1516 phylogenetic analysis. Molecular Biology and Evolution, 17: 540-552.

1517 Colless D.H. (1980). Congruence between morphometric and alloyzme data for Menidia species:
1518 a reappraisal. Systematic Zoology, 29:288–299.

1519 Crawford N.G., Faircloth B.C., McCormack J.E., Brumfield R.T., Winker K., Glenn T.C. (2012).
1520 More than 1000 ultraconserved elements provide evidence that turtles are the sister group
1521 of archosaurs. Biology Letters, 8: 783 LP-786.

1522 Degnan J. H., Rosenberg N. A. (2006). Discordance of species trees with their most likely gene
1523 trees. PLoS genetics, 2(5): e68.

1524 Delsuc F., Brinkmann H., Philippe H. (2005). Phylogenomics and the reconstruction of the tree of
1525 life. Nat Rev Genet, 6: 361-375.

1526 Edwards S.V., Liu, L., Pearl, D.K. (2007). High-resolution species trees without concatenation.
1527 Proceedings of the National Academy of Sciences, 104(14): 5936-5941.

1528 Emery C. (1911). Hymenoptera, Fam. Formicidae, Subfam. Ponerinae. In: Wytsman, P. Genera
1529 Insectorum. Fasc. 118. 124 pp.

1530 Faircloth B.C., Glenn T.C. (2012). Not all sequence tags are created equal: designing and
1531 validating sequence identification tags robust to indels. PLoS ONE, 7: e42543.

1532 Faircloth BC, McCormack JE, Crawford NG et al. (2012) Ultraconserved elements anchor
1533 thousands of genetic markers spanning multiple evolutionary timescales. Systematic
1534 Biology, 61, 717–726.

1535 Faircloth, B.C. (2013). Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality
1536 trimming. Available at: http://dx.doi.org/10.6079/J9ILL.

1537 Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. (2013). A Phylogenomic Perspective on the
1538 Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved
1539 Elements (UCEs). PLOS ONE 8: e65923.

1540 Faircloth B.C., Branstetter M.G., White N.D., Brady S.G. (2014). Target enrichment of
1541 ultraconserved elements from arthropods provides a genomic perspective on relationships
1542 among Hymenoptera. Molecular Ecology Resources 15: 489-501.

1543 Faircloth B.C., Branstetter M.G., White N.D., Brady S.G. (2015). Target enrichment of
1544 ultraconserved elements from arthropods provides a genomic perspective on relationships
1545 among Hymenoptera. Molecular Ecology Resources, 15:489–501.

1546 Feitosa R.M. (2011). Revisão taxonômica e análise filogenética de Heteroponerinae
1547 (Hymenoptera, Formicidae). Universidade de São Paulo.

1548 Feitosa, R.M. (2015). Estado da arte sobre a Filogenia, Taxonomia e Biologia de Heteroponerinae.
1549 In: As formigas Poneromorfas do Brasil. – pp. 33-43.

1550 Fisher S., Barry A., Abreu J., Minie B., Nolan J., Delorey T.M., Young G., Fennell T.J., Allen A.,
1551 Ambrogio L. (2011). A scalable, fully automated process for construction of sequence-
1552 ready human exome targeted capture libraries. Genome Biology, 12: R1.Gatesy J.,
1553 O'Grady P., Baker R.H. (1999). Corroboration among data sets in simultaneous analysis:
1554 hidden support for phylogenetic relationships among higher level artiodactyl taxa.
1555 Cladistics, 15: 271–313.

1556 Gatesy J., Baker R.H. (2005). Hidden likelihood support in genomic data: can forty-five wrongs
1557 make a right? Systematic Biology, 54:483–492.

1558 Gatesy J., Springer M.S. (2014). Phylogenetic analysis at deep timescales: unreliable gene trees,
1559 bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular
1560 Phylogenetics and Evolution, 80:231-66.

1561 Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L.
1562 et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference
1563 genome. Nature Biotechnology 29, 644–652.

1564 Heibl C. 2014. ips: Interfaces to Phylogenetic Software in R. Available from CRAN at
1565 https://CRAN.R-project.org/package=ips.

1566 Heled, J., & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data.
1567 Molecular biology and evolution, 27: 570-580.

1568 Heraty J., Ronquist F., Carpenter J.M., Hawks D., Schulmeister S., Dowling A.P., et al. (2011).
1569 Evolution of the hymenopteran megaradiation. Molecular Phylogenetics and Evolution,
1570 60:73–88.

1571 Hillis, D. M., Heath, T. A., & John, K. S. (2005). Analysis and visualization of tree space.
1572 Systematic Biology, 54(3), 471-482.

1573 Hintze J.L., Nelson R.D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. The
1574 American Statistician, 52:181–4.

1575 Hölldobler B., Wilson E.O. (2008). The Superorganism: The Beauty, Elegance, and Strangeness
1576 of Insect Societies. Norton, 320 p.

1577 Janicki J., Narula N., Ziegler M., Guenard, B., Economo E.P. (2016). Visualizing and interacting
1578 with large-volume biodiversity data using client-server web-mapping applications: The
1579 design and implementation of antmaps.org. Ecological Informatics, 32: 185-193.

1580 Johnson B.R., Borowiec M.L., Chiu J.C., Lee E.K., Atallah J., Ward P.S. (2013). Phylogenomics
1581 resolves evolutionary relationships among ants, bees, and wasps. Current Biology, 23:
1582 2058-2062.

1583 Katoh K., Asimenos G., Toh H. (2009). Multiple alignment of DNA sequences with MAFFT. In:
1584 Bioinformatics for DNA sequence analysis. Springer; 2009: 39–64.

1585 Keller R.A. (2000). Cladistics of the tribe Ectatommini (Hymenoptera: Formicidae): a reappraisal.
1586 Insect Systematics & Evolution, 31: 59-69.

1587 Keller R.A. (2011). A Phylogenetic Analysis of Ant Morphology (Hymenoptera: Formicidae) with
1588 Special Reference to the Poneromorph Subfamilies. Bulletin of the American Museum of
1589 Natural History, 355:1–90.

1590 Kumar S., Stecher G., Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis
1591 Version 7.0 for Bigger Datasets, Molecular Biology and Evolution, 33: 1870–1874.

1592 Lattke J.E. (1994). Phylogenetic relationships and classification of Ectatommine ants
1593 (Hymenoptera, Formicidae). Entomologica Scandinavica, 25: 105-119.

1594 Lattke J. (1995). Revision of the Ant Genus *Gnamptogenys* in the New world (Hymenoptera:
1595 Formicidae). Journal of Hymenoptera Research, 4:137–93.

1596 Lattke J.E. (2004). A taxonomic revision and phylogenetic analysis of the ant genus
1597 *Gnamptogenys* Roger in Southeast Asia and Australasia (Hymenoptera: Formicidae:
1598 Ponerinae). University of California Publications, Entomology, 122: 1-266.

1599 Leebens-Mack J., Raubeson L.A., Cui L., Kuehl J.V., Fourcase M.H., Chumley T.W., Boore J.L.,
1600 Jansen R.K., dePamphilis C.W. (2005). Identifying the basal angiosperm node in
1601 chloroplast genome phylogenies: sampling one's way out of the Felsenstein Zone.
1602 Molecular Biology and Evolution, 22: 1948–1963.

1603 Lewis P.O. (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological
1604 Character Data. Systematic Biology, 50: 913–925.

1605 Liu, L. (2008). BEST: Bayesian estimation of species trees under the coalescent model.
1606 Bioinformatics, 24: 2542-2543.

1607 Liu, L., Yu, L., Pearl, D. K., & Edwards, S. V. (2009). Estimating species phylogenies using
1608 coalescence times among sequences. Systematic Biology, 58: 468-477.

1609 Liu L., Yu L., Edwards S.V. (2010). A maximum pseudo-likelihood approach for estimating
1610 species trees under the coalescent model. BMC Evolutionary Biology, 10:302.

1611 Liu L., Yu L. (2011). Estimating species trees from unrooted gene trees. Systematic Biology, 60:
1612 661-667.

1613 Maddison W.P. (1997). Gene trees in species trees. Systematic biology, 46: 523-536.

1614 McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. (2012)
1615 Ultraconserved Elements Are Novel Phylogenomic Markers that Resolve Placental
1616 Mammal Phylogeny when Combined with Species Tree Analysis. Genome Research, 22:
1617 746–754.

1618 McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T.
1619 (2013). A Phylogeny of Birds Based on Over 1,500 Loci Collected by Target Enrichment
1620 and High-Throughput Sequencing. PLoS ONE, 8: e54848.

1621 Mirarab, Siavash, Tandy Warnow. "ASTRAL-II: Coalescent-Based Species Tree Estimation with
1622 Many Hundreds of Taxa and Thousands of Genes.". Bioinformatics, 31: i44–i52.

1623 Moreau C.S., Bell C.D., Vila R., Archibald S.B., Pierce N.E. (2006). Phylogeny of the Ants:
1624 Diversification in the Age of Angiosperms. Science, 312:101–4.

1625 Moreau C.S., Bell C.D. (2013). Testing the museum versus cradle tropical biological diversity
1626 hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the

1627 ants. Evolution; 67:2240–57.

1628 Mossel E., Roch S. (2010). Incomplete lineage sorting: consistent phylogeny estimation from
1629 multiple loci. Computational Biology and Bioinformatics, 7: 166-171.

1630 Narechania A., Baker R.H., Sit R., Kolokotronis S.O., DeSalle R., Planet P.J. (2012). Random
1631 Addition Concatenation Analysis: A novel approach to the exploration of phylogenomic
1632 signal reveals strong agreement between core and shell genomic partitions in the
1633 cyanobacteria. Genome Biology and Evolution, 4:30–43.

1634 Neuwirth E. (2007). RColorBrewer: ColorBrewer palettes. R package version 1.0-2.

1635 Nei M., Kumar S. (2000) Molecular Evolution and Phylogenetics. Oxford University Press, New
1636 York.

1637 Nettel-Hernanz A., Lachaud J., Fresneau D., López-Muñoz R.A., Poteaux C. (2015).
1638 Biogeography, cryptic diversity, and queen dimorphism evolution of the Neotropical ant
1639 genus Ectatomma Smith, 1958 (Formicidae, Ectatomminae). Organisms Diversity &
1640 Evolution, 15: 543-553.

1641 Nye T.M.W. (2008). Trees of trees: an approach to comparing multiple alternative phylogenies.
1642 Systematic Biology, 57:785–94.

1643 Ouellette G.D., Fisher B.L., Girman D.J. (2006). Molecular systematics of basal subfamilies of
1644 ants using 28S rRNA (Hymenoptera: Formicidae). Molecular Phylogenetics and
1645 Evolution, 40: 359-369.

1646 Pagel M. (1999). Inferring the historical patterns of biological evolution. Nature, 401:877.

1647 Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., et al. (2017).
1648 Evolutionary History of the Hymenoptera. Current Biology, 27:1013–8.

1649 Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G, et al.
1650 (2011). Resolving difficult phylogenetic questions: Why more sequences are not enough.
1651 PLoS Biology, 9(3): e1000602.

1652 Phillips M.J., Delsuc F., Penny D. (2004). Genome-scale phylogeny and the detection of
1653 systematic biases. Molecular Biology and Evolution, 21:1455–8.

1654 Pilgrim E.M., Von Dohlen C.D., Pitts J.P. (2008). Molecular phylogenetics of Vespoidea indicate

paraphyly of the superfamily and novel relationships of its component families and subfamilies. Zoologica Scripta, 37:539–60.

Puritz J.B., Addison J.A., Toonen R.J. (2012). Next-generation phylogeography: a targeted approach for multilocus sequencing of non-model organisms. PLoS One, 7: e34241.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rabeling C., Brown J.M., Verhaagh M. (2008). Newly discovered sister lineage sheds light on early ant evolution. Proceedings of the National Academy of Sciences, 105:14913–7.

Rannala B., Yang Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics, 164: 1645-1656.

Revell L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3: 217–223.

Robinson D.F., Foulds L.R. (1981). Comparison of phylogenetic trees. Mathematical Biosciences, 53:131–147.

Romiguier J., Cameron S.A., Woodard S.H., Fischman B.J., Keller L., Praz C.J. (2016). Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. Molecular Biology and Evolution, 33: 670–678.

Ruane S., Raxworthy C.J., Lemmon A.R., Lemmon E.M., Burbrink F.T. (2015). Comparing species tree estimation with large anchored phylogenomic and small Sanger-sequenced molecular datasets: an empirical study on Malagasy pseudoxyrhophiine snakes. BMC Evolutionary Biology, 15:221.

Sayyari E., Mirarab S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. Molecular Biology and Evolution, 33:msw079.

Schluter D., Price T., Mooers A.O., Ludwig D. (1997). Likelihood of ancestor states in adaptive radiation. Evolution, 51: 1699-1711.

Schliep K.P. (2011). phangorn: phylogenetic analysis in R. Bioinformatics, 27: 592-593.

Schmidt C.A. (2013). Molecular phylogenetics and taxonomic revision of ponerine ants (Hymenoptera: Formicidae: Ponerinae). Zootaxa, 3647:201–50.

1683 Seo T.K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data.
1684 Molecular Biology and Evolution, 25:960–71.

1685 Simmons M.P., Goloboff P.A. (2014). Dubious resolution and support from published sparse
1686 supermatrices: The importance of thorough tree searches. Molecular Phylogenetics and
1687 Evolution, 78:334–48.

1688 Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. (2013). Target Capture and
1689 Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at
1690 Shallow Evolutionary Time Scales. Systematic Biology, 63: 83-95.

1691 Springer M.S., Gatesy J. (2015). The gene tree delusion. Molecular Phylogenetics and Evolution,
1692 94:1–33.

1693 Stamatakis A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of
1694 large phylogenies. Bioinformatics, 30:1312–3.

1695 Takahata N., Satta Y., Klein J. (1995). Divergence time and population size in the lineage leading
1696 to modern humans. Theoretical Population Biology, p. 198–221.

1697 Tamura K., Dudley J., Nei M., Kumar S. (2007). MEGA4: Molecular Evolutionary Genetics
1698 Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

1699 Taylor R.W. (1980). Notes on the Russian endemic ant genus Aulacopone Arnoldi (Hymenoptera:
1700 Formicidae). Psyche, 86: 353-361.

1701 Townsend J.P. (2007). Profiling Phylogenetic Informativeness. Systematic Biology, 56: 222–231.

1702 Veevers J.J., Powell C.M., Roots S.R. (1991). Review of seafloor spreading around Australia. I.
1703 synthesis of the patterns of spreading. Australian Journal of Earth Sciences, 38:373–89.

1704 Ward P.S. (1980). A systematic revision of the *Rhytidoponera impressa* group (Hymenoptera:
1705 Formicidae) in Australia and New Guinea. Australian Journal of Zoology, 28: 341-365.

1706 Ward P.S. (1984). A revision of the ant genus *Rhytidoponera* (Hymenoptera: Formicidae) in New
1707 Caledonia. Australian Journal of Zoology, 32: 131-175.

1708 Ward P.S. (2007). Phylogeny, classification, and species-level taxonomy of ants (Hymenoptera:
1709 Formicidae). Zootaxa, 563:549–63.

1710 Ward P.S., Brady S.G. (2003). Phylogeny and biogeography of the ant subfamily
1711 Prionomyrmecinae (Hymenoptera: Formicidae). Invertebrate Systematics, 17:361–86.

1712 Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. (2010). Phylogeny and biogeography of
1713 dolichoderine ants: Effects of data partitioning and relict taxa on historical inference.
1714 Systematic Biology, 59:342–62.

1715 Ward P.S. (2014). The Phylogeny and Evolution of Ants. Annual Review of Ecology, Evolution,
1716 and Systematics, 45: 23-43.

1717 Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. (2015). The evolution of myrmicine ants:
1718 phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae).
1719 Systematic Entomology, 40:61–81.

1720 Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. (2015). The evolution of myrmicine ants:
1721 Phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae).
1722 Systematic Entomology, 40:61–81.

1723 Ward P.S., Fisher B.L. (2016). Tales of dracula ants: The evolutionary history of the ant subfamily
1724 Amblyoponinae (Hymenoptera: Formicidae). Systematic Entomology, 41:683–93.

1725 Ward P.S., Blaimer B.B., Fisher B.L. (2016). A revised phylogenetic classification of the ant
1726 subfamily Formicinae (Hymenoptera: Formicidae), with resurrection of the genera
1727 *Colobopsis* and *Dinomyrmex*. Zootaxa, 4072:343–57.

1728 Welch A.J., Collins K., Ratan A., Drautz-Moses D.I., Schuster S.C., Lindqvist C. (2016). The
1729 quest to resolve recent radiations: Plastid phylogenomics of extinct and endangered
1730 Hawaiian endemic mints (Lamiaceae). Molecular Phylogenetics and Evolution, 99:16–33.

1731 Woodburne M.O., Case J. (1996). Dispersal, vicariance, and the Late Cretaceous to early tertiary
1732 land mammal biogeography from South America to Australia. Journal of Mammal
1733 Evolution, 3:121–61.

1734 Wong K.M., Suchard M.A., Huelsenbeck J.P. (2008). Alignment uncertainty and genomic
1735 analysis. Science, 319: 473–476.

1736 Xi Z., Liu L., Davis C.C. (2015). Genes with minimal phylogenetic information are problematic
1737 for coalescent analyses when gene tree estimation is biased. Molecular Phylogenetics and
1738 Evolution, 92:63–71.

1739

Anexo I

Supplementary material S1. Summary of UCE capture statistics.

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Acanthoponera_goeldii_GPC01	2325	1365295	587.2236559	3.333486098	201	2207	595	12
Acanthoponera_goeldii_GPC04	2175	838231	385.3935632	2.178881231	201	1856	380	4
Acanthoponera_minor_EX1582	2315	1787685	772.2181425	4.931133152	201	1728	792	356
Acanthoponera_minor_MIN01	2220	2191347	987.0932432	6.974907157	203	5102	1026	1205
Acanthoponera_minor_MIN02	2287	2031065	888.0913861	6.160588632	203	5483	915	842
Acanthoponera_mucronata_GPC02	2096	836744	399.2099237	2.405261323	201	1146	391	2
Acanthoponera_mucronata_GPC05	2252	1884038	836.6065719	5.346027987	201	2085	863	601
Acanthoponera_mucronata_GPC06	2114	878525	415.5747398	2.524019576	201	934	411	0
Acanthoponera_mucronata_GPC23	2136	1159535	542.8534644	3.742796592	202	1535	545	18
Acanthoponera_sp.n.B_GPC03	2225	912560	410.1393258	2.572463463	201	2151	401	4
Anillidris_sp_ANIFei21	858	512182	596.9487179	6.652585926	201	1238	597.5	20
Bothriomyrmex_sp_BotFei19	761	563633	740.6478318	10.37383867	201	2316	736	149
Cerapachys_sp_EX1579	2106	1669974	792.960114	6.258345008	204	3253	787	487
Dolichoderus_lamellosus_EX843	2288	2268927	991.6638986	6.438096576	209	2850	1033	1227
Eciton_burchellii_EX1580	2225	1897185	852.6674157	7.91859648	204	10416	865	736
Ectatomma_brunneum_ECTA16	2283	2480189	1086.372755	6.535441042	209	2453	1119	1486
Ectatomma_brunneum_CNIN2983	2297	2465281	1073.26121	7.171303536	201	6460	1115	1464
Ectatomma_brunneum_CNIN2985	2315	1583183	683.8803456	4.678789255	203	2326	679	166
Ectatomma_brunneum_CNIN2987	2206	1980940	897.9782412	6.375512688	201	3230	928	900
Ectatomma_edentatum_ECTA05	2246	2305209	1026.361977	6.902431965	201	5936	1069	1316
Ectatomma_edentatum_ECTA14	2330	1427959	612.8579399	3.775938407	202	1854	613	46
Ectatomma_gibbum_EX814	2371	1555342	655.9856601	4.359846743	215	4663	638	116
Ectatomma_goninion_ECTA11	2256	2307951	1023.027926	8.201355997	201	10994	1065.5	1324
Ectatomma_lugens_CNIN2988	2151	775286	360.4304974	4.151023897	201	8166	348	1
Ectatomma_lugens_ECTA15	2375	1655673	697.1254737	5.14079728	201	6539	694	143
Ectatomma_muticum_ECTA04	2319	1502650	647.9732643	4.162101874	201	1478	651	87
Ectatomma_opaciventre_CNIN2995	2353	2133333	906.6438589	5.606350225	201	2441	919	886
Ectatomma_opaciventre_ECTA09	2259	2422028	1072.168216	7.107954701	202	2864	1108	1420
Ectatomma_permagnum_ECTA03	1611	516951	320.8882682	2.325472582	201	1830	308	3
Ectatomma_planidens_ECTA19	150	35687	237.9133333	3.139613916	201	432	229	0
Ectatomma_ruidum_ECTA02	2354	1772041	752.7786746	4.765081672	202	2372	755	306
Ectatomma_ruidum_ECTA08	183	49276	269.2677596	5.270921944	201	669	249	0
Ectatomma_ruidum_ECTA12	2235	1710256	765.2152125	5.076898374	202	2146	778	349
Ectatomma_ruidum_ECTA13	2274	2465566	1084.241865	7.105649267	204	2479	1131	1466
Ectatomma_ruidum_ECTA18	2290	2613295	1141.176856	7.220991519	203	2695	1188	1597
Ectatomma_ruidum1_CNIN2074	2225	2100812	944.1851685	6.261711938	201	2203	979	1043

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Ectatomma_ruidum2_CNIN2066	2295	2477401	1079.47756	6.743793768	205	2739	1087	1432
Ectatomma_ruidum2_CNIN2068	2261	2488529	1100.632021	7.191951324	201	5527	1146	1514
Ectatomma_ruidum2_CNIN2070	2278	2551565	1120.089991	7.229412523	205	2607	1159	1524
Ectatomma_ruidum3_CNIN2075	2275	2499481	1098.672967	7.116340188	203	2282	1136	1503
Ectatomma_ruidum3_CNIN2078	2301	2306149	1002.237723	6.411879701	201	3676	1039	1269
Ectatomma_ruidum4_CNIN2088	2289	2462542	1075.81564	6.964961349	201	3316	1117	1454
Ectatomma_tuberculatum_CNIN2998	2295	2489933	1084.938126	7.057936998	210	2367	1104	1439
Ectatomma_tuberculatum_CNIN3005	2293	2368910	1033.105102	7.458007043	202	7987	1063	1313
Ectatomma_tuberculatum_ECTA07	2332	1828370	784.035163	5.075536947	204	2588	795	385
Formica_podzolica_EX1583	2223	1947215	875.9401709	6.083163522	202	1911	911	818
Gnamptogenys_acuminata_GAB162	1548	488373	315.4864341	2.123296445	201	864	300	0
Gnamptogenys_acuminata_GAB165	2261	1294033	572.3277311	3.702529134	201	1392	576	12
Gnamptogenys_acuminata_GAB61	2257	1681015	744.8006203	5.057442666	201	1651	759	322
Gnamptogenys_albiclava_GAB188	2298	2245667	977.2267189	6.890657951	201	2295	997.5	1142
Gnamptogenys_alfaroi_EX1310	2222	1959091	881.6791179	7.222816708	202	4207	897	852
Gnamptogenys_alfaroi_EX1332	2244	1642670	732.0276292	5.871980399	201	5632	735	305
Gnamptogenys_ammophila_GAB134	2231	1949710	873.9175258	6.126193009	203	1928	897	797
Gnamptogenys_ammophila_GAB77	2270	2079855	916.2356828	6.424734217	201	2609	934	928
Gnamptogenys_annulata_EX1274	2215	2052888	926.8117381	6.642728046	207	4029	962	995
Gnamptogenys_annulata_EX1284	2232	2039141	913.593638	6.272595995	201	2580	940	936
Gnamptogenys_annulata_EX1286	2236	1952319	873.1301431	6.248803532	202	3224	894.5	790
Gnamptogenys_annulata_EX1300	2250	2137772	950.1208889	7.091450887	202	2173	970	1054
Gnamptogenys_annulata_EX1334	2221	1827051	822.625394	5.944191834	201	1882	832	631
Gnamptogenys_annulata_EX1342	2272	2049403	902.0259683	6.642564636	201	1809	925	941
Gnamptogenys_annulata_GAB107	2223	1938914	872.2060279	6.671181556	201	5534	887	776
Gnamptogenys_annulata_GAB142	2170	1741737	802.6437788	5.909284948	201	3163	824	542
Gnamptogenys_annulata_GAB167	1567	435432	277.8761966	1.438476876	201	740	268	0
Gnamptogenys_annulata_GAB23	2264	1854653	819.1930212	5.318230832	204	2215	835	577
Gnamptogenys_annulata_GAB30	2265	1236375	545.8609272	3.325401209	201	1365	553	8
Gnamptogenys_annulata_GAB94	2302	1981767	860.8892268	5.783276628	203	2075	871	736
Gnamptogenys_atrata_GAB182	2295	2134998	930.2823529	6.368525731	202	5754	948	980
Gnamptogenys_banksi_EX1325	2253	1807069	802.072348	5.624407609	205	1834	813	528
Gnamptogenys_banksi_EX1343	2297	2126558	925.7979974	6.328405911	202	2178	946	984
Gnamptogenys_banksi_EX1344	2279	1669399	732.5138219	4.74051271	201	1936	739	250
Gnamptogenys_binghamii_EX1345	2258	1953631	865.204163	6.501107863	203	2561	872	804
Gnamptogenys_bispinosa_EX1320	2295	2124883	925.8749455	6.441775562	201	4284	943	978
Gnamptogenys_bispinosa_EX1331	2273	1721310	757.2855257	5.125003782	202	3202	771	335
Gnamptogenys_bisulca_EX1326	2250	1635883	727.0591111	4.917141562	202	2084	741	250
Gnamptogenys_bisulca_GAB151	2182	815537	373.7566453	3.11883082	201	5570	366	1

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_bisulca_GAB43	2249	1804072	802.1662961	5.217175151	201	2778	826	482
Gnamptogenys_boliviensis_GAB57	2229	1867972	838.0314042	5.962690525	204	3767	853	697
Gnamptogenys_boliviensis_GAB98	2212	1670542	755.2179024	5.536483702	203	3085	769	374
Gnamptogenys_brunnea_GAB152	1957	679212	347.0679612	2.149787546	201	1874	336	1
Gnamptogenys_concinna_EX1346	2223	1542851	694.040036	4.726293588	201	1644	707	166
Gnamptogenys_concinna_GAB168	2212	1266665	572.6333635	3.984939687	201	1491	576	18
Gnamptogenys_concinna_GAB60	1438	1102162	766.4547983	7.696473114	205	1870	775	324
Gnamptogenys_continua_GAB127	2262	1967430	869.7745358	6.264399327	201	5662	885.5	766
Gnamptogenys_continua_GAB137	2267	1211160	534.256727	3.941476706	201	5359	535	12
Gnamptogenys_continua_GAB33	2308	1407283	609.7413345	4.048434826	201	1675	604	69
Gnamptogenys_continua_GAB53	2200	2173493	987.9513636	6.982917443	202	2097	1025	1172
Gnamptogenys_continua_GAB93	2239	1672731	747.0884323	5.133811688	202	1795	756	347
Gnamptogenys_coxalis_EX1347	2219	1629672	734.4173051	5.729723673	201	2620	734	384
Gnamptogenys_crenaticeps_GAB187	2315	1995273	861.8889849	5.91529771	204	1918	880	779
Gnamptogenys_cribrata_EX1294	2261	2350767	1039.702344	6.989140548	202	2348	1070	1328
Gnamptogenys_enodis_GAB172	1977	659894	333.7855336	1.945541838	201	1316	325	1
Gnamptogenys_enodis_GAB198	1984	648711	326.9712702	1.863629814	201	996	315	0
Gnamptogenys_ericae_GAB149	2264	1958463	865.0454947	6.029891431	203	3441	885.5	759
Gnamptogenys_ericae_GAB81	2290	2309105	1008.342795	6.323047121	203	2108	1035	1251
Gnamptogenys_ericae_GAB82	2290	2141324	935.0759825	6.410601688	202	2221	957	1033
Gnamptogenys_extra_GAB150	1818	540781	297.4592959	1.585711065	201	786	287	0
Gnamptogenys_fontana_EX1348	2200	1938484	881.1290909	7.444725045	201	6847	898.5	865
Gnamptogenys_gr.coxalis_EX1293	2198	1917613	872.4353958	6.738105078	201	2126	889	840
Gnamptogenys_gr.mordax_GAB128	2213	1225238	553.6547673	3.620183285	202	1441	555	17
Gnamptogenys_gr.rastrata_GAB124	2223	1081691	486.5906433	2.916841798	202	1334	488	8
Gnamptogenys_gr.rastrata_GAB157	2269	1924971	848.3785809	5.956200706	201	1823	860	728
Gnamptogenys_gr.rastrata_GAB176	2262	1079346	477.1644562	2.849948718	201	1254	478	4
Gnamptogenys_gr.rastrata_GAB19	2254	1512137	670.8682343	4.308528283	201	2291	681	98
Gnamptogenys_gr.rastrata_GAB71	2283	1922380	842.0411739	5.450802651	202	2182	862	649
Gnamptogenys_gr.rastrata_GAB97	1626	479782	295.0688807	1.678990668	201	827	283.5	0
Gnamptogenys_gracilis_GAB196	2201	2123359	964.7246706	6.406148937	207	2044	1005	1120
Gnamptogenys_haenschi_EX1292	2275	2386866	1049.171868	7.400688238	201	3574	1076	1339
Gnamptogenys_haenschi_EX1317	2241	1462709	652.7037037	4.359961181	202	1505	660	77
Gnamptogenys_haenschi_EX1350	2244	2003570	892.8565062	6.872371665	203	2344	904	891
Gnamptogenys_haenschi_GAB143	2125	846595	398.3976471	2.370197843	201	1025	393	2
Gnamptogenys_haenschi_GAB161	2271	1459030	642.4614707	4.195576394	201	1497	652	66
Gnamptogenys_haesnchi_GAB169	2219	1131682	509.9963948	3.266750018	202	1342	509	5
Gnamptogenys_hartmani_EX1351	2243	2032538	906.169416	6.86283484	201	2495	914	932
Gnamptogenys_hartmani_EX1352	2272	2075580	913.5475352	6.657449163	205	2758	933	935

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_hartmani_EX1353	2302	1810989	786.7024327	6.505679819	203	8877	789	459
Gnamptogenys_helisa_EX1354	2146	1797549	837.6276794	6.657394748	202	2168	850	699
Gnamptogenys_horni_EX1309	2220	2253256	1014.98018	8.375870686	202	8575	1034	1201
Gnamptogenys_horni_GAB111	2198	886827	403.4699727	2.408249899	201	978	398	0
Gnamptogenys_horni_GAB156	2179	2029310	931.3033502	6.985478352	204	2006	957	983
Gnamptogenys_horni_GAB20	2279	2043767	896.7823607	6.013445048	201	4605	915	847
Gnamptogenys_horni_GAB66	2332	1910636	819.3121784	4.847561129	201	2166	825	475
Gnamptogenys_horni_GAB70	2283	1964353	860.4261936	5.973019755	201	4985	873	730
Gnamptogenys_interrupta_GAB193	2251	2140052	950.7116837	7.066353646	204	7116	970	1053
Gnamptogenys_JTL001_EX1209	2325	2196024	944.5264516	6.56928406	201	2309	947	1020
Gnamptogenys_JTL001_EX1212	2257	2212407	980.2423571	7.47194784	201	2520	1000	1129
Gnamptogenys_JTL001_EX1214	2216	2109745	952.0509928	7.381786316	209	3456	957	1025
Gnamptogenys_JTL001_EX1216	2234	2162475	967.9834378	7.854221596	202	5746	981	1072
Gnamptogenys_JTL001_EX1221	2305	1998153	866.8776573	6.169114459	203	2063	864	749
Gnamptogenys_JTL001_EX1275	2226	2154332	967.804133	7.354254595	201	2890	983.5	1074
Gnamptogenys_JTL001_EX1288	2242	1969577	878.4910794	6.010757387	201	2488	896.5	787
Gnamptogenys_JTL001_EX1301	2212	2051789	927.5718807	7.00364927	201	2744	942.5	956
Gnamptogenys_JTL001_EX1355	2220	2054536	925.4666667	7.150940276	201	2640	946	983
Gnamptogenys_JTL001_EX1356	2273	2071630	911.4078311	6.954649334	201	3077	928	975
Gnamptogenys_JTL001_EX1357	2262	2193940	969.9115827	7.054314544	207	2645	989	1091
Gnamptogenys_JTL002_EX1207	2265	2260787	998.1399559	7.189757871	205	2517	1022	1193
Gnamptogenys_JTL002_EX1210	2255	2160527	958.1050998	6.705352927	201	2089	978	1071
Gnamptogenys_JTL002_EX1211	2256	2162073	958.3656915	7.146428039	202	5013	975	1065
Gnamptogenys_JTL002_EX1213	2212	2128364	962.1898734	7.116867814	202	2162	976.5	1042
Gnamptogenys_JTL002_EX1215	2219	2158847	972.8918432	7.521261681	202	4771	981	1070
Gnamptogenys_JTL002_EX1264	2235	2192811	981.1234899	7.036435675	201	3896	995	1112
Gnamptogenys_JTL002_EX1270	2173	1918808	883.0225495	6.7436735	203	2396	896	814
Gnamptogenys_JTL002_EX1273	2203	2017781	915.9241943	6.843024146	201	3508	932	922
Gnamptogenys_JTL002_EX1291	2234	2200710	985.0984781	7.279505937	203	3194	1010.5	1145
Gnamptogenys_JTL002_EX1313	2230	2242120	1005.434978	8.378224814	203	9502	1021	1177
Gnamptogenys_JTL002_EX1339	2265	1735019	766.0128035	5.330589345	203	2786	779	369
Gnamptogenys_JTL002_EX1358	2282	2283613	1000.706836	7.207183264	204	2948	1002	1149
Gnamptogenys_JTL002_EX1359	2295	1829076	796.9830065	5.503094097	203	1903	811	531
Gnamptogenys_JTL002_EX1360	2284	2280199	998.3358144	7.192626888	201	3916	1010.5	1175
Gnamptogenys_JTL005_EX1361	2277	1985159	871.8309179	6.19824406	201	2308	885	771
Gnamptogenys_JTL008_EX1208	2321	1853713	798.6699698	5.391469899	209	2280	809	514
Gnamptogenys_JTL008_EX1268	2291	1754945	766.0170231	5.155154504	202	2012	773	369
Gnamptogenys_JTL009_EX1362	2288	2121807	927.3631993	6.945370394	201	2728	931	941
Gnamptogenys_JTL010_EX1363	2300	2056497	894.1291304	6.248977943	203	2688	914	891

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_JTL011_EX1302	2222	2073035	932.9590459	7.096229533	203	2556	942.5	976
Gnamptogenys_JTL011_EX1364	2253	1872424	831.0803373	6.290501465	203	4981	832	653
Gnamptogenys_JTL011_EX1365	2245	2168730	966.0267261	6.965471957	202	3098	992	1099
Gnamptogenys_JTL011_EX1366	2249	2159643	960.2681192	6.904488627	201	2138	982	1085
Gnamptogenys_JTL011_EX1367	2249	1921754	854.4926634	6.149862734	202	2615	855	700
Gnamptogenys_JTL013_EX1368	2286	1828608	799.9160105	5.665127497	205	2064	806	533
Gnamptogenys_JTL013_GAB138	2257	1714648	759.7022596	5.229779878	202	1892	770	367
Gnamptogenys_JTL013_GAB159	2212	1802523	814.8838156	6.210134421	203	2125	828.5	622
Gnamptogenys_JTL014_EX1369	2216	1940158	875.5225632	6.513005208	204	2698	886.5	802
Gnamptogenys_JTL015_EX1370	2216	2050984	925.534296	6.854629777	201	2546	932.5	942
Gnamptogenys_JTL016_EX1272	2116	1613758	762.6455577	6.135107958	201	1734	776	472
Gnamptogenys_JTL016_EX1371	2241	1758087	784.5100402	6.379455191	201	7227	789	488
Gnamptogenys_JTL018_EX1372	2303	2430770	1055.479809	7.139963193	202	2314	1079	1356
Gnamptogenys_JTL019_EX1312	2217	2041663	920.9124944	6.733729423	201	2216	956	988
Gnamptogenys_JTL019_GAB110	2258	1798846	796.6545616	5.608480755	203	4113	806	458
Gnamptogenys_kempfi_GAB171	2341	1381492	590.1290047	3.746559801	203	1720	588	47
Gnamptogenys_kempfi_GAB24	2249	2102832	935.0075589	6.651427567	202	2556	950	988
Gnamptogenys_lavra_GAB108	2203	1737189	788.5560599	5.539825149	201	2159	810	469
Gnamptogenys_lavra_GAB148	2257	1721874	762.9038547	5.358977894	203	1866	772	408
Gnamptogenys_leiolabia_EX1349	2201	2058751	935.3707406	7.573612798	201	2020	947	1007
Gnamptogenys_lucaris_GAB69	2200	1191572	541.6236364	3.601590718	202	1752	537	10
Gnamptogenys_macretes_GAB185	2273	1536104	675.8046634	4.694410451	201	2470	680	135
Gnamptogenys_mecotyle_EX1336	2291	1960066	855.5504147	5.81208875	202	2425	873	703
Gnamptogenys_mecotyle_EX1373	2265	1946330	859.3068433	5.984140541	202	2491	872	763
Gnamptogenys_mecotyle_EX1374	2268	2148209	947.1820988	6.615088874	201	3974	968	1047
Gnamptogenys_mecotyle_EX1375	2250	2283896	1015.064889	6.957450083	201	2476	1048.5	1240
Gnamptogenys_mecotyle_EX1376	2295	2136494	930.9342048	6.12815982	201	2939	960	1012
Gnamptogenys_mecotyle_EX1377	2266	2111137	931.6579876	6.741335445	201	5526	946	980
Gnamptogenys_mediatrix_GAB146	2336	1936612	829.0291096	5.458101186	204	1876	833.5	608
Gnamptogenys_mediatrix_GAB173	2282	1106746	484.9894829	2.874861719	205	1458	481.5	7
Gnamptogenys_menadensis_EX1378	2242	2040899	910.3028546	6.924057684	202	3000	925	939
Gnamptogenys_mina_GAB195	2355	1603877	681.0518047	5.535549477	202	8551	662	146
Gnamptogenys_minuta_EX1206	2256	2011563	891.650266	6.850341523	201	2284	895.5	856
Gnamptogenys_minuta_EX1222	2278	1935792	849.7769974	6.835038563	201	5570	854.5	757
Gnamptogenys_minuta_EX1226	2282	1804471	790.7410167	5.92532372	203	3163	788	561
Gnamptogenys_minuta_EX1228	2281	1838605	806.0521701	5.828880801	201	2317	819	583
Gnamptogenys_minuta_EX1230	2272	2013106	886.0501761	7.415343859	201	9039	897.5	850
Gnamptogenys_minuta_EX1238	2267	2190362	966.1940891	7.748109792	201	6097	970	1076
Gnamptogenys_minuta_EX1242	2251	1928753	856.8427366	6.70883616	201	2409	862	770

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_minuta_EX1245	2259	1954493	865.2027446	6.063540722	201	2222	878	773
Gnamptogenys_minuta_EX1247	2266	2235946	986.7369815	7.76567747	203	6112	989.5	1115
Gnamptogenys_minuta_EX1251	2256	1849814	819.9530142	6.711937688	201	8551	829	572
Gnamptogenys_minuta_EX1253	2237	2154881	963.2905677	7.782979053	201	6466	962	1026
Gnamptogenys_minuta_EX1255	2241	2001300	893.038822	6.942175116	201	3612	899	841
Gnamptogenys_minuta_EX1263	2263	2203707	973.7989395	7.51624254	201	2790	990	1111
Gnamptogenys_minuta_EX1265	2256	1959482	868.5647163	6.433366736	201	2317	876.5	799
Gnamptogenys_minuta_EX1269	2224	1946041	875.0184353	6.863747596	201	2252	881.5	822
Gnamptogenys_minuta_EX1279	2232	1950681	873.9610215	6.780156577	203	3148	880.5	836
Gnamptogenys_minuta_EX1303	2221	2078457	935.8203512	7.416660554	201	3606	942	969
Gnamptogenys_minuta_EX1304	2223	2060947	927.1016644	10.13769691	204	16070	935	954
Gnamptogenys_minuta_EX1311	2249	1897214	843.5811472	6.283504233	201	2390	841	712
Gnamptogenys_minuta_EX1315	2244	1900607	846.9728164	7.000877663	201	5890	858.5	708
Gnamptogenys_minuta_EX1322	2242	1963865	875.9433541	6.631503587	204	3910	878	815
Gnamptogenys_minuta_EX1329	2238	1698598	758.9803396	5.86895869	201	2727	761	462
Gnamptogenys_minuta_EX1335	2237	1656863	740.6629414	5.682951453	202	2340	745	393
Gnamptogenys_minuta_EX1379	2226	1810858	813.5031447	6.259947199	201	2299	823.5	627
Gnamptogenys_minuta_GAB122	2184	1295499	593.1771978	5.330353356	201	8446	593	24
Gnamptogenys_minuta_GAB180	1962	671582	342.2945974	2.153368106	201	1721	334	3
Gnamptogenys_minuta_GAB67	2196	1964052	894.3770492	6.701390825	205	2206	910	852
Gnamptogenys_moelleri_GAB02	2241	2038275	909.5381526	6.523818765	201	5169	938	938
Gnamptogenys_moelleri_GAB100	2201	1401569	636.7873694	3.958316913	202	1605	652	41
Gnamptogenys_moelleri_GAB102	1663	590050	354.8105833	2.575922204	201	920	340	0
Gnamptogenys_moelleri_GAB139	2242	1864320	831.5432649	5.455572856	201	2315	855	607
Gnamptogenys_moelleri_GAB140	2238	1980997	885.1639857	6.083541241	201	1732	911.5	845
Gnamptogenys_moelleri_GAB179	2230	1513154	678.5443946	4.183196545	201	1673	697.5	73
Gnamptogenys_moelleri_GAB25	2226	2074013	931.7219227	6.083189821	202	3245	961	968
Gnamptogenys_moelleri_GAB74	2262	2353698	1040.538462	7.051550883	205	2671	1066	1288
Gnamptogenys_moelleri_GAB83	2278	1867915	819.9802458	5.245735992	204	2029	839	533
Gnamptogenys_petiscapa_GAB194	2236	1769802	791.5035778	5.309305385	203	2685	798	417
Gnamptogenys_pleurodon_GAB132	2311	1876512	811.9913457	4.915856161	201	2984	825	426
Gnamptogenys_pleurodon_GAB27	1647	496285	301.3266545	1.775319551	201	724	287	0
Gnamptogenys_pleurodon_GAB31	1780	521127	292.7679775	1.608993996	201	847	282	0
Gnamptogenys_pleurodon_GAB46	2213	2170971	981.0081338	6.733511861	211	2155	1012	1147
Gnamptogenys_pleurodon_GAB76	2211	1939371	877.14654	6.163230414	201	2277	902	825
Gnamptogenys_porcata_EX1217	2307	1663221	720.9453836	4.779179505	201	1916	722	238
Gnamptogenys_porcata_EX1218	2246	1435667	639.2105966	4.471319715	204	1627	641	92
Gnamptogenys_porcata_EX1235	2206	1742925	790.0838622	5.588534068	202	1998	807	486
Gnamptogenys_porcata_EX1240	2208	1658610	751.1820652	6.503595624	201	8572	758	347

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_porcata_EX1244	2236	1824973	816.1775492	5.735641906	201	4731	838	541
Gnamptogenys_porcata_EX1254	2255	1839907	815.9232816	6.11281239	201	4261	831	608
Gnamptogenys_porcata_EX1259	2246	2015853	897.530276	6.512375374	201	3978	917	866
Gnamptogenys_porcata_EX1280	2204	1836732	833.3629764	6.150906494	201	1913	858.5	662
Gnamptogenys_porcata_EX1282	2262	1674331	740.1993811	4.969477178	202	2298	755	285
Gnamptogenys_porcata_EX1285	2246	1331505	592.833927	4.031809263	201	1797	594	31
Gnamptogenys_porcata_EX1290	2276	1499410	658.7917399	4.655237862	201	1771	649.5	141
Gnamptogenys_porcata_EX1296	2253	1923921	853.9374168	7.337439785	205	7787	857	749
Gnamptogenys_porcata_EX1299	2224	1870512	841.057554	6.520305403	203	2413	859.5	733
Gnamptogenys_porcata_EX1314	2239	1708811	763.2027691	5.724912968	201	3613	766	421
Gnamptogenys_porcata_EX1328	2220	1759516	792.5747748	5.86800287	201	1878	815	534
Gnamptogenys_porcata_EX1338	2230	1582432	709.6107623	5.089079799	202	1661	717	230
Gnamptogenys_porcata_EX1380	2230	1895039	849.7932735	6.384104611	201	2030	871	755
Gnamptogenys_porcata_EX1381	2251	1648025	732.1301644	5.377772413	201	2153	740	314
Gnamptogenys_porcata_EX1382	2237	1696326	758.3039785	5.595100388	202	1904	769	431
Gnamptogenys_porcata_EX1383	2225	1441654	647.934382	4.67674292	201	2256	652	95
Gnamptogenys_porcata_GAB153	1294	332828	257.2086553	1.315314184	201	723	248	0
Gnamptogenys_posteropsis_EX1384	2205	1743970	790.9160998	6.677224492	201	2586	782	583
Gnamptogenys_rastrata_GAB174	2197	1761466	801.7596723	5.714033545	202	1672	822	566
Gnamptogenys_rastrata_GAB72	2257	1091349	483.5396544	3.066465944	201	1532	480	4
Gnamptogenys_regularis_EX1298	2271	2476147	1090.333333	7.70024662	209	2659	1137	1414
Gnamptogenys_regularis_EX1307	2154	2125472	986.7558032	7.326562323	201	3017	1030.5	1130
Gnamptogenys_regularis_EX1385	2293	1982374	864.5329263	6.050640835	203	1941	878	789
Gnamptogenys_regularis_EX1386	2235	1880057	841.1888143	5.555710763	201	1895	857	650
Gnamptogenys_regularis_GAB129	1611	517785	321.405959	2.124283466	201	896	307	0
Gnamptogenys_reinchenspergeri_GAB79	2170	2125357	979.4271889	7.030884763	202	3883	1020	1133
Gnamptogenys_relicta_GAB03	2274	2188231	962.2827617	6.844662419	201	3746	983	1101
Gnamptogenys_relicta_GAB18	2262	1789755	791.2267905	5.212782275	203	1805	803	471
Gnamptogenys_simulans_EX1219	2249	1812833	806.0618052	6.215156612	203	1900	803	603
Gnamptogenys_simulans_EX1223	2232	1929645	864.5362903	6.905288552	203	3024	867.5	770
Gnamptogenys_simulans_EX1227	2222	2015655	907.1354635	7.296338668	201	5617	919	912
Gnamptogenys_simulans_EX1234	2232	1997721	895.0362903	7.20456222	201	5068	901.5	886
Gnamptogenys_simulans_EX1236	2237	1839059	822.1095217	6.306439728	202	2057	816	649
Gnamptogenys_simulans_EX1243	2251	1673865	743.6095069	5.962973761	201	5503	734	384
Gnamptogenys_simulans_EX1257	2223	1973514	887.7705803	6.975795572	203	2118	888	841
Gnamptogenys_simulans_EX1260	2272	2240470	986.1223592	7.5981586	203	3222	982	1094
Gnamptogenys_simulans_EX1262	2281	1737163	761.5795704	5.449562074	203	2538	757	401
Gnamptogenys_simulans_EX1267	2222	1807157	813.3019802	6.069166223	203	1927	819.5	607
Gnamptogenys_simulans_EX1276	2238	1918310	857.1537087	6.792783789	202	3537	854	778

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_simulans_EX1277	2204	1871766	849.2586207	6.839197613	201	2274	847.5	727
Gnamptogenys_simulans_EX1330	2222	1665578	749.5850585	5.978297317	202	3732	751	426
Gnamptogenys_simulans_EX1340	2247	1703806	758.2581219	5.597469517	201	2054	753	423
Gnamptogenys_simulans_EX1387	2276	1782831	783.3176626	5.967696947	202	2706	779.5	523
Gnamptogenys_simulans_EX1388	2277	1664470	730.992534	6.006128409	201	6934	735	327
Gnamptogenys_solomonensis_GAB190	2269	2370652	1044.800353	8.004396383	201	2554	1073	1283
Gnamptogenys_sp_GAB16	2177	1284507	590.0353698	4.330466221	201	2059	589	43
Gnamptogenys_sp_GAB26	2171	1762090	811.6490097	6.249248477	201	1856	820	602
Gnamptogenys_sp.06_GAB78	2224	1697353	763.1982914	5.628550879	201	2045	775.5	403
Gnamptogenys_sp.n.C_GAB155	2016	726764	360.4980159	2.260119104	201	2238	354	1
Gnamptogenys_sp001_GAB199	2225	2088503	938.6530337	8.779082115	201	14813	966	1007
Gnamptogenys_stellae_EX1389	2278	1894611	831.6992976	5.887718263	206	1812	851	669
Gnamptogenys_striatula_EX1248	2250	2158147	959.1764444	6.188476583	211	2940	988	1084
Gnamptogenys_striatula_EX1287	2254	1893062	839.8677906	5.362079259	201	3480	861	606
Gnamptogenys_striatula_EX1295	2225	2159433	970.5316854	6.649929604	202	2231	1007	1123
Gnamptogenys_striatula_EX1316	2251	2166788	962.5890715	6.663815426	203	3661	993	1110
Gnamptogenys_striatula_EX1324	2249	2019274	897.8541574	6.489688464	201	2205	922	914
Gnamptogenys_striatula_EX1390	2220	2175080	979.7657658	6.975996875	204	2426	1013	1151
Gnamptogenys_striatula_EX1391	2229	2022467	907.3427546	6.196499506	201	2387	929	910
Gnamptogenys_striatula_EX1392	2261	1924700	851.2605042	5.5352223	202	1852	867	679
Gnamptogenys_striatula_EX1393	2251	2061264	915.710351	6.160780869	204	1738	937	955
Gnamptogenys_striatula_EX1394	2280	1629292	714.6017544	4.741762118	202	1665	722	224
Gnamptogenys_striatula_GAB01	2207	1884627	853.9315813	6.156138757	201	5515	877	722
Gnamptogenys_striatula_GAB10	2206	2185630	990.7660925	6.361337785	201	2717	1032	1197
Gnamptogenys_striatula_GAB105	2257	1444607	640.0562694	4.3061397	201	1643	635	92
Gnamptogenys_striatula_GAB106	2243	1971673	879.0338832	5.767757062	201	2594	907	786
Gnamptogenys_striatula_GAB11	2236	1467741	656.4136852	4.297086765	201	1638	666	68
Gnamptogenys_striatula_GAB145	2257	1560420	691.369074	4.30629795	201	1730	709	100
Gnamptogenys_striatula_GAB147	2274	1462755	643.2519789	4.488300548	201	1469	646.5	94
Gnamptogenys_striatula_GAB21	2228	2103418	944.0834829	6.705616919	201	2055	975.5	1060
Gnamptogenys_striatula_GAB50	2228	1703166	764.4371634	4.950031306	201	1616	785	345
Gnamptogenys_striatula_GAB56	2251	1404158	623.7929809	4.018405297	201	1624	632	46
Gnamptogenys_striatula_GAB73	2307	1766521	765.72215	5.150762219	201	4314	774	285
Gnamptogenys_striatula_GAB75	2261	1295348	572.9093322	3.58162194	201	1758	576	20
Gnamptogenys_strigata_EX1220	2275	1762393	774.6782418	5.822568407	202	5923	783	395
Gnamptogenys_strigata_EX1225	840	236116	281.0904762	2.243000326	201	806	270	0
Gnamptogenys_strigata_EX1232	2239	1930397	862.169272	6.791018914	201	6387	875	714
Gnamptogenys_strigata_EX1233	2256	1956857	867.4011525	5.918687097	202	3128	885	780
Gnamptogenys_strigata_EX1237	2235	1722485	770.6868009	5.487395091	202	2692	784	395

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_strigata_EX1239	2204	2028564	920.4010889	6.848295104	201	1828	941.5	968
Gnamptogenys_strigata_EX1241	2225	2012437	904.4660674	6.58984274	201	2705	922	910
Gnamptogenys_strigata_EX1250	2210	1978833	895.3995475	6.406798983	203	2116	924	872
Gnamptogenys_strigata_EX1256	2250	1698123	754.7213333	5.39510548	201	2191	763	357
Gnamptogenys_strigata_EX1258	2178	1076674	494.3406795	3.27658181	202	1375	487.5	8
Gnamptogenys_strigata_EX1261	2196	2074667	944.7481785	6.737114029	201	2210	978	1052
Gnamptogenys_strigata_EX1266	2194	1797419	819.2429353	5.539498807	204	2309	841	559
Gnamptogenys_strigata_EX1271	2166	1752403	809.0503232	6.131399627	201	4901	826	547
Gnamptogenys_strigata_EX1278	2129	1761765	827.5082198	6.066309279	201	2250	837	626
Gnamptogenys_strigata_EX1283	2231	1923775	862.2926939	6.163846492	206	3093	868	740
Gnamptogenys_strigata_EX1297	2185	2051358	938.8366133	6.828630857	201	2046	972	1015
Gnamptogenys_strigata_EX1305	2171	1878983	865.4919392	6.551174195	201	2040	883	798
Gnamptogenys_strigata_EX1318	2223	1611666	724.9959514	5.096872744	201	1875	731	286
Gnamptogenys_strigata_EX1319	2216	1685352	760.5379061	5.636627497	201	2062	774	410
Gnamptogenys_strigata_EX1321	2225	1783149	801.4152809	6.415146988	202	5908	804	546
Gnamptogenys_strigata_EX1323	2247	1632061	726.328883	5.231829709	202	2475	732	293
Gnamptogenys_strigata_EX1327	2226	1618744	727.1985624	5.45271423	202	1814	732.5	320
Gnamptogenys_strigata_EX1337	2241	1626186	725.6519411	5.254933156	201	2131	733	285
Gnamptogenys_strigata_EX1341	2169	1521903	701.6611342	5.23840952	203	1882	697	234
Gnamptogenys_strigata_EX1395	2194	1763982	804.0027347	5.991808782	203	2587	813	571
Gnamptogenys_strigata_EX1396	2209	1626825	736.4531462	5.500352806	204	1823	744	356
Gnamptogenys_strigata_EX1397	2215	1378968	622.5589165	4.335812355	201	2912	622	51
Gnamptogenys_strigata_GAB178	2221	1421043	639.8212517	4.591524142	202	1831	641	91
Gnamptogenys_sulcata_EX1398	2256	2130891	944.543883	7.38471421	203	7567	957	1027
Gnamptogenys_sulcata_EX1399	2224	2168029	974.8331835	7.508800654	201	2336	991	1096
Gnamptogenys_sulcata_EX1400	2237	2172968	971.3759499	7.863664582	201	2740	986	1095
Gnamptogenys_sulcata_EX1401	2252	1803595	800.8858792	6.152573147	202	4918	808	555
Gnamptogenys_sulcata_GAB164	2182	1034357	474.0407883	3.407373769	201	1500	462	10
Gnamptogenys_sulcata_GAB177	2227	2247781	1009.331388	7.574283436	202	2429	1031	1193
Gnamptogenys_sulcata_GAB48	2148	812419	378.2211359	3.247544316	201	5355	362	4
Gnamptogenys_tornata_EX1205	2306	2347402	1017.954033	6.986773584	213	4049	1034.5	1233
Gnamptogenys_tornata_EX1224	2276	2224850	977.526362	6.805647479	201	2406	1000	1141
Gnamptogenys_tornata_EX1229	2218	1995888	899.8593327	6.27986482	201	2503	914	872
Gnamptogenys_tornata_EX1231	2316	2125943	917.9373921	5.949252928	209	2494	927.5	931
Gnamptogenys_tornata_EX1246	2321	2349749	1012.386471	7.012784274	207	4007	1022	1226
Gnamptogenys_tornata_EX1249	2274	2119365	931.9986807	6.258591474	208	2370	952	1006
Gnamptogenys_tornata_EX1252	2313	2276794	984.3467358	6.99854447	208	4812	1000	1157
Gnamptogenys_tornata_EX1281	2322	2310164	994.9026701	6.736665578	201	3732	1019	1219
Gnamptogenys_tornata_EX1289	2202	1974720	896.7847411	6.940010107	203	4651	902	853

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Gnamptogenys_tornata_EX1306	2265	2510628	1108.445033	8.109184929	206	3220	1138	1395
Gnamptogenys_tornata_EX1333	2293	2043318	891.111208	6.347918169	201	2120	904	899
Gnamptogenys_tornata_EX1402	2260	2175188	962.4725664	7.49213449	201	2210	975.5	1079
Gnamptogenys_tornata_EX1403	2205	2272487	1030.606349	8.275886169	202	2730	1061	1210
Gnamptogenys_tornata_EX1404	2244	1965581	875.9273619	6.988407683	201	2215	885	847
Gnamptogenys_tornata_EX1405	2306	2142755	929.2085863	7.309137323	201	2440	922	987
Gnamptogenys_tornata_EX1406	2227	2429387	1090.878761	8.79157506	201	3605	1123	1338
Gnamptogenys_tornata_EX1407	2256	2166903	960.5066489	7.438193062	202	2873	978.5	1083
Gnamptogenys_tornata_EX1408	2307	1875501	812.9609883	6.24714741	201	1828	807	630
Gnamptogenys_tornata_GAB166	2327	1478046	635.1723249	4.130380738	202	1864	643	68
Gnamptogenys_tortuolosa_GAB101	2245	1924863	857.4	6.218288787	202	1810	873	777
Gnamptogenys_tortuolosa_GAB15	2269	1973054	869.5698546	5.994325345	201	2218	881	781
Gnamptogenys_tortuolosa_GAB160	2178	1096752	503.5592287	3.307762296	204	1492	500	6
Gnamptogenys_tortuolosa_GAB163	2337	1446506	618.9584938	3.975865495	203	3833	621	44
Gnamptogenys_tortuolosa_GAB32	2237	1133570	506.7367009	3.699312603	201	3909	501	12
Gnamptogenys_treta_GAB184	2119	952681	449.5899009	2.817757036	201	1997	453	5
Gnamptogenys_triangularis_EX1308	2288	2380827	1040.571241	7.103386741	207	2716	1070	1335
Gnamptogenys_triangularis_GAB123	2087	853421	408.9223766	2.457640685	201	1864	404	1
Gnamptogenys_volcano_EX1409	2206	1471829	667.193563	4.952927516	201	1720	671	159
Gnamptogenys_wilsoni_GAB154	1860	568100	305.4301075	1.644650017	201	1015	298	1
Heteroponera_brounii_GPC07	2246	1964001	874.4439003	5.774792547	203	2329	906	817
Heteroponera_brounii_GPC08	2230	2035283	912.6829596	6.138749445	203	3853	940.5	936
Heteroponera_carinifrons_GPC09	2277	1341887	589.322354	3.856800445	201	1860	592	35
Heteroponera_dentinodis_HET07	833	906476	1088.206483	15.30495817	201	3257	1142	518
Heteroponera_dentinodis_HET08	831	880808	1059.937425	14.04496863	204	3338	1126	516
Heteroponera_dentinodis_HET09	686	269080	392.244898	4.519326123	201	970	380	0
Heteroponera_dentinodis_HET10	908	933780	1028.39207	15.01763637	202	2945	1059.5	506
Heteroponera_dentinodis_HET11	2309	1541942	667.7964487	4.086473944	204	3743	677	58
Heteroponera_dolo_GPC10	2261	2306819	1020.264927	7.232398992	202	7979	1059	1306
Heteroponera_dolo_GPC11	2290	1970184	860.3423581	5.976970448	201	6627	876	669
Heteroponera_imbelis_GPC12	1816	591750	325.8535242	2.048679953	201	1134	313	1
Heteroponera_inermis_GPC13	1953	650984	333.3251408	1.839686812	201	1036	325	1
Heteroponera_mayri_COD03	868	976618	1125.135945	15.61167888	202	3389	1166	559
Heteroponera_mayri_COD21	690	231537	335.5608696	3.623770627	201	851	321	0
Heteroponera_mayri_HET02	940	850882	905.193617	16.58897343	201	10562	899.5	387
Heteroponera_mayri_HET04	876	1019623	1163.953196	17.05649135	201	3372	1231.5	578
Heteroponera_mayri_HET12	2325	1303291	560.5552688	3.419392781	201	2227	558	20
Heteroponera_mayri_HET27	697	239547	343.6829268	3.699084982	201	773	325	0
Heteroponera_mayri_HET27Pool8	840	485719	578.2369048	8.159334714	201	1543	556.5	41

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Heteroponera_mayri_HET43	2354	1775466	754.2336449	4.39972576	201	2004	758.5	235
Heteroponera_mayri_HET46	808	439350	543.75	7.599952068	202	1447	512.5	24
Heteroponera_mayri_HET48	2337	1529104	654.3020967	4.302248324	206	4066	658	67
Heteroponera_mayri_HET50	2290	1007764	440.0716157	2.55200102	201	1388	432	6
Heteroponera_microps_GPC14	2286	2148871	940.0135608	5.548205945	211	3149	971	1033
Heteroponera_microps_GPC15	2218	2097815	945.8137962	6.229718648	202	2547	978.5	1049
Heteroponera_monticola_GPC16	2321	2057291	886.3813012	5.857447771	203	2109	911	879
Heteroponera_panamensis_PAN01	2156	807697	374.627551	2.171004751	201	2061	365.5	1
Heteroponera_panamensis_PAN04	828	812898	981.7608696	13.59860467	201	2537	1031.5	438
Heteroponera_panamensis_PAN05	2233	2386684	1068.824004	7.57317921	202	2332	1109	1381
Heteroponera_panamensis_PAN06	2321	1724234	742.8841017	4.406426649	206	2098	753	208
Heteroponera_panamensis_PAN08	795	611488	769.1672956	9.356275941	203	2389	790	131
Heteroponera_relicta_GPC17	2271	1808583	796.3817701	5.039051674	203	3394	815	436
Heteroponera_relicta_GPC18	2195	979903	446.4250569	2.891334415	201	1318	435	4
Heteroponera_relicta_GPC19	2177	974576	447.6692696	3.040384046	201	1188	433	7
Heteroponera_rhodopygea_GPC20	2329	1488733	639.2155432	3.795222086	204	1816	653	43
Heteroponera_robusta_GPC21	2274	2315155	1018.098065	6.481013773	201	3200	1051	1309
Heteroponera_sp.n.E_GPC22	1632	450741	276.1893382	1.498256614	201	837	262	0
Lasius_sitiens_EX1585	2224	2232011	1003.602068	7.687386769	201	2724	1033	1196
Megalomyrmex_silvestrii_EX821	2201	2103183	955.5579282	6.936822693	201	2602	985	1073
Myrmecia_varians_EX1564	2292	3047085	1329.443717	9.941137471	206	11893	1350	1822
Myrmelachista_joycei_EX819	1931	1397315	723.6224754	6.504747658	201	1687	729	361
Myrmicinae_sp_EXT01	578	184271	318.8079585	3.323396419	202	628	307	0
Philidris_sp_EX1581	2284	2399647	1050.633538	7.33433881	201	2091	1098	1363
Pseudomyrmex_gracilis_EX853	2134	1814478	850.2708529	7.01114925	201	2954	846	674
Rhytidoponera_acanthoponeroides_RHY14	2178	1058514	486.0027548	3.340006682	202	3680	485	4
Rhytidoponera_cerastes_RHY35	1923	731753	380.5267811	2.481246568	201	878	371	0
Rhytidoponera_cristata_RHY36	2150	1568405	729.4906977	5.094092217	201	3730	753	224
Rhytidoponera_croesus_RHY17	2313	2003145	866.0376135	5.037359728	202	1820	889	673
Rhytidoponera_dubia_RHY01	2342	1755059	749.3847139	4.568700268	201	2088	760	252
Rhytidoponera_eremita_RHY26	2209	969406	438.8438207	2.835190598	201	2613	433	3
Rhytidoponera_foreli_RHY05	2372	2016056	849.9392917	4.839090953	202	1929	858	540
Rhytidoponera_fulgens_RHY13	1783	532943	298.9024117	1.696204415	201	845	286	0
Rhytidoponera_haeckeli_RHY18	2347	1733432	738.5734981	4.520010105	203	4091	748	158
Rhytidoponera_incisa_RHY27	2282	1360976	596.3961437	3.508310794	201	2098	607	14
Rhytidoponera_kurandensis_RHY09	2358	2250857	954.5619169	6.173990112	204	2538	973.5	1087
Rhytidoponera_laticeps_RHY24	2299	1627564	707.9443236	4.334425769	203	1908	728	141
Rhytidoponera_metallica_RHY06	2300	1616366	702.7678261	4.054625588	201	1653	717	100
Rhytidoponera_numeensis_RHY12	2325	1610350	692.6236559	3.864772844	201	1696	714	80

Species_Extraction code	Contigs	Total bp	Mean length	95 CI length	Min length	Max length	Median legnth	Contigs >1kb
Rhytidoponera_punctata_RHY28	553	133347	241.1338156	1.941755899	201	673	231	0
Rhytidoponera_purpurea_RHY16	2293	1409229	614.5787178	3.73286873	203	1797	621	30
Rhytidoponera_reflexa_RHY21	2306	1752572	760.0052038	4.968539508	201	1701	779.5	332
Rhytidoponera_sp.2_GAB192	2322	1562123	672.7489233	4.245152954	202	1925	680	119
Rhytidoponera_sp.26_RHY33	1845	580434	314.598374	1.713115978	201	841	307	0
Rhytidoponera_sp.8_GAB191	2244	2555534	1138.829768	7.325885693	206	2920	1172.5	1530
Rhytidoponera_sp.9_GAB189	2325	2101604	903.9156989	5.99115622	202	2815	931	919
Rhytidoponera_sp.9_RHY03	2341	1899674	811.4797095	5.042850413	206	4374	834	446
Rhytidoponera_sp.D_RHY04	2332	1653118	708.8842196	3.939192016	202	1888	729	112
Rhytidoponera_sp.F_RHY20	2165	1164001	537.6448037	3.383133162	201	1454	548	5
Rhytidoponera_sp.FER_RHY30	2277	1303052	572.267018	3.457571583	201	3289	583	12
Rhytidoponera_sp.G_RHY02	2311	1374376	594.7105149	3.823561406	201	2099	601	31
Rhytidoponera_sp.X_RHY37	2224	1577796	709.4406475	4.747286057	201	1654	731	175
Rhytidoponera_versicolor_RHY11	2339	1615847	690.8281317	7.499797378	201	15750	699	71
Rhytidoponera_victoriae_RHY08	2319	1942851	837.7968952	5.143454421	202	2633	864	595
Stenamma_alas_EX832	2333	1963865	841.7766824	4.814055388	204	1949	875	604
Tapinoma_sessile_EX1584	2223	2071572	931.8812416	6.325025765	203	2381	971	1019
Tetramorium_boltoni_EX842	2283	2210214	968.1182654	5.918702458	205	2374	995	1133
Typhlomyrmex_major_TYP01	2144	991686	462.5401119	3.078369511	201	1734	459.5	10
Typhlomyrmex_pusillus_TYP03	1919	644322	335.7592496	1.942362737	201	912	325	0
Typhlomyrmex_pusillus_TYP07	2146	1972411	919.110438	7.134002609	201	1947	948	939
Typhlomyrmex_rogenhoferi_TYP04	2208	1616324	732.0307971	6.485030084	204	9691	733.5	276
Typhlomyrmex_rogenhoferi_TYP05	2198	1529449	695.8366697	4.870067013	201	1873	696	183
Typhlomyrmex_rogenhoferi_TYP06	2122	1782595	840.0541942	6.649923542	201	1816	867	699

1742

1743 **Anexo II**

1744 **Supplementary material S2.** Taxon distribution matrix used in the biogeographic analyses.

Species_Extraction code	Region	Species_Extraction code	Region
Anillidris_ANIFei21	Neotropical	*Gnamptogenys_helisa_EX1354*	Indo-Australian
Bothriomyrmex_sp_BotFei19	Neotropical	*Gnamptogenys_JTL001_EX1357*	Neotropical
Ectatomma_zapotecum_ECTA02	Neotropical	*Gnamptogenys_JTL005_EX1361*	Neotropical
Ectatomma_permagnum_ECTA03	Neotropical	*Gnamptogenys_JTL009_EX1362*	Neotropical
Ectatomma_muticum_ECTA04	Neotropical	*Gnamptogenys_JTL010_EX1363*	Neotropical
Ectatomma_edentatum_ECTA05	Neotropical	*Gnamptogenys_JTL013_EX1368*	Neotropical
Ectatomma_tuberculatum_ECTA07	Neotropical	*Gnamptogenys_JTL014_EX1369*	Neotropical
Ectatomma_ruidum_ECTA08	Neotropical	*Gnamptogenys_JTL015_EX1370*	Neotropical
Ectatomma_opaciventre_ECTA09	Neotropical	*Gnamptogenys_JTL016_EX1371*	Neotropical
Ectatomma_goninion_ECTA11	Neotropical	*Gnamptogenys_JTL018_EX1372*	Neotropical
Ectatomm_ruidum_ECTA12	Neotropical	*Gnamptogenys_menadensis_EX1378*	Indo-Australian
Ectatomma_morgani_ECTA14	Neotropical	*Gnamptogenys_posteropsis_EX1384*	Indo-Australian
Ectatomma_lugens_ECTA15	Neotropical	*Gnamptogenys_stellae_EX1389*	Neotropical
Ectatomma_brunneum_ECTA16	Neotropical	*Gnamptogenys_sulcata_EX1398*	Neotropical
Ectatomma_planidens_ECTA19	Neotropical	*Gnamptogenys_sulcata_EX1399*	Neotropical
Gnamptogenys_minuta_EX1206	Neotropical	*Gnamptogenys_tornata_EX1408*	Neotropical
Gnamptogenys_JTL002_EX1207	Neotropical	*Gnamptogenys_volcano_EX1409*	Neotropical
Gnamptogenys_JTL008_EX1208	Neotropical	*Myrmecia_varians_EX1564*	Australasian
Gnamptogenys_porcata_EX1217	Neotropical	*Cerapachys_sp_EX1579*	Neotropical
Gnamptogenys_simulans_EX1219	Neotropical	*Eciton_burchellii_EX1580*	Neotropical
Gnamptogenys_strigata_EX1250	Neotropical	*Philidris_sp_EX1581*	Indo-Australian
Gnamptogenys_striatula_EX1287	Neotropical	*Formica_podzolica_EX1583*	Neartic
Gnamptogenys_gr.coxalis_EX1293	Indo-Australian	*Tapinoma_sessile_EX1584*	Neartic
Gnamptogenys_cribrata_EX1294	Indo-Australian	*Lasius_sitiens_EX1585*	Neartic
Gnamptogenys_regularis_EX1298	Neotropical	*Ectatomma_gibbum_EX814*	Neotropical
Gnamptogenys_JTL011_EX1302	Neotropical	*Myrmelachista_joycei_EX819*	Neotropical
Gnamptogenys_triangularis_EX1308	Neotropical	*Megalomyrmex_silvestrii_EX821*	Neotropical
Gnamptogenys_JTL019_EX1312	Neotropical	*Stenamma_alas_EX832*	Neotropical
Gnamptogenys_bispinosa_EX1320	Neotropical	*Tetramorium_boltoni_EX842*	Afrotropical
Gnamptogenys_strigata_EX1323	Neotropical	*Dolichoderus_lamellosus_EX843*	Neotropical
Gnamptogenys_mecotyle_EX1336	Neotropical	*Pseudomyrmex_gracilis_EX853*	Neotropical
Gnamptogenys_banksi_EX1344	Neotropical	*Gnamptogenys_moelleri_GAB02*	Neotropical
Gnamptogenys_binghamii_EX1345	Indo-Australian	*Gnamptogenys_relicta_GAB03*	Neotropical
Gnamptogenys_coxalis_EX1347	Indo-Australian	*Gnamptogenys_striatula_GAB145*	Neotropical
Gnamptogenys_fontana_EX1348	Indo-Australian	*Gnamptogenys_mediatrix_GAB146*	Neotropical

Gnamptogenys_leiolabia_EX1349	Indo-Australian	*Gnamptogenys_lavra_GAB148*	Neotropical
Gnamptogenys_hartmani_EX1353	Neotropical	*Gnamptogenys_tortuolosa_GAB15*	Neotropical
Species_Extraction code	**Region**	**Species_Extraction code**	**Region**
Gnamptogenys_extra_GAB150	Neotropical	*Gnamptogenys_boliviensis_GAB98*	Neotropical
Gnamptogenys_brunnea_GAB152	Neotropical	*Acanthoponera_goeldii_GPC01*	Neotropical
Gnamptogenys_wilsoni_GAB154	Neotropical	*Acanthoponera_sp.n.B_GPC03*	Neotropical
Gnamptogenys_sp.n.C_GAB155	Neotropical	*Acanthoponera_mucronata_GPC05*	Neotropical
Gnamptogenys_haesnchi_GAB169	Neotropical	*Heteroponera_brounii_GPC08*	Australasian
Gnamptogenys_kempfi_GAB171	Neotropical	*Heteroponera_carinifrons_GPC09*	Neotropical
Gnamptogenys_mediatrix_GAB173	Neotropical	*Heteroponera_dolo_GPC10*	Neotropical
Gnamptogenys_sulcata_GAB177	Neotropical	*Heteroponera_imbelis_GPC12*	Australasian
Gnamptogenys_atrata_GAB182	Australasian	*Heteroponera_inermis_GPC13*	Neotropical
Gnamptogenys_macretes_GAB185	Australasian	*Heteroponera_microps_GPC14*	Neotropical
Gnamptogenys_crenaticeps_GAB187	Australasian	*Heteroponera_monticola_GPC16*	Neotropical
Gnamptogenys_albiclava_GAB188	Australasian	*Heteroponera_relicta_GPC17*	Australasian
Rhytidoponera_sp.9_GAB189	Australasian	*Heteroponera_rhodopygea_GPC20*	Australasian
Gnamptogenys_solomonensis_GAB190	Australasian	*Heteroponera_robusta_GPC21*	Neotropical
Rhytidoponera_sp.8_GAB191	Australasian	*Heteroponera_sp.n.E_GPC22*	Neotropical
Rhytidoponera_sp.2_GAB192	Australasian	*Heteroponera_dentinodis_HET11*	Neotropical
Gnamptogenys_interrupta_GAB193	Neotropical	*Heteroponera_mayri_HET12*	Neotropical
Gnamptogenys_petiscapa_GAB194	Neotropical	*Acanthoponera_minor_MIN02*	Neotropical
Gnamptogenys_mina_GAB195	Neotropical	*Myrmicinae_sp_EXT01*	Neotropical
Gnamptogenys_gracilis_GAB196	Neotropical	*Heteroponera_panamensis_PAN04*	Neotropical
Gnamptogenys_enodis_GAB198	Neotropical	*Rhytidoponera_dubia_RHY01*	Australasian
Gnamptogenys_sp001_GAB199	Neotropical	*Rhytidoponera_sp.G_RHY02*	Australasian
Gnamptogenys_annulata_GAB23	Neotropical	*Rhytidoponera_sp.9_RHY03*	Australasian
Gnamptogenys_bisulca_GAB43	Neotropical	*Rhytidoponera_sp.D_RHY04*	Australasian
Gnamptogenys_pleurodon_GAB46	Neotropical	*Rhytidoponera_foreli_RHY05*	Australasian
Gnamptogenys_sulcata_GAB48	Neotropical	*Rhytidoponera_metallica_RHY06*	Australasian
Gnamptogenys_acuminata_GAB61	Neotropical	*Rhytidoponera_victoriae_RHY08*	Australasian
Gnamptogenys_lucaris_GAB69	Neotropical	*Rhytidoponera_kurandensis_RHY09*	Australasian
Gnamptogenys_horni_GAB70	Neotropical	*Rhytidoponera_versicolor_RHY11*	Australasian
Gnamptogenys_rastrata_GAB72	Neotropical	*Rhytidoponera_numeensis_RHY12*	Australasian
Gnamptogenys_pleurodon_GAB76	Neotropical	*Rhytidoponera_fulgens_RHY13*	Australasian
Gnamptogenys_ammophila_GAB77	Neotropical	*Rhytidoponera_acanthoponeroides_RHY14*	Australasian
Gnamptogenys_sp.06_GAB78	Neotropical	*Rhytidoponera_purpurea_RHY16*	Australasian
Gnamptogenys_reinchenspergeri_GAB79	Neotropical	*Rhytidoponera_croesus_RHY17*	Australasian
Gnamptogenys_ericae_GAB82	Neotropical	*Rhytidoponera_haeckeli_RHY18*	Australasian
Gnamptogenys_continua_GAB93	Neotropical	*Rhytidoponera_sp.F_RHY20*	Australasian
Gnamptogenys_annulata_GAB94	Neotropical	*Rhytidoponera_reflexa_RHY21*	Australasian

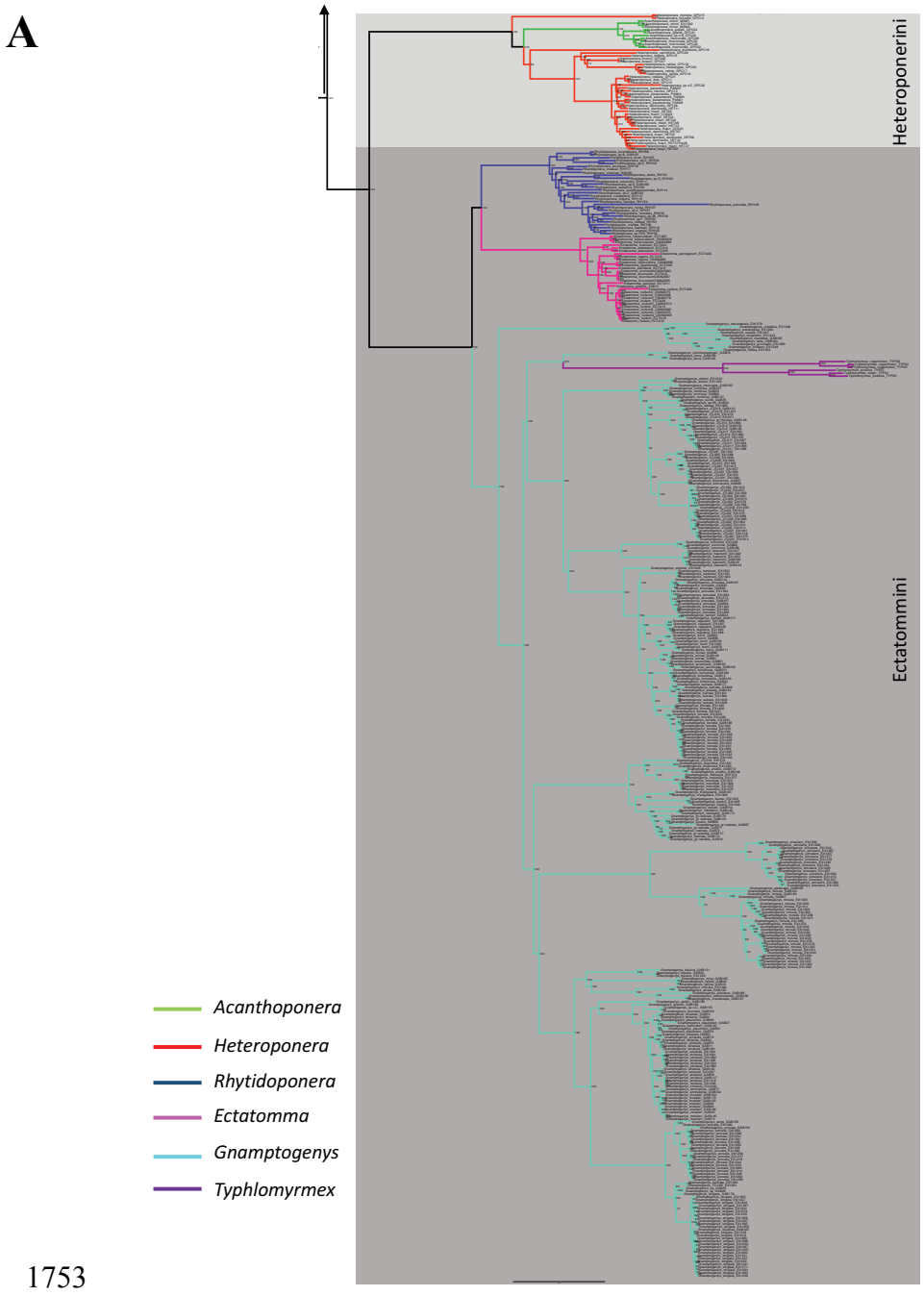
1745
1746

Species_Extraction code	Region
Rhytidoponera_laticeps_RHY24	Australasian
Rhytidoponera_eremita_RHY26	Australasian
Rhytidoponera_incisa_RHY27	Australasian
Rhytidoponera_punctata_RHY28	Australasian
Rhytidoponera_sp.FER_RHY30	Australasian
Rhytidoponera_sp.26_RHY33	Australasian
Rhytidoponera_cerastes_RHY35	Australasian
Rhytidoponera_cristata_RHY36	Australasian
Rhytidoponera_sp.X_RHY37	Australasian
Typhlomyrmex_major_TYP01	Neotropical
Typhlomyrmex_rogenhoferi_TYP04	Neotropical
Typhlomyrmex_rogenhoferi_TYP06	Neotropical
Typhlomyrmex_pusillus_TYP07	Neotropical

1747

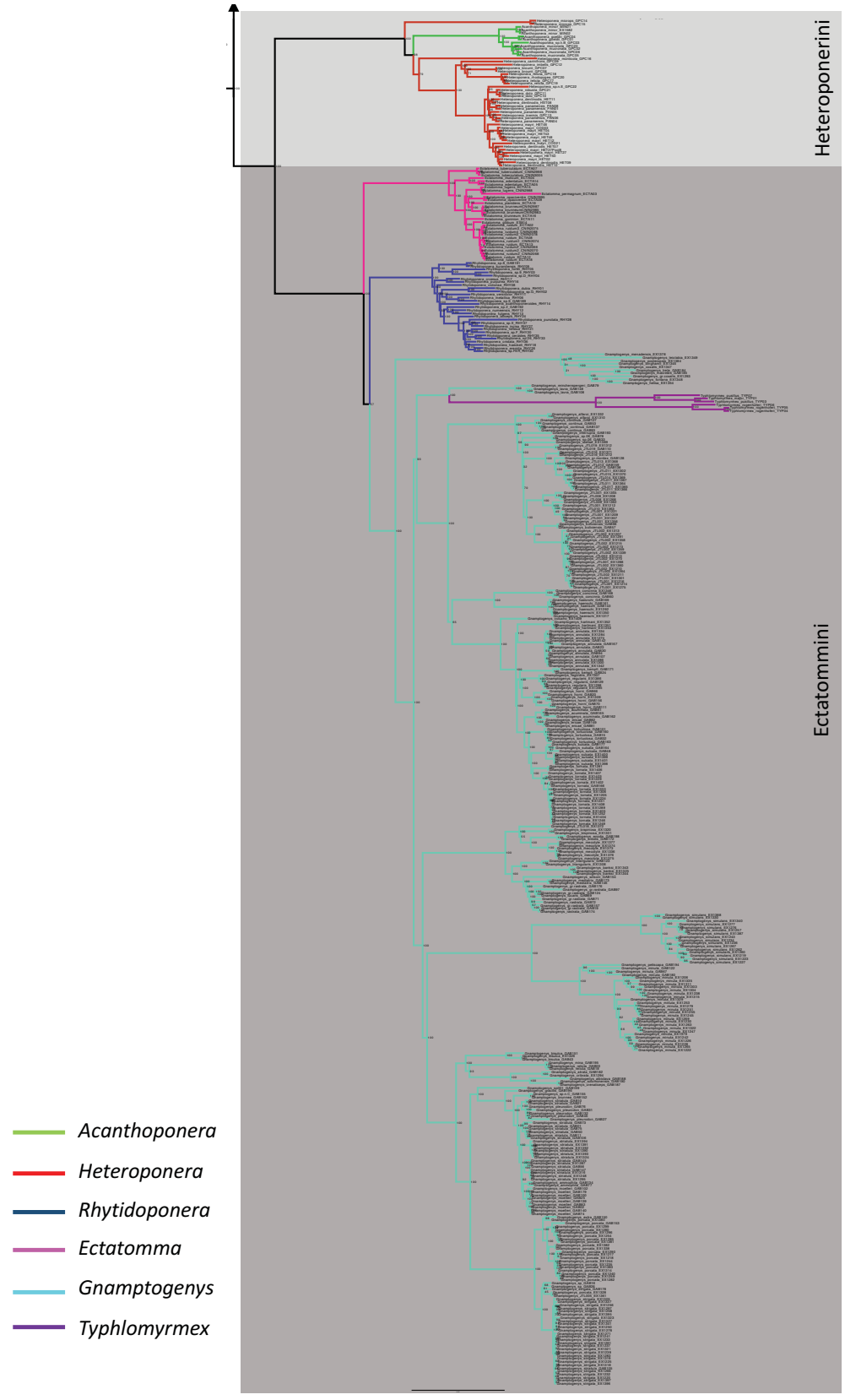
1748 **Anexo III**

1749 **Supplementary material S3.** Phylogenetic trees from analyses not illustrated in the main text. A)
1750 RAxML bootstrap tree for 90p-matrix data set; B) RAxML bootstrap tree for 95p-matrix; C)
1751 Species tree with local posterior probability values for the 1603 UCE loci in the 90p-pruned data
1752 set.



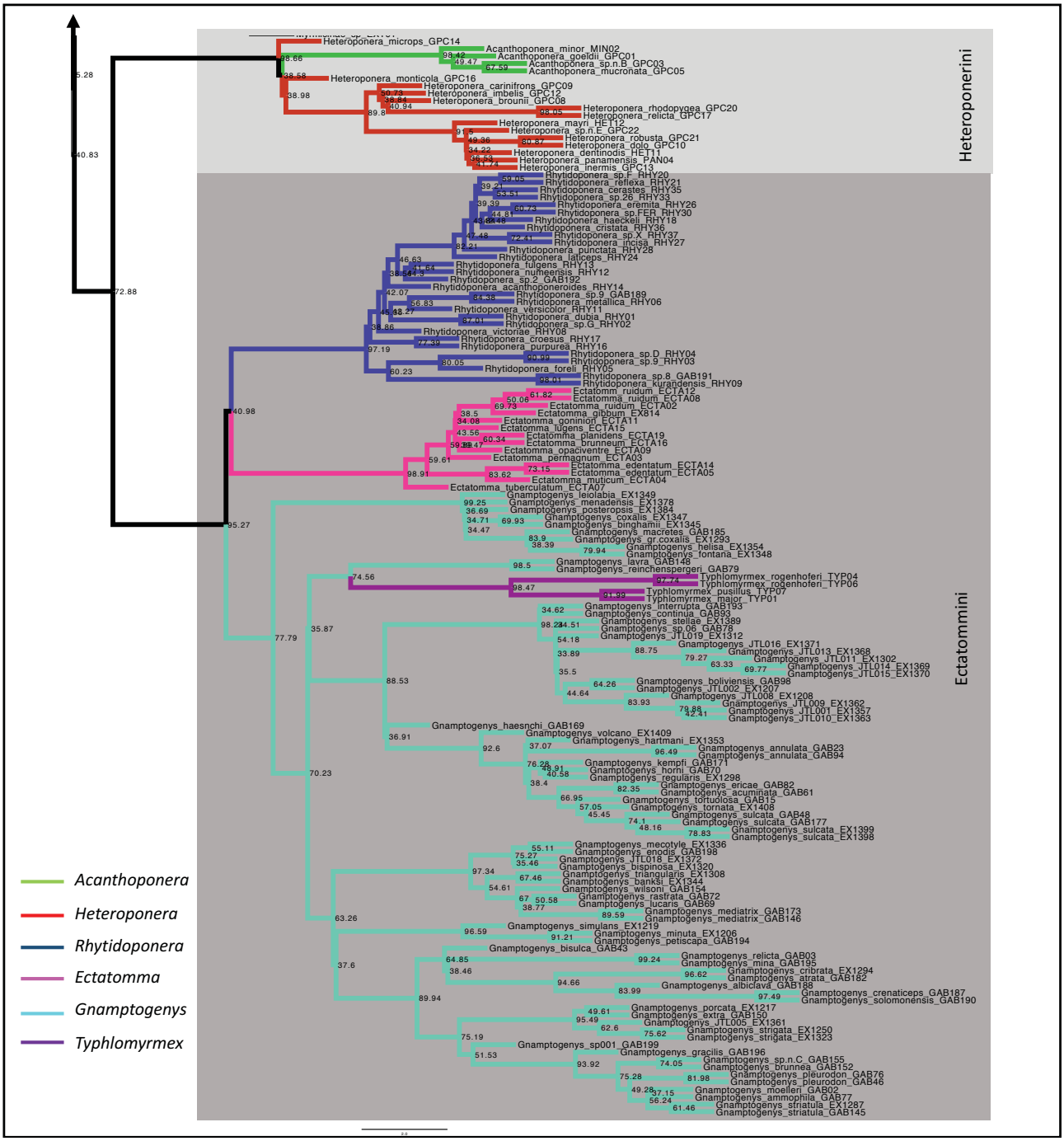
1753

B


Heteroponerini
Ectatommini
Acanthoponera
Heteroponera
Rhytidoponera
Ectatomma
Gnamptogenys
Typhlomyrmex

1754

C


Heteroponerini
Ectatommini
Acanthoponera
Heteroponera
Rhytidoponera
Ectatomma
Gnamptogenys
Typhlomyrmex