Evolution of the aquatic beetle family Noteridae (Coleoptera: Adephaga): integrating methods to examine patterns of diversification across temporal scales.

By

Stephen Baca

M.A. University of Kansas, 2015

B.Sc., University of New Mexico, 2012

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chair: Andrew E.Z. Short

_____

Bruce Lieberman

_____

Paulyn Cartwright

_____

Rob Moyle

_____

Paul Seldon

Date Defended: 28th June 2021

The dissertation committee for Stephen M. Baca certifies that this is the
approved version of the following dissertation:

Evolution of the aquatic beetle family Noteridae (Coleoptera: Adephaga): integrating methods to examine patterns of diversification across temporal scales.

_____

Chair: Andrew E.Z. Short

Date Approved: 7th October 2021

Abstract

Parsing the evolutionary patterns and processes that underly beetle diversity has been the subject of countless investigations. With an estimated crown age of 300 million years and millions of species, beetle evolution has unfolded over broad temporal and spatial scales. My dissertation focuses on the evolution of the aquatic beetle family Noteridae (Coleoptera: Adephaga). By integrating a wide variety of methods, I investigate patterns of noterid evolution across hundreds of millions of years of evolutionary time, from major lineages over 250 million years old, to the level of species and populations, potentially less than 1 million years old.

Adephaga has been the subject of many phylogenetic investigations using a variety of data types, without strong consensus on the relationships of its aquatic members. We conducted the first phylogenomic study of Adephaga with ultraconserved elements (UCEs). UCEs are captured by targeted, probe-based, enrichment and subsequent next-gen sequencing, and are known for effective phylogenetic reconstructions at both deep and shallow time scales. Capturing over 300 loci, we showed that aquatic adephagans, or 'Hydradephaga', are not monophyletic and further proved the efficacy of UCEs for coleopteran phylogenomics. As Adephaga is the parent suborder of Noteridae, this chapter is provided a contextual and methodological basis for my dissertation research.

Within Noteridae, the subfamily of minute beetles, Notomicrinae, presents a partial Gondwanan distribution, occupying Oceania, Indomalaya and the New World. My dissertation research shows that several species in the largest genus of the group, *Notomicrus,* are yet undescribed, and the subfamily displays a wide range of habitat preference among species with varying degrees of specificity. To investigate these patterns, we constructed a Sanger-based

dataset of five markers and robust sampling of Notomicrinae, with a strong emphasis on New World species of *Notomicrus*. We used this to reconstruct the phylogeny of Notomicrinae and estimate divergence times for a biogeographic reconstruction. We recovered a monophyletic *Notomicrus*, divided into five species groups of Notomicrus, and affirmed the existence of several undescribed species. The crown age of *Notomicrus* is estimated to be ca. 72 million years (ma), with Old and New World clades reciprocally monophyletic and separating in a time-span consistent with Gondwanan vicariance. The results of the phylogenetic reconstruction are then used to guide the taxonomic circumscription of the five species groups of *Notomicrus.* Provided is a key and diagnosis to each species group and *insertae sedis* species. We additionally performed the first of four reviews of the four New World groups, and review the *josiahi* group, complete with a description of a new species, *N. interstinctus* Baca and Short, 2021.

The phylogenetic reconstruction and taxonomic research of Notomicrinae showed that there exists a large species complex within *Notomicrus*, namely, the *traili* complex. Individuals of this complex can be found throughout South American and north into Mexico and the West Indies, with often overlapping distributions among phylogenetically distinct groups. To examine the evolution of this group and guide taxonomy, we designed a tailored noterid UCE probe set. We captured data from a robust population level sampling of the traili complex (45 individuals) and generated UCE data matrices at different levels of completeness and trimming regimes. With up to UCE 1,252 loci, we reconstructed the phylogeny of the complex. We recovered two distinct species *N. gracilipes* Sharp, 1882 and a tentative new species as successive sisters to the greater complex. In most analyses the complex split into four main clades with extensive and often overlapping ranges throughout south America. We discuss the taxonomic and

biogeographic implications of this reconstruction and establish a foundation for future research

in the *traili complex.*

Acknowledgments

Foremost, I would like to thank my advisor, Dr. Andrew Short. Andrew has been endless source of support and wisdom over the years and his guidance was fundamental in the development of my research aspirations. From conception to fruition, Andrew made my investigations possible. He has shared in my concerns, found solutions, and always encouraged me in my pursuits. I'm sure I owe him a special thanks for his patience also. I feel that Andrew has always held my best interests at heart and has had my back in all aspects of my studies. It has been wonderful being a member of Andrew's lab and I cannot thank him enough. I very much look forward to our future as colleagues and friends. Thank you for everything, Andrew.

I would also like to extend a big thanks to the members of my committee: Bruce Lieberman, Paulyn Cartwright, Rob Moyle and Paul Selden. I very much appreciate the time and support you all have offered during my studies. Your help and advice were indispensable and your perspectives greatly bolstered my understanding and broadened my horizons. You truly showcased the best of the collaborative and welcoming aspect of our research community here at KU. I wish you all the best in your pursuits and thank you wholeheartedly for your help in mine.

To the members of the Short lab, past and present, Crystal, Grey, Manu, Marianna, Taro, Jennifer, Rachel Neff, Rachel Glynn, and all others, thank you for your help and support.

I must extend a huge amount of gratitude to my long-time mentor Grey Gustafson. At the University of New Mexico, Grey took me under his wing, encouraged my interest, and helped me get into grad school. Since then, Grey has become my academic hero and one of my best friends. I can say without a doubt that wouldn't be here if it wasn't for him. He continues in his mentorship role, sharing his own success and knowledge along the way and leaving an ever-present influence on my research (and music tastes). I am both proud and lucky to consider

myself Grey's friend and I don't think I could ever repay him for all his help. I don't doubt that this marks the first of many wonderful decades of friendship and collaboration, to which I very much look forward. Thank you, Grey. Skål, brother!

To Emmanuel 'Manu' Toussaint, I also extend a huge thanks. As a post-doc in the Short lab, Manu stoked my interest in working with molecular data. I owe the foundation of my knowledge of molecular phylogenetics to him, he lit the path that scaffolds my dissertation. Beyond him being an inspiring influence, I am lucky to count him among my closest friends. Despite an ocean of separation, I know there are many good years ahead. See you soon, brother.

Huge thanks to Alana Alexander for all the help and guidance with my leap into phylogenomics and UCEs. There is no doubt that I would have been lost without your help!

Special thanks also to Kelly Miller, for giving me a shot at research in his lab. He brought to fruition a life-long dream and set my trajectory in noterid research. Thanks, man!

I must also acknowledge my co-authors and colleagues; especially Mario Toledo, Bruno A.C Guimarães and Michael Manuel, I am happy to be part of our noterid community!

The friends I made here played a major part of this fulfilling experience. Many have come and gone in my tenure, but I sincerely treasure the time we had together. Thank you for all your support, and for being there for me as colleagues and friends. To Fernando, Kaylee, Emily, Kaila, Lucas (Klicka), Rene, Lucas (Decicco) and Matthew, thank you especially. To the Ento-crew, Laura, Marianna, Jennifer, Mabel, Manu, Grey and Baruch, thank you for being my family away from home. I miss you all very much and I can't wait to see you again. And special thanks to my roommates: Jack, Fernando, Jacob, Joe, Luke, Luke, Luke and Lucas. I miss you and I threw out the stuff you left at the house (:-*).

To my partner, Taryn, thank you for your love, support, and patience during all this. I could not have wished for a better partner and I can't wait for our next adventure, near or far. I love you very much.

The community in KU-EEB and the KUNHM, you are some of the finest people I have ever met. Special thanks to Mark Holder, Rich Glor, and Paul Hime for always making time to chat about my projects. Huge thanks also to the BIMOL group, you were a constant source of inspiration and help. And thanks also to the KU-SACNAS chapter, for helping build a small but welcoming community. Thanks especially to Lynn Villafuerte for always being there for us. And the biggest thanks to Aagje Ashe, Dorothy Johanning, Ryan Zeigler, Lori Schlenkler, Jamie Keeler and Teri Chambers for keeping me in line.

And my family. To my mom and dad, thank you for all your love and support throughout these years. And thank you for nurturing my curiosity and wonder for the natural world and for allowing me to be who I am. I could not have asked for better parents and I am very proud to be your son. I love you very much. To my sisters and brother, Amanda, Noelle and Guy, thank you for all of the wonderful years and for always being a part of home. Finally, to my extended family, my aunties, uncles, all my wonderful cousins, and especially my grandparents, Grandma Hazel and Grandpa Archie, Nana and Pappy, thank you for your love, support, and kindness. I love you all very much.

Thank you also to the KU-NHM molecular lab, and the KU-CRC for providing facilities and resources for conducting my work.

Table of Contents

Introduction

Aquatic beetles

With ca. one million described species, insects comprise over half of all known organisms on the planet, and current estimates place the total diversity at ca. 7 million (Stork, 2018). Within insects, order Coleoptera includes nearly 500,000 described species, with this estimated to represent only a third of total beetle diversity (Stork, 2018). Originating 250 to over 300 million years ago (Toussaint et al., 2017; Zhang et al., 2018), beetles have evolved to occupy countless ecosystems and niches therein. While mostly terrestrial, beetles have independently invaded aquatic ecosystems at least eight times over the course of their evolutionary history (Bilton et al., 2019), with some fully aquatic lineages estimated to have originated as early as the middle Permian (e.g. suborder Myxophaga, 270ma; Toussaint et al., 2017). Other groups, such as the Hydrophilidae, have seen repeated transitions between aquatic and terrestrial habitats. Currently, there are 13,000 aquatic beetle species described among over 30 families and three of the four suborders (Short et al., 2018).

Aquatic beetles have long held a cult following of dedicated professionals and amateur workers and hobbyists. Though often exhibiting partiality, collectors are usually familiar with the greater diversity of aquatic beetles as many are often collected out of the same environments. Species of all three suborders and several families therein often occur simultaneously in habitats such as ponds or marshes, or even in some of the more specialized habitats such as hygropetric seeps. As a result, some aquatic lineages are among the best collected and documented in Coleoptera (e.g. Dytiscidae, estimated 85% of total species described, Short, 2018). This coupled with their wide breadth of ecological specialization, biotic interactions, and utility as

bioindicators (Bilton et al., 2019), make aquatic beetles excellent candidates for examining evolutionary patterns and processes at a wide range of scales.

Adephaga

The beetle suborder Adephaga includes nearly half of all known aquatic beetle diversity with ca. 6000 species distributed among several families (Short, 2018). Often collectively referred to as 'Hydradephaga', these aquatic adephagan families encompass the most charismatic and well-known aquatic beetle lineages in their ranks, including the diving beetles (family Dytiscidae) and whirligigs (family Gyrinidae). They also include some enigmatic families such as the cliff beetles (family Aspidytidae) or the comb-clawed cascade beetles (family Meruidae).

Ecologically, 'hydradephagans' are largely predatory or scavengers, preying on small invertebrates (though the larger species can capture small vertebrates) or scavenging. Their habitat preferences are wide ranging. While common in more familiar aquatic environments such as ponds and streams and those associated habitats, they also occupy a variety of more unusual habitats. These include hygropetric seeps or flows, brackish estuaries, or even underground aquifers. Gyrinidae, found primarily in ponds and streams, is of note here as the entire family is specialized to life on the water's surface, the sole extant lineage of any aquatic beetle to occupy this aquatic niche. Thus, while the suborder contains diversity that is in itself charismatic, hydradephagans are of further interest to entomologists for their many evolutionary paths and processes, some unique among aquatic Coleoptera.

Noteridae Thomson

Noteridae, sometimes referred to as the "burrowing water beetles", has shared some of the popularity of its cousins in the diving beetle family Dytiscidae. Noteridae is superficially

very similar to Dytiscidae and was historically treated as a dytiscid subfamily (Miller, 2009; Baca, Toussaint, et al., 2017) before Bertrand's (1928) investigation of adephagan larvae suggested it be classified as its own family. Following this, Noteridae slowly began to be recognized as its own family, which would later be confirmed by phylogenetic investigations finding that families such as Amphizoidae, Aspidytidae and Hygrobiidae were actually closer relatives to Dytiscidae than Noteridae (e.g. Balke et al., 2008; Mckenna et al., 2015; Toussaint et al., 2016). This appears to have led to the neglect of the family by other aquatic beetle workers focused on the diversity of Dytiscidae (Miller, 2009). It is only recently, in part following the systematic review of Miller (2009), that Noteridae has seen a revived interest especially in taxonomy (e.g. Miller, 2009, 2013; Baca et al., 2014; Manuel, 2015; Guimarães & Ferreira-Jr, 2019 and see Toledo, 2004; 2008; 2010)

Currently, Noteridae comprises 18 genera and nearly 300 species. As the diversity of Noteridae is documented, patterns relating to their ecologies and distributions beg questions of evolutionary processes. With respect to ecology, most noterids, are found in open, marshy habitats with abundant aquatic vegetation. However, several species or even entire genera have evolved to occupy more specialized habitats such as streams (*Liocanthydrus;* see Baca et al., 2014, and others), hygropetric seeps (*Tonerus*, Miller, 2009; species of *Notomicrus*, Baca et al., 2018; Guimarães & Ferreira-Jr, 2019), or subterranean aquifers and caves (*Phreatodytes*, Uéno, 1957; Kato et al., 2010, *Speonoterus,* Spangler, 1996). The sister to Noteridae, Meruidae also occupies hygropetric seeps. As such understanding the drivers of ecological evolution and how it relates to the diversity of Noteridae offers rich investigative fodder.

Noteridae also provides an excellent system to study historical biogeography at a global, continental scale. Though comprising less than 300 known species, the family is cosmopolitan,

distributed widely throughout the tropics and into some artic regions of the world and is estimated to have a crown age of ca. 110 ma (Toussaint et al., 2017). Thus, Noteridae spans many spatial and temporal scales, and in a group with a 'manageable' amount diversity (Coleoptera being one of the most diverse clades on earth). Several genera or species groups show biogeographic patterns that appear to correlate with historical geography at various scales (e.g. Gondwanan distributions; Pleistocene refugia) and may yet prove useful for testing biogeographic processes of broader interest.

It is with these sorts of evolutionary questions that I pursued my doctoral research. As a general framing of my dissertation, I sought to integrate various methods to (1) document the diversity of Noteridae and (2) understand the drivers of that diversity. As such my research includes a broad span of often linked investigations, ranging from morphology-based taxonomy (e.g. Baca et al., 2014; Baca & Toledo, 2015; Baca & Short, 2018, 2021; Chapter 3) , phylogenetics and phylogenetic methods testing (Baca et al., 2017b, Fig. X), divergence time estimation and biogeographic reconstruction (Baca & Short, 2020; Chapter 2) and finally UCE-based phylogenomics, both deep time (Baca, et al., 2017a; Gustafson et al., 2020; Baca et al, in review) and shallow-scale (Chapter 4). The following overview of my dissertation chapters summarizes the contexts, questions, and methods of my chapters, as well as any updates since their publications. These are presented in order of evolutionary scale, from investigations at deep evolutionary time to shallow, population level relationships. As is almost always the case, these chapters and resulting publications are not the end of these investigative threads, only steppingstones or a landing from which further questions are pursued.

Overview of dissertation chapters

Chapter 1: Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'.

The first chapter of my dissertation is contextual with regards to the evolution of Noteridae. Published in 2017 (Baca., et al., 2017b), it investigates the phylogeny of Noteridae's parent suborder, Adephaga. Comprising over 50k species , Adephaga is the second largest suborder of Coleoptera. It contains several aquatic lineages, collectively referred to as 'Hydradephaga', as well as the terrestrial 'Geadephaga', which houses nearly 90% of known adephagan diversity among only a few families. As detailed in Chapter 1, Adephaga has been subject to many studies and which has thereby resulted in many conflicting phylogenetic hypotheses (and held opinions thereof). Of particular interest has been the monophyly of 'Hydradephaga'. This was debated as there is morphological evidence that suggests the paraphyly of these aquatic lineages with respect to Geadephaga. This would imply multiple historic invasions of aquatic habitats, or a single, nested, invasion of land. Chapter 1 addresses this conflict and uses UCEs to reconstruct the phylogeny of major Adephagan lineages.

We recovered a paraphyletic 'Hydradephaga', consistent with the most robust investigations using morphology (but few of the molecular-based investigations). As the first published *in vitro* phylogenomic investigation within Coleoptera, we additionally proved the efficacy of UCEs for investigations of Coleoptera. The study was not without its limitations. First, we used the general Coleoptera 1.1Kv1 probe set (Faircloth, 2017), designed using genomes from a separate beetle suborder (Polyphaga) with which our resulting matrix was much smaller than other UCE studies at the time (e.g. in Hymenoptera, Blaimer et al., 2016; Branstetter et al., 2016). Second, our taxon sampling was limited, with only 20 total samples, including outgroups. Nonetheless, our recovered relationships, especially with respect to a

paraphyletic 'Hydradepahga', have since been corroborated by independent phylogenomic investigations (McKenna et al., 2019; Vasilikopoulos et al., 2019; see also Cai et al., 2020) and subsequent investigations of morphology (Beutel et al., 2020). Additionally, I have taken part in follow-up UCE investigations of Adephaga. The first, Gustafson et al., (2020), used a probe set tailored to Adephaga (Adephaga 2.9Kv1 probe set) to capture UCE loci from a greatly expanded taxon sampling (55 samples total), extending to the subfamily and tribe level. This effort resulted in an data alignment nearly three times the size of that used in Chapter 1 and corroborated the relationships recovered therein. The second, Baca., et al (in review) further expanded taxon sampling by harvesting UCEs *in silico* from a combination of whole genomic, transcriptomic, and UCE-enriched reads with the tailored Adephaga 2.9Kv1 probe set. This dataset was used to (1) address specific relationships in the aquatic superfamily Dytiscoidea (Noteridae's parent clade); (2) conduct the first divergence time estimate with a focus on Adephaga (using any data type); and (3) further support the exonic nature of UCEs in Coleoptera with a very high success of UCE extraction from RNAseq reads. In this vein, Chapter 1 is already somewhat outdated, however it served as a foundational study to coleopteran phylogenomics and was fundamental in my dissertation research.

Publication: Baca, S. M., Alexander, A., Gustafson, G. T. & Short A. E. Z. (2017). Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'. *Systematic Entomology*, 42, 786–795.

Chapter 2: Molecular Phylogeny of the Notomicrine Water Beetles (Coleoptera:

Noteridae) Reveals Signatures of Gondwanan Vicariance and Ecological Plasticity

Published in 2020 (Baca & Short, 2020), Chapter 2 focuses on the systematics and biogeography

of the minute noterid subfamily Notomicrinae. These very small beetles (1.2 – 1.8 mm) show a

significant ecological variation across its three genera, occupying a variety of habitats, spanning

ponds, swamps, streams, hygropetric seeps and terrestrial leaf litter. Two of the three genera are

eyeless, inhabiting subterranean aquifers and caves in Japan (*Phreatodytes*) and Indonesia

(*Speonoterus*), respectively. The genus *Notomicrus* contains most of the diversity of this group

(18 of the 27 notomicrine species) and can be found in Indomalaya, Oceania, and the New

World, with most diversity known from South America. Using a Sanger dataset of five markers,

and species-level sampling, the chapter (1) investigated the biogeographic history of the partial

Gondwanan distribution of the group; (2) examined the patterns of evolution regarding the wide

variety of habitat preference; and (3) established a phylogenetic foundation for a needed

taxonomic revision of New World *Notomicrus*. With phylogenetic analyses, divergence time

estimates and ancestral range reconstructions, we found that Old and New World members of

Notomicrus are reciprocally monophyletic, with a timing and pattern of evolution consistent with

Gondwanan vicariance. Further, we show that that habitat preference is relatively plastic across

the phylogeny, suggesting the potential nesting of the eyeless genus *Speonoterus* within

Notomicrus. The phylogenetic reconstructions recovered five major species groups within

Notomicrus and showed there is indeed several undescribed species as supported by morphology.

These latter findings would serve as the basis for the taxonomic work of Chapter 3. Additionally,

we included sampling of a suspected species complex, the *N. traili* complex, which would serve

as the first step in the phylogenomic investigation of Chapter 4.

Publication: Baca, S. M., & Short, A. E. Z. (2020). Molecular Phylogeny of the Notomicrine Water Beetles (Coleoptera: Noteridae) Reveals Signatures of Gondwanan Vicariance and Ecological Plasticity. *Insect Systematics and Diversity*, *4*(4). https://doi.org/10.1093/isd/ixaa015

Chapter 3: Review of the New World *Notomicrus* Sharp (Coleoptera: Noteridae) I: Circumscription of species groups and review of the *josiahi* group with description of a new species from Brazil

Recently published (Baca and Short, 2021), Chapter 3 uses the phylogenetic estimate of Chapter 2 to guide the delimitation of species groups of *Notomicrus* and review the first of these, the *josiahi* group, including the description of a new species. In this chapter we provide a diagnostic key to the five species groups and two *insertae sedis* species. A diagnosis of each species group is presented, accompanied by a list of each group's species, notes on identification and taxonomic comments. Finally, the *josiahi* group is reviewed. Composed of two species, the review includes a redescription of *N. josiahi* Miller, 2013 and the description of a new species, *N. interstinctus* Baca and Short, 2021. Included are habitus images and illustrations of diagnostic characters for the species groups and each of the *josiahi* species. This chapter is the first of four reviews of New World *Notomicrus*. While originally intended to be a single cohesive revision, my research and our lab's field work uncovered ca. 30 tentative species, vs. the 14 described as of 2015. These not including additional undescribed species that might result from investigation of the *traili* complex (Chapter 4). As such, circumscribing the species groups establishes a modular base to continue taxonomic work. My research for this chapter also yielded the

description of the first known hygropetric seep-dwelling species within Notomicrinae: *Notomicrus petrareptans* Baca and Short, 2018.

Publication: Baca, S. M., & Short, A. E. Z. (2021). Review of the New World Notomicrus Sharp (Coleoptera, Noteridae) I: Circumscription of species groups and review of the josiahi group with description of a new species from Brazil. *ZooKeys*, *1025*, 177–201.

Chapter 4: Shallow scale phylogenomics with Ultraconserved elements parse relationships and inform taxonomy in the *Notomicrus traili* species complex.

Chapter 4 remains the only yet unpublished chapter of my dissertation. As shown in the phylogenetic reconstructions of Chapter 2 (Baca and Short, 2020), the *Notomicrus traili* group includes a complex distributed throughout South America and extending into Mexico and the Antilles. Multiple phylogenetically distinct clades share overlapping distributions and individuals attributable to described species (e.g. cf. *N. traili* Sharp, 1882 or cf. *N. gracilipes* Sharp, 1882) were recovered as polyphyletic (Chapter 2). As such, this group requires investigation at the species to population level, making it a prime candidate for the application of phylogenomics to help guide taxonomy. We chose UCEs for this (versus RAD-seq, e.g.) as the divergence time estimate of Chapter 2 (Baca and Short, 2020) recovered a crown age of 13mya for the *traili* group, and thus we sought a reduced genomic representation method (i.e. non-whole genome) that would span evolutionary time scales and provide sufficient data for the adequate application to a variety of methods, such as coalescent-based analyses, to conduct our investigation.

Gustafson et al, (2020) showed the benefits of a tailored UCE probe set with respect to data capture. Here we used short-read sequencing of four noterid genomes to design a noterid-specific UCE probe set (Noteridae 3.4Kv1). With population-level taxon sampling of 45 *traili* group taxa from across its range we reconstructed the phylogeny of the *traili* group with multiple data trimming and data completeness criteria. Our phylogenetic estimates show that there are indeed several phylogenetically distinct clades within the traili group. For taxonomic implications, no reconstruction indicated the need to synonymize any of the four species in this group. Instead, it appears that additional undescribed species are present, and will require careful investigation of morphology. The evolutionary patterns recovered are consistent with repeated cycles of diversification and range expansion. While elucidating at present, further work is required to adequately inform the taxonomy of the group. With data phasing, haplotype and genetic clustering methods we intend to examine the underlying genetic structure of the group to search for evidence of gene-flow among populations and potentially apply these data to a species-delimitation framework. While the chapter is taxonomy-focused the results show intriguing spatial patterns. This chapter will serve as the foundation for further biogeographic analyses, allowing us to test the influence of the dynamic historical geography on the evolution of the *traili* group.

References

Baca, S. M., Alexander, A., Gustafson, G. T., & Short, A. E. Z. (2017). Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga.' *Systematic Entomology*, *42*(4), 786–795. https://doi.org/10.1111/syen.12244

Baca, S. M., Gustafson, G. T., Alexander, A. M., Gough, H. M., Toussaint., E. A. F. (In Review). Integrative phylogenomics reveals a Permian origin of Adephaga beetles. *Systematic Entomology.*

Baca, S. M., Gustafson, G. T., Toledo, M., & Miller, K. B. (2014). Revision of the Neotropical burrowing water beetle genus Liocanthydrus Guignot (Coleoptera: Noteridae: Noterinae: Noterini) with the description of two new species. *Zootaxa*, *3793*, 231–246. https://doi.org/10.11646/zootaxa.3793.2.3

Baca, S. M., & Short, A. E. Z. (2018). Notomicrus petrareptans sp. n., a new seep-dwelling species of Noteridae from Suriname (Coleoptera: Adephaga). *Zootaxa*, *4388*(2), 182–190. https://doi.org/10.11646/zootaxa.4388.2.2

Baca, S. M., & Short, A. E. Z. (2020). Molecular Phylogeny of the Notomicrine Water Beetles (Coleoptera: Noteridae) Reveals Signatures of Gondwanan Vicariance and Ecological Plasticity. *Insect Systematics and Diversity*, *4*(4). https://doi.org/10.1093/isd/ixaa015

Baca, S. M., & Short, A. E. Z. (2021). Review of the New World Notomicrus Sharp (Coleoptera, Noteridae) I: Circumscription of species groups and review of the josiahi group with description of a new species from Brazil. *ZooKeys*, *1025*, 177–201. https://doi.org/10.3897/zookeys.1025.60442

Baca, S. M., & Toledo, M. (2015). Canthysellus Baca and Toledo (Coleoptera: Noteridae: Noterini), a New Genus of Burrowing Water Beetle from South America. *The Coleopterists Bulletin*, *69*(3), 477–488. https://doi.org/10.1649/0010-065X-69.3.477

Baca, S. M., Toussaint, E. F. A., Miller, K. B., & Short, A. E. Z. (2017). Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on data partitioning strategies. *Molecular Phylogenetics and Evolution*, *107*, 282–292. https://doi.org/10.1016/j.ympev.2016.10.016

Balke, M., Ribera, I., Beutel, R., Viloria, A., Garcia, M., & Vogler, A. P. (2008). Systematic placement of the recently discovered beetle family Meruidae (Coleoptera: Dytiscoidea) based on molecular data. *Zoologica Scripta*, *37*(6), 647–650. https://doi.org/10.1111/j.1463-6409.2008.00345.x

Beutel, R. G., Ribera, I., Fikáček, M., Vasilikopoulos, A., Misof, B., & Balke, M. (2020). The morphological evolution of the Adephaga (Coleoptera). *Systematic Entomology*, *45*(2), 378–395. https://doi.org/10.1111/syen.12403

Bilton, D. T., Ribera, I., & Short, A. E. Z. (2019). Water Beetles as Models in Ecology and Evolution. *Annual Review of Entomology*, *64*(1), 359–377. https://doi.org/10.1146/annurev-ento-011118-111829

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*(8), e0161531. https://doi.org/10.1371/journal.pone.0161531

Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., Gates, M. W., Kula, R. R., & Brady, S. G. (2016). Phylogenomic Analysis of Ants, Bees and Stinging Wasps: Improved Taxon Sampling Enhances Understanding of Hymenopteran Evolution. *BioRxiv*, 068957. https://doi.org/10.1101/068957

Cai, C., Tihelka, E., Pisani, D., & Donoghue, P. C. J. (2020). Data curation and modeling of compositional heterogeneity in insect phylogenomics: A case study of the phylogeny of

Dytiscoidea (Coleoptera: Adephaga). *Molecular Phylogenetics and Evolution*, *147*, 106782. https://doi.org/10.1016/j.ympev.2020.106782

Guimarães, B. a. C., & Ferreira-Jr, N. (2019). Two new species and new records of Notomicrus Sharp, 1882 (Coleoptera: Noteridae) from Brazil. *Zootaxa*, *4629*(2), 263–270. https://doi.org/10.11646/zootaxa.4629.2.8

Gustafson, G. T., Baca, S. M., Alexander, A. M., & Short, A. E. Z. (2020). Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Systematic Entomology*, *45*(3), 552–570. https://doi.org/10.1111/syen.12413

Kato, M., Kawakita, A., & Kato, T. (2010). Colonization to aquifers and adaptations to subterranean interstitial life by a water beetle clade (Noteridae) with description of a new Phreatodytes species. *Zoological Science*, *27*(9), 717–722. https://doi.org/10.2108/zsj.27.717

Manuel, M. (2015). The genus Notomicrus in Guadeloupe, with description of three new species (Coleoptera: Noteridae). *Zootaxa*, *4018*(4), 506–534. https://doi.org/10.11646/zootaxa.4018.4.2

McKenna, D. D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D. J., Donath, A., Escalona, H. E., Friedrich, F., Letsch, H., Liu, S., Maddison, D., Mayer, C., Misof, B., Murin, P. J., Niehuis, O., Peters, R. S., Podsiadlowski, L., Pohl, H., … Beutel, R. G. (2019). The evolution and genomic basis of beetle diversity. *Proceedings of the National Academy of Sciences*, *116*(49), 24729–24737. https://doi.org/10.1073/pnas.1909655116

Mckenna, D. D., Wild, A. L., Kanda, K., Bellamy, C. L., Beutel, R. G., Caterino, M. S., Farnum, C. W., Hawks, D. C., Ivie, M. A., Jameson, M. L., Leschen, R. a. B., Marvaldi, A. E., Mchugh, J. V., Newton, A. F., Robertson, J. A., Thayer, M. K., Whiting, M. F., Lawrence, J. F., Ślipiński, A., … Farrell, B. D. (2015). The beetle tree of life reveals that Coleoptera survived end-Permian

mass extinction to diversify during the Cretaceous terrestrial revolution. *Systematic Entomology*, *40*(4), 835–880. https://doi.org/10.1111/syen.12132

Miller, K. B. (2009). On the systematics of Noteridae (Coleoptera: Adephaga: Hydradephaga): Phylogeny, description of a new tribe, genus and species, and survey of female genital morphology. *Systematics and Biodiversity*, *7*(2), 191–214. https://doi.org/10.1017/S1477200008002946

Miller, K. B. (2013). Notomicrus josiahi , a new species of Noteridae (Coleoptera) from Venezuela. *Zootaxa*, *3609*(2), 243–247. https://doi.org/10.11646/zootaxa.3609.2.11

Short, A. E. Z. (2018). Systematics of aquatic beetles (Coleoptera): Current state and future directions. *Systematic Entomology*, *43*(1), 1–18. https://doi.org/10.1111/syen.12270

Spangler, P. J. (1996). *Four new stygobiontic beetles (Coleoptera: Dytiscidae; Noteridae; Elmidae)*. *10*, 20.

Stork, N. E. (2018). How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annual Review of Entomology*, *63*(1), 31–45. https://doi.org/10.1146/annurev-ento-020117-043348

Toussaint, E. F. A., Beutel, R. G., Morinière, J., Jia, F., Xu, S., Michat, M. C., Zhou, X., Bilton, D. T., Ribera, I., Hájek, J., & Balke, M. (2016). Molecular phylogeny of the highly disjunct cliff water beetles from South Africa and China (Coleoptera: Aspidytidae). *Zoological Journal of the Linnean Society*, *176*(3), 537–546. https://doi.org/10.1111/zoj.12332

Toussaint, E. F. A., Seidel, M., Arriaga-Varela, E., Hájek, J., Král, D., Sekerka, L., Short, A. E. Z., & Fikáček, M. (2017). The peril of dating beetles. *Systematic Entomology*, *42*(1), 1–10. https://doi.org/10.1111/syen.12198

Vasilikopoulos, A., Balke, M., Beutel, R. G., Donath, A., Podsiadlowski, L., Pflug, J. M., Waterhouse, R. M., Meusemann, K., Peters, R. S., Escalona, H. E., Mayer, C., Liu, S., Hendrich, L., Alarie, Y., Bilton, D. T., Jia, F., Zhou, X., Maddison, D. R., Niehuis, O., & Misof, B. (2019). Phylogenomics of the superfamily Dytiscoidea (Coleoptera: Adephaga) with an evaluation of phylogenetic conflict and systematic error. *Molecular Phylogenetics and Evolution*, *135*, 270–285. https://doi.org/10.1016/j.ympev.2019.02.022

Zhang, S.-Q., Che, L.-H., Li, Y., Dan Liang, Pang, H., Ślipiński, A., & Zhang, P. (2018). Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. *Nature Communications*, *9*(1), 205. https://doi.org/10.1038/s41467-017-02644-4

Chapter 1: Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'

Abstract

The beetle suborder Adephaga has been the subject of many phylogenetic reconstructions utilizing a variety of data sources and inference methods. However, no strong consensus has yet emerged on the relationships among major adephagan lineages. Ultraconserved elements (UCEs) have proved useful for inferring difficult or unresolved phylogenies at varying timescales in vertebrates, arachnids and Hymenoptera. Recently, a UCE bait set was developed for Coleoptera using polyphagan genomes and a member of the order Strepsiptera as an outgroup. Here, we examine the utility of UCEs for reconstructing the phylogeny of adephagan families, in the first in vitro application a UCE bait set in Coleoptera. Our final dataset included 305 UCE loci for 18 representatives of all adephagan families except Aspidytidae, and two polyphagan outgroups, with a total concatenated length of 83,547 bp. We inferred trees using maximum likelihood analyses of the concatenated UCE alignment and coalescent species tree methods (astral ii, ASTRID, svdquartets). Although the coalescent species tree methods had poor resolution and weak support, concatenated analyses produced well-resolved, highly supported trees. Hydradephaga was recovered as paraphyletic, with Gyrinidae sister to Geadephaga and all other adephagans. Haliplidae was recovered as sister to Dytiscoidea, with Hygrobiidae and Amphizoidae successive sisters to Dytiscidae. Finally, Noteridae was recovered as monophyletic and sister to Meruidae. Given the success of UCE data for resolving phylogenetic relationships within Adephaga, we suggest the potential for further resolution of relationships within Adephaga using UCEs with improved taxon sampling, and by developing Adephaga-specific probes.

Introduction

Advances in phylogenomic methods have allowed for large multilocus datasets to be assembled with increased efficiency, cost and quality (Faircloth et al., 2012; Meiklejohn et al., 2016). One such method for generating large multilocus datasets is the targeted capture and enrichment of ultraconserved elements (UCEs; (Crawford et al., 2012, 2015; Faircloth et al., 2012; McCormack et al., 2013; Smith et al., 2014; Gilbert et al., 2015; Moyle et al., 2016; Oswald et al., 2016; Alexander et al., 2017)Faircloth et al., 2012) – areas of the genome that are strongly conserved (i.e. invariant) among divergent taxa (Bejerano, 2004; Faircloth et al., 2012). Ultraconserved elements allow for effective alignment of the more variable (and phylogenetically informative) UCE-flanking regions across divergent taxa (Faircloth et al., 2012), with comparable or potentially decreased frequencies of homoplasy versus other types of marker (Meiklejohn et al., 2016). UCEs have resolved both deep and shallow relationships of various taxa, primarily vertebrates (Crawford et al., 2012, 2015; Meiklejohn et al., 2016; McCormack et al., 2013; Smith et al., 2014; Gilbert et al., 2015; Moyle et al., 2016; Oswald et al., 2016; Alexander et al., 2017). To date, UCE studies of arthropods have been limited to arachnids (Starrett et al., 2017) and Hymenoptera (Blaimer et al., 2015, 2016; Faircloth et al., 2015; Branstetter et al., 2016, 2017). Recently, Faircloth, (2017) designed a Coleoptera-specific UCE probe set consisting of 13,674 baits targeting 1,172 UCE loci. This was developed using seven genomes within the beetle suborder Polyphaga, and an additional genome from the order Strepsiptera as an outgroup (Faircloth, 2017). Here, we use this new UCE bait set to infer the phylogeny of Adephaga and demonstrate the potential of UCEs for phylogenetic studies of Coleoptera.

Comprising ~45,000 extant species (Slipinski et al., 2011), Adephaga is the second largest beetle suborder. Members occupy a broad range of habitats, often exhibiting high degrees

of specialization for their respective ecologies. The terrestrial adephagans (families Carabidae,

Trachypachidae), often referred to as 'Geadephaga', represent the majority of the diversity of

Adephaga with ~40 000 species (Slipinski et al., 2011). Although less species-rich, the aquatic

adephagan families, or 'Hydradephaga', have attracted interest in their aquatic adaptations; while

generally common in ponds, streams and pools, these beetles can also be found in an array of

specialized habitats, such as hygropetric seeps (Ribera et al., 2002b; Spangler & Steiner, 2005;

Miller, 2009), underground caves (Spangler, 1996) and subterranean aquifers (Young &

Longley, 1976; Larson & LaBonte, 1994; Watts et al., 2007). A few members of the diving

beetle family Dytiscidae have even shifted to terrestrial habitats (Toussaint, Hendrich, et al.,

2016).

The phylogeny of Adephaga has been investigated extensively utilizing various sources

of data and methods, and while there is strong support for the monophyly of the suborder (e.g.

Ribera et al., 2002a; Beutel et al., 2006; Hunt et al., 2007; Lawrence et al., 2011; Mckenna et al.,

2015) there remains no strong consensus on the relationships of its major lineages (Fig. 1.1). Of

particular interest is the monophyly of Hydradephaga with respect to the terrestrial families, and

its bearing on the ancestral condition of Adephaga (terrestrial versus aquatic). Cladistic analyses

of morphology have repeatedly suggested the unique surface-swimming aquatic family

Gyrinidae (whirligig beetles) to be sister to all adephagans, aquatic and terrestrial (e.g. Beutel &

Roughley, 1988; Beutel et al., 2006, 2013; Maddison et al., 2009; Dressler et al., 2011).

However, the first molecular-based investigations of Adephaga (Shull et al., 2001; Ribera et al.,

2002a) recovered both Hydradephaga and Geadephaga as reciprocally monophyletic sister

groups. This has also been corroborated by recent studies focused on the phylogeny of all

Coleoptera (Hunt et al., 2007; Lawrence et al., 2011; Mckenna et al., 2015).

The phylogenetic relationships of Dytiscoidea (families Aspidytidae, Hygrobiidae, Amphizoidae, Dytiscidae, Noteridae and Meruidae) have been intensely investigated using both morphological and molecular sequence data. Despite this, a consensus has yet to be reached (Balke et al., 2005; Balke et al., 2008; Beutel et al., 2006, 2013; Dressler et al., 2011; Toussaint, et al., 2016b). The placement of Meruidae and the monophyly of Noteridae have also been addressed by several studies (e.g. Beutel et al., 2006; Balke et al., 2008; Alarie et al., 2011; Dressler et al., 2011; Short et al., 2012). It has been suggested that Meruidae could be considered a member of Noteridae, but most studies, whether based on morphology or molecular sequence data, recover Meruidae as sister to a monophyletic Noteridae and member of Dytiscoidea. With no consensus on its relationships, and the proven utility of UCEs in resolving difficult phylogenies, Adephaga is an excellent candidate for testing the utility of UCEs for phylogenetic inference within Coleoptera.
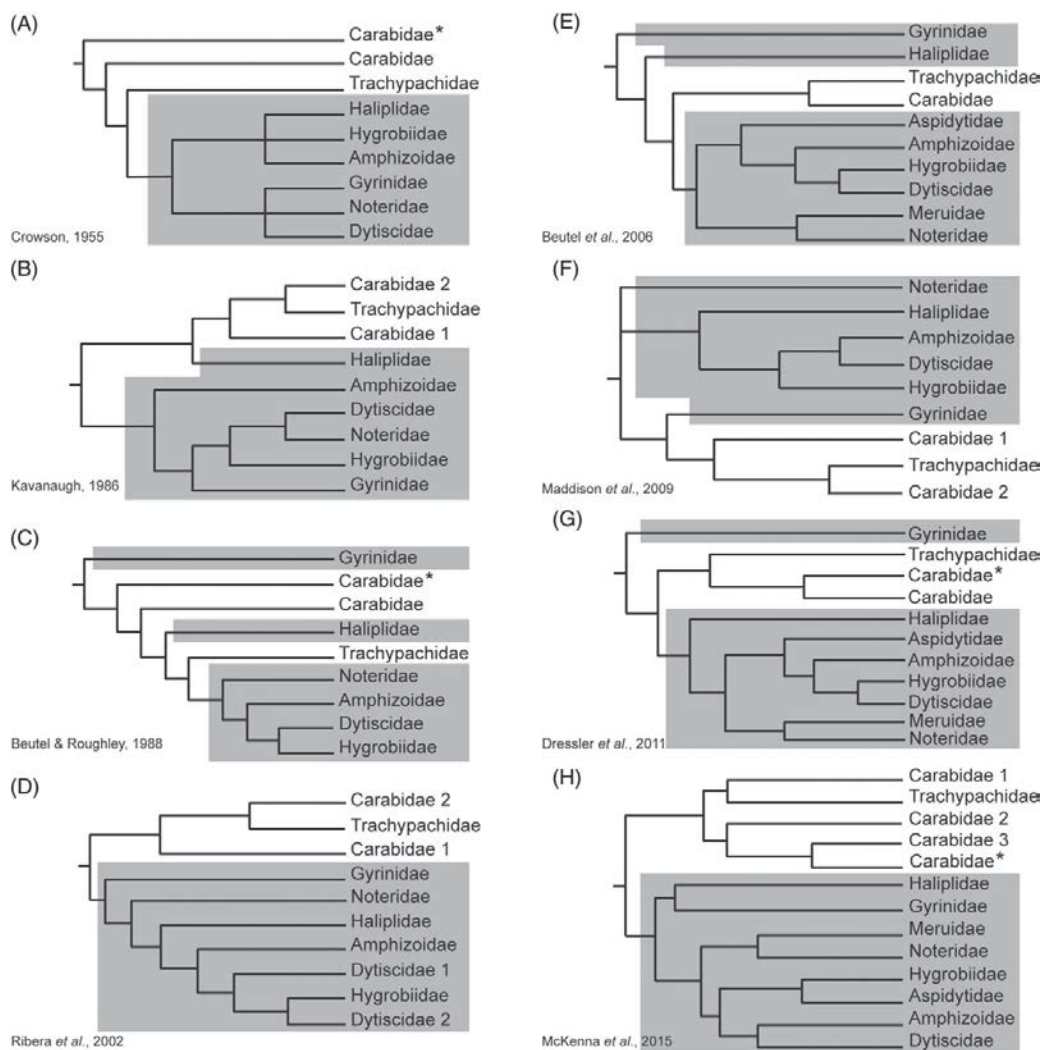
Figure 1.1. Previously proposed relationships of adephagan families. Grey boxes indicate members of Hydradephaga. (A) Precladistic relationships hypothesized by Crowson (1955) based on morphology. (B) Cladistic relationships hypothesized by Kavanaugh (1986) based on morphology. (C) Cladistic relationships hypothesized by Beutel & Roughley (1988) based on morphology. (D) Tree from Ribera et al. (2002a) based on parsimony analysis of 18S rRNA. (E) Tree from Beutel et al. (2006) based on parsimony analysis of morphology. (F) Tree from Maddison et al. (2009) based on Bayesian analysis of three nuclear genes. (G) Tree from Dressler et al. (2011) based on parsimony analysis of morphology. (H) Tree from McKenna et al. (2015) based on Bayesian analysis of eight nuclear genes. Carabids with asterisk (*) are Rhysodidae.

Materials and methods

Taxon sampling and DNA extraction

We chose representatives of all major lineages of Adephaga (Table 1.1), targeting each family of Hydradephaga with the exception of Aspidytidae, as specimens of this family were unavailable at the time. We selected two polyphagan individuals (Staphylinidae, Georissidae) as

outgroups, resulting in a total dataset of 20 specimens. Specimens were less than 10 years old and fixed and stored in >95% ethanol. DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Ger- many) following the manufacturer's protocols. Extractions were quantified using a Qubit (Life Technologies, Carlsbad, CA, U.S.A.) system and for each sample, 50, 100 or 500 ng of DNA was used as input for library preparation, with four exceptions: specimens of *Meru* (*n* = 2), *Hygrobia* (*n* = 1) and *Notomicrus* (*n* = 1). As the quantity of DNA was limited for these samples, all available DNA (ca. 10 – 25 ng) was used (Table 1.1). All specimen vouchers are deposited in the Snow Entomology Collection at the University of Kansas, except for *Hygrobia* Hgsp906, which is deposited in the Museum of Southwestern Biology, University of New Mexico.

Sequence capture and data processing

Library preparation and targeted enrichment of UCE loci were carried out with the following protocols: libraries were prepared using the Kapa HyperPlus Library Preparation Kit KR1145-v2.15 (Kapa Biosystems, Wilmington, MA, U.S.A.) following the manufacturer's instructions with five minor modifications. First, genomic DNA was enzymatically fragmented for 5 min and 30 s. Second, half-volume reactions were used. Third, universal iTru y-yoke adapters (Glenn et al., 2016) were ligated onto the genomic DNA. Fourth, following adapter ligation, one post-ligation clean-up was performed using 1:1 ratio of Sera-Mag Speedbeads (Thermo-Scientific, Waltham, MA, U.S.A.; prepared according to Glenn et al., 2016). Finally, a polymerase chain reaction (PCR) at 95°C for 45 s, followed by 14 cycles of 98°C for 30 s, 60°C for 30 s, 72°C for 30 s, then 72°C for a 5 min final extension was executed using iTru5 and iTru7 primers to produce dual-indexed Illumina TruSeqHT compatible libraries (Glenn et al., 2016). Libraries were verified on a 1.5% agarose gel. Two samples (SLE829b, SLE894) were

treated with an additional 18-cycle PCR, due to low concentration following the previous PCR.

Libraries were quantified with a Qubit 2.0 system (Life Technologies) and 42 ng from each library was added to create pools of 12. Library pools were then enriched using Coleoptera 5K version 1 baits (Faircloth, 2017) from MYcroar- ray (Ann Arbor, MI, U.S.A.) following their mybaits protocol v. 3.01 with a 24 h incubation. A PCR was performed on enriched libraries using a KAPA HiFi Hotstart PCR Kit at 98° C for 3 min, followed by 15 cycles of 98∘C for 20 s, 60° C for 30 s, 72° C for 1 min, then 72° C for a 5 min final extension, using P5 and P7 primers. Subsequent product was cleaned and size-selected for fragments <300 bp with speed- beads using a 1:1 ratio, quantified with Qubit and pooled with other unrelated samples (total = 96 dual-indexed samples) prior to sequencing on an Illumina HiSeq 3000 to produce paired-end reads of 151 bases. Samples were demultiplexed with Illumina software bcl2fastq.

Phylogenetic analyses

We assembled our UCE dataset as described at https://github .com/laninsky/UCE_processing_steps, primarily using the phyluce 1.5 software package for Python 2.7 (Faircloth, 2016). Adaptor contamination was trimmed using cutadapt 1.12 (Martin, 2011). Reads were then merged using pear (Zhang et al., 2014) and assembled into contigs with trinity v. 2.0.6 (Grabherr et al., 2011). To avoid including multiple isoforms of the same locus in the dataset, output contigs were filtered to only the longest isoforms. Contigs were matched to the Coleoptera v1 probeset (Faircloth, 2017) and contigs containing UCEs were aligned using mafft v. 7.130b (Katoh & Standley, 2013), through phyluce. Ambiguously aligned or divergent regions were removed with gblocks 0.91 (Talavera & Castresana, 2007) with decreased stringency in the b2 parameter (minimum number of sequences for a flank position). Settings for b1, b2, b3 and b4 parameters were 0.5, 0.5, 8 and 10, respectively. A 50% complete data

matrix was constructed (allows up to 50% missing taxa at each locus). To evaluate the impact of allowing missing taxa on total alignment length, we also constructed 60%, 70%, 80%, 90% and 100% complete matrices (Tables S1.1, S1.2). Phylogenetic analyses We conducted phylogenetic analyses using both maximum likelihood (ML) analyses of the concatenated dataset and species tree methods (coalescent-based methods) to accommodate for incomplete lineage sorting (ILS; Maddison, 1997; Edwards, 2009; Mirarab et al., 2014a; 2014b; Vachaspati & Warnow, 2015). For concatenated analyses, we first analysed the unpartitioned dataset using phyluce and raxml 8.2.4 (Stamatakis, 2014), under a GTRCAT model of evolution, with 100 bootstrap replicates to assess support. For analyses of the partitioned concatenated dataset, we used phyluce to generate phylip alignments of each locus, then cloudforest 2.0.1 (Crawford & Faircloth, 2014) to estimate models of evolution for each locus. Using phyluce and custom r scripts (Alexander, 2016), loci were concatenated into data partitions by their respective models of evolution as estimated by cloudforest. The partitioned dataset was then subjected to ML analyses using raxml 8.2.4 (Stamatakis, 2014), with 100 bootstrap replicates to assess support. A bootstrap ≥70% was considered as strong support (Hillis & Bull, 1993) Species tree inference was conducted using astral ii (S. Mirarab et al., 2014; Siavash Mirarab & Warnow, 2015) and astrid (Vachaspati & Warnow, 2015), following the UCE processing steps of Alexander (2016), as astral and astrid have been shown to be effective and efficient at analysing large datasets of unrooted gene trees (Vachaspati & Warnow, 2015). astral and astrid species trees were constructed from gene trees generated using raxml and cloudforest. Support was assessed by annotating the astral and astrid phylogenies with local posterior probabilities (PPs; Sayyari & Mirarab, 2016). A PP ≥0.95 was considered as strong support (Erixon et al., 2003). An additional coalescent species tree was generated using svdquartets (Chifman & Kubatko, 2014)

as implemented in paup* v. 4 (Swofford, 2003). Unlike the summary species-tree methods described above, svdquartets addresses ILS without the need to generate gene trees by estimating unrooted species trees for each quartet of species, for each single nucleotide polymorphism in the alignment. Support for the final species tree was assessed by n=500 bootstrap replicates.

Table 1.1. Taxon sampling and summary metrics for UCE data matrix.

Clade	Family	Taxon	Voucher ID	ng DNA	Reads	Contigs	UCEs	UCEs 50%	Mean (SD) coverage ×	% missing data
Polyphaga	Georissidae	*Georissus* sp.	SLE36	500	7,332,792	15,215	397	193	25.5(16.7)	48.32
	Staphylinidae	*Platydracus* sp.	SLE1012	500	1,619,816	20,068	475	209	22.6 (13.1)	40.06
'Geadephaga'	Carabidae	*Cicindela* sp.	SLE993	500	1,194,914	19,036	309	197	20.8 (16.3)	39.80
	Trachypachidae	*Trachypachus inermis*	SLE991	500	1,025,894	14,740	361	216	23.3 (14.7)	31.54
	Amphizoidae	*Amphizoa insolens*	SLE990	500	1,375,322	16,838	353	208	31.1 (24.7)	34.80
	Dytiscidae	*Platynectes* sp.	SLE956	500	1,361,936	22,145	339	211	25.6 (18.2)	34.93
	Dytiscidae	*Aglymbus* sp.	SLE935	500	1,406,940	18,939	281	190	11.9 (9.0)	44.48
	Dytiscidae	*Rhantus calidus*	SLE1006	100	1,358,542	12,985	383	229	39.5 (33.1)	25.43
	Gyrinidae	*Spanglerogyrus albiventris*	SLE992	100	1,587,440	23,923	368	216	36.0 (27.0)	31.84
	Gyrinidae	*Gyretes* sp.	SLE996	500	1,039,130	12,809	352	199	33.2 (24.4)	37.60
'Hydradephaga'	Haliplidae	*Peltodytes* sp	SLE994	500	1,188,484	17,445	363	207	48.6 (37.1)	34.10
	Haliplidae	*Haliplus* sp	SLE997	500	2,681,424	46,801	426	233	77.7 (66.4)	25.50
	Hygrobiidae	*Hygrobia* sp.	Hgsp906	10	1,875,968	35,745	402	236	15.7 (9.9)	26.63
	Meruidae	*Meru phyllisae*	SLE901	21.5	3,039,232	16,568	395	213	89.2 (16.7)	31.01
	Meruidae	*Meru phyllisae*	SLE998	22.5	2,890,644	16,914	373	161	78.3 (63.2)	46.85
	Noteridae	*Hydrocanthus dibilis*	SLE829	100	7,069,210	20,836	406	234	23.0 (14.1)	26.62
	Noteridae	*Noterus japonicus*	SLE699	500	8,815,384	13,276	352	199	21.8 (16.1)	38.60
	Noteridae	*Liocanthydrus bicolor*	SLE681	100	6,191,174	15,454	239	172	5.4 (4.4)	52.12
	Noteridae	*Notomicrus* sp.	SLE894	25	3,052,800	17,528	390	218	51.6 (46.6)	31.40
	Noteridae	*Suphisellus pereirai*	SLE840	100	7,069,210	16,597	195	138	4.7 (4.6)	65.94

Metrics include the number of reads, contigs, total number of UCEs recovered; *UCEs 50%*: number of UCEs present in 50% matrix, *Mean (SD) coverage* ×: mean coverage (×) and standard deviation of UCE loci per sample in 50% matrix; *% missing data*: % missing data (bp) in 50% matrix. The counts of UCE loci and coverage for the 50% data matrix were obtained using the following code: https://github.com/laninsky/reference_aligning_to_established_loci/tree/master/phase_everyone.

Results

UCE data

For the adephagan taxa, we recovered a mean of 3,158,813 paired-end reads, 19,693 contigs and 346 UCE loci per sample, with a median of 362 UCE loci and a range of 195–426 loci (Table 1.1). Our final 50% data matrix used in all phylogenetic analyses comprised 305 loci with a total alignment length of 83 547 bp, a mean of 204 UCE loci with a median of 208 UCE

loci, a range of 138–236 loci per sample, and a mean coverage of 34.8× per sample. The average locus length was 376 bp, with a range of 76–922 bp. The construction of data matrices allowing for less missing data showed drastic decreases in the number of loci and alignment length (Table S1.1). For example, increasing the completeness threshold to 60% removed nearly 100 loci from the alignment. A 100% complete matrix yielded an alignment of a single locus. Assembled sequence data and raw sequence reads have been deposited in GenBank and GenBank SRA, respectively (BioProject accession PRJNA379181; BioSample accessions SAMN06603286–SAMN06603309).

Phylogenetic analyses

Maximum likelihood analyses of the concatenated dataset, both partitioned and unpartitioned, recovered well-resolved trees with identical relationships and strong bootstrap support at nearly all nodes (Fig. 1.2, Table 1.2; Figures S1.1, S1.2). Hydradephaga was recovered as paraphyletic, with Gyrinidae sister to a monophyletic Geadephaga and the remaining Hydradephagan families. Haliplidae was recovered as sister to Dytiscoidea including Meruidae. Dytiscoidea was recovered as monophyletic with strong support, with Hygrobiidae sister to Amphizoidae+ Dytiscidae, together sister to Meruidae+ Noteridae. Noteridae was recovered as monophyletic and sister to Meruidae with strong support.
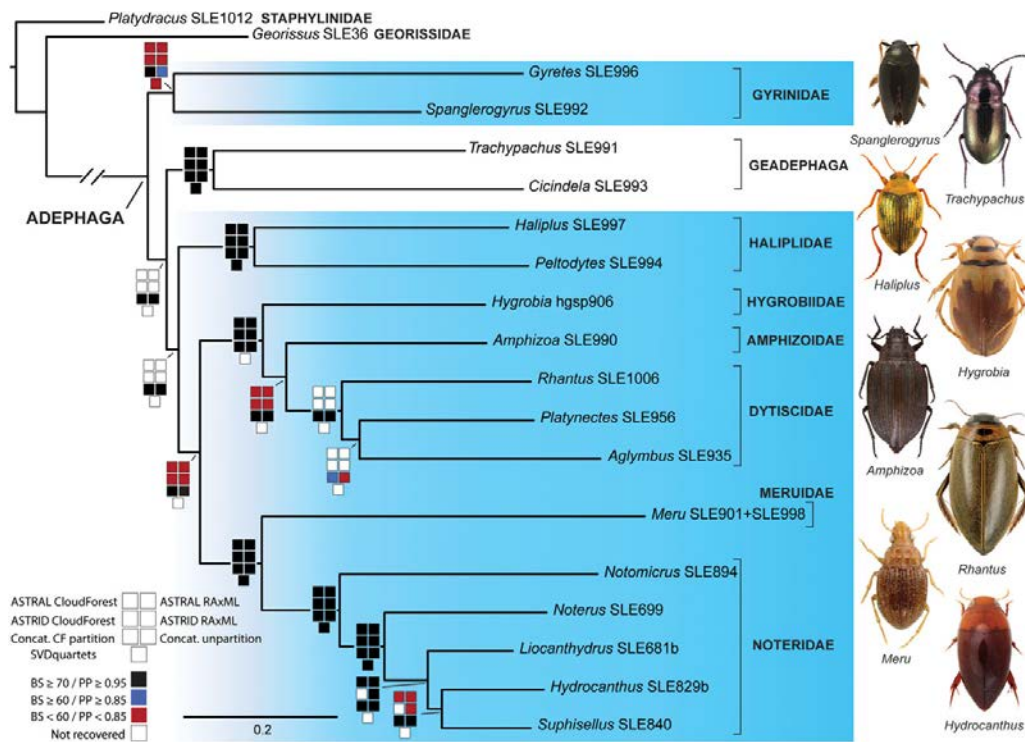
Figure 1.2. Phylogenetic relationships of the Adephaga based on ultraconserved element (UCE) data. The tree was recovered by maximum likelihood (ML) analysis of the full concatenated dataset of 305 UCE loci with the cloudforest partitioning scheme. Boxes at each node indicate analysis type, with recovered support denoted by box color. The key at bottom left provides the analysis type and support value colour scheme for boxes at nodes. Boxes for ML analyses of the concatenated datasets and svdquartets indicate recovered bootstrap values (BS). Boxes for astral and astrid analyses indicate recovered posterior probability (PP). Nodes with a BS ≥70 or a PP ≥0.95 are considered strongly supported (Hillis & Bull, 1993; Erixon et al., 2003). The 'Hydradephagan' families are highlighted in blue. Image credits: *Trachypachus gibbsii* is from Bousquet (2012); *Haliplus ruficollis* is by Udo Schmidt, used with permission; *Amphizoa lecontei* is by Guy Hanley, used with permission; *Meru phyllisae* is by Harold Schillhammer, used with permission. Other images by authors or from free media commons. [Colour figure can be viewed at wileyonlinelibrary.com].

Coalescent species tree methods (astral, astrid, svdquartets) recovered poorly resolved trees with weak support (Fig. 1.2, Table 1.2, Figures S1.3–S1.7). However, Hydradephaga was still recovered as paraphyletic across these methods. In addition, astrid and astral corroborated the relationships recovered by the concatenated analyses for Dytiscoidea, with Noteridae recovered as monophyletic and Meruidae its sister, with high posterior support (Fig. 1.2, Table 1.2; Figures S1.3–S1.6), and Hygrobiidae as sister to Amphizoidae+ Dytiscidae (albeit, with weak support Fig. 1.2, Table 1.2). svdquartets +paup* recovered a variant, but poorly supported tree (Table 1.2; Figure S1.7). Varied topologies were recovered among all coalescent species

tree analyses with respect to relationships among the major adephagan lineages (Geadephaga, Gyrinidae, Haliplidae, Dytiscoidea); however, no differences in comparison with the concatenated topologies were strongly supported.

Table 1.2. Summary of support recovered by analyses of ultraconserved element (UCE) data for select clades of Adephaga. CA-ML, concatenated maximum likelihood analysis. Bootstrap support is presented for CA-ML and svdquartets analyses. Posterior probability is presented for astrid and astral analyses.

Clade	CA-ML Unpartitioned	CA-ML Partitioned	ASTRID RaxML	ASTRID Cloudforest	ASTRAL RAxML	ASTRAL CloudForest	SVD quartets
Adephaga	100	100	1.00	1.00	1.00	1.00	100
Geadephaga'	100	100	0.99	0.99	0.97	0.97	87
Hydradephaga	-	-	-	-	-	-	-
Haliplidae + Dytiscoidea	89	98	-	-	0.36	-	-
Dytiscoidea	100	100	0.36	0.35	0.70	0.35	-
Amphizoidae + Dytiscidae	95	99	0.84	0.81	0.84	0.81	-
Noteridae + Meruidae	100	100	1.00	1.00	1.00	1.00	100
Noteridae	100	100	1.00	1.00	1.00	1.00	100

Discussion

Performance of analyses

Our concatenated analyses outperformed coalescent methods (astral, astrid, svdquartets) in producing well-resolved trees. Poor performance in coalescent methods can be attributed to gene-tree discordance, causes of which include ILS (Maddison, 1997; Edwards, 2009), lack of phylogenetic signal among gene trees and missing data (Thomson et al., 2008; Edwards, 2009; Gatesy & Springer, 2014; Springer & Gatesy, 2014; Xi et al., 2015; Edwards et al., 2016; Meiklejohn et al., 2016; Moyle et al., 2016). It has been shown that species tree methods such as astrid and astral are susceptible to errors in gene tree estimation (Edwards et al., 2016; Hosner et al., 2016; Meiklejohn et al., 2016; Springer & Gatesy, 2016). While individual loci can be problematic in this regard (Hosner et al., 2016; Meiklejohn et al., 2016), Hosner et al. (2016) further showed that missing data can also cause significant errors in species tree

estimation, especially when including taxa with only partially captured UCE loci (i.e. taxa with

shorter contigs for a given locus). As UCE sequences are obtained via targeted capture methods,

contig length of a given locus may vary significantly among samples (Hosner et al., 2016;

Streicher et al., 2016). Missing data was prevalent in our dataset (Tables 1, S1.1, S1.2), and our

50% complete dataset contained considerably fewer data than other published UCE datasets

(e.g. Faircloth et al., 2012; Blaimer et al., 2015; Gilbert et al., 2015; Moyle et al., 2016;

Alexander et al., 2017), suggesting the potential for this to impact our coalescent analyses. In

contrast, a growing body of literature has shown that there is a positive trade-off in constructing

a larger data matrix by allowing for the inclusion of loci with missing taxa while performing

concatenated analyses (e.g. Hosner et al., 2016; Streicher et al., 2016). Therefore, the amount of

missing data included in our analyses probably explains the difference in resolution between

our concatenated and coalescent analyses. Missing data may also have played a role in the

recovery of weakly supported nodes in the concatenated analyses. No adephagan genomes were

publicly available for use in the development of the Coleoptera bait set (Faircloth, 2017). This

potentially limited the number of loci recovered among our samples due to ascertainment bias.

The deep timescale covered by our analyses in comparison with studies on other taxa utilizing

UCEs (crown age estimate of Adephaga, ~200–250 million years ago; Mckenna et al., 2015;

Toussaint et al., 2017) may also have played a role in influencing the amount of missing data

among our samples. Under the assumption that missing data influenced the performance of our

coalescent-based analyses, we expect that the most pragmatic solution would be to decrease the

amount of missing data in our dataset and broaden our sampling of UCE loci, perhaps by

designing Adephaga-specific UCE probes based on the loci recovered in this study and/or novel

genomic and transcriptomic sequencing. The development of taxon-specific baits has proved

effective in improving sequence capture success and reducing missing data (Branstetter et al., 2017). Increasing the number of captured UCE loci would also grant flexibility in filtering problematic loci for estimating input gene trees in coalescent methods.

Relationships within Adephaga

The results of our analyses show the potential of UCEs for resolving relationships within Adephaga, adding support for the paraphyly of Hydradephaga, relationships within Dytiscoidea, and reaffirming Meruidae as sister to Noteridae. Our concatenated analyses produced well-resolved trees with robust support, and the recovery of the paraphyletic relationship of Hydradephaga with respect to Geadephaga agrees with several previous studies (Rolf G. Beutel et al., 2006, 2008, 2013; D. R. Maddison et al., 2009; Dressler et al., 2011). In addition, recovering Gyrinidae as sister to the rest of Adephaga, and Geadephaga and Haliplidae as successive sisters to Dytiscoidea, agrees with Dressler et al. (2011) and Beutel et al. (2013). This contrasts with Ribera et al. (2002a), in their investigation of Adephaga and recent investigations of the phylogeny of Coleoptera based on morphology (Lawrence et al., 2011) and molecules (Shull et al., 2001; Hunt et al., 2007; McKenna et al., 2015), which found a monophyletic Hydradephaga sister to Geadephaga. These studies may have been influenced by limited data [one marker in both Shull et al. (2001) and Ribera et al. (2002a) and three markers in Hunt et al. (2007)], or the a priori constraint of relationships within analyses (McKenna et al., 2015, in part). The morphology-based investigation of Lawrence et al. (2011) was robust (516 characters), but the authors noted the potential issues of character polarization and homoplasy that arises when attempting to infer a phylogeny for a group as diverse as Coleoptera. We believe the present investigation casts further doubt on the monophyly of Hydradephaga, and highlights the need for further investigation. In addition to the paraphyly of Hydradephaga,

another long-standing area of phylogenetic uncertainty within Adephaga involves the families of Dytiscoidea: Aspidytidae, Amphizoidae, Hygrobiidae and Dytiscidae. Most previous research recovered Hygrobiidae as sister to Dytiscidae (e.g. Balke et al., 2005; Beutel et al., 2006, 2008, 2013; Dressler et al., 2011; McKenna et al., 2015; cf. Toussaint et al., 2016b). In all analyses except svdquartets (Figure S1.7), we recovered Amphizoidae and Dytiscidae as sister, with Hygrobiidae sister to both these lineages (Fig. 1.2, Table 1.2). This placement agrees with the phylogenetic estimate of Balke et al. (2008). The phylogeny of Dytiscoidea will require further investigation with increased sampling, especially the addition of representatives of Aspidytidae. Our recovery of Noteridae as monophyletic and the sister relationship of Noteridae and Meruidae corroborate previous investigations (Beutel et al., 2006; Balke et al., 2008; Dressler et al., 2011; Short et al., 2012; McKenna et al., 2015; Toussaint et al., 2016b; Baca et al., 2017). The relationships recovered here have several evolutionary implications. Foremost among them, the nested placement of Geadephaga within the aquatic adephagans suggests that an aquatic lifestyle was the ancestral condition of the suborder. However, an alternative hypothesis could be a terrestrial ancestor of Adephaga, with aquatic lifestyles evolving multiple times, rather than the terrestrial lifestyle evolving secondarily. As noted by as noted by Shull et al. (2001), Beutel et al. (2006), and Maddison et al. (2009), many authors fail to reject this alternative hypothesis. The diversity of aquatic lifestyles among the major aquatic lineages further adds to the uncertainty about the ancestral condition of Adephaga (Shull et al., 2001; Ribera et al., 2002). Despite this uncertainty, our results corroborate previous research suggesting that there were several transitions in ecology and swimming behaviour among these lineages (Shull et al., 2001; Ribera et al., 2002a).

Conclusion

Here, we explored the utility of UCEs and the Coleoptera 5k version 1 bait set in estimating the phylogeny of the major lineages of Adephaga. Our concatenated analyses produced robust support for several key relationships, including the paraphyly of Hydradephaga, with Gyrinidae sister to all other adephagans, the sister relationship of Hygrobiidae to Amphizoidae+ Dytiscidae, and the monophyly of Noteridae. To further elucidate relationships of Adephaga using UCEs, it would be desirable to design Adephaga-specific UCE baits, or coleopteran baits with increased representation for Adephaga, to simultaneously broaden the number of targeted loci and reduce the amount of missing data. In addition, expanded taxon sampling within Dytiscoidea (particularly the inclusion of Aspidytidae) and Geadephaga, as well as a more diverse set of outgroups, would enhance future analyses. However, despite our relatively limited taxon sampling and number of loci, we found that UCEs show promising utility to resolve difficult phylogenetic relationships in Adephaga and Coleoptera.

Acknowledgments

References

Alarie, Y., Short, A. E. Z., Garcia, M., & Joly, L. (2011). Larval Morphology of Meruidae (Coleoptera: Adephaga) and Its Phylogenetic Implications. *Annals of the Entomological Society of America*, *104*(1), 25–36. https://doi.org/10.1603/AN10054

Alexander, A. (2016) UCE_Processing_Steps. URL https://github.com/laninsky/UCE_processing_steps/ [accessed on 27 December, 2016].

Alexander, A. M., Su, Y.-C., Oliveros, C. H., Olson, K. V., Travers, S. L., & Brown, R. M. (2017). Genomic data reveals potential for hybridization, introgression, and incomplete lineage sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. *Evolution*, *71*(2), 475–488. https://doi.org/10.1111/evo.13133

Baca, S. M., Toussaint, E. F. A., Miller, K. B., & Short, A. E. Z. (2017). Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on data partitioning strategies. *Molecular Phylogenetics and Evolution*, *107*, 282–292. https://doi.org/10.1016/j.ympev.2016.10.016

Balke, M., Ribera, I., & Beutel, R. G. (2005). The systematic position of Aspidytidae, the diversification of Dytiscoidea (Coleoptera, Adephaga) and the phylogenetic signal of third codon positions. *Journal of Zoological Systematics and Evolutionary Research*, *43*(3), 223–242. https://doi.org/10.1111/j.1439-0469.2005.00318.x

Balke, M., Ribera, I., Beutel, R., Viloria, A., Garcia, M., & Vogler, A. P. (2008). Systematic placement of the recently discovered beetle family Meruidae (Coleoptera: Dytiscoidea) based on molecular data. *Zoologica Scripta*, *37*(6), 647–650. https://doi.org/10.1111/j.1463-6409.2008.00345.x

Bejerano, G. (2004). Ultraconserved Elements in the Human Genome. *Science*, *304*(5675), 1321–1325. https://doi.org/10.1126/science.1098119

Beutel, R. G., & Roughley, R. E. (1988). On the systematic position of the family Gyrinidae (Coleoptera: Adephaga). *Journal of Zoological Systematics and Evolutionary Research*, *26*(5), 380–400. https://doi.org/10.1111/j.1439-0469.1988.tb00324.x

Beutel, R. G., Balke, M., & Steiner, W. E. (2006). The systematic position of Meruidae (Coleoptera, Adephaga) and the phylogeny of the smaller aquatic adephagan beetle families. *Cladistics*, *22*(2), 102–131. https://doi.org/10.1111/j.1096-0031.2006.00092.x

Beutel, R. G., Ribera, I., & Bininda-Emonds, O. R. P. (2008). A genus-level supertree of Adephaga (Coleoptera). *Organisms Diversity & Evolution*, *7*(4), 255–269. https://doi.org/10.1016/j.ode.2006.05.003

Beutel, R. G., Wang, B., Tan, J.-J., Ge, S.-Q., Ren, D., & Yang, X.-K. (2013). On the phylogeny and evolution of Mesozoic and extant lineages of Adephaga (Coleoptera, Insecta). *Cladistics*, *29*(2), 147–165. https://doi.org/10.1111/j.1096-0031.2012.00420.x

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*(1), 271. https://doi.org/10.1186/s12862-015-0552-5

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*(8), e0161531. https://doi.org/10.1371/journal.pone.0161531

Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., Gates, M. W., Kula, R. R., & Brady, S. G. (2016). Phylogenomic Analysis of Ants, Bees and Stinging Wasps: Improved Taxon Sampling Enhances Understanding of Hymenopteran Evolution. *BioRxiv*, 068957. https://doi.org/10.1101/068957

Branstetter, M. G., Longino, J. T., Ward, P. S., & Faircloth, B. C. (2017). Enriching the ant tree of life: Enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. *Methods in Ecology and Evolution*, *8*(6), 768–776. https://doi.org/10.1111/2041-210X.12742

Chifman, J., & Kubatko, L. (2014). Quartet Inference from SNP Data Under the Coalescent Model. *Bioinformatics*, *30*(23), 3317–3324. https://doi.org/10.1093/bioinformatics/btu530

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*(5), 783–786. https://doi.org/10.1098/rsbl.2012.0331

Crawford, N. G., Parham, J. F., Sellas, A. B., Faircloth, B. C., Glenn, T. C., Papenfuss, T. J., Henderson, J. B., Hansen, M. H., & Simison, W. B. (2015). A phylogenomic analysis of turtles. *Molecular Phylogenetics and Evolution*, *83*, 250–257. https://doi.org/10.1016/j.ympev.2014.10.021

Dressler, C., Ge, S.-Q., & Beutel, R. G. (2011). Is Meru a specialized noterid (Coleoptera, Adephaga)? *Systematic Entomology*, *36*(4), 705–712. https://doi.org/10.1111/j.1365-3113.2011.00585.x

Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution; International Journal of Organic Evolution*, *63*(1), 1–19. https://doi.org/10.1111/j.1558-5646.2008.00549.x

Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., Zhong, B., Wu, S., Lemmon, E. M., Lemmon, A. R., Leaché, A. D., Liu, L., & Davis, C. C. (2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. *Molecular Phylogenetics and Evolution*, *94*, 447–462. https://doi.org/10.1016/j.ympev.2015.10.027

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics (Oxford, England)*, *32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, *8*(9), 1103–1112. https://doi.org/10.1111/2041-210X.12754

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources*, *15*(3), 489–501. https://doi.org/10.1111/1755-0998.12328

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning Multiple Evolutionary Timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Gatesy, J., & Springer, M. S. (2014). Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution*, *80*, 231–266. https://doi.org/10.1016/j.ympev.2014.08.013

Gilbert, P. S., Chang, J., Pan, C., Sobel, E. M., Sinsheimer, J. S., Faircloth, B. C., & Alfaro, M. E. (2015). Genome-wide ultraconserved elements exhibit higher phylogenetic informativeness than traditional gene markers in percomorph fishes. *Molecular Phylogenetics and Evolution*, *92*, 140–146. https://doi.org/10.1016/j.ympev.2015.05.027

Glenn, T. C., Nilsen, R. A., Kieran, T. J., Sanders, J. G., Bayona-Vásquez, N. J., Finger, J. W., Pierson, T. W., Bentley, K. E., Hoffberg, S. L., Louha, S., García-De León, F. J., Del Río-

Portilla, M. A., Reed, K. D., Anderson, J. L., Meece, J. K., Aggrey, S. E., Rekaya, R., Alabady, M., Bélanger, M., … Faircloth, B. C. (2016). *Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext)* [Preprint]. Genomics. https://doi.org/10.1101/049114

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Hillis, D. M., & Bull, J. J. (1993). An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology*, *42*(2), 182–192. https://doi.org/10.1093/sysbio/42.2.182

Hosner, P. A., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2016). Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes). *Molecular Biology and Evolution*, *33*(4), 1110–1125. https://doi.org/10.1093/molbev/msv347

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O. S., Wild, R., Hammond, P. M., Ahrens, D., Balke, M., Caterino, M. S., Gómez-Zurita, J., Ribera, I., Barraclough, T. G., Bocakova, M., Bocak, L., & Vogler, A. P. (2007). A Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation. *Science*, *318*(5858), 1913–1916. https://doi.org/10.1126/science.1146954

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Larson, D. J., & LaBonte, J. R. (1994). Stygoporus oregonensis, a New Genus and Species of Subterranean Water Beetle (Coleoptera: Dytiscidae: Hydroporini) from the United States. *The Coleopterists Bulletin*, *48*(4), 371–379.

Lawrence, J. F., Ślipiński, A., Seago, A. E., Thayer, M. K., Newton, A. F., & Marvaldi, A. E. (2011). Phylogeny of the Coleoptera Based on Morphological Characters of Adults and Larvae. *Annales Zoologici*, *61*(1), 1–217. https://doi.org/10.3161/000345411X576725

Maddison, D. R., Moore, W., Baker, M. D., Ellis, T. M., Ober, K. A., Cannone, J. J., & Gutell, R. R. (2009). Monophyly of terrestrial adephagan beetles as indicated by three nuclear genes (Coleoptera: Carabidae and Trachypachidae). *Zoologica Scripta*, *38*(1), 43–62. https://doi.org/10.1111/j.1463-6409.2008.00359.x

Maddison, W. P. (1997). Gene Trees in Species Trees. *Systematic Biology*, *46*(3), 523–536. https://doi.org/10.1093/sysbio/46.3.523

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

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*(1), e54848. https://doi.org/10.1371/journal.pone.0054848

Mckenna, D. D., Wild, A. L., Kanda, K., Bellamy, C. L., Beutel, R. G., Caterino, M. S., Farnum, C. W., Hawks, D. C., Ivie, M. A., Jameson, M. L., Leschen, R. a. B., Marvaldi, A. E., Mchugh, J.

V., Newton, A. F., Robertson, J. A., Thayer, M. K., Whiting, M. F., Lawrence, J. F., Ślipiński,

A., … Farrell, B. D. (2015). The beetle tree of life reveals that Coleoptera survived end-Permian

mass extinction to diversify during the Cretaceous terrestrial revolution. *Systematic Entomology*,

40(4), 835–880. https://doi.org/10.1111/syen.12132

Meiklejohn, K. A., Faircloth, B. C., Glenn, T. C., Kimball, R. T., & Braun, E. L. (2016).

Analysis of a Rapid Evolutionary Radiation Using Ultraconserved Elements: Evidence for a Bias

in Some Multispecies Coalescent Methods. *Systematic Biology*, *65*(4), 612–627.

https://doi.org/10.1093/sysbio/syw014

Miller, K. B. (2009). On the systematics of Noteridae (Coleoptera: Adephaga: Hydradephaga):

Phylogeny, description of a new tribe, genus and species, and survey of female genital

morphology. *Systematics and Biodiversity*, *7*(2), 191–214.

https://doi.org/10.1017/S1477200008002946

Mirarab S, Bayzid MS, & Warnow T. (2014). Evaluating Summary Methods for Multilocus

Species Tree Estimation in the Presence of Incomplete Lineage Sorting. *Systematic Biology*,

65(3), 366–380. https://doi.org/10.1093/sysbio/syu063

Mirarab, S., Reaz, R., Bayzid, Md. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014).

ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*(17), i541–

i548. https://doi.org/10.1093/bioinformatics/btu462

Mirarab, S., & Warnow, T. (2015). ASTRAL-II: Coalescent-based species tree estimation with

many hundreds of taxa and thousands of genes. *Bioinformatics*, *31*(12), i44–i52.

https://doi.org/10.1093/bioinformatics/btv234

Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D.,

Travers, S. L., Brown, R. M., & Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea

triggered the global songbird radiation. *Nature Communications*, *7*(1), 12709.

https://doi.org/10.1038/ncomms12709

Nick Crawford, & Brant Faircloth. (2014). *CloudForest: Bug fix in parallel bootstrapping*.

Zenodo. https://doi.org/10.5281/zenodo.12259

Oswald, J. A., Harvey, M. G., Remsen, R. C., Foxworth, D. U., Cardiff, S. W., Dittmann, D. L.,

Megna, L. C., Carling, M. D., & Brumfield, R. T. (2016). Willet be one species or two? A

genomic view of the evolutionary history of Tringa semipalmata. *The Auk*, *133*(4), 593–614.

https://doi.org/10.1642/AUK-15-232.1

Ribera, I., Beutel, R. G., Balke, M., & Vogler, A. P. (2002). Discovery of Aspidytidae, a new

family of aquatic Coleoptera. *Proceedings of the Royal Society of London. Series B: Biological*

Sciences, *269*(1507), 2351–2356. https://doi.org/10.1098/rspb.2002.2157

Ribera, I., Hogan, J. E., & Vogler, A. P. (2002). Phylogeny of Hydradephagan Water Beetles

Inferred from 18S rRNA Sequences. *Molecular Phylogenetics and Evolution*, *23*(1), 43–62.

https://doi.org/10.1006/mpev.2001.1080

Sayyari, E., & Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support

from Quartet Frequencies. *Molecular Biology and Evolution*, *33*(7), 1654–1668.

https://doi.org/10.1093/molbev/msw079

Short, A. E. Z., Alarie, Y., García, M., & Joly, L. J. (2012). Are noterids specialised meruids

(Coleoptera, Adephaga)? A reply to Dressler et al. *Systematic Entomology*, *37*(3), 417–419.

https://doi.org/10.1111/j.1365-3113.2012.00626.x

Shull, V. L., Vogler, A. P., Baker, M. D., Maddison, D. R., & Hammond, P. M. (2001).

Sequence Alignment of 18S Ribosomal RNA and the Basal Relationships of Adephagan Beetles:

Evidence for Monophyly of Aquatic Families and the Placement of Trachypachidae. *Systematic Biology*, *50*(6), 945–969. https://doi.org/10.1080/106351501753462894

Slipinski, S. A., Leschen, R. a. B., & Lawrence, J. F. (2011). Order Coleoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. *Zootaxa*, *3148*(1), 203–208. https://doi.org/10.11646/zootaxa.3148.1.39

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology*, *63*(1), 83–95. https://doi.org/10.1093/sysbio/syt061

Spangler, P. J. (1996). *Four new stygobiontic beetles (Coleoptera: Dytiscidae; Noteridae; Elmidae)*. *10*, 20.

Spangler, P. J., & Steiner, W. E. (2005). A new aquatic beetle family, Meruidae, from Venezuela (Coleoptera: Adephaga). *Systematic Entomology*, *30*(3), 339–357. https://doi.org/10.1111/j.1365-3113.2005.00288.x

Springer, M. S., & Gatesy, J. (2014). Land plant origins and coalescence confusion. *Trends in Plant Science*, *19*(5), 267–269. https://doi.org/10.1016/j.tplants.2014.02.012

Springer, M. S., & Gatesy, J. (2016). The gene tree delusion. *Molecular Phylogenetics and Evolution*, *94*(Pt A), 1–33. https://doi.org/10.1016/j.ympev.2015.07.018

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Starrett, J., Derkarabetian, S., Hedin, M., Bryson, R. W., McCormack, J. E., & Faircloth, B. C. (2017). High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Molecular Ecology Resources*, *17*(4), 812–823. https://doi.org/10.1111/1755-0998.12621

Streicher, J. W., Schulte, J. A., & Wiens, J. J. (2016). How Should Genes and Taxa be Sampled for Phylogenomic Analyses with Missing Data? An Empirical Study in Iguanian Lizards. *Systematic Biology*, *65*(1), 128–145. https://doi.org/10.1093/sysbio/syv058

Swofford, D.L. (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Talavera, G., & Castresana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, *56*(4), 564–577. https://doi.org/10.1080/10635150701472164

Thomson, R. C., Shedlock, A. M., Edwards, S. V., & Shaffer, H. B. (2008). Developing markers for multilocus phylogenetics in non-model organisms: A test case with turtles. *Molecular Phylogenetics and Evolution*, *49*(2), 514–525. https://doi.org/10.1016/j.ympev.2008.08.006

Toussaint, E. F. A., Beutel, R. G., Morinière, J., Jia, F., Xu, S., Michat, M. C., Zhou, X., Bilton, D. T., Ribera, I., Hájek, J., & Balke, M. (2016). Molecular phylogeny of the highly disjunct cliff water beetles from South Africa and China (Coleoptera: Aspidytidae). *Zoological Journal of the Linnean Society*, *176*(3), 537–546. https://doi.org/10.1111/zoj.12332

Toussaint, E. F. A., Hendrich, L., Escalona, H. E., Porch, N., & Balke, M. (2016). Evolutionary history of a secondary terrestrial Australian diving beetle (Coleoptera, Dytiscidae) reveals a lineage of high morphological and ecological plasticity. *Systematic Entomology*, *41*(3), 650–657. https://doi.org/10.1111/syen.12182

Toussaint, E. F. A., Seidel, M., Arriaga-Varela, E., Hájek, J., Král, D., Sekerka, L., Short, A. E. Z., & Fikáček, M. (2017). The peril of dating beetles. *Systematic Entomology*, *42*(1), 1–10. https://doi.org/10.1111/syen.12198

Vachaspati, P., & Warnow, T. (2015). ASTRID: Accurate Species Trees from Internode Distances. *BMC Genomics*, *16*(10), S3. https://doi.org/10.1186/1471-2164-16-S10-S3

Watts, C. H. S., Hancock, P. J., & Leys, R. (2007). A stygobitic Carabhydrus Watts (Dytiscidae, Coleoptera) from the Hunter Valley in New South Wales, Australia. *Australian Journal of Entomology*, *46*(1), 56–59. https://doi.org/10.1111/j.1440-6055.2007.00585.x

Xi, Z., Liu, L., & Davis, C. C. (2015). Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution*, *92*, 63–71. https://doi.org/10.1016/j.ympev.2015.06.009

Young, F. N., & Longley, G. (1976). A New Subterranean Aquatic Beetle from Texas (Coleoptera: Dytiscidae-Hydroporinae). *Annals of the Entomological Society of America*, *69*(5), 787–792. https://doi.org/10.1093/aesa/69.5.787

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics (Oxford, England)*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

<u>Chapter 2</u>: Molecular Phylogeny of the Notomicrine Water Beetles (Coleoptera: Noteridae) Reveals Signatures of Gondwanan Vicariance and Ecological Plasticity

Abstract

Notomicrinae (Coleoptera: Noteridae) is a subfamily of minute and ecologically diverse aquatic beetles distributed across the Southeast Asia, Oceania, and the Americas. We investigate the evolution of Notomicrinae and construct the first species-level phylogeny within Noteridae using five nuclear and mitochondrial gene fragments. We focus on the genus *Notomicrus* Sharp (Coleoptera: Noteridae), sampling 13 of the 17 known *Notomicrus* species and an additional 11 putative undescribed species. We also include *Phreatodytes haibaraensis* Uéno (Coleoptera: Noteridae). Datasets are analyzed in Maximum Likelihood and Bayesian frameworks. With these, we 1) estimate divergence times among notomicrine taxa and reconstruct the biogeographical history of the group, particularly testing the hypothesis of Gondwanan vicariance between Old World and New World *Notomicrus*; 2) additionally, we assess ecological plasticity within Notomicrinae in the context of the phylogeny; and 3) finally, we test the monophyly of tentative species groups within *Notomicrus* and place putative new taxa. We recover a monophyletic Notomicrinae, with *Phreatodytes* sister to *Notomicrus*. We estimate the crown age of Notomicrinae to be ca. 110 Mya. The crown age of *Notomicrus* is recovered as ca. 75 Mya, there diverging into reciprocally monophyletic Old and New World clades, suggesting Gondwanan vicariance. Our phylogenetic estimate indicates a strong degree of ecological plasticity within Notomicrinae, with habitat switching occurring in recently diverging taxa. Finally, we recover five main species groups in *Notomicrus*, one Old World, Four New World, with tentative affirmation of the placement of undescribed species.

Introduction

The adephagan aquatic beetle family Noteridae consists of 18 genera (with *Suphisellus* Sensu

lato) and >270 species (Nilsson 2011, Baca et al. 2017b). Though not rich in diversity relative to

other beetles, noterids can be found in aquatic habitats around the world. An increased

interest in this family in the last decade (e.g., Miller 2009, Gomez and Miller 2013, Baca and

Toledo 2015, Manuel 2015, Baca et al. 2017b, Guimaraes and Ferreira Jr 2019) has revealed

novel diversity and relationships among lineages. However, the mode and tempo of noterid

evolution has seen little investigation. For example, the biogeographical history of Noteridae has

never been reconstructed despite repeated phylogenetic patterns of deep sistergroup relationships

between Old and New World taxa (Baca et al. 2017b). Also, while noterids show a wide range of

ecological preferences, ecological evolution has not been explored in a phylogenetic framework.

And finally, the species level diversity remains poorly known. To date, no attempt has been

made to investigate noterid phylogeny at the species level. Within Noteridae, the combination of

New World-Old World geographical divergences, ecological variability, and poorly known

relationships are especially pronounced in the subfamily Notomicrinae. Notomicrinae (sensu

Baca et al. 2017b) comprises 26 described species in three genera: *Phreatodytes* Ueno

(Coleoptera: Noteridae; 7 species), *Speonoterus* Spangler (Coleoptera: Noteridae; monotypic),

and *Notomicrus* Sharp (Coleoptera: Noteridae; 17 species). All notomicrine taxa are small to

minute, the largest not surpassing 2 mm in length, and present a broad range of ecological

preference among genera and species (Baca and Short 2018, Guimarães and Ferreira Jr 2019,

pers. obs.). *Phreatodytes* and *Speonoterus* are blind (eyeless) genera known from subterranean

environments, with *Phreatodytes* inhabiting aquifers (Ueno 1957, 1996; Kato et al.

2010) and *Speonoterus* collected from a small puddle in a shallow cave (Spangler 1996).

Notomicrus, with the bulk of notomicrine diversity, occurs in a variety of habitats, including the

margins of detrital pools, ponds, swamps, streams, and hygropetric seeps with varying specificity among species (Baca and Short 2018, Guimarães and Ferreira Jr 2019). Recent sampling with Winkler traps has yielded a *Notomicrus* species inhabiting terrestrial leaf litter in South America (pers. obs., unpublished data).

In terms of biogeography, *Phreatodytes* and *Speonoterus* are restricted in their distributions, with *Phreatodytes* known only from aquifers in Japan and *Speonoterus* solely from the type locality, a cave in Indonesia. *Notomicrus* ranges from Indomalaysia to Australia and Oceania, and the Americas, with most species (14 of 17) described from the Neotropics (Nilsson 2011, Manuel 2015, Baca and Short 2018, Guimaraes and Ferreira Jr 2019). With respect to this distribution, Baca et al. (2017b) recovered New and Old World *Notomicrus* as reciprocally monophyletic. This presents a pattern similar to those in other aquatic beetles (Bukontaite et al. 2014, Gustafson and Miller 2017, Toussaint and Short 2017), wherein the apparent mutual isolation among Old and New World clades was suggested to be produced by the breakup of Gondwana.

Recent investigations using molecular data (Kato et al. 2010, Toussaint et al 2016, Baca et al. 2017b, Gustafson et al. 2020) have consistently recovered *Phreatodytes* and *Notomicrus* as monophyletic sisters to the rest of Noteridae, with Baca et al. (2017b) subsuming Phreatodytinae into the Notomicrinae. Yet unclear is the relationship of *Speonoterus* to these other two genera. Morphological data suggest that it is sister to or nested within *Notomicrus* (Spangler 1996, Miller 2009), although this remains to be tested with molecular data. Within *Notomicrus*, Baca et al. (2017b) recovered reciprocally monophyletic Old and New World groups but with limited sampling. To date, species level investigations of Notomicrinae have focused almost exclusively on taxonomy (Kato et al. 2010, Toledo 2010, Miller 2013, Manuel 2015, Baca and Short 2018,

Guimaraes and Ferreira Jr 2019). These works are indicative of our shortcomings in knowledge in the diversity in Notomicrinae, with Toledo (2010) and personal observations further suggesting significant lumped or undescribed diversity. The lack of a comprehensive species-level phylogeny leaves questions of the historical biogeography, ecological evolution, and taxonomy of Notomicrinae without a strong foundation for investigation.

Here, we construct a comprehensive molecular phylogenetic estimate of Notomicrinae, with an emphasis on relationships among species within *Notomicrus*. We use this phylogeny to estimate divergence times and reconstruct ancestral ranges. With these estimates, we 1) investigate the biogeographical history of Notomicrinae and test the hypothesis that the disjunction of Old and New World taxa is due to Gondwanan vicariance, 2) weigh relationships against habitat preference as a test of ecological plasticity, and 3) investigate relationships among and within suspected species groups, and among putative new species. The latter will provide a scaffold for future taxonomic work within *Notomicrus*, including forthcoming revision of New World taxa (Baca and Short, in preparation).

Materials and Methods

Taxon Sampling

We were able to include a single species of *Phraetodytes* (*Phraetodytes haibaraensis* Kato et al. 2010). Within *Notomicrus*, we sampled 13 of the 17 described species, in addition to 11 putative new species, for a total of 25 in-group species (Table S2.1). *Speonoterus*, known only from the four type specimens, was not included. Our initial phylogeny included 47 in-group samples (including *Phreatodytes*) and 7 outgroups, including genera of the noterid subfamily Noterinae, and families Amphizoidae, Dytiscidae, and Meruidae, for a total of 54 samples ('Full'

dataset; Fig. 2.2, Table 2.1). This was used for detailed investigation of *Notomicrus*

relationships. For divergence time estimation and biogeographical analyses, we followed our

reconstructions using the 'Full' data set and reduced sampling to individual terminals for each

putative species, with a total of 32 samples including outgroups ('Trimmed' dataset). In the case

of *Notomicrus tenellus*, we included terminals across its wide distribution, as our initial

reconstructions suggest that multiple species are subsumed under this name (see also Toledo

2010). In the case of the *traili* complex, we included a representative for each larger clade within

the group, as recovered by analyses of our 'Full' dataset (Table S2.1). We also removed

Amphizoidae from the outgroups. We studied type material of 13 known species to accurately

identify described versus undescribed species where needed (Table S2.3.

DNA Extraction and Data Collection

DNA was extracted from whole beetles, stored in ≥95% ethanol, using a Qiagen DNeasy kit

(Qiagen, Hilden, Germany), following the manufacturers protocols. As notomicrines are very

small, extraction modifications were added to maximize DNA yield, following several of the

recommendations of Cruaud et al. (2019). In brief, 1) beetle abdomens were separated from the

thorax; 2) each sample was lysed overnight with periodic vortexing; and 3) DNA was eluted with

heated elution buffer AE (56°C), with a 15-min incubation period prior to centrifuging. Detailed

methods are provided in Supplementary Methods.

We used the primers listed in Table 2.1 in polymerase chain reactions for amplification of

the following mitochondrial and nuclear gene fragments: cytochrome oxidase subunit 1 mtDNA

(CO1), carbamoylphosphate synthetase nDNA (CAD), histone 3 nDNA (H3), 16S mtrDNA, and

28S rDNA; see Supp Methods (online only) for details. Amplicons were diluted where necessary

and sent to GENEWIZ-Boston, MA (Brooks Automation Inc., Chelmsford, MA) for Sanger

sequencing.

Table 2.1. List of gene fragments and primers used in this study.

Gene	Location	Primer	Direction	Sequence	Reference
COI	mitochondrial	Jerry	Forward	CAACAYTTATTTTGATTTTTTGG	Simon et al. 1994
COI	mitochondrial	Pat	Reverse	ATCCATTACATATAATCTGCCATA	Simon et al. 1994
CAD	nuclear	CD439F	Forward	TTCAGTGTACARTTYCAYCCHGARCAYAC	Wild & Maddison (2008)
CAD	nuclear	CD688R	Reverse	TGTATACCTAGAGGATCDACRTTYTCCATRTTRCA	Wild & Maddison (2008)
H3	nuclear	H3aF	Forward	ATGGCTCGTACCAAGCAGACGGC	Colgan et al. 1998
H3	nuclear	H3aR	Reverse	ATATCCTTGGGCATGATGGTGAC	Colgan et al. 1998
16S	mitochondrial	LR-N-13398	Forward	CGCCTGTTTAACAAAAACA	Simon et al. 1994
16S	mitochondrial	LR-J-12887	Reverse	CCGGTCTGAACTCAGATCACGT	Simon et al. 1994
28S	nuclear	NLF184-21	Forward	ACCCGCTGAAYTTAAGCATAT	Van der Auwera et al.1994
28S	nuclear	LS1041R	Reverse	TACGGACRTCCATCAGGGTTTCCCCTGACTTC	Maddison 2008

Alignment and Phylogenetic Analyses

Sequence data were assembled, aligned, and concatenated with Geneious 10.2.2

(Biomatters, http://www.geneious.com). Trace files were visually inspected and refined before

generating contigs. Data for each fragment were aligned using MAFFT v7.450 (Katoh and

Standley 2013) with default settings. Alignments were visually inspected and protein coding

alignments were translated to amino acid reading frames to verify absence of stop codons.

Individual fragment alignments were then trimmed and concatenated. The 'Full' alignment of 54

samples was 3,422 bp in length; the 'Trimmed' alignment of 32 samples was 3,382 bp in length.

To search for optimal models for phylogenetic inference, alignments were partitioned a

priori by alignment features, following Baca et al. (2017b); respective alignments were divided

into subsets of individual gene fragments with protein coding genes (COI, CAD, H3) each

further divided by codon position, for a total of 11 input partitions for each alignment (Table

S2.2). Models were searched using IQ-Tree 2 (v.2.0-rc1, Minh et al. 2020) with the -m

TESTMERGE or -m TESTMERGEONLY functions which implement the *greedy* search

algorithm of PartitionFinder (Lanfear et al. 2012) and allow the merging of input partitions to

increase model fit and reduce the risk of over-parameterization.

For Bayesian inference, models were restricted to those available in the inference

software using the -mset option; however, this was done only to recover optimal merging of

partitions. For Maximum Likelihood (ML) inference, we used IQ-Tree 2 using the -m

TESTMERGE option (model search followed by phylogenetic inference), with edge-linked

partitions (-p or -spp). Support was assessed using UFBoot 2 Ultrafast bootstrapping (Hoang et

al. 2018) with 1,000 replicates (-B or -bb 1000). We considered a UFBoot bootstrap nodal

support of 95% or higher to be strong support for a recovered clade. Bayesian inference (BI)

analyses were conducted on both the 'Full' and 'Trimmed' alignments in BEAST2 2.6.1

(Bouckaert et al. 2019) via the CIPRES Science Gateway (Miller et al. 2010), with the.xml file

prepared in BEAUti 2 (Bouckaert et al. 2019). We used the merged partitioning scheme

recovered by the TESTMERGEONLY option in IQ-Tree 2. We unlinked site models across

partitions but linked clock models and trees. We did not use the models recovered by IQ-Tree for

site models. Instead we implemented bModelTest (Bouckaert et al. 2017), which uses a Bayesian

reversible jump to sample among different substitution models (including site heterogeneity and

invariant site selection), with mutation rate estimated. This eliminates potential bias in using

models recovered by ML methods (Bouckaert et al. 2017). Analyses were run with an

uncalibrated strict clock, under a Yule tree model with an exponential birthrate prior (initial =

1.0, ± infinite bounds), with a random initial tree. Analyses consisted of two independent runs of

50 million generations each, with trace and trees logged every 2,500 generations. Log files were

together viewed in Tracer 1.6 (Rambaut et al. 2018, https://beast.community/tracer) to assess

convergence and ensure sufficient sampling of parameters, with an effective sample size (ESS)

value of 200 considered as the threshold for good convergence. Tree files were combined and annotated with LogCombiner 2.6.2 and TreeAnnotator 2.6.2 (https://beast.community/). Posterior probability values of ≥0.95 were considered strong support for a given clade.

Divergence Time Estimation

Divergence times were estimated using the Trimmed sampling alignment (with sampling reduced to species' representatives, see above) in BEAST2, with.xml infiles generated using BEAUti 2. As above, we used the partitioning scheme recovered by IQ-tree, with unlinked site models, and linked clock and tree models. For all partition site models, we used bModelTest with mutation rate estimated. In lieu of fixing the tree topology, we implemented the recovered tree from the uncalibrated analyses above as the starting tree. We implemented four different combinations of clock and tree models to test the effect on divergence times: strict versus uncorrelated lognormal relaxed clock models and Yule versus Birth Death tree models (Table 2.2). We did not include the effect of unlinked clock models as initial analyses resulted in poor MCMC convergence. Exponential birthrate (Yule and Birth Death) and deathrate (Birth Death) tree prior distributions were, respectively, used across all analyses. For both Strict clock rate and uncorrelated lognormal relax clock mean, we used a uniform prior distribution, with an initial value of 0.01, and bounded from 0.00001 to 1.0 to prevent numerical instability in the analysis.

There are currently no described fossil Noteridae; consequently, we used secondary node calibrations. Using the ages recovered by Toussaint et al. (2017b), we implemented MRCA node calibrations on Meruidae + Noteridae (191.0 Mya; 95% HPD = 167.2, 216.0) and Noteridae (153.896 Mya; 95% HPD = 124.6, 183.3). Nodes were calibrated using a normal prior distribution, with the mean set to the respective mean ages of Toussaint et al. (2017b) and Sigma

values set so that the 95% prior distribution density quantiles matched as near as possible the

respective height posterior densities intervals (95% HPD) of Toussaint et al. (2017b).

All analyses were run on CIPRES and consisted of two independent runs of 100 million

generations with traces logged and trees sampled every 5,000 generations. Logs were imported

into Tracer 1.6 to assess analysis convergence, with an effective sample size (ESS) value of 200

considered the threshold for good convergence. Recovered tree files were then combined in

LogCombiner and respective Maximum Clade Credibility (MCC) trees were generated with

TreeAnnotator, with mean node heights, 95% HPD intervals, and 20% burn-ins. The fit of the

four model combinations implemented were compared using marginal likelihoods estimated with

Nested Sampling (Skilling 2006, Maturana Russel et al. 2019), see detailed explanation in Supp

Methods (online only). We set up the NS runs by manually editing the BEAUti generated.xml

infiles. Our NS runs consisted of the following parameters: chain length = 20,000, particle count

= 12, subchain length = 10,000. To select the best fit model combination, we calculated Bayes

Factors (BF, Kass and Raftery 1995) using the recovered marginal likelihood estimates (MLEs),

calculated as a ratio of the marginal likelihoods: 2 log(BF) = 2 × (MLE1 ‑ MLE2). A BF ≥ 10

was used as the threshold for strong support of a given model combination (Kass and Raftery

1995).

Table 2.2. Prior model combinations used for divergence time estimates in BEAST2. Included are marginal likelihood estimates (MLE) with standard deviations (SD MLE) and subsequent Bayes Factors (= 2log(BF)) recovered via Nested Sampling in BEAST2.

Clock Model	Tree Model	MLE	SD MLE	(Bayes Factor) = 2log(BF)			
				vs. (ST)(Y)	vs. (ST)(BD)	vs. (RC)(Y)	vs. (RC)(BD)
Strict (ST)	Yule (Y)	-26475.6364	4.949	0	-302.7412	-356.5470	-311.2354
Strict (ST)	Birth Death (BD)	-26324.2658	4.989	**302.7412**	0	-53.8058	-8.4942
Relaxed UCLN (RC)	Yule (Y)	**-26297.3629**	5.680	**356.547**	**53.8058**	0	**45.3116**
Relaxed UCLN (RC)	Birth Death (BD)	-26320.0187	5.544	**311.2354**	**8.4942**	-45.3116	0

Ancestral Range Reconstruction

Biogeographical inference was conducted with the BioGeoBEARS (Matzke 2018) package in R

v. 3.5.0 (R Core Team 2018). This package reconstructs ancestral ranges in Bayesian and

likelihood frameworks, offering models (and model testing) of various combinations of

anagenetic and cladogenetic evolutionary processes, including, dispersal, extinction, vicariance,

sympatry, and others. Analyses were conducted under the Dispersal Extinction Cladogenesis

(DEC) and Dispersal Vicariance Analysis (DIVALIKE) models. We also ran these analyses with

the jump dispersal (j) parameter (Matzke 2013). This models founder-event cladogenesis,

allowing a daughter lineage to inhabit a new range, whereas its sister inherits the ancestral range.

While the inclusion of the j parameter has shown to increase model likelihood (Matzke 2014,

2018), it was found to be flawed by Ree and Sanmartin (2018). As the imposed models of range

evolution do not account for time, the probability of time-dependent anagenetic processes of

range expansion via dispersal and extinction are underestimated, and the contribution of cladogenetic events to the likelihood artificially increases. The inclusion of the *j* parameter further exacerbates this, first by adding an additional cladogenetic process of range evolution, and second by *j* being a free cladogenetic parameter, that, at high values, can over-power any probability of nonjump events (Ree and Sanmartin 2018). Because of the potential effect on inference, and the likely inappropriateness of comparing models with *j* to those without (Ree and Sanmartin 2018), we included the *j* parameter as strictly exploratory in model comparison.

For the analyses, we used the MCC time-calibrated tree from the above BI analysis preferred by BF of marginal likelihoods. Outgroups were manually pruned to so that only in-group notomicrine taxa (*Phreatodytes + Notomicrus*) were included. We included the following biogeographical regions in our analyses: Oriental, O; Oceania, A; Nearctic, N; and South America/Neotropical, S. These encompass all regions inhabited by known Notomicrinae. For disambiguation, Oceania includes Australia and Polynesia, roughly separated from the Oriental region by the Wallace line; the oriental region here includes Japan in our coding; Nearctic is restricted to North America, excluding south and eastern Mexico; Neotropical includes South and Central America and the Antilles. We followed the time-slice, dispersal rate, and adjacency and areas allowed scheme of Toussaint et al. (2017a; excluding the African and Palearctic regions) as our focal clade age and range were very similar; see Supp Methods (online only) and Toussaint et al. (2017a), for complete details.

Results

Phylogenetic Analyses

The ML partitioning/models search in IQ-tree recovered eight and seven partitions for the full (54 samples) and reduced (32 samples) datasets, respectively; recovered partitioning schemes are

listed in Supp Table 2 (online only). The recovered ML and BI analyses of the full alignment resulted in well-resolved trees (Figs. 2.2, Supp Figs. S2.1–S2.13), with most relationships recovered with high support from both ultrafast bootstrap values in ML (UFB ≥ 95) and Posterior probability in BI (PP ≥ 0.95), with only some of the more shallow nodes recovered with poor support. Noteridae was recovered as monophyletic with high support, with Meruidae as its sister. Within noterids, Noterinae and Notomicrinae were recovered as monophyletic sister clades (UFB = 99; PP = 0.99), as were the notomicrine genera *Phreatodytes* and *Notomicrus* (UFB = 100; PP = 1.00). *Notomicrus* was recovered as monophyletic (UFB = 95, PP > 0.99), with the New World (UFB = 100; PP > 0.99) and Old World (UFB = 100; PP > 0.99) *Notomicrus* reciprocally monophyletic. Within the Old World 'tenellus' species group, *N. punctulatus* Fauvel, 1903, an endemic of New Caledonia, was recovered as sister to a monophyletic (UFB = 100; PP > 0.99) *N. tenellus* (Clark 1863) Sensu *lato*. Within *N. tenellus*, the Australian sample was found as sister to the Sumatran and Fijian samples (UFB = 100; PP > 0.99). The New World species were split into four major clades. *Notomicrus josiahi* Miller 2013 and an undescribed species, *N.* sp.1 were recovered as sister to all other New World *Notomicrus* (BF = 95; PP > 0.99) (hereafter 'josiahi' group). Next, *Notomicrus sharpi* Balfour-Browne 1939, *Notomicrus nanulus* (LeConte 1863)*, Notomicrus chailliei* Manuel 2015, and *Notomicrus femineus* Manuel 2015 were found nested within a clade (UFB = 100; PP > 0.99; hereafter 'nanulus' group) among several putative undescribed species. While ML and BI analyses recovered the same relationships within the *nanulus* clade, with high support recovered for the sister relationship of *N. sharpi* and *N. nanulus* (UFB = 100; PP = 100 and for an undescribed Peruvian species *N.* sp.10 sister (UFB = 99; PP > 0.99) to *N. chailliei* and *N. femineus* (UFB = 100;), other nodes
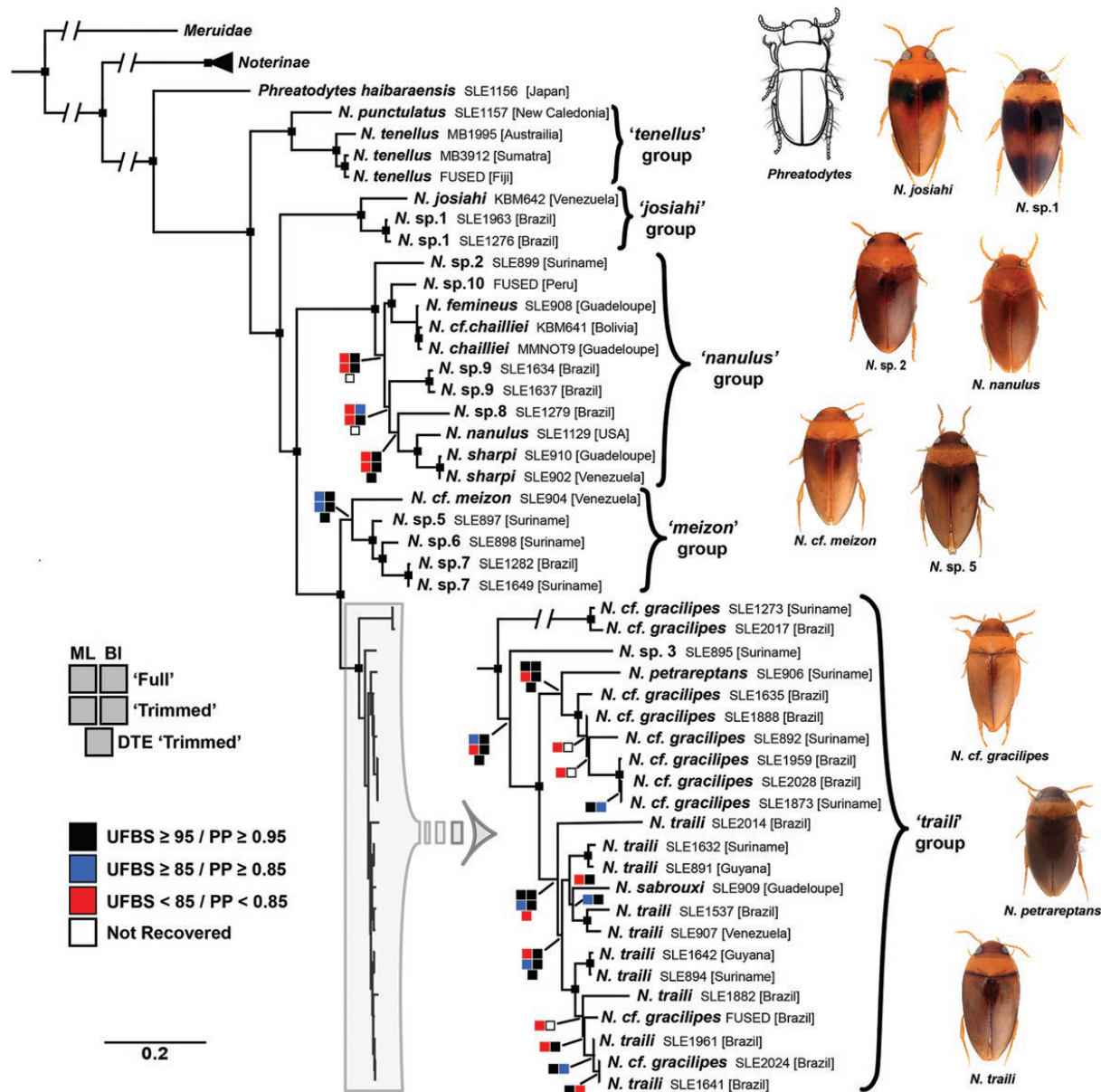
Figure 2.1. Phylogeny of Notomicrinae recovered by Maximum Likelihood analysis of 'Full' dataset in IQ-Tree2 with species groups indicated. Nodes with single black boxes indicate strong support in all analyses. Variable support across analyses is indicated by colored boxes as indicated by keys on left. Top key indicates respective analysis and dataset, Maximum Likelihood (ML) on left, Bayesian inference (BI) on right; 'Full' and 'Trimmed' indicate dataset; and lowest box indicates support recovered by divergence time estimate (DTE 'Trimmed') in BEAST2 with preferred prior models. Lower key indicates support by color for UltraFastBoot support (ML) and Posterior probability (PP) values. Inset depicts 'traili' branch expanded to show structure.

within this group were poorly supported by ML analyses (Figs. 2.2, Fig S2.1, S2.3, and S2.4).

Several undescribed species and a species attributable to *N. meizon* Guimaraes and Ferreira Jr.

2019 were found in a species group (hereafter '*meizon'* group; UFB = 87; PP > 0.99) sister to

the monophyletic 'N. *traili*' group (UFB = 100; PP > 0.99). Within the *traili* group,

specimens attributable to *N. gracilipes* Sharp 1882 and another undescribed putative species (*N.* sp.3) were found as successive sisters to the greater complex. The complex splits into two greater

clades, with additional supported relationships within these clades, but all with very short

internodes and no immediate geographic correlation. Specimens attributable to known species

(besides *N. traili* Sharp 1882) were found nested in various places among this complex, with

specimens attributable to *N. gracilipes* recovered as nonmonophyletic, in various clades. The

recovered topologies from reduced phylogenetic reconstructions reflect those of the full

alignment reconstruction, with similar support.

Divergence Time Estimation and Biogeographical Analyses

Bayes Factors calculated with marginal likelihoods recovered via Nested Sampling

analyses recovered the Relaxed Clock Log Normal clock model in tandem with the Yule tree

model (RCY; Ln = −26,295.78, SD = 5) as the preferred model combination for divergence time

estimation (Table 2.2). The combined MCC tree recovered by the two RCY runs were used in

the downstream BioGeoBears Analyses (Fig. 2.2). This recovered tree reflected the same

relationships recovered by ML and BI analyses above, except within the nanulus group, where

sisters to the clades well supported in the ML tree were recovered in different positions, but

without strong support (Figs. 2.1, S2.3–2.13). In BioGeoBears Analyses, the DEC and

DIVALIKE models recovered very similar ancestral reconstructions. Likelihoods of these

models were similar, though the DIVALIKE model was preferred in the AICc calculation, which

weighs model fit against the number of parameters, and notably by comparison of AICc model

weights, which are indicative of relative model probability based on this calculation (Table 2.4).

The likelihoods of the +j versions of these models were slightly higher for the respective

analyses, but AICc scores and AICc model weight comparisons still preferred models without

the j parameter (Table 2.4).

With the DIVALIKE BioGeoBears analysis, we recovered the crown age of

Notomicrinae in the early Cretaceous (110 Mya), with an ancestral range across a very large
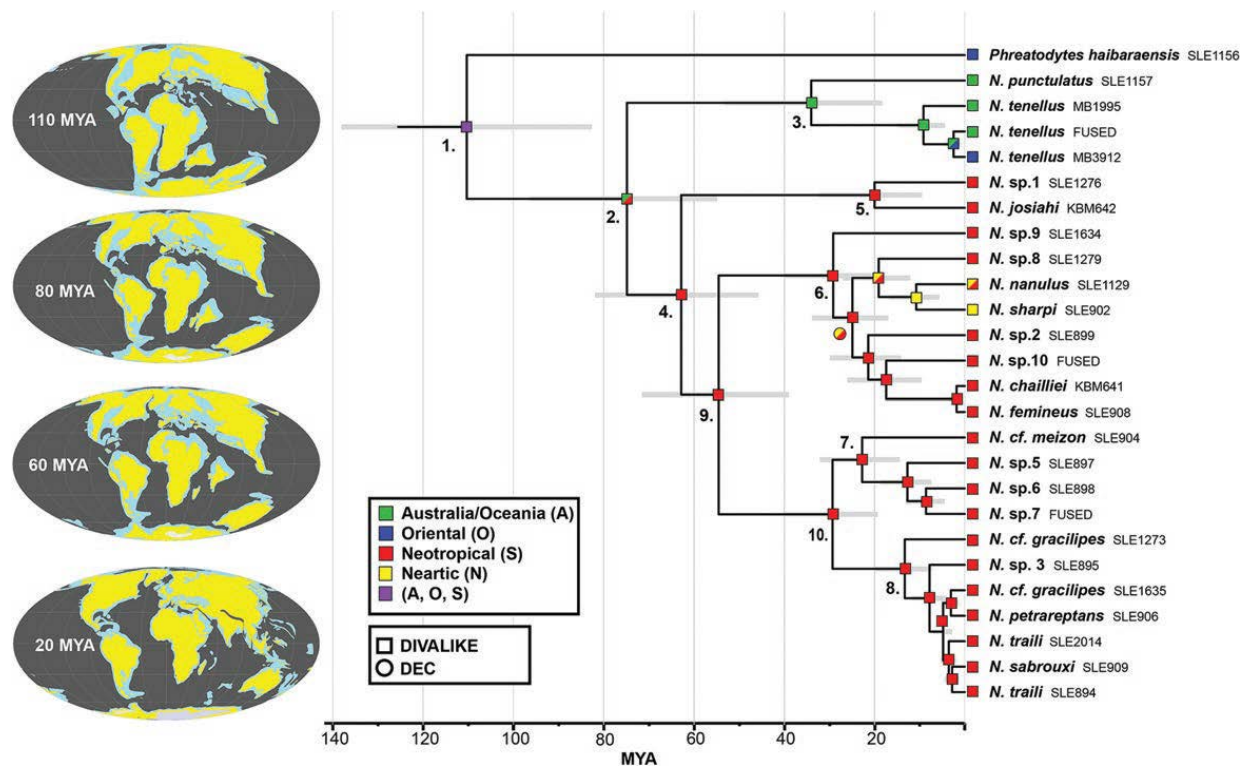
Figure 2.2. Divergence time estimate and ancestral range reconstruction of Notomicrinae. Ancestral ranges recovered by DIVALIKE analysis in BioGeoBears with preferred divergence time estimate recovered by BEAST 2. Boxes on nodes indicate most probable ancestral area states with color correlating to left key; circles indicate differing result recovered from DEC analysis. Numbers on nodes correlate to Table 3. Maps indicate continental landmasses at respective ages, reproduced from Cao et al. (2017) with G-Plates (Müller et al. 2018).

region including the South America, Oceania and Oriental regions. From this, the

Phreatodytes lineage eventually became restricted to the Oriental region, presently only known

from Japan, whereas the ancestor to all Notomicrus, Old and New World, inherited a distribution

including Australia and South America prior to the upper Cretaceous (Notomicrus crown age =

74.81 Mya). From there, near or after the KT boundary (ca. 75–63 Mya), Notomicrus separated

into two clades via apparent vicariance (Figs. S2.18–S2.19), one to inherit the Australian

(Oceania) area, the other inheriting South America. Following this split, the Old-World N.

tenellus group remained in Oceania, with the crown age of the N. punctulatus and N. tenellus

lineages ('tenellus' group) at the Eocene-Oligocene boundary (ca. 34 Mya; Table 2.4). Finally,

the group experienced a recent dispersal to the Oriental region during the upper Neogene (ca. 9–

2.5 Mya, Fig. 2), before dividing into the present-day lineages from Sumatra and Fiji. The New

World lineage diversified almost entirely in South America/Neotropics, with major clades

Table 2.3. Mean ages of divergence time estimates recovered by respective prior model combinations in BEAST2. 95% HPD intervals in parentheses beneath mean ages. Model combination preferred by Bayes Factors and used in downstream ancestral range reconstructions are in bold. Node numbers correlate to Fig. 2.

Node	Clade	**RC Yule**	RC BD	Strict Yule	Strict BD
1.	Notomicrinae	**110.28** (82.54,138.04)	108.13 (80.46, 138.09)	97.98 (78.50, 117.33)	97.62 (78.61, 117.81)
2.	*Notomicrus*	**74.81** (54.78, 96.42)	71.89 (51.72, 92.50)	66.97 (56.32, 77.54)	66.50 (56.13, 77.23)
3.	Old World *Notomicrus.* 'tenellus' group	**34.00** (18.22, 53.28)	31.92 (16.45, 49.29)	27.42 (21.93, 32.94)	27.1225 (21.67, 32.76)
4.	New World *Notomicrus*	**62.78** (45.65,81.94)	59.91 (43.18, 78.58)	66.97 (56.322, 77.55)	60.30 (50.73, 69.95)
5.	'*josiahi*' group	**19.63** (9.41, 32.30)	18.75 (9.09, 30.42)	18.05 (14.18, 21.97)	17.96 (13.00, 21.83)
6.	'*nanulus*' group	**29.13** (19.87, 39.91)	27.32 (18.24, 37.07)	25.82 (21.33, 30.50)	25.52 (21.04, 30.10)
7.	'*meizon*' group	**22.71** (14.34, 32.12)	21.18 (13.29, 29.86)	21.64 (17.52, 25.90)	21.39 (17.29, 25.65)
8.	'*traili*' group	**13.27** (7.14, 19.90)	12.40 (6.86, 18.69)	12.89 (10.02, 25.79)	12.70 (9.99, 15.71)
9.	'nanulus' + ('meizon','traili')	**54.49** (38.89, 71.52)	51.84 (36.85, 68.76)	55.40 (46.33, 64.43)	54.96 (46.23, 64.28)
10.	'meizon' + 'traili'	**29.26** (19.26, 40.39)	27.34 (17.95, 37.82)	26.22 (21.52, 31.04)	25.92 (21.22, 30.67)

experiencing ancestral lineage diversification during the Paleocene ('josiahi' + ['nanulus',

'meizon', 'traili'] crown age = ca. 63 Mya; 'nanulus' + ['meizon', 'traili'] crown age = ca. 54

Mya) and Oligocene ('meizon' – 'traili' crown age = 0.29 Mya) (Fig 2.2, Table 2.3).

Diversification in the Neotropics continued across these clades from the upper Paleogene

through the Neogene (ca. 29–2.5 Mya; Fig. 2.2). Within the *nanulus* group, there was an inferred

dispersal into North America during the early Miocene (Fig. 2.2), with the group subsequently

occupying North and South America before dividing into respective North and South American

lineages. *Notomicrus sharpi* and *N. nanulus* shared an ancestor restricted to North America

during the upper Miocene (ca. 20‑11 Mya), with present day *N. nanulus* remaining in North

America, and populations of *N. sharpi* dispersing back into South America/Neotropics. The DEC

model recovered identical ancestral ranges, except within the *nanulus* group, where dispersal to

North America was recovered to occur in an earlier common ancestor (Fig. 2.2), before dividing

into the ancestral ranges described above. See Table 2.3 for summary of means and 95% HPD of

major divergence times.

Discussion

Performance of Analyses

Both ML and BI phylogenetic analyses of the ‘Full’ and ‘Trimmed’ datasets,

recovered nearly identical topologies with high support. The difference in the number of subsets

in the partitioning schemes of the datasets was not surprising given that 1) taxon sampling,

including outgroups, was drastically reduced in the latter and consequently 2) so was the relative

amount of missing data (i.e., most of the taxa in the reduced dataset had representation for most

or all of the targeted genetic markers versus some taxa having only partial representation across

markers in the full dataset). Despite using the same starting tree across all time-calibrated

analyses, there were differences in topologies within the *nanulus* clade between the Uncorrelated

Relaxed clock analyses (RCY and RCBD (Supp Figs. 6‑9 [online only]) from other BI and ML

analyses (Figs. 2.1, S2.3–S2.5 and S2.10–S2.13). Time-calibrated analyses under a Strict clock

model (both Yule and BD) recovered identical topologies to uncalibrated BI analyses and ML

analyses, with high posterior support (Figs. S2.1‑S2.5 and S2.10–S2.13 [online only]). This is

presumed to be the effect of clock model choice, which is known to affect posterior probabilities in phylogenetic estimates, causing differences between best trees recovered by strict versus relaxed clocks (Drummond 2006). Clock model choice (RC vs Strict; Table 2) also appears to have had a more profound effect on recovered ages, with relatively negligible differences across tree models (Yule vs BD; Table 2.3, Figs. S2.6‑S2.13). The effect on ages is most notable in the deeper nodes of the trees, with more recent ages increasingly convergent at shallow levels. These suspected effects of clock model choice are not particularly surprising as variable influence of clock and tree prior models on phylogenetic and divergence time inference is known (e.g., Drummond et al. 2006, Brown and Yang 2011, Condamine et al. 2015). In any event, the conflicting relationships were recovered with low support. In addition, with the coarseness of our investigation, these differences in topology and ages may not have had a strong effect on downstream ancestral range reconstructions, and we are confident in our model choice given the resulting Bayes Factor preference (Table 2.2). Nested Sampling analyses appear to have performed well, with multiple replications yielding similar marginal likelihoods and standard deviation estimates. While marginal likelihood estimates for the StrictBD and RCBD model treatments did not significantly differ, these were both bettered by the RCY treatment used in ancestral range reconstruction.

BioGeoBears analyses under the DEC and DIVALIKE models produced similar results, with the only notable difference recovered in the *nanulus* group, where the DEC model recovered a range expansion to North America in an earlier ancestor than in the DIVALIKE reconstruction (Figs. S2.14‑S2.21). Similarly, the +*j* versions of the analyses varied little from those without the +*j* jump dispersal parameter. Analyses without +*j* were preferred by AICc and AICc weights. Following Ree and Sanmartin’s (2018) critique of model comparisons in

BioGeoBears, and especially the *j* parameter, it is potentially inappropriate to rely heavily on these metrics to select a preferred model. However, with DIVALIKE also being arguably the most intuitive and parsimonious reconstruction, there is little ground to reject the resulting preferred analysis for one of the others. Furthermore, given the overall similarity of reconstructions across models (Figs. 2.2, S2.14‑S2.21), the nuanced differences among models at shallower nodes are of little consequence to the primary hypothesis being tested: Gondwandan vicariance in *Notomicrus*.

Table 2.4. Biogeographic models used in ancestral range reconstruction in BioGeoBears, with recovered parameter values and metrics from model comparisons. LnL = log likelihood, # params = number of parameters, d = dispersal parameter, e = extinction parameter, j = founder (jump dispersal) parameter. Bold indicates best values for likelihood, Akaike Information Cri- terion corrected (AICc) score, and AICc model weight.

Model	LnL	# params	d	e	j	AICc	AICc weight
DEC	-16.72	2	0.0036	1.0e-12	0	37.97	0.16
DEC+J	-15.63	3	0.0019	1.0e-12	0.024	38.34	0.13
DIVALIKE	-15.67	2	0.0053	1.0e-12	0	**35.86**	**0.45**
DIVALIKE+J	**-14.93**	3	0.0026	1.0e-12	0.028	36.95	0.26

Systematics

Notomicrinae

The recovered topologies across all analyses were consistent with those recovered in previous molecular phylogenetic estimates incorporating relevant notomicrine sampling, including Sanger sequenced (Kato et al. 2010, Toussaint et al. 2016, Baca et al. 2017b) and Ultraconserved elements (Gustafson et al. 2020) datasets.While these previous studies did not sample extensively within Notomicrinae, they did recover *Phreatodytes* as sister to *Notomicrus* rather than sister to all other Noteridae, as has been suggested by some previous morphological studies (e.g., Belkaceme 1991; Beutel 1997; Miller et al. 2009; Dressler 2011; Beutel et al. 2006,

2020). The source of this conflict likely lies in the confounding effect of extreme morphological adaptation of *Phreatodytes* to stygobitic environments. Similar cases have been documented in diving beetles, where convergent evolution to subterranean environments has been shown to complicate the inference of relationships based on morphology alone (e.g., Leys et al. 2003, and see Miller and Bergsten 2016:13). As discussed by Baca et al. (2017b) and Beutel et al. (2020), the molecular reconstruction of the noterid phylogeny implies several secondary losses and character reversals in the family. Given the congruence of molecular datasets, and as suggested by Baca et al. (2017b), a reassessment of noterid morphology is warranted see Baca et al. (2017b) and Beutel et al. (2020) for more detailed discussions of these data source conflict.

The relationship of the cave-dwelling *Speonoterus* to other notomicrines remains untested with molecular data. The morphological similarity suggests *Speonoterus* is sister to *Notomicrus*, though it is also possible that *Speonoterus* is nested within *Notomicrus* (further discussed below). In any case, this relationship will be difficult to resolve without molecular data.

Notomicrus

Our results are congruent with Baca et al (2017b), recovering reciprocally monophyletic New and Old World clades of *Notomicrus*. Within the Old World *tenellus* group, our recovered relationships support the status of *N. punctulatus* as a valid species (see Toledo 2010), sister to *N. tenellus* Sensu *lato*. It has been suspected that the broadly distributed *N. tenellus*, with several junior synonyms and records spanning from mainland Malaysia, south to Australia, and east to French Polynesia (Toledo 2010, Nilsson 2011), likely encompasses multiple species (Toledo 2010). Our results support this, and perhaps the disparity in New versus Old World *Notomicrus* diversity is not as great as current classification indicates. However, our study＇s Old World sampling is too sparse to address the taxonomy of this group in the appropriate scope.

The New World *Notomicrus* relationships recovered in the full tree (Figs. 2.1;S2.1 and S2.2) supports the validity of most described species and reveals the extent of undescribed diversity. With samples selected on merits of diverging morphology, our analyses also grant support to our tentative species hypotheses. This will find particular utility in revising New World members of *Notomicrus* (in prep). In that respect, the phylogeny also shows that certain morphological characters are indeed diagnostic and predictive of relationships, especially among major lineages. Male tarsal claws, genitalia, and microreticulation were especially indicative, with the ability to also diagnose most species, as shown previously (Young 1978, Manuel 2015, Baca and Short 2018, Guimaraes and Ferreira Jr 2019). That said, a comprehensive review of *Notomicrus* morphology is also beyond the scope of this paper and will be the subject of forthcoming revisions of the group.

The recovered trees of the 'Full' dataset support the status of most described species and several tentative new ones, but they depict a large complex within the *traili* group (Fig. 2.1). The successive sisters to the *traili* complex (a species attributable to *N. gracilipes* and an undescribed species, *N.* sp.3) are much more clearly demarcated genetically. Within the complex we recover genetic structure, with generally well- to moderately supported larger clusters with short internodes. The morphological variation is subtle among members of the group, e.g., in the genitalia or pattern of punctation (see Young 1978, Manuel 2015, Baca and Short 2018), and identifying any outstanding morphological signal among the supported clusters will require close investigation. This difficulty is here exemplified in the scattered placement of specimens attributable to *N. gracilipes* following Young's (1978) description and comparison to the holotype. In identification, particular note was paid to patterns of punctation, which here indicates that some diagnostic characters may be unreliable within the *traili* group. Recovered

relationships indicate that the *traili* complex likely contains multiple species, with the complex experiencing repeated diversification and range expansion in the Neotropics. This is evidenced by the fact that individuals within the clusters range variably across South America, and in many cases, members of these clusters occur sympatrically, or at least in close enough proximity to assume contact, but maintain genetic and phylogenetic distinction. This complex will require further investigation before taxonomic action is taken. Given the issues of phylogenetic discordance at short internodes (Degnan and Rosenberg 2006, Liu and Edwards 2009), a phylogenomic approach, e.g., with Ultraconserved elements, may be the most appropriate for parsing relationships within the *traili* complex.

Historical Biogeography

The ancestral range reconstruction suggests that notomicrines occupied all included areas except North America during or prior to the Cretaceous (ca. 110 Mya). Notomicrines were inferred to be present in the Oriental Region, which at the time (110 Mya) was part of greater Eurasia (Seton et al. 2012). With that, and the currently available data limited to extant taxa, it is possible they were more widespread than their present-day distributions would indicate. At that time, Africa would have only just begun to separate from South America (Seton et al. 2012, Toussaint et al. 2017b), with Antarctica still linking South America and Australia (Fig. 2.2; Seton et al. 2012). The presence of Notomicrinae in the Oriental region would suggest either a previous long dispersal event, or more likely, a distribution including Africa and western Eurasia (present day Palearctic), with subsequent local extinction.

Both DIVALIKE and DEC reconstructions recovered the *Notomicrus* ancestral range in the upper Cretaceous (up to ca. 75 Mya) to include Australia/Oceania and South America, before splitting via apparent vicariance (Figs. 2.2; S2.14-S2.21). Given that Antarctica linked these two

continents, ancestral *Notomicrus* also likely inhabited Antarctica before the continued breakup of Gondwana more completely isolated these lineage

s in the early to mid-Paleogene. Such a pattern is seen in other organisms (e.g., Bukontaite et al. 2014, Kim and Farrell 2015, Toussaint and Gillet 2018, Gustafson and Miller 2017, Toussaint and Short 2017), and as our reconstructions suggest (Figs. 2.2; Figs. 2.14‑2.21), this is consistent with vicariance via the breakup of Gondwana.

The sister relationship between the New Caledonia endemic *N. punctulatus* and the *N. tenellus* complex is of interest with respect to South Pacific biogeography. A formerly popular hypothesis stated that New Caledonia acted as an 80-Mya-old refuge to many lineages after Zealandia split from Australia in the Cretaceous (Toussaint et al. (2017c) and citations therein). However, there is compelling geological evidence that suggests that Zealandia was completely submerged during the Oligocene, for up to 20 Mya, until the Eocene-Oligocene boundary (37 Mya), and that the ancestors to New Caledonian endemics dispersed to the island after its emergence (Cruaud et al. 2012, Grandcolas 2017, Condamine et al. 2017, Toussaint et al. 2017c, Vicente et al. 2017, but see Heads 2019). Our estimated divergence of *N. punctulatus* from *N. tenellus* (34.5 Mya) is more in agreement with this latter hypothesis, though finer-scale further study is needed to test this in Notomicrinae. New World *Notomicrus* was inferred to have diversified almost exclusively in South America/Neotropics, which began soon after their separation from the ancestral Old World lineage (Figs. 2.2; 2.14‑S2.21). The variety of ecological preferences in *Notomicrus* and dynamic paleogeography of South America (Hoorn and Wessenlingh, 2011) offer several modes by which this diversification could have occurred (e.g., Chazot et al. 2018, Gillett require a more fine-scale investigation. Only one lineage, that within the *N. nanulus* clade, was inferred to have expanded its range into North America, before

spitting into respective South and North American lineages. *Notomicrus sharpi* has since expanded back into the Neotropics following a split with *N. nanulus*. Taken alone, *N. sharpi* is a species of phylogeographic interest given its wide distribution from northern South America, through the Antilles and Central America into North America. The coarseness of our analyses, further confounded by the conflict in timing of dispersal to North America between the DIVALIKE and DEC model, prevents us from speculating on the route by which these dispersals occurred.

Ecological Evolution

Phreatodytes is arguably the most ecologically specialized group of Notomicrinae (and Noteridae), inhabiting subterranean aquifers, and is sister to *Notomicrus* with common ancestry terminating in the Cretaceous (110 Mya). However, our recovered tree shows strong habitat variation among species of *Notomicrus*, often despite relatively recent common ancestry, i.e., habitat evolution within a lineage is not relegated to ancient evolutionary events, but may be quite derived. A prime example is that of *N. petrareptans*, endemic to Suriname and specific to hygropetric seeps. This transition occurred very recently, within the past 3 Mya, from the *traili* complex which largely inhabits lentic habitats (e.g., detrital pools in forests or nonflowing stream beds, though occasionally lotic habitats). Another example is the undescribed *N.* sp.2 of the *nanulus* group. While some specimens attributable to this species are found in lentic forest pools, the largest series of this species are from dry leaf litter samples taken via Winkler traps, and notably not from previously submerged areas, e.g., dry stream beds (pers. obs.). This species exhibits morphological characters than might be attributable to this shift, such as shorter appendages, reduction in eye size, and reduction of natatory setae. Finally, and also belonging to

the *nanulus* clade which is found almost exclusively in lentic habitats, is the undescribed N. sp.8, which is known only from lotic environments (pers. obs.).

In aggregate with our analyses, these ecological observations show that habitat transitions are common in Notomicrinae, and can happen relatively rapidly. This may also explain the close morphological affinity of *Speonoterus* to *Notomicrus*. It is possible that this genus made the transition to caves more recently than might be assumed, perhaps even after the New World–Old World split (ca. 75 Mya; Fig. 2.2). This possibility is further supported by Leys et al. (2003), who showed that the many transitions to the arguably more extreme stygobitic habitats occurred very recently (5‑10 Mya) within diving beetles. In that vein, a 40-Mya gap exists between the separation of the New and Old World clades and subsequent diversification in the *tenellus* group (34 Mya), leaving much of the evolutionary history in this region shrouded in mystery. This pattern is likely the effect of coarse sampling and extinction, but with such rapid ecological evolution evidenced in the group and morphological similarity with *Notomicrus* it is reasonable to suspect that *Speonoterus* is actually a derived *Notomicrus* rather than its assumed sister. This hypothesis, however, requires formal phylogenetic testing.

Conclusions

We have reconstructed the first species-level phylogeny within the family Noteridae. We found support for the current circumscription of the Notomicrinae, which includes *Notomicrus*, *Phreatodytes*, and presumably *Speonoterus*, though samples of the latter were unavailable for our investigation. We revealed that *N. tenellus* as currently delimited is a composite of numerous Old World species. Further, we identified substantial undescribed diversity in the New World and a large species complex in the *N. traili* group that requires further resolution. Our divergence time estimates coupled with ancestral range reconstruction reveal an early Cretaceous origin of

Notomicrinae followed by signatures of Gondwanan vicariance during the late Cretaceous. And finally, the recovered relationships indicate a high amount of ecological plasticity to be investigated. In sum, our current study provides a robust scaffold on which further investigations of Notomicrinae may build. With this we show again that Noteridae provides a model system in which to test widely applicable evolutionary hypotheses.

Acknowledgments

Data Availability

All sequence data used in this study has been deposited in GenBank. Newly generated data are available under GenBank accession numbers MW035848‑MW035912, MW041925‑MW041967 and MW043797‑MW043864. See Supp Table 1 (online only) for full list of samples and corresponding accession numbers.

References

Baca, S. M., & Short, A. E. Z. (2018). Notomicrus petrareptans sp. n., a new seep-dwelling species of Noteridae from Suriname (Coleoptera: Adephaga). *Zootaxa*, 4388(2), 182–190.

Baca, S. M., & Toledo, M. (2015). Canthysellus Baca and Toledo (Coleoptera: Noteridae: Noterini), a new genus of burrowing water beetle from South America. *Coleopterists Bulletin*, 69, 477–488.

Baca, S. M., Alexander, A., Gustafson, G. T. & Short A. E. Z. (2017a). Ultraconserved elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly of 'Hydradephaga'. *Systematic Entomology*, 42, 786–795.

Baca, S. M., Toussaint, E. F. A., Miller, K. B., & Short A. E. Z. (2017b). Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on partitioning strategies. *Molecular Phylogenetics and Evolution*, 107, 282–292.

Balfour-Browne, J. (1939). A contribution to the study of the Dytiscidae—I (Coleoptera, Adephaga). *The Annals and Magazine of Natural History*, 3(11), 97–114.

Belkaceme, T. (1991). Skelet und muskulatur des kopfes und thorax von Noterus laevis Sturm.ein beitrag zur morphologie und phylogenie der Noteridae (Coleoptera: Adephaga). *Stuttgarter Beiträge zur Naturkunde* A, 462, 1–94.

Beutel, R. G. (1997). ber Phylogenese und evolution der Coleoptera (Insecta). Abhandlungen des Naturwissenschaftlichen, Goecke & Evers, Vereins in Hamburg, Germany, 31, 1–164.

Beutel, R. G., Balke, M., & Steiner, W. E. (2006). The systematic position of Meruidae (Coleoptera, Adephaga) and the phylogeny of the smaller aquatic adephagan beetle families. *Cladistics*, 22, 122–131.

Beutel, R. G., Ribera, I. G., Fikáček, M., Vasilikopoulos, A., Misof, B., & Balke, M. (2020). The morphological evolution of the Adephaga (Coleoptera). *Systematic Entomology*, 45(2), 378–395.

Bouckaert, R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17(1), 42.

Bouckaert, R., Vaughan, T., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., & De Maio, N. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4): e1006650.

Brown, B. P., & Yang, Z. (2011). Rate variation and estimation of divergence times using strict and relaxed clocks. *BMC Evolutionary Biology*, 11(1), 271.

Bukontaite, R., Miller, K. B., &Bergsten, J. 2014. The utility of CAD in recovering Gondwanan vicariance events and the evolutionary history of Aciliini (Coleoptera: Dytiscidae). *BMC Evolutionary Biology*, 14(1), 1–18.

Cao, W., Zahirovic, S., Flament, N., Williams S., Golonka, J., & Müller, R. D. (2017). Improving global paleogeography since the late Paleozoic using paleobiology. *Biogeosciences*, 14(23), 5425–5439.

Chazot, N., De-Silva, D. L., Willmott, K. R., Freitas, A. V., Lamas, G., Mallet, J., Giraldo, C. E., Uribe, S., & Elias, M. (2018). Contrasting patterns of Andean diversification among three diverse clades of Neotropical clearwing butterflies. *Ecology and Evolution*, 8(8), 3965–3982.

Clark, H. (1863). Descriptions of new East-Asiatic species of Haliplidae and Hydroporidae. *The Royal Entomological Society London*, 1(3), 417–428.

Colgan, D. J., McLauchlan, A., Wilson, G. D. F., Livingston, S. P., Edgecombe, G. D., Macaranas, J., Cassis, G., & Gray M. R. (1998). Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Australian Journal Zoology*, 46, 419–437.

Condamine, F. L., Nagalingum, N. S., Marshall, C. R., & Morlon, H. (2015). Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in Bayesian molecular dating. *BMC Evolutionary Biology*, 15, 65.

Condamine, F. L., Leslie, A. B., & Antonelli, A. (2017). Ancient islands acted as refugia and pumps for conifer diversity. *Cladistics*. 33(1), 69–92.

Cruaud, A., Jabbour-Zahab, R., Genson, G., Ungricht, S. & Rasplus, J. Y. (2012). Testing the emergence of New Caledonia: fig wasp mutualism as a case study and a review of evidence. *PLoS One*, 7(2), e30941.

Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., Huber, J., Polaszek, A. & Rasplus. J. Y. (2019). Optimized DNA extraction and library preparation for minute arthropods: application to target enrichment in chalcid wasps used for biocontrol. *Molecular Ecology Resources*, 19(3), 702–710.

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

Dressler, C., Ge, S.-Q., & Beutel, R. G. (2011). Is Meru a specialized noterid (Coleoptera, Adephaga)? *Systematic Entomology*, *36*(4), 705–712. https://doi.org/10.1111/j.1365-3113.2011.00585.x

Drummond, A. J., Ho, S. Y., Phillips, M. J. & A. Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4(5), e88.

Fauvel, A. (1903). Faune analytique des coléoptères de la Nouvelle-Calédonie. *Revue d'Entomologie*, 22, 203–379.

Gillett, C. P., & Toussaint, E. F. (2020). Macroevolution and shifts in the feeding biology of the New World scarab beetle tribe Phanaeini (Coleoptera: Scarabaeidae: Scarabaeinae). Biological Journal of the Linnean Society, 130(4), 661–682.

Gómez, R. A., & Miller, K. B. (2013). *Prionohydrus*, a new genus of Noterini Thomson (Coleoptera: Noteridae) from South America with three new species and its phylogenetic considerations. *Annals of the Entomological Society of America*, 106, 1–4.

Grandcolas, P. (2017). Ten false ideas about New Caledonia biogeography. *Cladistics*. 33(5), 481–487.

Guimarães, B. A., & Ferreira Jr., N. (2019). Two new species and new records of Notomicrus Sharp, 1882 (Coleoptera: Noteridae) from Brazil. *Zootaxa*. 4629(2): 263–270.

Gustafson, G. T., & Miller K. B. (2017). Systematics and evolution of the whirligig beetle tribe Dineutini (Coleoptera: Gyrinidae: Gyrininae). *Zoological Journal of the Linnean Society*, 181(1), 118–150.

Gustafson, G. T., Baca, S. M., Alexander, A. M. & Short. A. E. Z. (2020). Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Systematic Entomology*, 45, 552–570.

Heads, M. (2019). Recent advances in New Caledonian biogeography. *Biological Reviews*. 94(3), 957–980.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522.

Hoorn, C., & Wesselingh, F. (2011). Amazonia: landscape and species evolution: a look into the past. John Wiley and Sons, Hoboken, NJ, USA.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773–795.

Kato, M., Kawakita, A. & Kato, T. (2010). Colonization to aquifers and adaptations to subterranean interstitial life by a water beetle clade (Noteridae) with description of a new Phreatodytes species. *Zoological Science*, 27, 717–722.

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kim, S. I., & Farrell, B. D. (2015). Phylogeny of world stag beetles (Coleoptera: Lucanidae) reveals a Gondwanan origin of Darwin's stag beetle. *Molecular Phylogenetics and Evolution 86*, 35–48.

Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution 29(6)*, 1695–1701.

LeConte, J. L. (1863). New species of North American Coleoptera. Part 1. Smithsonian Miscellaneous Collect, 6(167), 1–92.

Leys, R., Watts, C. H., Cooper, S. J. & Humphreys, W. F. (2003). Evolution of subterranean diving beetles (Coleoptera: Dytiscidae Hydroporini, Bidessini) in the arid zone of Australia. *Evolution. 57(12)*, 2819–2834.

Liu, L., & Edwards, S. V. (2009). Phylogenetic analysis in the anomaly zone. *Systematic Biology*, 58(4), 452–460.

Maddison, D. R. (2008). Systematics of the North American beetle subgenus Pseudoperyphus (Coleoptera: Carabidae: Bembidion) based on morphological, chromosomal and molecular data. *Annals of the Carnegie Museum*, 77, 147–193.

Manuel, M. (2015). The genus Notomicrus in Guadeloupe, with description of three new species (Coleoptera: Noteridae). *Zootaxa*, 4018 (4), 506–534.

Maturana Russel, P., Brewer, B. J., Klaere, S., & Bouckaert, R. R. (2019). Model selection and parameter inference in phylogenetics using Nested Sampling. *Systematic Biology*, 68(2): 219–233.

Matzke, N. J. (2013). Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing. *Frontiers of Biogeography*, 5(4), 242–248.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, 63(6), 951–970.

Matzke, N. J. (2018). BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts. version 1.1.1, published on GitHub. https://github.com/nmatzke/BioGeoBEARS. Accessed 24.vii.2020.

Miller, K. B. (2009). On the systematics of Noteridae (Coleoptera: Adephaga: Hydradephaga): phylogeny, description of a new tribe, genus and species, and survey of female genital morphology. *Systematics and Biodiversity*, 7, 191–214.

Miller, K. B. (2013). Notomicrus josiahi, a new species of Noteridae (Coleoptera) from Venezuela. *Zootaxa*, 3609(2), 243–247.

Miller, K. B., & Bergsten., J. (2016). Diving beetles of the world: systematics and biology of the Dytiscidae. John Hopkins University Press, Baltimore, MD.

Miller, M. A., Pfeiffer, W. & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees, pp. 1–8. In 2010 gateway computing environments workshop (GCE), 14 November 2010, Institute of Electrical and Electronics Engineers, New Orleans, LA.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A. & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*. 37(5): 1530–1534.

Müller, R. Cannon, D., J. Qin, X., Watson, R. J., Gurnis, M. Williams, S., Pfaffelmoser, T., Seton, M., Russell, S. H. J. & Zahirovic, S. (2018). GPlates: building a virtual Earth through deep time. *Geochemistry Geophysics Geosystems*. 19(7): 2243–2261.

Nilsson, A. N. (2011). A World Catalogue of the family Noteridae, or the Burrowing water beetles (Coleoptera: Adephaga). Version 16.VIII.2011. 1–54 pp. Downloaded from: http://www2.emg.umu.se/projects/biginst/andersn/WCN/WCN_20110816.pdf. Accessed 29.i.2016.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.Rproject.org/

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*. 67(5): 901.

Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal Biogeography*. 45(4). 741–749.

Sánchez-Herrera, M., Beatty, C. D., Nunes, R. Salazar, C. & Ware, J. L. (2020). An exploration of the complex biogeographical history of the Neotropical banner-wing damselflies (Odonata: Polythoridae). *BMC Evolutionary Biology*. 20(1): 1–14.

Seton, M., R. Müller, S. Zahirovic, C. Gaina, T. Torsvik, G. Shephard, A. Talsma, M. Gurnis, M. Turner & Chandler, M. (2012). Global continental and ocean basin reconstructions since 200 Ma. *Earth Science Reviews*. 113: 212–270.

Sharp, D. (1882). On aquatic carnivorous Coleoptera or Dytiscidae. *Scientific Transactions of the Royal Dublin Society*. 2: 179–1003.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*. 87: 651–701.

Skilling, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Analysis*. 1(4): 833–859.

Spangler, P. J. (1996). Four new stygobiontic beetles (Coleoptera: Dytiscidae; Noteridae; Elmidae). *Insecta Mundi*. 10: 241–259.

Toledo, M. (2010). Noteridae: Review of the species occurring east of the Wallace line (Coleoptera), pp. 195–236. In M. A. Jäch, and M. Balke (eds.), Water Beetles of New Caledonia (part1): monographs of Coleoptera Vol. 3. Zoologische-Botanische Gesellshaft, Vienna, AT.

Toussaint, E. F., & Gillett, C. P. (2018). Rekindling Jeannel's Gondwanan vision? Phylogenetics and evolution of Carabinae with a focus on Calosoma caterpillar hunter beetles. *Biological Journal of the Linnean Society*, 123(1): 191–207.

Toussaint, E. F., & Short, A.E. (2017). Biogeographic mirages? Molecular evidence for dispersal-driven evolution in Hydrobiusini water scavenger beetles. *Systematic Entomology*. 42(4): 692–702.

Toussaint, E. F. A., Beutel, R. G., Morinière, J. Jia, F., Xu, S., Michat, M. C., Zhou, X., Bilton, D. T., Ribera, I., Hájek, J., & Balke, M. (2016). Molecular phylogeny of the highly disjunct cliff water beetles from South Africa and China (Coleoptera: Aspidytidae). *Zoological. Journal of the Linnean Society*. 176(3): 537–546.

Toussaint, E. F. A., Bloom, D. & A. E. Z. Short. (2017a). Cretaceous West Gondwana vicariance shaped giant water scavenger beetle biogeography. *Journal of Biogeography*. 44(9): 1952–1965.

Toussaint, E. F. A., Seidel, A., Arriaga-Varela, E., Hájek, J., Kral, D., Sekerka, L., & Fikáček, M. (2017b). The peril of dating beetles. Syst. Entomol. 42(1): 1–10. Toussaint, E. F. A., R. Tänzler, M. Balke & Riedel, A. (2017c). Transoceanic origin of microendemic and flightless New Caledonian weevils. *Royal Society Open Science*. 4(6): 160546.

Uéno, S.-I. (1957). Blind aquatic beetles of Japan, with some accounts of the fauna of Japanese subterranean waters. *Archiv fuer Hydrobiologie Stuttgart*. 53: 250–296.

Uéno, S.-I. (1996). New phreatobiontic beetles (Coleoptera, Phreatodytidae and Dytiscidae) from Japan. *Journal of the Speleological Society Japan*. 21: 1–50.

Van der Auwera, G., Chapelle, S., & De Wächter, R. (1994). Structure of the large ribosomal subunit RNA of Phytophthora megasperma, and phylogeny of oomycetes. *FEBS letters*. 338: 133–136.

Vicente, N., Kergoat, G. J., Dong, J., Yotoko, K., Legendre, F., Nattier, R. & Robillard T. (2017). In and out of the Neotropics: historical biogeography of Eneopterinae crickets. *Journal of Biogeography*. 44(10): 2199–2210.

Wild, A. L. & Maddison, D. R. (2008). Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. *Molecular Phylogenetics and Evolution*. 48(3): 877–891.

Young, F. N. (1978). The New World species of the water-beetle genus Notomicrus (Noteridae). *Systematic Entomology* 3(3): 285–293.

Chapter 3: Review of the New World *Notomicrus* Sharp (Coleoptera: Noteridae) I: Circumscription of species groups and review of the *josiahi* group with description of a new species from Brazil

Stephen M. Baca and Andrew Edward Z. Short

Abstract

The New World species of the minute aquatic beetle genus *Notomicrus* Sharp compose a much greater diversity than their Old World congeners, with 14 of the 17 known *Notomicrus* species occurring in the Neotropics. A recent phylogenetic study recovered four primary New World species groups and found that there are a number of undescribed species across all of these main lineages. Here, we provide a taxonomic key to these New World species groups, including two described species that we currently do not place in any group ("*incertae sedis*" species), complete with images and illustrations of diagnostic characters and taxonomic notes including a list of known species in each group. This work provides a scaffold for further planned taxonomic revisions within the genus. In addition, we review the first of the four New World groups, the *josiahi* species group and describe one new taxon, *N. interstinctus* sp. n. from northern Brazil. Provided are descriptions, habitus images and illustrations of diagnostic characters.

Key words

aquatic beetles, taxonomy, new species, South America, Brazil.

Introduction

Notomicrus Sharp is the most speciose genus of the minute aquatic beetle subfamily Notomicrinae (Coleoptera: Noteridae). Its distribution spans Indomalaya, Oceania and the New World, though the majority of *Notomicrus* diversity occurs in the Neotropics (14 of 17 described species). *Notomicrus* species occupy a wide range of habitats, including the margins of ponds, streams, marshes and swamps, drying stream beds, forest pools, hygropetric habitats and terrestrial leaf litter. Some species present a high specificity in their habitat preference, while

others are found to be more generalists (Baca and Short 2020; personal observation). This ecological plasticity is a quality of the subfamily Notomicrinae as a whole. Both of the other notomicrine genera are subterranean specialists: *Phreatodytes* Uéno from aquifers in Japan and *Speonoterus* Spangler, a monotypic genus known only from a single collection in a shallow cave in Indonesia. *Speonoterus* appears to be a very close relative of *Notomicrus*, with *Phreatodytes* being their sister-group (Baca et al. 2017; Baca and Short 2020). It has been speculated that, given the plasticity of the habit preferences of *Notomicrus* and the aforementioned morphological similarity, *Speonoterus* may represent a specialized *Notomicrus* species (Baca and Short 2020). These relationships remain to be tested with molecular sequence data as *Speonoterus* is known only from the few specimens of the type series (Spangler 1996).

The taxonomic history of *Notomicrus* is uncomplicated at the genus level. *Notomicrus nanulus* (LeConte, 1863) and *Notomicrus tenellus* (Clark, 1863) are the only species which were described before the establishment of the genus by Sharp (1882a). Classification at this level has since been stable.

The monophyly of *Notomicrus* has been previously supported, with the Old and New World clades each also being found to represent reciprocally monophyletic lineages (Baca et al. 2017; Baca and Short 2020). These studies have also revealed that, unsurprisingly, there remain many undescribed species in the genus, especially in South America. This is further indicated by recent descriptions of new Neotropical species (Miller 2013, Manuel 2015; Baca & Short 2018; Guimarães & Ferreira-Jr 2019). Together, these works have greatly strengthened our understanding of notomicrine diversity and evolution. However, in effect, Young's (1978) benchmark revision of the New World *Notomicrus* now includes just over half of the currently

described species, amplifying the need for a comprehensive treatment of the group, especially with more diversity remaining to be described.

The species-level phylogenetic reconstruction of Baca and Short (2020 placed heavy emphasis on New World *Notomicrus* diversity. They recovered New World *Notomicrus* as diverging into four clades, the *josiahi*, *nanulus*, *meizon* and *traili* species groups, reciprocally supported by morphological characters. As such, Baca and Short (2020) provide an appropriate scaffold for taxonomic treatment of the groups.

Here, we (1) diagnose and provide a taxonomic key to the four primary species groups of New World *Notomicrus*. As part of this objective, we review morphological characters of importance, illustrate diagnostic characters and provide habitus images of exemplar species, taxonomic notes and a list of known species and references for each group. We then (2) present the first of four species-level revisionary works of New World *Notomicrus* by reviewing the *josiahi* species group. Included are a diagnosis of the group, a re-description of *N. josiahi* Miller, 2013 and a description of a new species from Brazil.

Materials and methods

Observations and measurements

Specimens were observed and measured using an Olympus SZX7 stereomicroscope. The microscope was equipped with 10x eyepieces, a DF PL 2x-4 objective (16–112x magnification) and a calibrated ocular micrometer. Genitalia and tarsal claws were relaxed in hot water and dissected. Dissections were placed in glycerine on glass slides for observation. For additional observations and images of the prosternal process, aedeagi and tarsal claws, selected specimens were cleared in a warm 10% potassium hydroxide (KOH) solution and periodically checked

multiple times an hour. Once desired elimination of soft tissue was achieved, specimens were thoroughly rinsed in DI (deionized) water. In some cases, DNA voucher specimens were used for observation and imaging of structures as the lysing process also dissolves soft tissue, effectively clearing the specimen and negating the need to damage additional specimens.
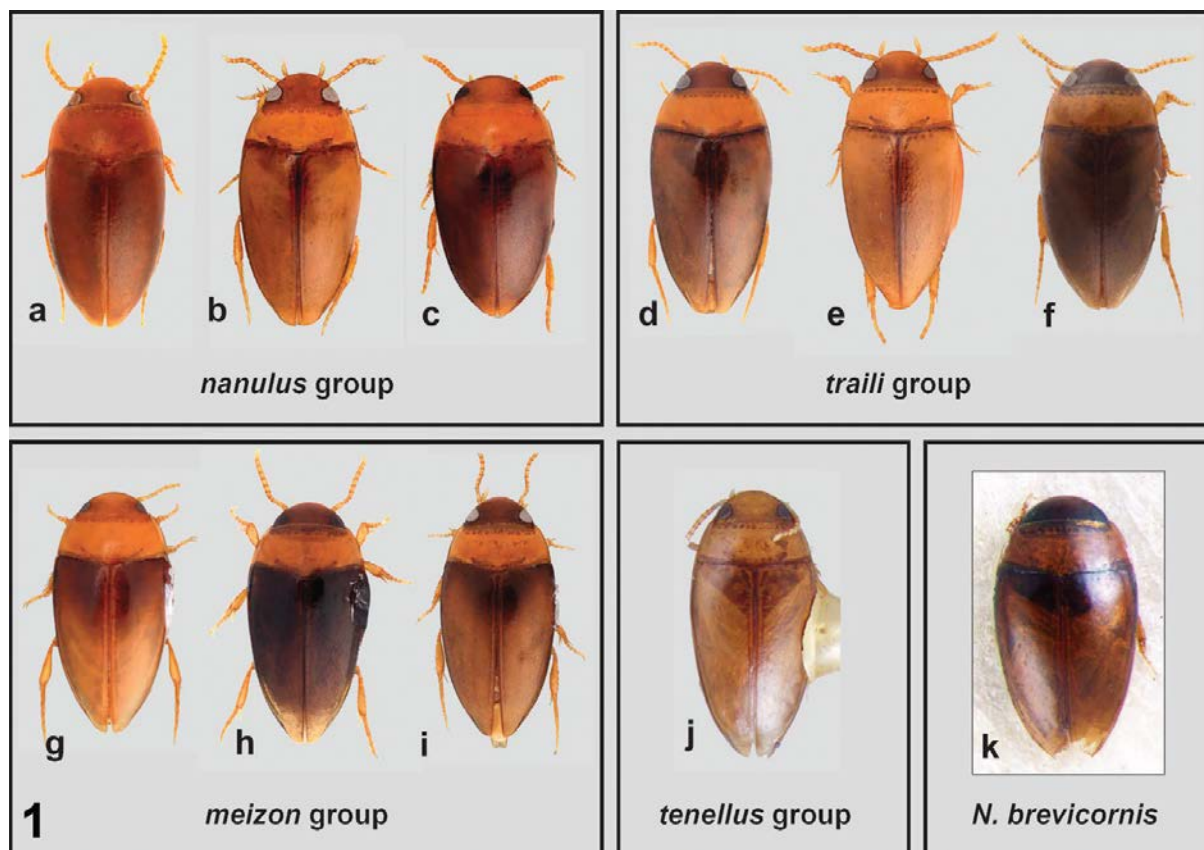
Figure 3.1 *Notomicrus* species groups. Dorsal habitus of representative *Notomicrus* species: **a** *N. nanulus*, **b** *N. sharpi*, **c** *N.* sp., **d** *N.* cf. *traili*, **e** *N.* cf. *gracilipes*, **f** *N. petrareptans*, **g** *N.* sp., **h** *N.* sp nr. *malkini*, **I**, *N.* sp. nr. *meizon*, **j** *N. tenellus*, **k** *N. brevicornis* male syntype.

Images and illustrations

Dorsal habitus images were obtained with a Visionary Digital microphotography system equipped with an Infinity K2 microscope using a 5X objective and Helicon Focus imaging software. Photos were aligned and stacked using CombineZP (www. hadleyweb.pwp.blueyonder.co.uk) and refined in Adobe Photoshop. Ventral images and images of structures to be used for illustrations were taken with an Olympus DP72 camera system

attached to either an Olympus SZX16 stereomicroscope with an SDF PLAPO 1xPF or 2xPF objective or an Olympus BX51 compound microscope with an UPlanFLN 40x oil immersion objective. The digital images were then stacked as above, with structures traced using Adobe Illustrator. Prolegs, prosterna, noterid platforms and male genitalia were imaged with the aforementioned stereomicroscope imagining system; illustrations were traced from these images. Male genitalia were placed in a depression slide with a drop of KY jelly and the remainder of the depression was filled with ethanol (EtOH). The KY jelly maintains its viscosity so that genitalia will hold its position for imaging. The EtOH eliminates obscuring refraction. Tarsal claws were imaged on the compound microscope. The fifth (V) pro- and metatarsomeres with tarsal claws were placed on a flat slide with EtOH and a cover slip was applied and glycerine was then used the seal the outside of the slip. The lower surface tension of the EtOH allows the cover slip to press on the claws, flattening them against the slide.

Terminology

Descriptive terminology follows previous works (e.g. Manuel 2015; Baca and Short 2018). *Noterid platform*. In *Notomicrus*, the noterid platform is formed by the raised projections of the inner metacoxal lamellae.
Genitalia and appendages. Following Miller and Nilsson (2003), genitalia and appendages are described in their fundamental homologous positions.

List of Depositories

INPA: Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil (N. Hamada)
MIZA: Museo del Instituto de Zoología Agrícola, Maracay, Venezuela (L. Joly)

MSB: Museum of Southwestern Biology, University of New Mexico (K. Miller)

NHM: Natural History Museum, London, UK (M. Barclay, C. Taylor)

SEMC: Snow Entomological Collection, University of Kansas, Lawrence, KS (A. Short)

USNM: U.S. National Museum of Natural History, Smithsonian Institution, Washington, DC (C.

Micheli).
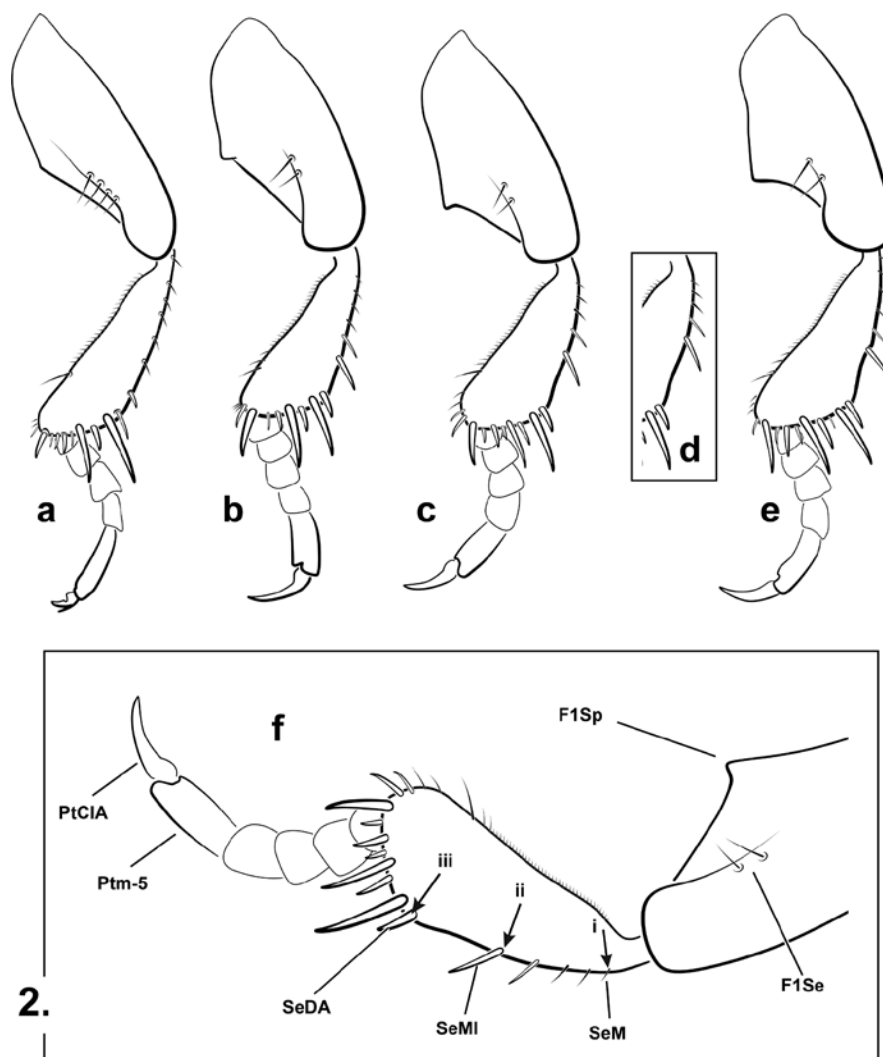
Figure 3.2. Representative prolegs of *Notomicrus* species groups (left proleg, anterior aspect); **a** *josiahi* group (*N. josiahi*); **b** *nanulus* group (*N. nanulus*); **c** *meizon* group (*Notomicrus* sp.); **d** *meizon* group, alternative setal spacing of dorsal (outer) protibial margin (*Notomicrus* sp.), **e** *traili* group (*N.* cf. *traili*); **f** detail of structures of importance. F1Se = setae of anteroventral margin of profemur; F1Sp = protuberance of posteroventral margin of profemur; PtCA = anterior protarsal claw; Ptm-5 = protarsomere V; SeDA = first robust seta of dorsoapical angle; SeMl = robust seta at mid-length of anterodorsal margin of protibia; SeM = First seta of anterodorsal margin of protibia; arrows indicate points for relative lengths (see key): i = anteroapical angle, ii = robust seta at mid-length, iii = first seta of marginal row.

Structures of Taxonomic Importance for diagnoses of *Notomicrus* species

   Size

The total body length of *Notomicrus* species ranges between ca. 1.0 mm and 1.8 mm. Following

Young (1978), size can, in combination with other characters, be very helpful in species

determination. Size is quantified in terms of (1) total length (TL), as measured from anterior

margin of head to apex of elytra, in dorsal aspect, (2) total length without head (TLPn), measured

from medial anterior margin of pronotum to elytral apex (this is included to provide a consistent

length measurement, as the degree to which the head is ventrally reflexed can affect the TL

measurement) and (3) greatest width (GW), as measured transversely at the widest point of the

beetle. Means of the measurements for each species, with standard deviations (SD) of the mean

are also presented. Ratios of TL and GW are given as a way of quantifying the shape of the body

outline. Means of the measurements for each species, with standard deviations (SD) of the mean

for TL are also presented. Ratios of TL and GW are given as a way of quantifying the shape of

the body outline.

   Color

Most species of *Notomicrus* present dorsal coloration as varying shades from brown to yellow.

However, individuals of some species present specific color patterns among sclerites. For

example, some species appear bicolorous, with the elytra and head darker brown and contrasting

against a lighter colored pronotum, for example, *N. traili* Sharp, 1882 (Figs. 3.1d and f). Other

species, such as *N. nanulus* (LeConte, 1863) (Figs. 3.1a–c), are more uniformly brown, with little

contrast between elytra, head and pronotum. Others still, such as *N. josiahi* Miller, 2013, present

elytra with dark areas distinctly contrasting against lighter areas and/or may have a notable

iridescent sheen (Fig. 3.5). Color patterns of the ventral surface can also be helpful in delimiting

species. Color is best used in conjunction with other characters, as many species share similar

coloration. Intraspecific variation is often present, with individuals appearing relatively lighter or

darker in color, this variation being additionally present between mature and teneral individuals.

Punctation. Elytral punctation can be very helpful in diagnosing species of *Notomicrus*. Many

species differ in the relative coarseness, density and patterns of punctation. Punctation should be

used in combination with other characters to diagnose species as this character often presents

similarly across multiple species.
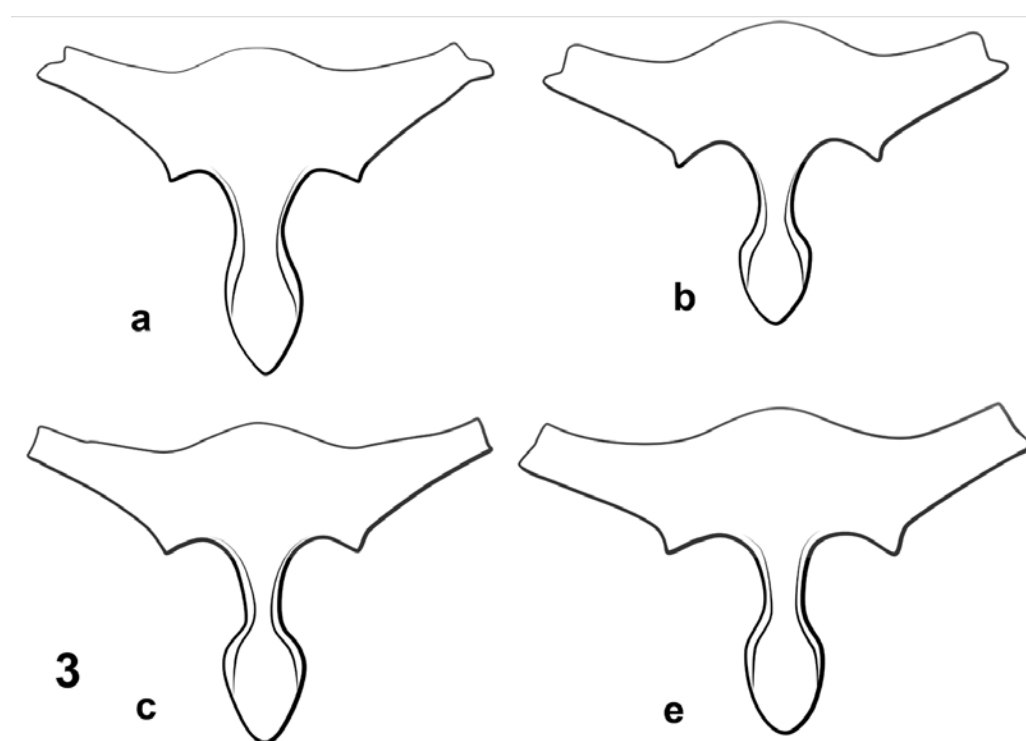
Figure 3.3. Representative prosterna of *Notomicrus* species groups; **a** *N. josiahi* group (*N. josiahi*); **b** *N. nanulus* group (*N. nanulus*); **c** *N. meizon* group (*N.* sp. nr. *malkini*); **d** *N. traili* group (*N.* cf. *traili*).

Microsculpture

External microsculpture in *Notomicrus* varies among species and, in combination with other

characters, can be helpful for diagnosis. In *Notomicrus*, the microsculpture consists of a

microreticulation, where a superficially impressed mesh of very fine lines or grooves creates

small cells. This is usually present on most external sclerites of the head, thorax, abdomen and

legs, though it may not be uniform across these sclerites in an individual (e.g. the microsculpture of the noterid platform often differs from that of the elytra). In particular, the degree of impression and size or density of the meshes can be characteristic for a species or group of species.
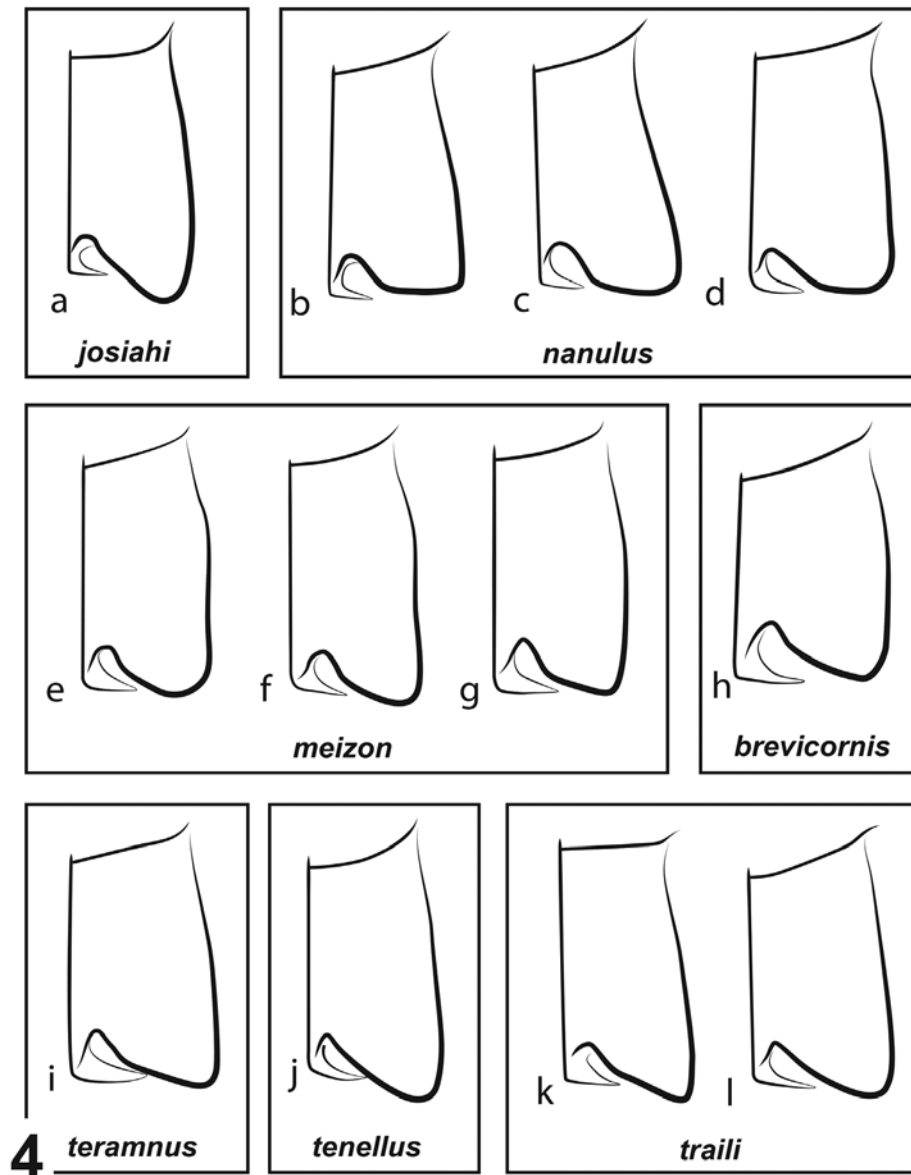
Figure 3.4. Representative noterid platforms of *Notomicrus* species groups (left side, ventral aspect). Names in boxes indicate species groups or species; **a** *N. josiahi*; **b** *N. nanulus*; **c** *N. sharpi*; **d** *N.* sp. (nr. *chailliei*); **e** *N.* sp. nr. *meizon*; **f** *N. meizon* (paratype); **g** *N.* sp. nr. *malkini*; **h** *N. brevicornis* (female syntype); **i** *N. teramnus* (female paratype); **j** *N. tenellus* (Indonesia); **k** *N. sabrouxi* (female paratype, sketched from Manuel (2015:518); **l** *N. petrareptans*.

Eye size

The size of the eyes, relative to the head capsule, can be very useful in identifying species. Here, the relative eye size is presented as a ratio of the greatest width of the head (HW) and interocular distance (EW). Measurements are taken from dorsal aspect, approximately at posterolateral margins of the eyes. Interocular distance is taken from the narrowest point between the eyes. The larger the eyes relative to the head capsule, the larger the ratio HW/EW, for example, *N. josiahi* HW/EW = 2.35–2.53, *N. petrareptans* Baca and Short, 2018 HW/EW = 1.65–1.73.

Prosternal process. The shape of the prosternal process was observed to be variable among some species and species groups of *Notomicrus* (Fig. 3.3). In particular, the shape of the apex and degree of constriction between the procoxae can be diagnostic in combination with other characters. For example, being acutely angled as in *N. josiahi* (Fig. 3.3a) or more rounded or blunt, as in *N. nanulus* (Fig. 3.3b).

Tarsal claws

The pro- and mesotarsal claws of males of *Notomicrus* show significant interspecific variation in size and shape. Following Young (1978), the shape of the claws, as well as the relative sizes of the anterior claws and posterior claws can be helpful in diagnosing species. Here we describe and illustrate the claws in lateral view (Figs. 3.8 and 3.9). The finer details of the claws' shape may be difficult to view without the use of a compound microscope. It should be noted that slide mounting the claws can variably alter the appearance compared to the *in situ* appearance under a stereomicroscope, this in part being due to their asymmetrical shape or the claws being slightly splayed on dried specimens. Characters of the tarsal claws are best used in combination with other characters.

Aedeagus

The aedeagus is especially helpful for diagnosing species. The median lobe should be observed from several angles as it tends to be asymmetric and an oblique orientation can give the appearance of a different shape. Despite relative reliability, the aedeagus is still best used in combination with other characters for identification. Many species, even those across species groups, can present very similar aedeagi. For example, the aedeagus of *N. interstinctus* sp. n. (Fig. 3.7) converges very closely on members of the *traili* group. Additionally, the males of some species are unknown, suggesting these lineages may only comprise females, for example, *N. femineus* Manuel, 2015.

Taxonomy

Genus *Notomicrus* Sharp, 1882

Type species. *N. brevicornis* Sharp, 1882. Designation by Guignot 1946:115.

Diagnosis

(1) Eyes present; (2) metacoxae and metaventrite fused, suture indistinct laterad of noterid platform; (3) noterid platform not extending anteriorly on to metaventrite; (4) protibia with loose rows of spines and setae, lacking large spur at apex and tight comb of small spines on distolateral margin and not expanded distally beyond protarsal insertion; (5) partial fusion of metafurca and metacoxae, not forming complete ring; (6) mid-gular apodeme absent (Beutel and Roughly 1987; Miller 2009); (7) female laterotergite short, posteriorly extending beyond base of gonocoxae (Miller 2009).

Remarks

As noted by Miller (2009) and others (Manuel 2015 and citations therein), the characters that

define *Notomicrus* are primarily plesiomorphic with the exception of the fusion of the metacoxae

and metaventrite. *Speonoterus* Spangler is also defined by the above character combination,

except absence of eyes. Spangler (1996) also noted that the distance from the anterior terminus

of the noterid platform (metacoxal lamellae) to the mesocoxal cavities is shorter in *Speonoterus,*

less than the width of the mesocoxal cavities, whereas in *Notomicrus,* this distance is greater than

the width of the mesocoxal cavities (See Spangler 1996; Manuel 2015). Notomicrine species are

all notably small (ca. 1.0–1.8 mm). Characters listed above without specific citation have been

more common in use for defining *Notomicrus* (e.g. Sharp 1882; Young 1978; Buetel and

Roughly 1987); see Miller (2009) and Manuel (2015) for details.

Key to species groups and *insertae sedis* species of *Notomicrus* Sharp

This key is intended to be used as a first step in identifying New World species of *Notomicrus*.

Identification of *Notomicrus* species can prove difficult for non-specialists, especially without

additional species in hand for comparisons. Diagnoses of the species groups of *Notomicrus* also

reflect this difficulty.

1   Size small, TL = 1.3 mm. Elytral punctation almost entirely indistinct, except discal row

and submargin of elytral suture with distinct punctures, with very fine scattered setose punctures

near lateral margins; elytral surface with microreticulation consisting of round, isodiametric

cells, somewhat scale-like in appearance. Head appendages short, antennomeres VI-X wider than

long; apical palpomeres distinctly bifurcate with enlarged sensory fields. Aedeagus as in Fig. 3.2

of Guimarães and Ferreira-Jr (2019); median lobe with large base and very large processes and hooks. Male pro- and mesotarsal claws short, anterior protarsal claw expanded at base. Known only from high elevation hygropetric habitats in Minas Gerais, Brazil. … *N. teramnus*

1'　　Size variable, ca. 1.2–1.9 mm. Elytral punctation and microsculpture variable. Antennomeres usually longer than wide; apical palpomeres variable. Median lobe of aedeagus without conspicuous hooks or large processes (e.g. Figs. 3.6 and 3.7). Male pro- and mesotarsal claws variable. Habitat preference variable. … 2.

2.　　Dorsal (outer) margin of protibia without notable robust seta at or near mid-length (Fig. 2a). Eyes large relative to head capsule (Fig 3.5), HW/EW ≥ 2.0. Elytra with notable contrasting dark and light colors (Figs. 3.5a and c). Profemur with > 3 distinct closely spaced setae on anteroventral margin (Fig. 3.2a); posteroventral margin of male profemur lacking notable protuberance, only weakly angled near mid-length (Fig. 3.2a); male protarsal claws very small, distinctly shorter than half the length of protarsomere V, anterior protarsal claw bifurcate, branching dorsally (Figs. 3.8 and 3.9). … *josiahi* **group**

2'　　Dorsal margin of protibia with a robust seta at or near mid-length (often two in females), at least as long as most dorsal seta on dorsoapical angle (Figs. 3.2b–f). Eyes smaller, HW/HW < 2.0. Elytra with or without contrasting colors. Profemur with > 3 closely-spaced setae on anteroventral margin; posteroventral margin of male profemur with notable protuberance or acute angle at ca. mid-length; male protarsal claws variable, anterior claw length almost always at least half the length of protarsomere V, almost always larger than female claws, sometimes bifurcate. … **3.**

3.　　Noterid platform with angles of posterior lobes squared or rounded (Figs. 3.4b–e). … **4.**

3'.　　Noterid platform with angles posterior lobes acutely angled (as in Figs. 3.4a and f–l)..…

5.

4.　　Elytral surface impunctate to weakly punctate, punctures usually inconspicuous and sporadic under normal magnification, except for discal series; microreticulation variably impressed, consisting of small, round, isodiametric cells, giving the appearance of small scales. Body form variable, but usually oblong, less attenuated posteriorly (Figs. 3.1a–c). Elytral color uniform, brown, sometimes shiny, not iridescent or only weakly so. … *nanulus* **group.**

4'　　Punctation distinctly present and often dense on posterior half of elytra, punctures finely to moderately impressed, bearing short setae, often extending on to anterior half of elytra; microreticulation variably impressed, consisting of fine mesh-like reticulation. Body form variable, but more elongate and attenuated posteriorly (Figs. 3.1g–i). Color variable, but elytra of mature specimens of most species with darker triangular area medially at base (Figs. 3.1g, i); in most species, dorsal surface very shiny and iridescent. … *meizon* **group** (in part).

5.　　Color uniformly brown. Elytral surface with microreticulation variably impressed, consisting of small, round, isodiametric cells, giving the appearance of fine scales, somewhat shiny, but never iridescent; punctation variable. Males with anterior protarsal claw bifurcate or branched (as in Fig. 3.8a). … **6**.

5'　　Color variable, uniform or bicolorous. Elytral surface with microreticulation variably impressed, consisting of fine mesh-like reticulation, sometimes iridescent. Males with protarsal claws never bifurcated or branched. … **7**.

6.　　Body form oblong, rounded posteriorly (as in Fig. 3.1l or similar to *nanulus* group, for example, Fig. 3.1d). Elytral surface weakly punctate. Median lobe in lateral view as in Fig. 3.12b. New World. … *N. brevicornis* **Sharp, 1882**.

6' Body form ovoid, more elongate, more attenuated posteriorly (as in Fig. 3.1J). Elytral surface weakly to moderately punctate. Median lobe different. Indomalaya and Oceania. … *tenellus* **group**

7. Protibia with robust seta of dorsal margin distinctly distad of half-length of outer margin, approximately at 2/3 margin length (Fig. 3.2e), distance between robust seta and dorsoapical angle distinctly shorter than distance between robust seta and first seta from protibial insertion. Dorsal coloration uniformly brown or bicolorous (Figs. 3.1d–f), with pronotum distinctly lighter than head and elytra. Elytral surface matte to somewhat shiny and iridescent …*traili* **group**.

7' Protibia with robust seta of dorsal margin approximately at half-length of outer margin, distance between robust seta and dorsoapical angle subequal to distance between robust seta and protibial insertion (Fig. 3.2d). Color variable, but elytra of mature specimens of most species with darker triangle medially at base (Figs. 3.1g and i). Most species with elytral surface very shiny, iridescent. … *meizon* **group** (in part).

Description of species groups

1. N. josiahi species group

Diagnosis. The *josiahi* group is diagnosed by the following combination of characters. Dorsal (outer) margin of protibia without notable robust seta at or near mid-length (Fig. 3.2a). Body form elongate, strongly, but regularly attenuated posteriorly. Eyes large relative to head capsule. Elytra with notable contrasting dark and light colors (Figs. 3.5 and 3.6); shiny and iridescent; microsculpture very fine. Prosternal process narrow (Fig. 3.3a). Protibiae elongate, with penultimate dorsal seta only slightly longer than others on dorsal margin (Fig. 3.2a); males with profemur lacking notable protuberance on posteroventral margin (Fig. 3.2a), only weakly

angulate at mid-length; protarsal claws very small, distinctly less than half the length of

protarsomere V (Figs. 3.2a; 3.8 and 3.9), not distinctly larger than female claws, anterior

protarsal claw bifurcate, with small dorsal spur (Figs. 3.8 and 3.9). The large eyes, elytral color

pattern and coloration and characters of the protibiae, make this species group easily

distinguishable from others. There are only two species known.
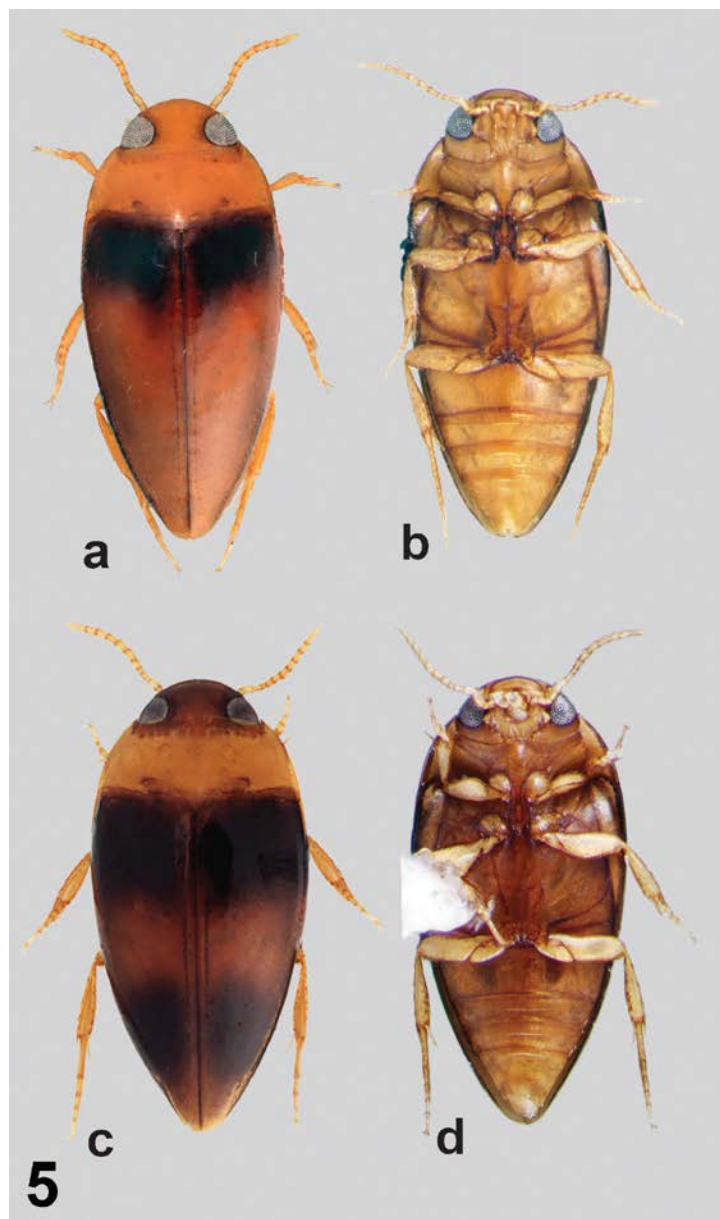
Figure 3.5. *Notomicrus josiahi* species group, dorsal and ventral habitus; **a, b** *Notomicrus josiahi* Miller, 2013; **c, d** *Notomicrus interstinctus* sp. nov.

Notomicrus josiahi Miller, 2013 (Figs. 3.2a; 3.3a; 3.4a; 3.5a, b; 3.6, 3.8)

Notomicrus josiahi Miller, 2013:244; Holotype: MIZA

Type locality. Venezuela, Amazonas State, Communidad Caño Gato, Rio Sipapo, 4º 58.838' N 67º 44.341' W.

Material examined. PARATYPES: "VENEZUELA: Amazonas State/ 4°58.845'N, 67°44.341'W, 100m/ Communidad Caño Gato on Rio/ Sipapo; sandy stream; 7.i.2006; AS-06-016; leg. A.E.Z. Short" [White label, typed print] (1 female ex. SEMC); "VENEZUELA: Amazonas State/ 4°58.845'N, 67°44.341'W, 100m/ Communidad Caño Gato on Rio/ Sipapo; 16.i.2009; leg. Short,/ Miller, Camacho, Joly, & Garcia/ VZ09-0116-01X; along stream" [White label, typed print] (1 male, 2 females ex. SEMC) All paratypes with white barcode label with the following numbers and "KUNHM-ENT": "SM0843570" "SM0831496" "SM0842848" "SM0843672"; all paratypes with "PARATYPE/ *Notomicrus josiahi*/ Miller, 2013" [Blue label with black border, typed print].

Other material: VENEZUELA: Amazonas State, 4°58.845'N, 67°44.341'W, 100m, Communidad Caño Gato on Rio Sipapo; 16.i.2009; leg. Short, Miller, Camacho, Joly, & Garcia/ VZ09-0116-01X; along stream (64 males and females ex. SEMC).

Measurements. TL = 1.46–1.69 mm (mean = 1.59 mm, SD. = 0.058, males = 1.46–1.69 mm, male mean = 1.58, SD. = 0.069, females = 1.55–1.68 mm, female mean = 1.62, SD. = 0.036); TLPn = 1.33–1.53 mm (mean = 1.44, SD. = 0.045, males = 1.33–1.49 mm, females = 1.43–1.53 mm); GW = 0.68–0.78 mm (mean = 0.74 mm, St. Dev. = 0.025, males = 0.68–0.78 mm, females = 0.72–0.78 mm); HW = 0.40–0.45 mm (mean = 0.42 mm, SD. = 0.014, males = 0.40–0.43 mm, females = 0.42–0.45 mm); EW = 0.16–0.19 mm (mean = 0.175 mm, SD. = 0.01, males = 0.16–

0.17 mm, females = 0.17–0.19 mm); TL/GW = 1.99–2.31 (mean = 2.16; SD = 0.070; males = 1.99–2.31, females = 2.13–2.22); HW/EW = 2.21–2.53 (mean = 2.39, SD = 0.083, males = 2.41–2.53, females = 2.21–2.44).

Diagnosis. *Notomicrus josiahi* can be diagnosed by the following combination of characters: (1) Size large TL = 1.46–1.69 mm; (2) elytron with strongly darkened region in anterior 1/3rd, contrasting against brownish-yellow of rest of elytron (Fig. 3.5a); (3) Eyes very large relative to head capsule (HW/EW= 2.21–2.53; males 2.41–2.53, females 2.21–2.37); (4) aedeagus as in Fig. 3.6, median lobe expanded on right side in dorsal or ventral aspect, weakly attenuated to apex from mid-length in lateral aspect, with apex curved dorsolaterally to the left, left lateral lobe with dense tuft of setae at apex, few setae along dorsal margin and sparse tuft near base; right lateral lobe with small tuft of setae at apex; (5) pro- and mesotarsal claws as in Fig. 3.8, anterior protarsal claw strongly bent, bifurcate, with slender spur originating on dorsal margin where curved (Fig. 8a), ventral margin strongly expanded ventrally near base.

Re-description. Males. Body elongate-oval, attenuated posteriorly (Fig. 3.5a), TL/GW = 1.99–2.31 lateral outline of elytra and pronotum continuous in dorsal aspect; regularly curved to head, posteriorly evenly attenuated to elytral apex from point of greatest width; widest point just posterior to humeral angles of elytra, as in Fig. 3.5a.

Color. Head, pronotum, venter and legs yellow; elytron dark brown to black in basal 1/3, darkened region extending posteriorly along elytral suture and contrasting against brownish-yellow color of posterior 2/3 of elytron (Fig. 5a); elytron with surface weakly iridescent. Venter and legs uniformly yellow (Fig. 3.5b).

Structures. Eyes very large relative to head capsule (HW/EW = 2.35–2.53); antennae with length greater than greatest width of head. Prosternal process narrow, not strongly constricted

between procoxae, with apex attenuated (Fig. 3.3a). Noterid platform with lateral margins subparallel (weakly convergent in posterior 2/3, convergent in anterior 1/3 (Fig. 3.4a); posterior lobes acute, angled, acutely rounded at apex. Profemur with loose comb of 3–5 stiff setae on anteroventral margin (Fig. 3.2a), posteroventral margin weakly angled at mid-length (Fig. 3.2a). Protibia elongate, dorsal and ventral margins weakly divergent distally in anterior aspect (Fig. 3.2a), anterodorsal margin with row of 6–7 stout setae, without distinctly larger seta near mid-length. Protarsi with adhesive discs on ventral surface of protarsomeres II and III, lacking disc on ventral surface of protarsomere I; protarsal claws as in Figs. 3.8a and b, subequal in length, small, length ca. 1/3 that of protarsomere V, anterior claw distinctly bifurcate in distal half, expanded basally, very sharply curved, posterior claw slender, weakly expanded basally, moderately curved. Mesotarsi with adhesive discs on ventral surface of protarsomere II only, lacking disc on ventral surface of protarsomere I; mesotarsal claws as in Figs. 3.8c and d, subequal in length, small, length slightly greater than that of protarsal claws, slender, weakly expanded at base and weakly curved.

Sculpture. Dorsal surface of head with microsculpture very weakly impressed, microreticulation very fine, meshes mostly indistinct; micropunctation nearly indistinct. Pronotum with microsculpture similar to that of head, microreticulation fine; with scattered punctation near base and lateral margin, lateral punctures moderately dense, some with very fine setae. Elytron with microsculpture weakly impressed, microreticulation very fine, nearly indistinct; with punctation sparse in anterior half, with fine punctures along lateral margin and along discal row, with very few to no punctures between discal row and elytral suture, punctate in posterior half, punctures fine, many with very fine setae; discal row composed of fine and irregularly scattered punctures, denser posteriorly, lateral row similar to discal row but more sparse; micropunctation present,

evenly scattered. Noterid platform and metaventrite surface with microsculpture weakly to moderately impressed, very fine, meshes of microreticulation nearly indistinct, cells transversely elongated.
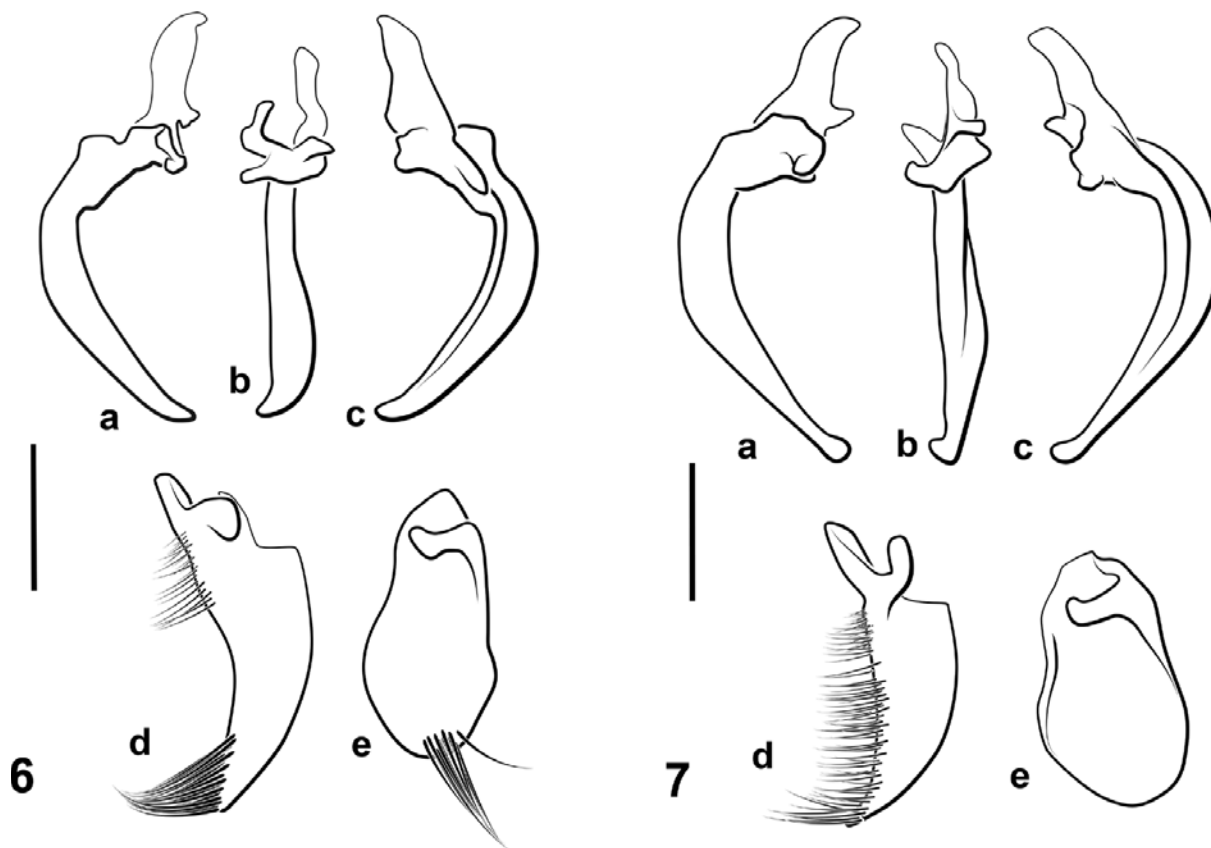
Figures 3.6, 3.7. Aedeagi of *josiahi* species group; **6** *N. josiahi*; **7** *Notomicrus interstinctus* sp. nov.; **a** median lobe, left lateral aspect; **b** median lobe, dorsal aspect; **c** median lobe, right lateral aspect; **d** left lateral lobe, medial surface/aspect; **e** right lateral lobe, medial surface/aspect. Scale bars: 100 μm

Aedeagus. Aedeagus as in Fig. 3.6. Median lobe in lateral aspect gradually curved from base to apex, dorsal and ventral margins subparallel, converging at apex; apex acute, sharply curved, in ventral aspect subapically expanded and curved to left (Figs. 3.4a–c). Left lateral lobe in lateral aspect elongate, curved dorsally, with dense tuft of setae at apex (Fig. 3.6d). Right lateral lobe in lateral aspect oval; apex rounded with small tuft of setae in apical cleft (Fig. 3.6e).

Females. As males, except eyes slightly smaller than in males (HW/EW females = 2.21–2.39); profemur with posteroventral margin smooth, lacking weak angle at mid-length; pro and

mesotarsomeres unmodified, slender, lacking adhesive discs; pro and mesotarsal claws unmodified, claws of respective tarsi subequal in length, slender, weakly curved.

Variation. As this species is known from only a single series, it is difficult to assess the degree of intraspecific variation. However, some variation was observed in the relative lightness or darkness in coloration of the individuals, with some brighter in color, more yellow, and others darker in color, more brownish yellow. The darkened region of the elytra also varied somewhat, occupying 1/4 to greater than 1/3 of the basal region of the elytron.

Differential Diagnosis. *Notomicrus josiahi* is among the most easily distinguished species of *Notomicrus* by the combination of the large eyes, color pattern, shape of male protarsal claws and of male aedeagus. Superficially, *N. josiahi* is similar to some species of the *N. meizon* group in color, wherein *N. meizon* Guimarães and Ferreira-Jr, 2019*, N. malkini* Young, 1978 and other undescribed species are also darkened at the base of the elytra. However, in *N. josiahi,* this darkened area is better defined with the posterior border less oblique, thus expanding more completely over the humeral angles of the elytron. More distinctly, *N. josiahi* differs from these and other species by the much larger eyes and bifurcate protarsal claws (in males), which to date, has only been observed in *N. interstinctus*, *N. brevicornis* and the *tenellus* group. Among all other species of *Notomicrus*, the aedeagus of *N. josiahi* is distinct, with the right lateral lobe rounded and bearing a small tuft of setae at apex, rather than without setae, as in all other neotropical species.

Distribution. Known only from Venezuela (Fig. 3.10).

Ecology. This species has been collected from only a single locality, from the margins of a small, sandy stream (Fig. 3.11a).

Notomicrus interstinctus **n. sp.** (Figs. 3.5c, d; 3.7, 3.9)

Type locality. BRAZIL: Amapá, Calcoene, 2.50019°, -50.97712°.

Material examined. HOLOTYPE, male: "BRAZIL: Amapá: Calcoene/ 2.50019°, -50.97712°; 5 m/ Colcoene (1 km W) on BR-156/ 22.vii.2018; leg. Short; Marshy/ savannah; BR18-0722-01A" [White label, typed print] "HOLOTYPE/ *Notomicrus/ interstinctus/* Baca and Short, 2020" [Red label, black border, typed print] (ex.INPA). PARATYPES: Same data as holotype, except with "PARATYPE/ *Notomicrus/ interstinctus/* Baca and Short, 2021" [Blue label, black borders, typed print] (4 males, 5 females exs. SEMC, INPA); BRAZIL: Amazonas, Manacapuru Municipality, -3.23037°, -60.64269°, 35 m, 9.vi.2017, leg. Benetti, margin of large marsh/river, lots of vegetation; BR17-0609-01A; with "PARATYPE/ *Notomicrus/ interstinctus/* Baca and Short, 2021" [Blue label, black borders, typed print] (3 males, 6 females exs. SEMC, INPA).

Measurements. TL = 1.50—1.63 (Holotype = 1.50 mm, mean = 1.56 mm, SD. = 0.045, males 1.50—1.63 mm, females 1.50–1.63 mm); TLPn = 1.38–1.48 (Holotype = 1.40 mm, mean = 1.42 mm, SD = 0.039, males = 1.40–1.45 mm, females = 1.38–1.48 mm); GW = 0.72–0.80 mm (Holotype = 0.72 mm, mean = 0.75 mm, SD. = 0.027, males = 0.72 mm–0.76 mm, females = 0.73–0.80 mm; HW = 0.41–0.45 mm (Holotype = 0.41 mm, mean = 0.43 mm, SD. = 0.013, males = 0.41–0.42 mm, females = 0.42–0.45 mm); EW = 0.18–0.22 mm (Holotype = 0.18 mm, mean = 0.20 mm, SD. = 0.013, males = 0.18–0.19 mm, females = 0.19–0.22 mm), TL/GW = 1.99–2.26 (Holotype = 2.08, mean = 2.07, SD. = 0.074, males = 2.06–2.26, females = 1.98–2.11); HW/EW = 2.04–2.33 (Holotype = 2.28, mean = 2.19, SD. = 0.088, males = 2.16–2.33, females = 2.04–2.26)

Diagnosis. *Notomicrus interstinctus* can be diagnosed by the following combination of characters: (1) Size large TL = 1.53–1.63 mm; (2) elytron dark with contrasting yellow band at

mid-length, apices yellow (Figs. 3.5c and d); (3) eyes very large relative to head capsule (HW/EW= 2.04–2.33; Fig. 3.5c); (4) aedeagus as in Fig. 3.7; median lobe not broadly expanded on right side in dorsal aspect, attenuated to apex in lateral aspect, with small, round apical club oriented laterally to left; left lateral lobe with row of setae along entire dorsal margin, only somewhat denser at apex; right lateral lobe glabrous, without small tuft of setae at apex; (5) protarsal claws as in Figs. 3.9a and b; anterior protarsal claw strongly bent, bifurcate, branching at mid-length, ventral margin somewhat expanded ventrally near base.
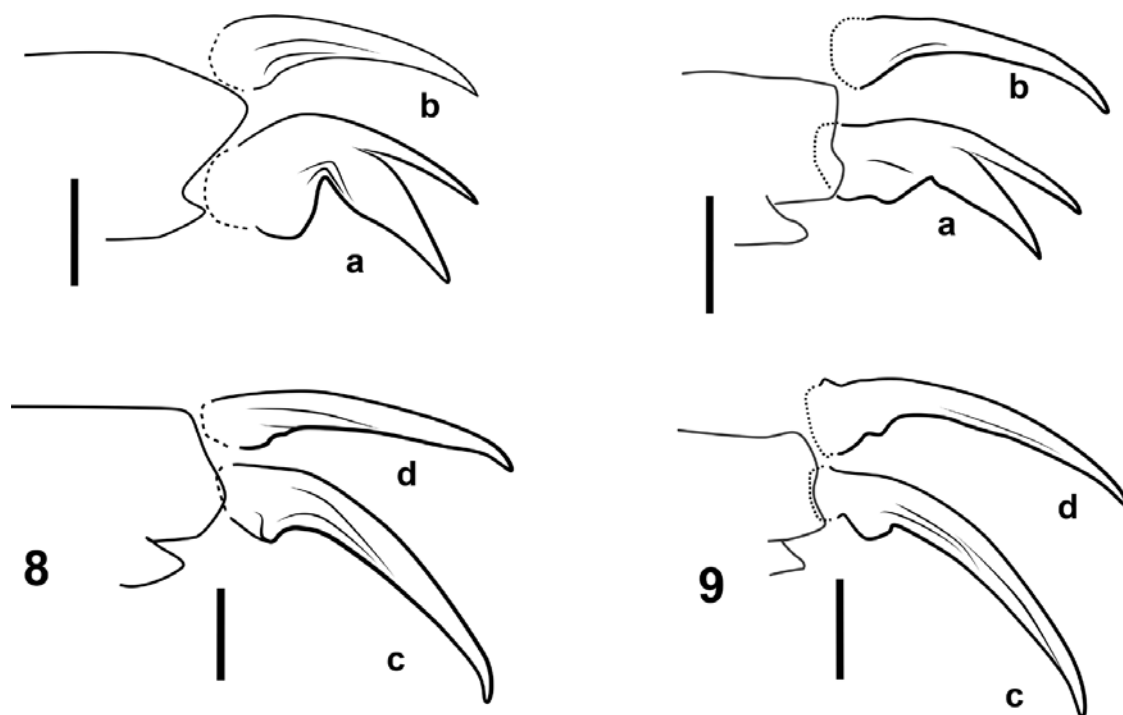
Figures 3.8, 3.9. Pro- and mesotarsal claws of *josiahi* species group; **8** *N. josiahi*; **9** *N. interstinctus* sp. nov.; **a** anterior protarsal claw; **b** posterior protarsal claw; **c** anterior mesotarsal claw; **d** posterior mesotarsal claw. All anterior aspect. Scale bars: 25 μm.

Description. Holotype. As described for *N. josiahi,* except the following. Size large, TL = 1.53 mm. Body very broad, elongate-oval, strongly attenuated posteriorly, TL/GW = 2.08; lateral outline of elytron evenly and gradually curved to apex from point of greatest width, as in Figs. 3.5c and d.

Color. Dorsal surface of head brown, lighter near clypeus; pronotum yellow; elytron dark, nearly black in anterior and posterior thirds, with lighter contrasting brownish-yellow transverse band near mid-length of elytron, elytral apex also lighter, brownish-yellow; elytron with surface moderately iridescent. Ventral surface of head and prosternum light brownish-yellow; rest of venter yellowish-brown; legs yellow.

Structures. Eyes large relative to head capsule (HW/EW = 2.28). Posterior lobes of noterid platform with angles acute, apices rounded (as in Figs. 3.3a and 3.5d). Pro- and mesotarsal claws as in Fig. 3.9.

Sculpture. Elytron with punctation as described in *N. josiahi,* but denser overall and less restricted to posterior half, with punctures along lateral margin and puncture rows more widely distributed and denser.

Aedeagus. Aedeagus as in Fig. 3.7. Median lobe in lateral aspect, strongly curved at base, distally weakly curved, nearly straight; dorsal and ventral margins subparallel to mid-length, then attenuated to apex; apex with small club, sharply bent dorsally and left; left lateral lobe in lateral aspect, elongate, dorsal margin curved with dense row of fine setae (Fig. 3.7d). Right lateral lobe in lateral aspect oblong, rounded distally.

Females. As males, except eyes slightly smaller (HW/EW = 2.04–2.16); profemur with posteroventral margin smooth, lacking weak angle at mid-length; pro- and mesotarsomeres unmodified, slender, lacking adhesive discs; pro- and mesotarsal claws unmodified, claws of respective tarsi subequal in length, slender, weakly curved.

Variation. The most notable variation was in size and color, with some specimens darker overall than others, with the yellow bands sometimes smaller.

Differential Diagnosis. *Notomicrus interstinctus* is easily distinguished by the combination of large eyes and elytral color pattern, darkened in anterior and posterior thirds with a yellow transverse band. This color pattern is unique among known species of *Notomicrus*. This species is also unusual in that it is one of the few known species (along with *N. brevicornis*, *N. josiahi* and members of the *tenellus* group), with males that present bifurcated anterior protarsal claws. The aedeagus, color pattern and more subtly the denser punctation, easily differentiates this species from *N. josiahi*. The aedeagus of *N. interstinctus* is similar to that of the *N. traili* group with the median lobe attenuated and the apex enlarged and bent in a left-dorsal direction, but other external characters, such as the color pattern, tarsal claws and large eyes, readily distinguish this species from the *traili* group. The elytral punctuation is somewhat similar to that of some members of the *N. meizon* group, being somewhat densely punctate posteriorly, with punctures fine, but the aforementioned combination of characters will differentiate *N. interstinctus* from these species as well.

Etymology. *Notomicrus interstinctus* **n. sp.** derives its name from the Latin adjective *interstinctus,* meaning checkered or variegated. This refers to the color pattern of the elytra of this species. It is treated as an adjective in the nominative singular.

Distribution. Known from northern Brazil, Amazonas and Amapá states (Fig. 3.10).

Ecology. The species seems to be a generalist in terms of habitat, but seems to prefer vegetated environments. It was collected from a very shallow open marshy area (Fig. 3.11c) in the Brazilian state of Amapá and the vegetated margins of a river in Amazonas (Fig. 3.11b).

Taxonomic comments. *Notomicrus interstinctus* appears very similar to specimens misidentified as *N. traili* Sharp, 1882 by Guimarães and Ferreira-Jr (2019). This was due to the similarities of the size, punctation and shape of the aedeagus. The records from this work would

potentially add to the distribution above, as most appear to be from the same regions of the

Amazon Basin as the Amazonas specimens. Verification will be needed to confirm these

individual records and these are not formally attributed to *N. interstinctus* here. Observations of

the lone female syntype of *N. traili* (NHM) indicate that the species is as described by Young

(1978), with males attributable to *N. traili*, appearing as in Fig. 3.2, with the head and elytra

brown, without a pattern.
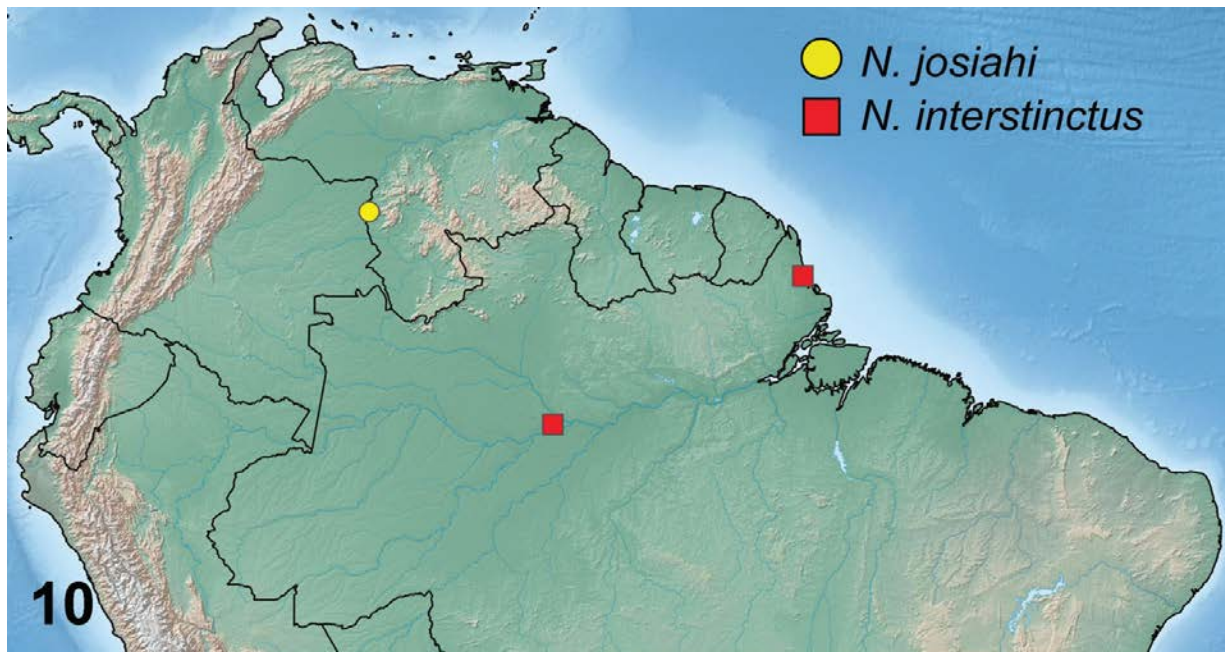
Figure 3.10. Distribution map of josiahi group species.
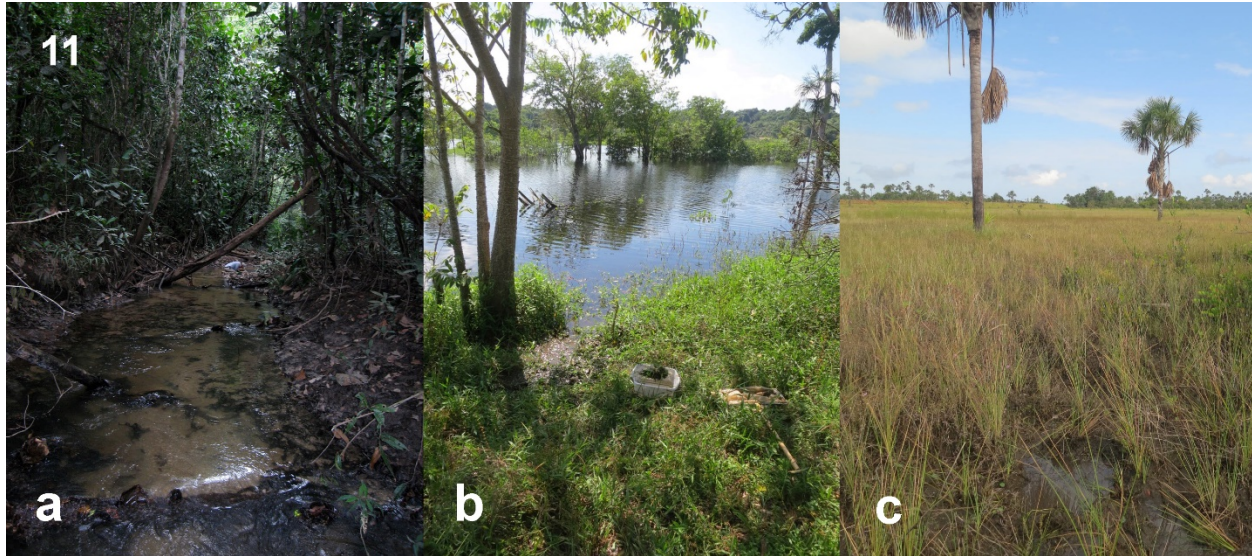
Figure 3.11. Habitats of *N. josiahi* group species: **a** Type locality of *N. josiahi*, Venezuela: Amazonas (collection code VZ09-0116-01A). Localities of *N. interstinctus*: **b** Brazil, Amazonas (collection code BR17-0609-01A) **c** type locality, Brazil, Amapá (collection code BR18-0722-01A).

2. *Notomicrus nanulus species group*

Diagnosis. Members of this species group are most easily identified by their (1) monotone brown elytral color (Figs. 3.1a–c); (2) rounded, oval body shape; (3) rounded posterior lobes of the noterid platform (Figs. 3.4b–d); and (4) by their coarser microsculpture, consisting of isodiametric cells, appearing scale-like, rather than as a finer mesh of transversally elongated cells. This latter character is best viewed in light reflecting off the elytra. Even in species with finer variants of the cell-like microsculpture (e.g. *N. sharpi*), there is no iridescence. Punctation is largely indistinct, except for the discal series and sometimes sporadic punctures posteriorly.

Members of the *N. nanulus* group present a combination of characters that are variably shared with *N. brevicornis*, *N. teramnus* and the members *N. tenellus* group. This pattern, in tandem with our phylogenetic understanding, for example, *tenellus* group being sister to all other *Notomicrus* (Baca and Short 2020), suggest that *N. nanulus* species are united by a combination of characters that are plesiomorphic at some level within the genus. All share similar

microsculpture consisting of isodiametric cells, often appearing scale-like to some degree. However, the *N. nanulus* group is distinguished from the *N. tenellus* group by the more rounded body outline (Figs. 3.1a–c and j) and from the *N. tenellus* group and *N brevicornis* by the shape of the noterid platform, with the *nanulus* group presenting posterior lobes that are rounded or squared (Figs. 3.4b–d). The rounded/squared lobes of the noterid platform character distinguish the *nanulus* group from *N. teramnus* (Fig. 3.4i) also, but more subtly. The *nanulus* group also typically presents a noterid platform with more longitudinally-elongated proportions than *N. brevicornis* (Fig. 3.4h). Males of the *N. nanulus* group present unbifurcated anterior protarsal claws, unlike *N. brevicornis* and the *N. tenellus* group. The aedeagi of known males of the *N. nanulus* group are easily distinguished; see Young (1978) and Manuel (2015).

Composition. *N. chailliei* Manuel, 2015; *N. femineus* Manuel, 2015; *N. huttoni* Young, 1978; *N. nanulus* (LeConte, 1863); *N. sharpi* Balfour-Browne, 1939.

Identification resources. Young (1978); Manuel (2015).

Remarks. Future work on this group may prove difficult as many species are collected with high ratios of females to males. An example was *N. femineus* Manuel, 2015, in which extensive collecting yielded only females, raising the possibility of parthenogenetic reproduction. Personal observations indicate that multiple undescribed species are represented by females only. We note that females of this group can be especially difficult to distinguish and are often misidentified as *N. brevicornis* (see comments on *N. brevicornis*, below). *Notomicrus teramnus* is a potential member of this species group based on color, shape and microsculpture, but is treated separately in the key, pending further investigation (see *insertae sedis* species below).

3. *Notomicrus meizon species group*

Diagnosis: Non-teneral specimens of this group tend to have the following combination of characters: (1) triangular pigmented area medially on the base of the elytra (Figs. 3.1g–i), similar to *N. josiahi,* but not as prominent and not always discernible in some populations or in teneral specimens; (2) dense, fine punctures bearing short setae on the posterior half of the elytra and sometimes extending far anteriorly (not as coarse as in members of the *traili* group); (3) microreticulation variably impressed, consisting of fine mesh-like reticulation; often iridescent; (4) posterior lobes of noterid platform with squared or rounded angles (Fig 3.4e); if posterior angles of noterid platform more acute (Figs. 3.4f and g), protibia presents robust seta of outer margin approximately at half-length of outer margin, distance between robust seta and dorsoapical angle subequal to distance between robust seta and first seta from protibial insertion (Fig. 3.2d).

Remarks. The *meizon* group is sometimes difficult to discern from the *traili* group, as the differences amongst diagnostic characters can be subtle. The darkened basal area of the elytra in the *meizon* group is helpful, but investigators may find great difficulty in diagnosing teneral members of this group, which often lack the pigmented triangular area on the elytra. It is important to note that this darkened area is really/truly pigmented, not just darker in appearance due to the folding of the wings under the elytra as often happens in lighter colored species. Fortunately, males of most individual species of the *meizon* group are easy to identify by their aedeagi in combination with other characters, such as tarsal claws. The male median lobes of the *meizon* group species are usually (but not always) very irregularly shaped (for example, see Young 1978 and Guimarães and Ferreira-Jr 2019). The aedeagus of most species of the *traili* group appear similar to Fig. 8, with a small club at apex, often hooked to the left (see Young

1978; Manuel 2015; Baca and Short 2018). Additionally, males of *meizon* species often present notable unequal lengths between the anterior and posterior protarsal claws. These are usually subequal in length in the *traili* group.

Composition: *N. malkini* Young, 1978; *N. meizon* Guimarães and Ferreira-Jr, 2019.

Identification resources: Young (1978); Guimarães and Ferreira-Jr (2019).

4. *Notomicrus traili species group*

Diagnosis. Non-teneral specimens of this group tend to have the following combination of characters: (1) lacking triangular pigmented area on the medial base of the elytra (Figs. 3.1d–f), lacking maculae; elytra with uniform shades of tan or brown; (2) irregular setose punctures in posterior half of elytral surface with density variable, increasingly coarse if extending on to anterior half of elytron; (3) microreticulation variably impressed, consisting of fine mesh-like reticulation; matte to shiny; elytral surface sometimes somewhat iridescent; (4) posterior lobes of noterid platform with angles acute (Figs. 3.4 k and l); (5) protibia as in Fig. 3.2e, with robust seta of outer margin at ca. 2/3 length of margin, distance between robust seta and dorsoapical angle distinctly less than distance between robust seta and first seta from protibial insertion.

Composition. *N. gracilipes* Sharp, 1882; *N. petrareptans* Baca and Short, 2018; *N. reticulatus* Sharp, 1882; *N. sabrouxi* Manuel, 2015; *N. traili* Sharp, 1882.

Identification resources. Young (1978); Manuel (2015); Baca and Short (2018).

Remarks. Species of the *traili* group are difficult to discern and constitute a widespread species complex (see Baca and Short 2020). Personal observations coupled with the phylogenetic reconstructions of Baca and Short (2020) show that the diagnostic power of the dorsal punctation (see Young 1978) is unreliable, with multiple clades within the complex sharing similar patterns

of punctation; for example, the pattern of punctation attributed to *N. gracilipes* by Young (1978) arises in multiple places within the complex. The group will require careful taxonomic investigation. The members of the *traili* group can be difficult to distinguish from those of the *meizon* group, but mature members lack a pigmented triangular area at the base of the elytra and most males of the *traili* group have similarly-shaped median lobes of the aedeagus, distinct from the *meizon* group. See notes in remarks of *meizon* group above.

5. Insertae sedis species

These species present characters combinations not found in other species groups. Both by presented character combination and even general *gestalt*, these are difficult to place with certainty. Molecular sequence data were unavailable for these species in the phylogenetic reconstruction of Baca and Short (2020). In particular, the species listed here both exhibit body shape, color, microsculpture and sparse punctation that would place them in the *N. nanulus* species group. However, in comparison with the *N. tenellus* species group, the sister to all New World taxa, several of these characters appear plesiomorphic in *Notomicrus*, making it difficult to discern their likely relatives from morphology alone.

N. brevicornis Sharp, 1882 (Figs. 3.1k, 3.12a)

Material examined. Syntypes: Male specimen on small rectangular card, "♂" is drawn around genitalia and other parts, prosternal process flanks the specimen. "Boa Sorta Nov./ Sahlberg 1850" [small rectangular label, handwritten], "Sharp Coll/ 1905-313" [small rectangular label, typed], "Notomicrus/ brevicornis Ind. typ./ D.S." [small rectangular label, handwritten] "SYN/ TYPE" [small circular label with blue border, printed] (ex. NHM); female specimen on

rectangular card, "S. America/ Brazil." [small rectangular label with blue line across, typed], "Sharp Coll/ 1905-313." Small rectangular label, typed], "Boa Sorta Nov./ Sahlberg 1850" [small rectangular label, handwritten], "Type 470/ Notomicrus/ brevicornis/ Boa Sorta" [rectangular label, handwritten], "SYN/ TYPE" [small circular label with blue border, printed], "TYPE" [small circular label with red border, printed], (ex. NHM); female specimen disarticulated on large card, "S. America/ Brazil." [small rectangular label with blue line across, typed], "Boa Sorta Nov./ Sahlberg 1850" [small rectangular label, handwritten], "Notomicrus/ brevicornis, Sharp./ Co-type." [rectangular label, handwritten], "SYN/ TYPE" [small circular label with blue border, printed], (ex. NHM); female specimen on small rectangular card, "Co-/ type" [small circular label with yellow border, printed], "S. America/ Brazil." [small rectangular label with blue line across, typed], "Sharp Coll/ 1905-313." Small rectangular label, typed], "Notomicrus/ brevicornis, Sharp./ Co-type." [rectangular label, handwritten] "SYN/ TYPE" [small circular label with blue border, printed] (ex. NHM). Note: this latter specimen also with small label "Not brevicornis/ maybe gracilipes?/ Manuel det. 2016". See notes below.

Remarks. *Notomicrus brevicornis* would otherwise appear to be a member of the *nanulus* group by the aforementioned characters. However, it differs by the more acute posterior angles of the noterid platform, a character shared with members of the *tenellus, josiahi* and *traili* groups. The male syntype presents a bifurcate anterior protarsal claw (as in fig. 3.8A), a character shared by the *josiahi* and *tenellus* species groups. With the Old World and New World taxa being reciprocally monophyletic (Baca and Short 2018) and the plesiomorphic appearance of these characters, we would speculate that this species is likely to be sister to one of the New World species groups.

Based on observation of the single male of the syntype series, it is suspected that Young (1978)

based his description, key and illustration of the aedeagus of *N. brevicornis* on the male of a

different species. First, the illustration in Young (1978) of the aedeagus of *N. brevicornis* does

not match that observed in the syntype. Second, Young (1978: 288–289) describes *N. brevicornis*

as being sexually dimorphic in elytral punctation, with males being more punctate than females.

However, as noted by Sharp (1882: 261), there is very little dimorphism observed between males

and females of the syntype series beyond characters of the tarsi. The punctation and sculpture are

very weakly dimorphic, both sexes being almost entirely impunctate, except for the weak discal

rows and a few scattered punctures near the elytral apex. The punctation is slightly less

impressed in females, with discal rows slightly less prominent. The relative difference of

punctation between the male and female syntypes of this species is so slight that splitting them

up in the key as did Young (1978:288, couplet 7) seems largely unnecessary, wherein the couplet

describing females of *N. brevicornis* also closely describes the male syntype (Young 1978: 288).

The specimens of the UMMZ, observed by Young, were not observed for this study, but the

stated differences by Young (1978) and the grouping of males of *N. brevicornis* with *N. malkini*

in Young's (1978: 288) key call the identity of the depicted male in Young (1978) into question.

Further adding to this suspicion is the fact that some male specimens attributable to *N. malkini* or

other undescribed species of the *meizon* group in the FSCA were identified as *N. brevicornis* by

Young (date of determination not recorded). For aiding in identification, we have included

images of the male syntype, labels and aedeagus (Figs. 3.1k and 3.12). One specimen of the

syntype series appears to be of a different species than the others; likely it is a member of the

traili species group. See last listed specimen and note in the examined syntype material above.

Personal observations show that many members of the *N. nanulus* group are misidentified as *N. brevicornis* in collections. This is no doubt due to the superficial similarities of *N. brevicornis* to members of the *nanulus* group and scarcity of males in the *nanulus* group. With that, there are likely inaccuracies in literature with respect to records and distributions.

Figure 3.12. Card mount, aedeagus and labels of male syntype of *N. brevicornis*; **a** *Notomicrus brevicornis* card mount, dorsal; **b** median lobe lateral aspect; **c** left lateral lobe, medial aspect; **d** right lateral lobe; **e** syntype labels.

N. teramnus Guimarães and Ferreira-Jr, 2019

Remarks. *Notomicrus teramnus* would also appear a member of the *nanulus* group, given the above-mentioned characters. An argument could be made that this is the case as it only appears

to differ in the shape of the posterior lobes of the noterid platform being more angular than most species in the *nanulus* group. This species otherwise appears to lack characters that would unite it with other species groups, though this will require examination and/or phylogenetic investigation. We abstain from placing it as member of the *nanulus* group as *N. teramnus* is known only from a high elevation hygropetric habitat, which may present confounding morphological specialization. Aedeagal morphology is not here considered to be indicative of a particular placement, but the very unusual morphology of the aedeagus of *N. teramnus* (see Guimarães and Ferreira-Jr 2019) further raises questions of placement.

Acknowledgements

References

Baca SM, Short AEZ (2018) *Notomicrus petrareptans* sp. n., a new seep-dwelling species of Noteridae from Suriname (Coleoptera: Adephaga). Zootaxa 4388(2): 182-190.

Baca SM, Short AEZ (2020) Molecular phylogeny of the notomicrine water beetles (Coleoptera: Noteridae) reveals signatures of Gondwanan vicariance and ecological plasticity. Insect Systematics and Diversity *4*(6), 4.

Baca SM, Toussaint EFA, Miller KB, Short AEZ (2017) Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on partitioning strategies. Molecular Phylogenetics and Evolution. 107: 282–292.

Balfour-Browne J (1939) A contribution to the study of the Dytiscidae. - I. (Coleoptera, Adephaga). The Annals and Magazine of Natural History (11) 3: 97–114.

Belkaceme T (1991) Skelet und muskulatur des kopfes und thorax von *Noterus laevis* Sturm.ein beitrag zur morphologie und phylogenie der Noteridae (Coleoptera: Adephaga). Stuttgarter Beiträge zur Naturkunde A 462: 1–94.

Beutel RG, Roughley RE (1987) On the systematic position of the genus *Notomicrus* Sharp (Hydradephaga, Coleoptera). Canadian Journal of Zoology 65(8), 1898-1905.

Clark H (1863) Descriptions of new East-Asiatic species of Haliplidae and Hydroporidae. T. Roy. Entomological Society of London (3) 1: 417–428.

Fauvel A (1903) Faune analytique des coléoptères de la Nouvelle-Calédonie. Revue d'Entomologie 22: 203-379.

Guimarães BA, Ferreira-Jr. N (2019) Two new species and new records of *Notomicrus* Sharp, 1882 (Coleoptera: Noteridae) from Brazil. Zootaxa 4629(2): 263–270.

LeConte JL (1863) New species of North American Coleoptera. Part1. Smithsonian Miscellaneous Collections 6(167): 1-92.

Manuel M (2015) The genus *Notomicrus* in Guadeloupe, with description of three new species (Coleoptera: Noteridae). Zootaxa. 4018 (4): 506–534.

Miller KB (2009) On the systematics of Noteridae (Coleoptera: Adephaga: Hydradephaga): phylogeny, description of a new tribe, genus and species, and survey of female genital morphology. Systematics and Biodiversity 7: 191–214.

Miller KB (2013) Notomicrus josiahi, a new species of Noteridae (Coleoptera) from Venezuela. Zootaxa 3609(2): 243–247.

Miller, K.B. & Nilsson, A.N. (2003) Homology and terminology: Communicating information about rotated structures in water beetles. Latissimus, 17, 1–4.

Nilsson AN (2011) A World Catalogue of the family Noteridae, or the Burrowing water beetles (Coleoptera: Adephaga). Version 16.VIII.2011, 1–54 pp. Downloaded from: http://www2.emg.umu.se/projects/biginst/andersn/WCN/WCN_20110816.pdf (accessed 1 June 2017).

Sharp D (1882) On aquatic carnivorous Coleoptera or Dytiscidae. Scientific Transactions of the Royal Dublin Society 2: 179–1003.

Toledo M (2010) Noteridae: Review of the species occurring east of the Wallace line (Coleoptera). In MA Jäch, and M Balke (eds.) Water Beetles of New Caledonia (part 1): monographs of Coleoptera Vo. 3. Zoologische-Botanische Gesellshaft, Vienna, AT, 195–236.

Young FN (1978) The New World species of the water-beetle genus Notomicrus (Noteridae). Systematic Entomology 3(3): 285–293.

Chapter 4: Shallow scale phylogenomics with Ultraconserved elements parse relationships and inform taxonomy in the *Notomicrus traili* species complex.

Abstract

The *Notomicrus traili* complex is distributed throughout South America and extends into Mexico and the West Indies. Previous research has shown that multiple phylogenetically distinct clades share overlapping distributions and individuals attributable to a given described species were recovered as polyphyletic (Chapter 2). As such, this group requires investigation at the species to population level, making it a prime candidate for the application of the phylogenomic capture of ultraconserved elements to examine patterns of evolution and guide taxonomy in the *traili* complex. We use short-read sequencing of four noterid genomes to design a noterid-specific UCE probe set (Noteridae 3.4Kv1). With this we captured UCE data from a population-level sampling of 45 *traili* group specimens from across the range of the complex. We reconstructed the phylogeny of the *traili* group with multiple data trimming and data completeness criteria, using concordance metrics to compare topologies recovered. Our phylogenetic estimates confirm there are several phylogenetically distinct clades within the *traili* group. No reconstruction indicated the need to synonymize any of the four species in this group. Instead, it appears that additional undescribed species are present, and will require careful investigation of morphology. The evolutionary patterns recovered are consistent with repeated cycles of diversification and range expansion.

Introduction

The advent of phylogenomic methods have allowed scientists to parse patterns of shallow scale evolution with unprecedented amounts of data and a wealth of methods spanning broadly many aspects of evolutionary theory. Evolutionary biologists now hold a greater understanding of the genetics underlying evolution and processes such as speciation than ever before. Further, these methods are becoming increasingly accessible, in both cost and ease of application. The targeted capture of Ultraconserved elements (UCEs; Faircloth et al., 2012) has provided systematists with a locus type that is broadly applicable to evolutionary questions across a wide breadth of evolutionary scales, deep (e.g. Faircloth et al., 2013; Branstetter et al., 2017; Gustafson et al., 2020) to shallow (e.g. Smith et al., 2014; Harvey et al., 2016; Manthey et al., 2016; Branstetter & Longino, 2019; Gueuning et al., 2020). Recent investigations have in particular shown the utility of UCEs for species delimitation in insects (e.g. Longino, 2019; Gueuning et al., 2020). As such, UCEs are an ideal tool for investigation of the *Notomicrus traili* species complex.

Notomicrus Sharp is a genus of minute aquatic beetles in the family Noteridae (Coleoptera: Adephaga), distributed across the New World, Oceania and Indomalaya. While most New World species present characters conducive to morphology-based species delimitation (e.g. Young, 1978; Manuel, 2015; Guimarães & Ferreira-Jr, 2019; Baca & Short, 2021), the *Notomicrus traili* species complex (hereafter "*traili* complex"; sensu Baca & Short, 2020:8) is a more difficult case. The complex is Neotropical in distribution, ranging from Mexico to Argentina and into the Antilles. Described species in the *traili* complex present subtle interspecific variation in diagnostic characters. Baca & Short (2020), in their molecular phylogenetic estimate, further showed that the characters used for morphological delimitation of

species in this group, such as patterns of dorsal punctation (Young, 1978; Manuel, 2015; Baca & Short, 2018), were sometimes homoplasious, and not consistent indicators of natural groups in this complex. Baca and Short (2020) demonstrated the need for more detailed investigation into the complex. They noted that a high degree of genetic structuring and overlapping distributions among phylogenetically distinct clusters of individuals presents a potential pattern of diversification followed by range expansion across the Neotropics.

As a first step in trying parse the evolutionary patterns and its drivers in the *traili* complex, I here use whole genome resequencing of select noterid taxa to design of a custom, noterid-specific UCE probe set. I then use the resulting captured sequences of UCEs from a broad geographic sampling of individuals within the *traili* complex to (1) gain a high-resolution understanding of the relationships within the *traili* complex and (2) inform taxonomy within the complex.

Material and methods

UCE probe set design (in brief)

Sampling

To design a noterid-specific probe set, we selected four species each representing one of four genera spanning disparate phylogenetic positions within the family (Baca et al., 2017b): *Neohydrocoptus, Liocanthydrus*, *Suphisellus*, and *Hydrocanthus* (Fig. 4.1). DNA was extracted from fresh specimens stored at -20 C in 95% EtOH with a Qiagen DNeasy kit (Qiagen, Hilden, Germany) following manufacturer's protocols modified by the recommendations of Cruaud et al., (2018) to maximize DNA yield. Extracted DNA was quantified on a Quantus flourometer (Promega, Maddison, WI, USA) using QuantiFlour dsDNA reagents. 500ng of DNA was

dehydrated and sent to RAPiD Genomics (Gainesville, FL, USA) for library prep and

sequencing. Library preparation was performed by RAPiD Genomics for Illumina sequencing

utilizing their high-throughput workflow with proprietary chemistry. Briefly, DNA is sheared to

a mean fragment length of 350bp, fragments are end-repaired, followed by incorporation of

unique dual-indexed Illumina adaptors and PCR enrichment. Samples are pooled equimolar and

sequenced using 2x150bp.
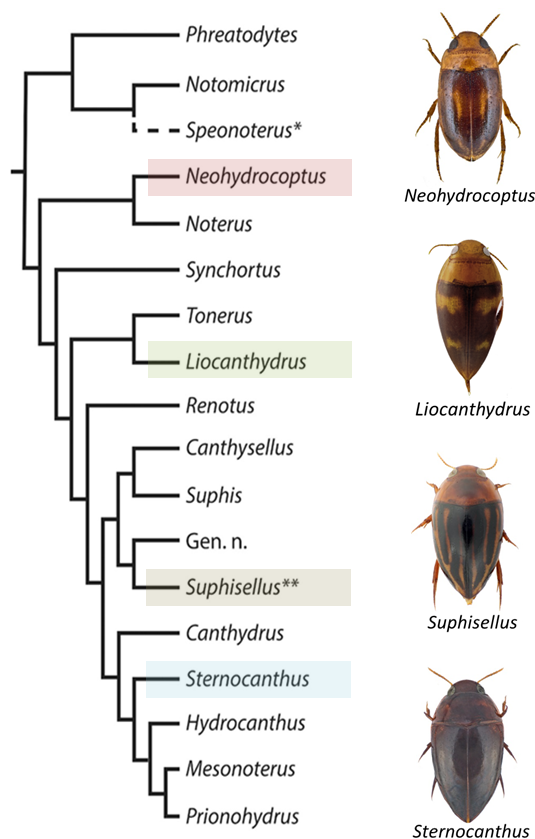
Figure 4.1. Genera sequenced for low-coverage genomes and used as models in probe design. Phylogenetic positions shown on summary tree from Baca et al., 2017b).

Data processing and assembly

Demultiplexed were iteratively trimmed for adapters and contamination with

Trimmomatic (Lohse et al., 2012) via IlluminaProcessor (Faircloth, 2013) and FastP (0.14.1

(Chen et al., 2018) with default settings. Trimmed reads were inspected for quality and adapter

contamination in the fastp outputs and via FastQC (Andrews, 2010), then assembled into contigs and scaffolded with SPAdes 3.13.1 (Bankevich et al., 2012), with default kmer settings (k = 21 – 77) and error correction. Assemblies were assessed by standard output metrics (e.g. scaffold size, N50, L50) calculated by QUAST (Gurevich et al., 2013) and by estimating genome completeness (%) via Benchmarking Universal Single Copy Orthologs (BUSCOs) using BUSCO 3.0 (Simão et al., 2015) with the endopterygota_odb9 ortholog database and the annotated *Tribolium castaneum* (Herbst) (Tribolium Genome Sequencing Consortium, 2008).

Probe set design was conducted with PHYLUCE 1.5 for Python 2.7 (Faircloth, 2016) following Gustafson et al., (2019). *Liocanthydrus* was selected as the base genome following *in silico* testing for the optimal base genome (See Gustafson et al., 2019). Following the Gustafson et al. (2019) finding that the average genetic distance of the base genome from others was negatively correlated with *in silico* data capture performance, we calculating the pairwise genetic distances of Sanger loci from previously sequenced individuals (Baca et al., 2017b; COI, CAD, H3, 16S, 18S, 28S) and from BUSCOs to test for this pattern. The output probe set was then filtered for potential paralogs using a custom R script (R Core Team; Alexander, 2018; Gustafson 2019) and BLAST (Altschul et al., 1990). To further reduce the probe set, probes were selected at random with a custom R script (Alexander 2019, further_whittling_random.R). This was merged with probes from the Adephaga-specific probeset (Adephaga 2.9kv1, Gustafson et al., 2019, 2020), filtered for markers that captured data *in silico* from all four noterid genomes and with probes that overlapped with the new noterid-specific probes. Finally, probes for the aforementioned Sanger markers "legacy markers" used in previous studies of Noteridae (Baca et al., 2017b; COI, CAD, H3, 16S, 18S, 28S) were added, resulting in the Noteridae 3.2kv1 probeset.

Table 4.1. Base genome selection for probe design. Average ranked pairwise genetic distance of each genome to others based on Sanger markers and BUSCOs and number of UCE loci and targeting probes in resultant probe set in base genome tests. Highest values in bold.

Taxon	Sanger markers	BUSCOs	UCE loci	Probes
Neohydrocoptus sp.	3.83	3.4	55,952	430k
Liocanthydrus bicolor	**1.83**	**1.4**	**67,783**	**517k**
Suphisellus puncticicollis	2.5	2.2	54,760	416k
Sternocanthus sp.	**1.83**	3	52,812	400k

UCE data capture in *traili* complex

Taxon sampling

We sampled 44 individuals within the *traili* complex (Table S4.1), in addition to three outgroup taxa representing the three other species groups of New World *Notomicrus* (see Baca et al, 2020; 2021): *N. josiahi* (Miller, 2013); *N. nanulus* (LeConte, 1863); *N.* sp. 7 (*malkini* group, Baca *et al*., 2020). Our sampling within the complex targeted (1) samples of individuals attributable to described species (including paratypes of *N. sabrouxi* Manuel, 2015 and *N. petrareptans* Baca and Short, 2018) or putative species (e.g. *N.* sp. 3, SLE895) and (2) the broadest and highest resolution geographic/population sampling available for these 'species'. Sampling was iterative and additionally guided by initial screening of COI via Sanger sequencing (see below). Sampling was emphasized in northern South America, particularly in the Guiana Shield, following these results.

UCE data capture

Samples were extracted and quantified as above in the UCE probe set design. For an initial phylogenetic screening of tentative samples, Sanger sequence data of the COI mtDNA

marker was captured and sequenced following the methods of Baca *et al.* (2020; ch.2). Between 10ng and 120ng of selected samples were dehydrated and sent to Rapid Genomics for Library prep and enrichment with the tailored Noteridae 3.2kv1 UCE probeset. The probeset was synthesized at 2X coverage (i.e. two copies of each probe) to attempt to maximize depth of sequencing coverage for UCE loci. Library preparation was performed by RAPiD Genomics for Illumina sequencing utilizing their high-throughput workflow with proprietary chemistry. Briefly, DNA is sheared to a mean fragment length of 350bp, fragments are end-repaired, followed by incorporation of unique dual-indexed Illumina adaptors and PCR enrichment. Sequence capture was performed using RAPiD Genomics proprietary chemistry and workflows. Briefly, fully constructed libraries are hybridized to 120bp probes, probe/DNA hybrids are captured on streptavidin beads, washed, and PCR amplified. Samples are pooled equimolar and sequenced using 2x150bp.

UCE data processing

UCE data were processed using PHYLUCE 1.5 (Faircloth, 2016) primarily following the workflows by Faircloth (https://phyluce.readthedocs.io/en/latest/tutorials/index.html) and Alexander, (https://github.com/laninsky/UCE_processing_steps). Raw reads were iteratively trimmed for adapter contamination and quality with Trimmomatic (Bolger et al., 2014) via the Illumiprocessor wrapper (Faircloth, 2013) and fastp (Chen et al., 2018). Read quality both before and after trimming was assessed via the fastp outputs and FastQC (Andrews, 2010). Reads were assembled into contigs using SPAdes (Bankevich et al., 2012) with default paired-end settings and default k values up to 77. Contigs were matched to UCE probes with a minimum percent identity of 80 followed by extraction of UCE loci. UCEs were aligned in PHYLUCE with MAFFT 7.215 (Katoh & Standley, 2013) with the "-- no trim" option, then internally and edge

trimmed via gblocks 0.91 (Talavera & Castresana, 2007). gblocks removes ambiguously aligned sites and gaps and trims alignment edges according to stringency thresholds. This is a common practice used to reduce alignment error or saturated sites. Though generally suggested for deeper evolutionary scales (Faircloth, 2016), visual inspection of select alignments showed potential benefit from internal trimming here. We used relaxed stringency parameters (b1 0.5, b2 0.5, b3 12, b4 7) to avoid removing phylogenetically informative sites while accounting for alignment error. To test the effect of gblocks, we also skipped this internal trimming step and assembled data sets with only alignment edges trimmed ("-- no trim" option omitted in MAFFT alignment). UCE data matrices for both treatments were then constructed at 70% and 90% taxon completeness (i.e. from UCE locus alignments with 70% and 90% representation of total taxon sampling). Missing data statistics for matrices were calculated via custom bash scripts.

Phylogenetic analyses

UCE alignments were concatenated for all four combinations of data trimming and taxon completeness treatments. Maximum likelihood analyses on concatenated datasets were conducted in IQ-TREE 2.0.3 (Minh et al., 2020b) under a GTR+G model of site evolution. Branch supports were assessed with 1000 replicates of ultrafast bootstrap approximation (UFboot; Hoang et al., 2018), with a support value ≥ 95 considered strong support for a given branch topology (Hoang et al., 2018).

Coalescent-based analyses were conducted in summary-tree framework with ASTRAL III 5.6.1 (Zhang et al., 2018), wherein a species tree is constructed from gene trees. To avoid phylogenetic error due to the sensitivity of coalescent-based analyses to missing data (Hosner et al., 2016), only the 90% complete matrices (edge trimmed and internal trimmed) were used. Gene trees were generated with IQ-TREE 2. Best-fit models of site evolution for each locus were

searched in ModelFinder (Kalyaanamoorthy et al., 2017) before subsequent tree inference to account for among-locus evolutionary heterogeneity. Branch support for the ASTRAL species tree was assessed by local posterior probability (Sayyari & Mirarab, 2016).

To assess concordance of our different phylogenetic constructions with their respective data matrices, we calculated gene and site concordance factors (Minh, et al., 2020b) in IQ-TREE 2. Concordance factors quantify the variance of topological support in the individual genes and sites of a dataset. Standard support metrics, such as the bootstrap, estimate the sampling variance of the data and thereby a relatively small difference in the number of supporting sites for a given branch topology can yield high or maximal support (Felsenstein, 1985; Lanfear, 2018), thus potentially inflating these values in large phylogenomic datasets (Lanfear, 2018; Minh et al., 2020). Gene concordance factor (gCF) values quantify the percentage of gene trees that are concordant with a given branch and thus gCF values range from 0% (no genes support branch) to 100% (all genes support branch). Site concordance factor (sCF) values are quartet-based, wherein branch support is quantified as the percentage of randomly sampled parsimony informative sites that support a given branch's topology vs. its other two quartet configurations. A sCF value less than ca. 33% for a given branch would indicate that maximum parsimony favors a different quartet configuration. Note that concordance factors are not intended to displace standard metrics of sampling variance to assess support, nor is there necessarily a threshold for strong support of concordance factors. They merely provide key insight into phylogenomic data sets by quantifying the support of the data for a given tree.

Results

Probe design

Whole genome sequence assembly stats for the four noterid taxa are given in Table 4.2, with an estimated average coverage depth of 7.5–9.2X and a BUSCOs estimated coverage breadth of 67%–85%. Ranked average pairwise genetic distances for Sanger loci and BUSCOs are shown in Table 4.1. *Liocanthydrus bicolor* was shown to have the lowest average genetic distance in both Sanger and BUSCO data. This corresponded with the results of the base genome tests in which candidate loci designed with *Liocanthydrus* as the base genome capture the most loci *in silico* where all model taxa were represented (Table 4.1).

Table 4.2. Stats summary of genome assemblies used in probe set design.

Taxon	Genome size (Megabases)	Capture completeness	Mean coverage (depth)	N50	L50
Neohydrocoptus sp.	350.8	79.3%	8.7X	7162	8720
Liocanthydrus bicolor	297.7	85.8%	9.2X	14047	4428
Suphisellus puncticicollis	383.9	69.3%	7.5X	4518	23916
Sternocanthus sp.	578.7	67.0%	8.2X	5178	22878

The final UCE probeset was designed with the *Liocanthydrus* serving as the base genome. The initial probe design yielded over 50,000 loci. After removal of potential paralogs, this was reduced to ca. 14,000 loci. With random UCE locus selection, the final Noteridae 3.4kv1 probe set targeted the following: 3,198 noterid-specific UCE loci, 11 UCE loci shared with the Adephaga 2.9kv1 probe set (Gustafson et al., 2019; 2020); 171 UCE loci from the Adephaga 2.9kv1 probe set that were captured *in silico* from all four noterid genomes, and five 'legacy' loci. This totaled 3,385 total loci targeted by 28,000 probes; with 2x probe enrichment, 56,000 probes were used in UCE data capture.

UCE data capture

We captured a total of 1,859 UCE loci (excluding 'legacy' markers) across all 47 taxa, with average of 1284.83 loci per taxon, ranging from 871 to 1,377 loci, with a total alignment

length of 1,796,501 bp (after edge trimming). Aligned locus length averaged 966.38bp and

ranged from 101bp to 8,496bp (see Table 4.3).

Table 4.3. Summary stats of UCE data matrices.

Matrix	UCE loci	Total length bp	Min. taxa	Mean taxa	Mean locus length bp	Max. locus length bp	Min. locus length bp	Informative sites	Mean informative sites
Total Edge Trim	1,865	1,796,501	3	31.9	966.4	8,496	101	122,147	65.7
Total gblocks	1,865	1,388,030	3	31.8	744.3	1,921	52	168,874	90.6
70CM-ET	1,252	1,574,620	32	43.3	1257.7	8,496	315	116,200	92.8
70CM-GB	1,252	1,188,199	32	43.3	948.28	1,921	354	159,016	126.9
90CM-ET	999	1,322,239	42	44.6	1323.65	8,496	382	97,877	98.0
90CM-GB	999	982,664	42	44.6	983.7	1,921	485	130,174	130.3

Matrix assemblies included a minimum of 32 (70%) and 42 (90%) of the 47 taxa (Table

4.3). The 70% complete edge-trimmed matrix (70% CM-ET) comprised 1,253 UCE loci with a

total concatenated alignment length of 1,574,620 bp. The mean length of aligned loci was

1257.68 bp, ranging from 315–8,496 bp. The 90% edge-trimmed matrix comprised 999 UCE

loci with a total concatenated length of 1,322,329 bp. The mean length of aligned loci was

1323.65 bp, ranging from 382–8,496 bp (Table 4.3). Missing data among different completeness

levels of the edge-trimmed matrices differed relatively little with the mean missing data per

taxon 13.0% and 11.1% in the 70% CM-ET and 90% CM-ET, respectively (Tables 4.4, S4.2).

Medians of missing data for these matrices were 8.3 and 6.2%. The decrease in mean and

average among these datasets appears to be largely driven by a decrease in missing data among

taxa already lower levels of missing data (Table S4.2). This pattern was also seen in the

prevalence of gaps, with very little change among completeness levels. Generally, the samples

with the largest proportion of missing data and gaps were outgroups and dried museum samples

(Table S4.2).

The removal of ambiguous sites, extensive gaps, and sites with poor taxon representation by gblocks reduced average alignment length significantly, even with relaxed parameters (Tables 4.4, S4.2) The 70% complete gblocks internal-trimmed matrix (70% CM-GB) comprised 1,253 UCE loci with a total length of 1,188,199 bp. The mean length of the aligned loci was 948.28 bp, ranging from 354–1,921 bp. The 90% complete gblocks internal-trimmed matrix (90% CM-GB) comprised 999 UCE loci with a total length of 982,664 bp. The mean length of the aligned loci was 983.65 bp, ranging from 485 bp to 1,921.

The gblocks trimmed datasets showed decreases in missing data compared to their edge-trimmed counterparts at both completeness levels, with mean missing data at 7.3% and 5.0% for the 90% CM-GB and 90% CM-GB datasets, respectively (Table 4.4). The greatest decrease was seen in samples with higher levels of missing data (Tables 4.4, S4.2), with the maximum values of the range greatly decreased. The gblocks matrices showed a similar trend to the edge-trimmed matrices in that levels of completeness seemed to have only a moderate effect (Tables 4.4, S4.2). The greatest effect of internal trimming was seen in the gblocks removal of gaps, with mean gaps in the 70%CMs reduced from 26.2% (median 27.9%) to 8.2% (median 6.0%) and in the 90% CMs from 27.2% (median 28.7%) to 8.6% (Median 5.9%) (Tables 4.4, S4.2).

Table 4.4. Summary of missing data statistics. Values indicate per sample means and medians. In the matrices, missing data was denoted by "?" and gaps by "-". Missing data stats for individual samples can be found in Table S4.2.

Matrix	Total length bp	Mean Missing % (range)	Median Missing %	Mean gap %	Median gap %	Mean missing + gap (range)	Median missing + gap
70CM-ET	1,574,620	13.0% (4.03, 74.4)	8.3%	26.2% (7.4, 29.1)	27.9%	39.2% (33.1, 81.8)	36.3%
70CM-GB	1,188,199	7.2% (2.8, 31.9)	5.2%	8.6% (1.8, 44.2)	6.0%	16% (5.4, 76.1)	11.2%
90CM-ET	1,322,239	11.1% (2.6, 73.2)	6.2%	27.2% (7.9, 30.1)	28.7%	38.3% (3.3, 81.1)	35.2%
90CM-GB	982,664	5.0% (1.2, 27.8)	3.2%	8.6% (1.7, 47.0)	5.9%	13.6 (3.8, 74.8)	9.3%

Performance of analyses

All phylogenetic analyses yielded well-resolved trees with strong support and mostly congruent relationships. Topological conflicts among analyses were largely restricted to the shallowest relationships within populations of densely sampled geographic areas, with most major clades and larger sub-clades/populations in the complex consistently recovered across analyses (Figs. 4.2, S4.1–S4.10). Most conflicts were recovered with poor support across analyses. There was a notable area of conflict among analyses along the backbone, oriented the relative positions of clades 2 and 3 (Figs. 4.2, S4.1–S4.10). Though clades 1–4 themselves were recovered as monophyletic with strong support in concatenated analyses, only the 90CM-GB recovered the placement of clade 2 with strong support (Fig. S4.7, S4.8). ASTRAL analyses recovered well-resolved trees with strong support at most nodes (Figs. 4.2, S4.9, S4.10), and nearly identical topologies with each other. However, these variably conflicted in some areas with the topologies of the concatenated analyses (Figs 4.2, S4.1–S4.10), notably with the monophyly of clade 2, and the placement of its putative constituent taxa, and the placement of clade 3 with respect to its position relative to clade 4.

Concordance Factor analyses in IQ-Tree showed that nearly all recovered relationships recovered the highest gene (gCF) and site concordance (sCF) factors verses the nearest neighbor interchanges (NNIs) for respective branches. Concordance factors ranged across trees from relatively low values for some of shallowest relationships, with gCF values dropping below 1.00 and sCF below 30 in population level relationships (Figs. S4.11–S4.17), to high values for the conspecific sister relationships (e.g. *N. petrareptans,* gCF > 85; sCF >90). Very few sCF values fell below the ~33% value expected for a random alignment of sufficient length (Figs. S4.11–S4.17).

Average gCF and sCF did not indicate preference for a distinct data analysis. However, concerning the conflict in backbone placements of clades 2 and 3, concordance factors did show a stronger preference for the topology recovered by the 70CM-ET and 90CM-ET analysis Figs. S4.11, S4.13, wherein clade 2 was found sister to clade 3 + clade 4. In the case of the 90CM-GB (Fig. S4.14), which recovered a conflicting topology to that of 70CM-ET and 90CM-ET with high support, concordance factors recovered higher values for the discordant (NNIs) topologies,

with higher CF values for the alternative topology congruent with the 70CM-ET and 90CM-ET

analyses (Tables S4.4–S4.9).
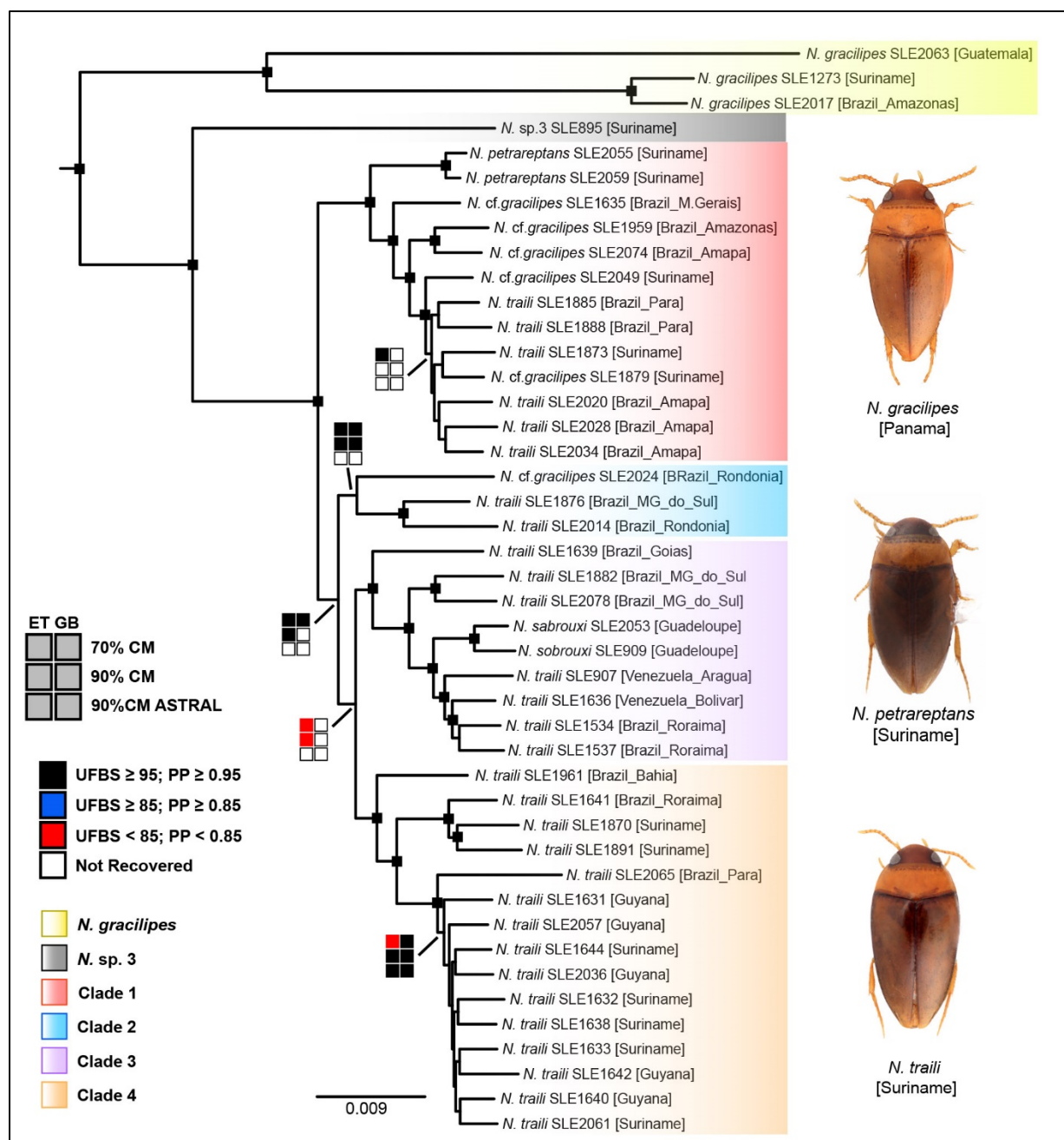
Figure 4.2. Phylogeny of the *Notomicrus traili* complex. Tree based on Maximum likelihood analysis of the 70%CM-ET dataset in IQ-Tree. Black boxes at nodes indicate high support for the depicted relationships across all analyses. 6-box configuration shows nodes with low support or conflicting topologies among analyses; legend at bottom left depicts corresponding analyses of boxes; color indicates support. ET = edge trim treatment; GB = gblocks treatment; 70% CM = 70% completeness matrix; 90%CM = 90% complete matrix.

Phylogenetics

All analyses recovered *N. gracilipes* and a tentative undescribed species, *N.* sp. 3, as successive sisters to the rest of the *traili* complex (Fig. 4.2). The main complex comprised four clades recovered across all concatenated analyses. Clade 1 is sister to the others. This includes *N. petrareptans*, a seep-dwelling species endemic to Suriname, sister to a clade extending from Minas Gerais, Brazil northeast to Para, Amapá, and Suriname. Though somewhat variable among analyses, there was generally a geographic correlate to the sub-population grouping in this and other clades in the tree (Figs. 4.2, 4.3). Clade 2 comprises three individuals from Western Brazil, Rondônia and Mato Grosso do Sul, with the two Rondônia individuals recovered as paraphyletic with respect to the individual from Mato Grosso do Sul (Fig. 4.3). Clade 3 consists of a complex extending from the Brazilian states of Goiás and Mato Grosso do Sul, north to northern Roraima (Brazil) and Venezuela. *Notomicrus sabrouxi*, from the lesser antillean island of Guadeloupe, was consistently found nested in this clade, sister to the Roraima and Venezuela populations. Clade 4 comprises several subclades/populations. A single individual from Bahia, Brazil is recovered as sister to the rest of the clade. A population from Roraima and Suriname and an individual from Para, just south of the Amazon River near the type locality of *N. traili*, are successive sisters to a densely sampled population from the rainforests of Guyana and Suriname.

The relative relationships among these four clades varied across analyses. Both edge-trimmed analyses (70CM-ET, 90CM-ET) recovered the same topology (Figs. 4.2, S4.11,S4.13) at this branch with clade 2 sister to clade 3 + clade 4, with poor support, while the gblocks internal-trimmed analyses (70CM-GB; 90CM-GB) recovered trees with variable positions of these clades (Figs 4.2; S4.12, S4.14). Notably, the 90CM-GB analysis recovered clade 2 sister to

clade 4, with strong support, and clade 1 and clade 3 as sisters, with poor support (Fig. S4.14).

ASTRAL analyses did not recover clade 2 as monophyletic, instead finding a single individual

sister to clade 2, with clade 3 sister to that grouping. The other two members of clade 2 were

recovered as monophyletic, sister to clade 1 (Fig. S4.9, S4.10).

Discussion

Performance of probe design and UCE data capture

The outcome of our probe design, with the high number of UCE loci identified (over 14k

before random reduction), was not unexpected. The blind-to-locus-identity nature of UCE locus

identification, which orients solely around the relative conservation in given portions of the

model genomes, coupled with the specificity (shallow evolutionary scale) of our probe design,

should produce far more putative UCE loci than in other systems bridging greater scales of

divergence in insects (e.g. Faircloth et al., 2015, Hymenoptera; Faircloth et al,. 2017, Coleoptera,

and others; Gustafson et al., 2019; 2020, Coleoptera: Adephaga). This specificity could have an

increased potential for recovering multiple UCEs from the same genomic locus, thus having

linked, congenic loci in the dataset (Van Dam et al., 2021). In future bait design, it would be

desirable to identify congenic loci during the bait design phase, or concatenate data of identified

congenic UCE loci. Van Dam et al., (2021), showed increased phylogenetic performance

implementing this practice, but the methods therein require an annotated genome which is

unavailable for Noteridae or its parent suborder Adephaga, thus limiting the number of loci that

can be identified and merged (see Baca et al., in review).

The UCE data capture of *Notomicrus* species and the *traili* complex with a tailored UCE

probe set exceeded previous investigations of Adephaga (Baca et al., 2017; in review; Gustafson

et al, 2020), even in absence of a *Notomicrus* model genome in probe design. Missing data, both in terms of per-locus taxon representation and missing sites within locus, was greatly reduced relative to these previous UCE studies (Table S4.2), even with tailored probe sets (Baca et al., 2017; in review; Gustafson et al., 2020). It is difficult to assess the contributing effect of 2x probe coverage in UCE enrichment (i.e. two copies of each probe) in the overall success of UCE data capture in *Notomicrus*, especially given the family-level focus at which we tailored our probe set. Overall, we showed that incorporating low coverage whole genome sequences for tailored probe set design can yield strong benefits for UCE data capture.

Phylogenetic performance

Concatenated phylogenetic reconstructions consistently recovered the same major clades, with relatively congruent relationships within those clades. The conflict in backbone relationships appears to be driven in part by the implementation of internal-trimming by gblocks, which significantly decreased the overall alignment length (Table 4.3, 4.4), even under the relaxed trimming regime. Given the relative conservation of UCE loci and the shallow evolutionary focus of the investigation, large areas of alignment ambiguity due to saturation would not be expected, so it is possible that alignment in MAFFT may have opened large gap sections which were removed by gblocks (Table 4.4, S4.2). The relatively greater genetic distance of the outgroups (e.g. *N. josiahi*) to the complex may have also opened large gaps due to indels or non-overlapping representation in loci. In that vein it is not surprising that outgroups and dried museum specimens showed the greatest levels of missing data (Table S4.2). With the *traili* samples closely related and composing the bulk of the dataset, one would expect a high degree of conservation not shared with outgroups, thus limiting their representation in the dataset. With degraded DNA, museum specimens are known to be recovered with more

fragmented data (Blaimer et al., 2016). A visual inspection of select loci showed the presence of

an apparent combination of these effects. By default, the PHYLUCE wrapper uses MAFFT

defaults, which calls on the FFT-NS-2 alignment algorithm, which sacrifices some accuracy for

speed. MAFFT does include other, more accurate algorithms, which might be more

accommodating to UCEs.

In the vein of gblocks, it appears that ASTRAL analyses were only minimally affected by

internal trimming. This perhaps shows that the decisive gene-trees were relatively unaffected by

internal trimming. With respect to backbone conflict, concordance factors at these nodes do not

strongly favor the recovered topologies, with gCF and sCF values being nearly equal, sometimes

slightly lower than the NNI discordance topologies (Tables S4.4–S4.9).

Phylogenetics of the traili complex

Relationships

We discuss relationships as recovered by the IQ-Tree analyses of the 70CM-ET dataset

(Fig 4.2) unless otherwise specified as we recovered topological congruence among the edge-

trimmed datasets at both levels of matrix completeness (70CM-ET and 90CM-ET) and

concordance factors favored these most over others. The recovered topology within the *traili*

complex is congruent in its deeper relationships with that of Baca and Short, 2020 (ch. 2).

Notomicrus gracilipes is found sister to the rest of the complex. The species itself appears to

extend from northern South America and into Guatemala, its which contains its type locality.

With no sampling of the intermediate areas it is difficult to assess whether these populations are

still in contact.

The relationships of the main complex are variably congruent with the results of Baca

and Short (2020), especially with respect to clade 1 in which *N. petrareptans* is recovered as

sister to a widespread group. While *N. petrareptans* is known only from a single hygropetric

seep habitat in Suriname, many individuals of this clade were collected from within largely open,

scrub and grassland areas, from both lentic and lotic environments. Aside from *N. petrareptans*,

the clade is widespread with disjunct populations (Fig 4.3). Though this apparent fragmentation

is likely exacerbated by disjunct sampling, members occupying the same region (e.g. individuals

from Amapá) are not monophyletic with respect to other regions. However, the short internodes

of the tree make it difficult to discern if the non-monophyly is due to genetic isolation of

multiple distinct populations, or if the tree is depicting patterns of genealogical inheritance (e.g.

via introgression) within a large population. Baca and Short (2020) also recovered clade 3, which

includes several disjunct populations from south-central Brazil (Goiás, Mato Grosso do Sul),

Guadeloupe, (lesser Antilles), and Venezuela and Roraima. It is difficult to assess to what extent

this is an effect of sampling, though individuals in other *traili* clades were collected in many

intermittent areas, including those near sampling localities of members of clade 2. Clades 2 and

4 were not recovered by Baca and Short, 2020 and while data resolution of UCEs versus Sanger

no doubt plays a role, preliminary trees based on COI data shows potential for the effect of

sampling also. Clade 2 is the smallest of the four major clades, comprising only individuals from

western Brazil. Clade 4 consisted of a single individual from Bahia sister to two distinct

Northern Amazon/Guiana Shield populations. Notably, the individual from Para (SLE2065), was

a dried museum specimen collected near the type locality of *N. traili*.

Biogeography

The recovered topology depicts a repeated pattern in which northern South American

populations have one or more successive south central or western Brazilian sisters. To a large

extent the biogeographic pattern depends on the topology of the tree. The backbone of the

preferred tree (70CM-ET, 90CM-ET; Figs 4.3, 4.4, S4.1, S4.3), with clade 2 sister to 3+4, would

be the expected pattern if these three clades' ancestral range was to the south. To some extent

this signal is still preserved in the other concatenated rearrangements, including in the ASTRAL

trees. In any case, the trees are consistent with a process of diversification followed by range

expansion, even as far north as Guatemala in the case of *N. gracilipes*.

With Baca and Short, 2020 estimating the crown age of the main complex at ca. 7.7 ma,

the evolution of this group would be subject to the dynamic environment of South American in

the late Miocene and Pliocene, where the Guiana and Amazon Shield regions provided

relatively stable, though potentially fragmented environments, and during the Pleistocene, where
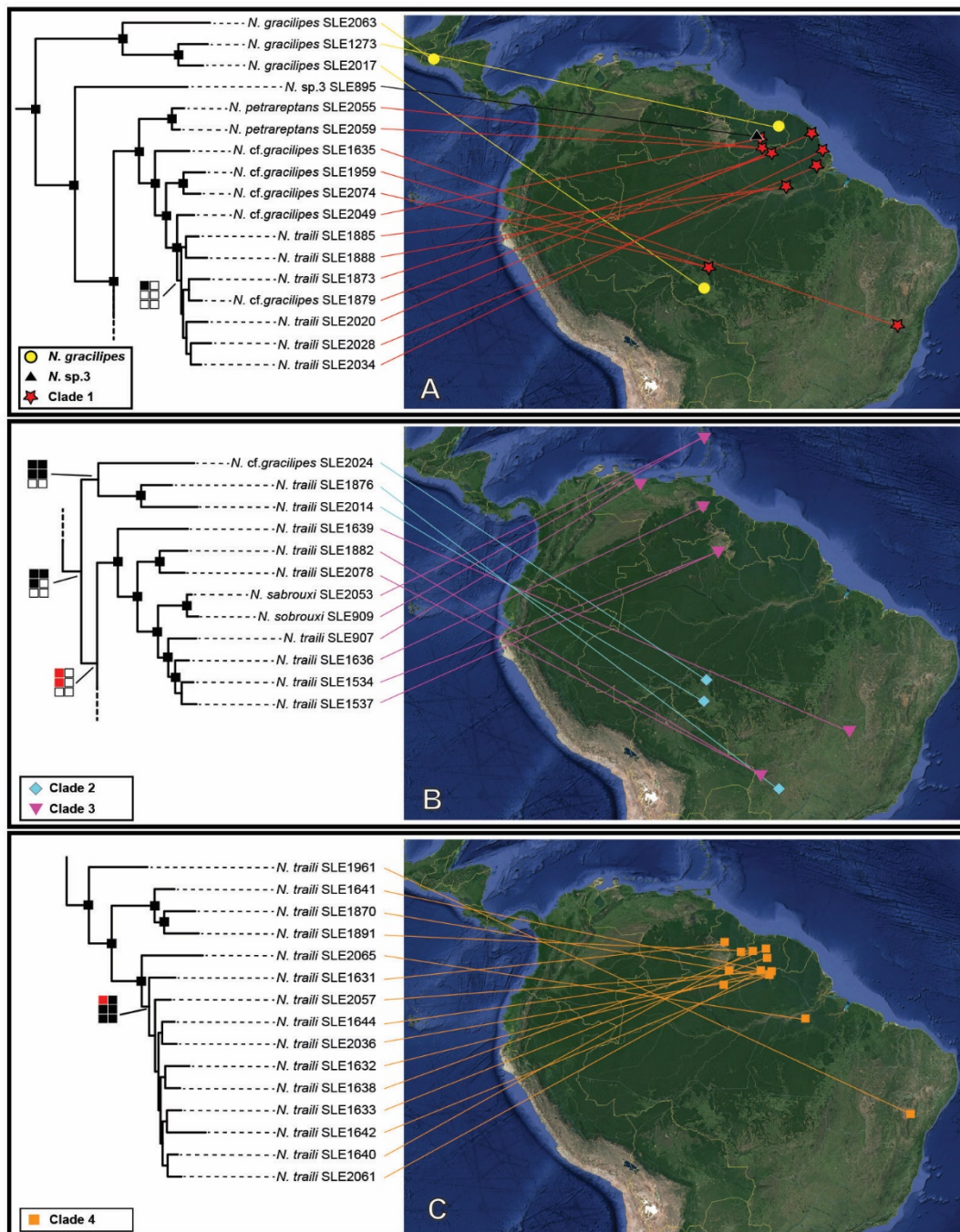
Figure 4.3. Clades of *N. traili* complex with samples mapped to source localities. Legend at bottom left indicates clade. Branch lengths have been altered to highlight relationships.

glaciation cycles drove repeated expansion and reduction of rainforest (Hoorn & Wesselingh, 2011; Baker et al., 2020), all potential drivers of the recovered phylogenetic patterns as inferred in other systems (Hoorn & Wesselingh, 2011; Baker et al., 2020 and citations therein; see also da

Rocha & Kaefer, 2019). However, a biogeographic investigation following divergence time estimation would be necessary to move beyond speculation of the directionality, tempo, or mode of evolution in this complex.

Taxonomic implications

The relationships recovered show that previous hypotheses of species' delimitations in this group (Young, 1978; Manuel, 2015; Baca and Short, 2018) are not consistent with the recovered phylogeny. The evolution of patterns of punctation provide a prime example. Historically, *Notomicrus gracilipes* was diagnosed from *N. traili* and others by presenting punctation along the anteromedial margin of the elytra, along the suture (Fig. 4.5). As Fig. 4.2 depicts, specimens attributable to *N. gracilipes* (specimens labeled *N.* cf. *gracilipes*) are polyphyletic, showing that this character is a poor indicator of species boundaries. Personal observation of the specimens also shows this character presents variably within clades. Thus, while *N. gracilipes* appears to be a valid species, sister to the rest of the *traili* group, the character formally used in identification is here shown unreliable. With other identifying characters such as in the aedeagus or male protarsal claws showing subtlety in how they vary

among individuals, careful observation will be required to find correlating patterns of character

state distributions across the traili phylogeny.
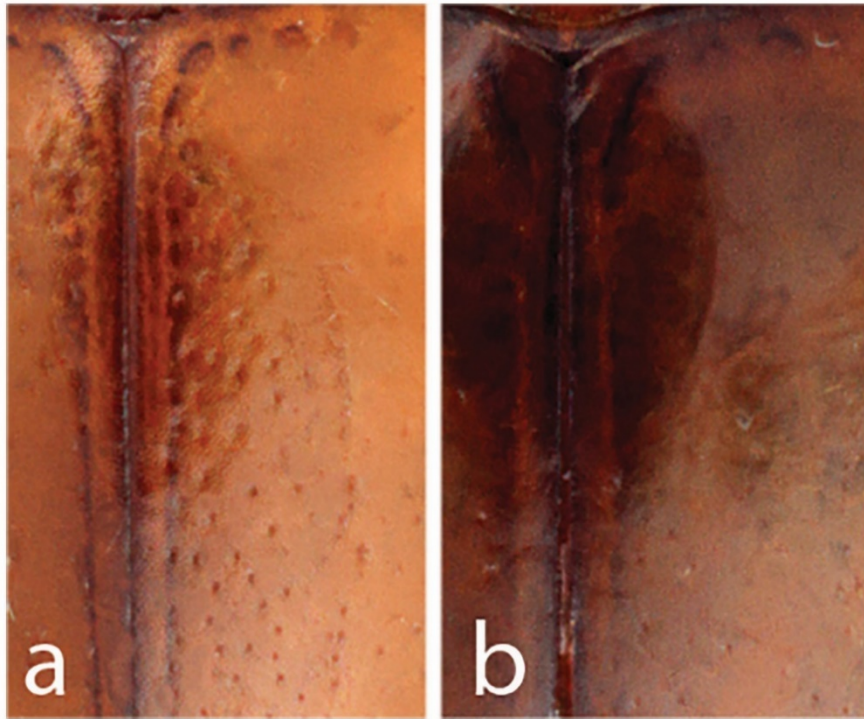
Figure 4.5. Detail of anteromedial portion of the elytra showing differing patterns of punctation historically used to delimit *traili* species. **a** *N. gracilipes* (Panama); **b** *N. traili* (Suriname).

The current investigation highlights the need for more rigorous testing and data-driven

species delimitation in the complex. Even then, parsing between species and populations may be

difficult. To the extent that can be inferred, the four currently valid species do not appear to

require synonymy. Instead, it appears there are likely six or more species in this clade. Should

further investigation support the intuitive appearance of more than the described species,

properly attributing the name *N. traili* Sharp, 1882 to a clade will require careful investigation.

The species was described from a single female type specimen (Sharp, 1882; Nilsson, 2011).

Females of *Notomicrus* lack the often rapidly evolving morphology of the aedeagus and

secondary sexual modifications of the protarsal claws presented by males and often useful in

morphological species delimitation. Observations of the female genitalia (Manuel, 2015;

personal observations) show potential for characterizing lineages at the genus level in *Notomicrus*, but these characters appear to be too conserved to aid reliably in species delimitation. A further complication will be whether to resurrect *N. grouvellei* Régimbart, 1895 from synonymy with *N. traili*. This species was described from "Brésil: Mato Grosso" (Régimbart, 1895:18), an already large Brazilian state, which at that time included Mato Grosso do Sul.

A tentative undescribed species *N.* sp. 3 is clearly distinct from other members of the complex, to date it is known only from a few samples, but presents a distinct aedeagal morphology making readily identifiable with dissection. Externally, it appears similar to *N. gracilipes* or other punctate members.

Future directions

Our UCE dataset provides a wealth of data to be applied to various methods and hypothesis testing. The most pertinent need is to limit alignment error by running MAFFT outside of the PHYLUCE wrapper in order to call a more arduous algorithm and adjust parameters. This will ensure greater confidence in downstream analyses and hopefully provide more stability in reconstructions. Downstream data processing will include phasing UCE data (Andermann et al., 2018) to retain information of heterozygous alleles and harvest SNPs using PHYLUCE and other software (e.g. PGDSpider, Lischer & Excoffier, 2012) and use these to examine genetic clustering and guide species delimitation analyses. Further, with the geographic signal of these analyses and the dynamic nature of South American geography, it is very desirable to include a biogeographic reconstruction.

Conclusion

With UCEs, we have successfully reconstructed phylogenies at the species and population level within the minute aquatic beetle genus *Notomicrus.* Our reconstructions show that current classification does appropriately delimit species in the *traili* complex, which by all appearances contains more species than currently described. Further investigation will be needed to make taxonomic classification congruent with the data of this group. Our investigation has further revealed a complex biogeographic pattern that suggests repeated diversification and range expansion, potentially due to Pleistocene glaciation cycles.

Overall, we (1) reinforce the findings of others (Gustafson et al., 2020) that tailored probe design greatly improves UCE data capture and (2) that UCE are useful in the investigative scope of the species and population levels. With a wealth of methodological possibilities to which this data can be applied, we are certain to gain more detailed understanding of *Notomicrus* diversity and South American evolutionary dynamics.

References

Alexander, A. (2018). comparing_monolithic_UCE_fastas v0.0. Available from: https://github.com/laninsky/comparing_monolithic_UCE_fastas

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andermann, T., Fernandes, A. M., Olsson, U., Töpel, M., Pfeil, B., Oxelman, B., Aleixo, A., Faircloth, B. C., & Antonelli, A. (2018). *Allele Phasing Greatly Improves the Phylogenetic Utility of Ultraconserved Elements* [Preprint]. Evolutionary Biology. https://doi.org/10.1101/255752

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available

854 online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. [accessed on 26

855 March 2021].

Baca, S. M., & Short, A. E. Z. (2018). Notomicrus petrareptans sp. n., a new seep-dwelling

species of Noteridae from Suriname (Coleoptera: Adephaga). *Zootaxa*, *4388*(2), 182–190.

https://doi.org/10.11646/zootaxa.4388.2.2

Baca, S. M., & Short, A. E. Z. (2020). Molecular Phylogeny of the Notomicrine Water Beetles

(Coleoptera: Noteridae) Reveals Signatures of Gondwanan Vicariance and Ecological Plasticity.

Insect Systematics and Diversity, *4*(4). https://doi.org/10.1093/isd/ixaa015

Baca, S. M., & Short, A. E. Z. (2021). Review of the New World Notomicrus Sharp (Coleoptera,

Noteridae) I: Circumscription of species groups and review of the josiahi group with description

of a new species from Brazil. *ZooKeys*, *1025*, 177–201.

https://doi.org/10.3897/zookeys.1025.60442

Baca, S. M., Alexander, A., Gustafson, G. T., & Short, A. E. Z. (2017a). Ultraconserved

elements show utility in phylogenetic inference of Adephaga (Coleoptera) and suggest paraphyly

of 'Hydradephaga.' *Systematic Entomology*, *42*(4), 786–795. https://doi.org/10.1111/syen.12244

Baca, S. M., Gustafson, G. T., Alexander, A. M., Gough, H. M., Toussaint., E. A. F. (In

Review). Integrative phylogenomics reveals a Permian origin of Adephaga beetles. *Systematic

Entomology*.

Baca, S. M., Toussaint, E. F. A., Miller, K. B., & Short, A. E. Z. (2017b). Molecular phylogeny

of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on data

partitioning strategies. *Molecular Phylogenetics and Evolution*, *107*, 282–292.

https://doi.org/10.1016/j.ympev.2016.10.016

Baker, P. A., Fritz, S. C., Battisti, D. S., Dick, C. W., Vargas, O. M., Asner, G. P., Martin, R. E., Wheatley, A., & Prates, I. (2020). Beyond Refugia: New Insights on Quaternary Climate Variation and the Evolution of Biotic Diversity in Tropical South America. In V. Rull & A. C. Carnaval (Eds.), *Neotropical Diversification: Patterns and Processes* (pp. 51–70). Springer International Publishing. https://doi.org/10.1007/978-3-030-31167-4_3

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

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*(8), e0161531. https://doi.org/10.1371/journal.pone.0161531

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., Gates, M. W., Kula, R. R., & Brady, S. G. (2017). Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. *Current Biology*, *27*(7), 1019–1025. https://doi.org/10.1016/j.cub.2017.03.027

Branstetter, M. G., & Longino, J. T. (2019). Ultra-Conserved Element Phylogenomics of New World Ponera (Hymenoptera: Formicidae) Illuminates the Origin and Phylogeographic History

of the Endemic Exotic Ant Ponera exotica. *Insect Systematics and Diversity*, *3*(1).

https://doi.org/10.1093/isd/ixz001

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ

preprocessor. *Bioinformatics*, *34*(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., Huber, J., Polaszek, A., &

Rasplus, J.-Y. (2018). *Optimised DNA extraction and library preparation for minute arthropods:*

Application to target enrichment in chalcid wasps used for biocontrol [Preprint]. Evolutionary

Biology. https://doi.org/10.1101/437590

da Rocha, D. G., & Kaefer, I. L. (2019). What has become of the refugia hypothesis to explain

biological diversity in Amazonia? Ecology and Evolution, 9(7), 4302–4309.

https://doi.org/10.1002/ece3.5051

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic

loci. *Bioinformatics (Oxford, England)*, *32*(5), 786–788.

https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015). Target enrichment of

ultraconserved elements from arthropods provides a genomic perspective on relationships among

H ymenoptera. *Molecular Ecology Resources*, *15*(3), 489–501. https://doi.org/10.1111/1755-

0998.12328

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn,

T. C. (2012). Ultraconserved Elements Anchor Thousands of Genetic Markers Spanning

Multiple Evolutionary Timescales. *Systematic Biology*, *61*(5), 717–726.

https://doi.org/10.1093/sysbio/sys004

Faircloth, B. C., Sorenson, L., Santini, F., & Alfaro, M. E. (2013). A Phylogenomic Perspective on the Radiation of Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *PLoS ONE*, *8*(6), e65923. https://doi.org/10.1371/journal.pone.0065923

Gueuning, M., Frey, J. E., & Praz, C. (2020). Ultraconserved yet informative for species delimitation: Ultraconserved elements resolve long-standing systematic enigma in Central European bees. *Molecular Ecology*, *29*(21), 4203–4220. https://doi.org/10.1111/mec.15629

Guimarães, B. a. C., & Ferreira-Jr, N. (2019). Two new species and new records of Notomicrus Sharp, 1882 (Coleoptera: Noteridae) from Brazil. *Zootaxa*, *4629*(2), 263–270. https://doi.org/10.11646/zootaxa.4629.2.8

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gustafson, G. T., Alexander, A., Sproul, J. S., Pflug, J. M., Maddison, D. R., & Short, A. E. Z. (2019). Ultraconserved element (UCE) probe set design: Base genome and initial design parameters critical for optimization. *Ecology and Evolution*, *9*(12), 6933–6948. https://doi.org/10.1002/ece3.5260

Gustafson, G. T., Baca, S. M., Alexander, A. M., & Short, A. E. Z. (2020). Phylogenomic analysis of the beetle suborder Adephaga with comparison of tailored and generalized ultraconserved element probe performance. *Systematic Entomology*, *45*(3), 552–570. https://doi.org/10.1111/syen.12413

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology*, *65*(5), 910–924. https://doi.org/10.1093/sysbio/syw036

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Hoorn, C., & Wesselingh, F. (2011). *Amazonia: Landscape and Species Evolution: A Look into the Past*. John Wiley & Sons.

Hosner, P. A., Faircloth, B. C., Glenn, T. C., Braun, E. L., & Kimball, R. T. (2016). Avoiding Missing Data Biases in Phylogenomic Inference: An Empirical Study in the Landfowl (Aves: Galliformes). *Molecular Biology and Evolution*, *33*(4), 1110–1125. https://doi.org/10.1093/molbev/msv347

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

LeConte JL (1863) New species of North American Coleoptera. Part1. Smithsonian Miscellaneous Collections 6(167): 1-92.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M., & Usadel, B. (2012). RobiNA: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Research*, *40*(W1), W622–W627. https://doi.org/10.1093/nar/gks540

Manthey, J. D., Campillo, L. C., Burns, K. J., & Moyle, R. G. (2016). Comparison of Target-Capture and Restriction-Site Associated DNA Sequencing for Phylogenomics: A Test in Cardinalid Tanagers (Aves, Genus: *Piranga*). *Systematic Biology*, *65*(4), 640–650. https://doi.org/10.1093/sysbio/syw005

Manuel, M. (2015). The genus Notomicrus in Guadeloupe, with description of three new species (Coleoptera: Noteridae). *Zootaxa*, *4018*(4), 506–534. https://doi.org/10.11646/zootaxa.4018.4.2

Miller, K. B. (2013). Notomicrus josiahi , a new species of Noteridae (Coleoptera) from Venezuela. *Zootaxa*, *3609*(2), 243–247. https://doi.org/10.11646/zootaxa.3609.2.11

Minh, B. Q., Hahn, M. W., & Lanfear, R. (2020). New Methods to Calculate Concordance Factors for Phylogenomic Datasets. *Molecular Biology and Evolution*, *37*(9), 2727–2733. https://doi.org/10.1093/molbev/msaa106

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

R Core Team. (2015). R: A language and environment for statistical computing. http://www.R-project.org/. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Sayyari, E., & Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. *Molecular Biology and Evolution*, *33*(7), 1654–1668. https://doi.org/10.1093/molbev/msw079

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology*, *63*(1), 83–95. https://doi.org/10.1093/sysbio/syt061

Talavera, G., & Castresana, J. (2007). Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, *56*(4), 564–577. https://doi.org/10.1080/10635150701472164

Tribolium Genome Sequencing Consortium. (2008). The genome of the model beetle and pest Tribolium castaneum. *Nature*, *452*(7190), 949–955. https://doi.org/10.1038/nature06784

Van Dam, M. H., Henderson, J. B., Esposito, L., & Trautwein, M. (2021). Genomic Characterization and Curation of UCEs Improves Species Tree Reconstruction. *Systematic Biology*, *70*(2), 307–321. https://doi.org/10.1093/sysbio/syaa063

Young, F. N. (1978). The New World species of the water-beetle genus Notomicrus (Noteridae)*. *Systematic Entomology*, *3*(3), 285–293. https://doi.org/10.1111/j.1365-3113.1978.tb00121.x

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 153. https://doi.org/10.1186/s12859-018-2129-y

Supplementary materials

Chapter 1.

Supplementary Figures
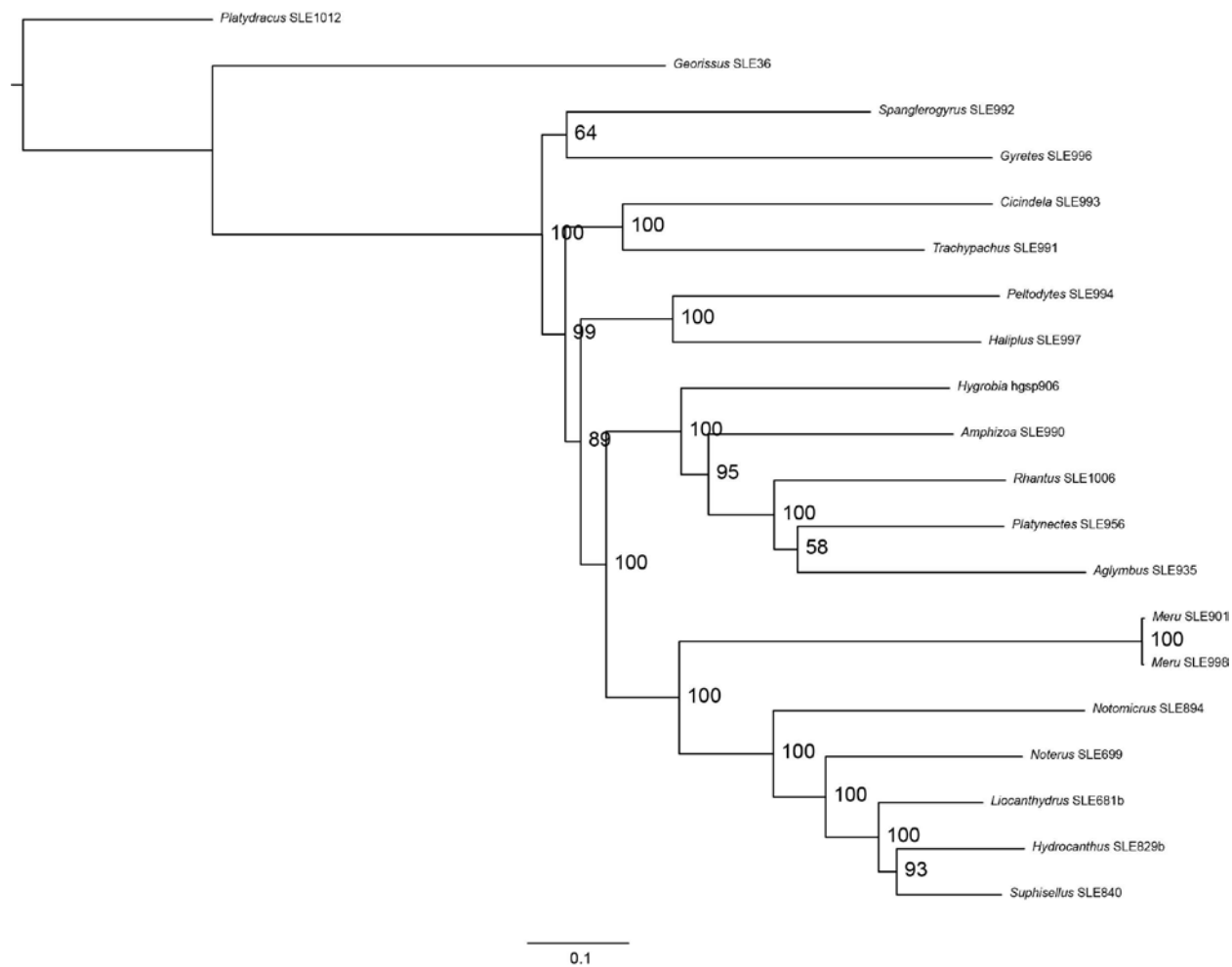
Figure S1. Tree recovered by Maximum Likelihood analysis of unpartitioned concatenated UCE dataset. Values at nodes indicate bootstrap support.
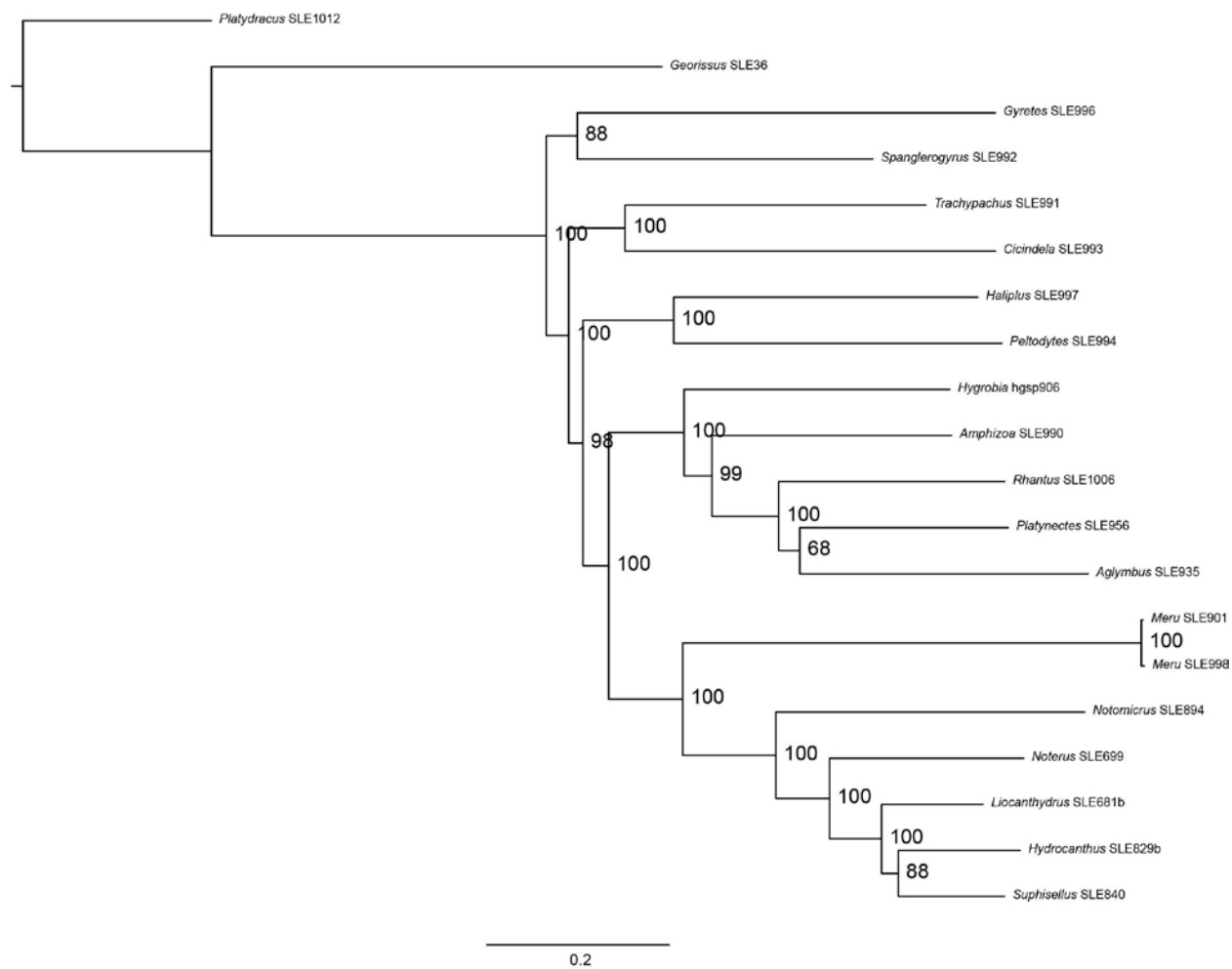
Figure S2. Tree recovered by Maximum Likelihood analysis of concatenated dataset partitioned using CloudForest. Values at nodes indicate bootstrap support.
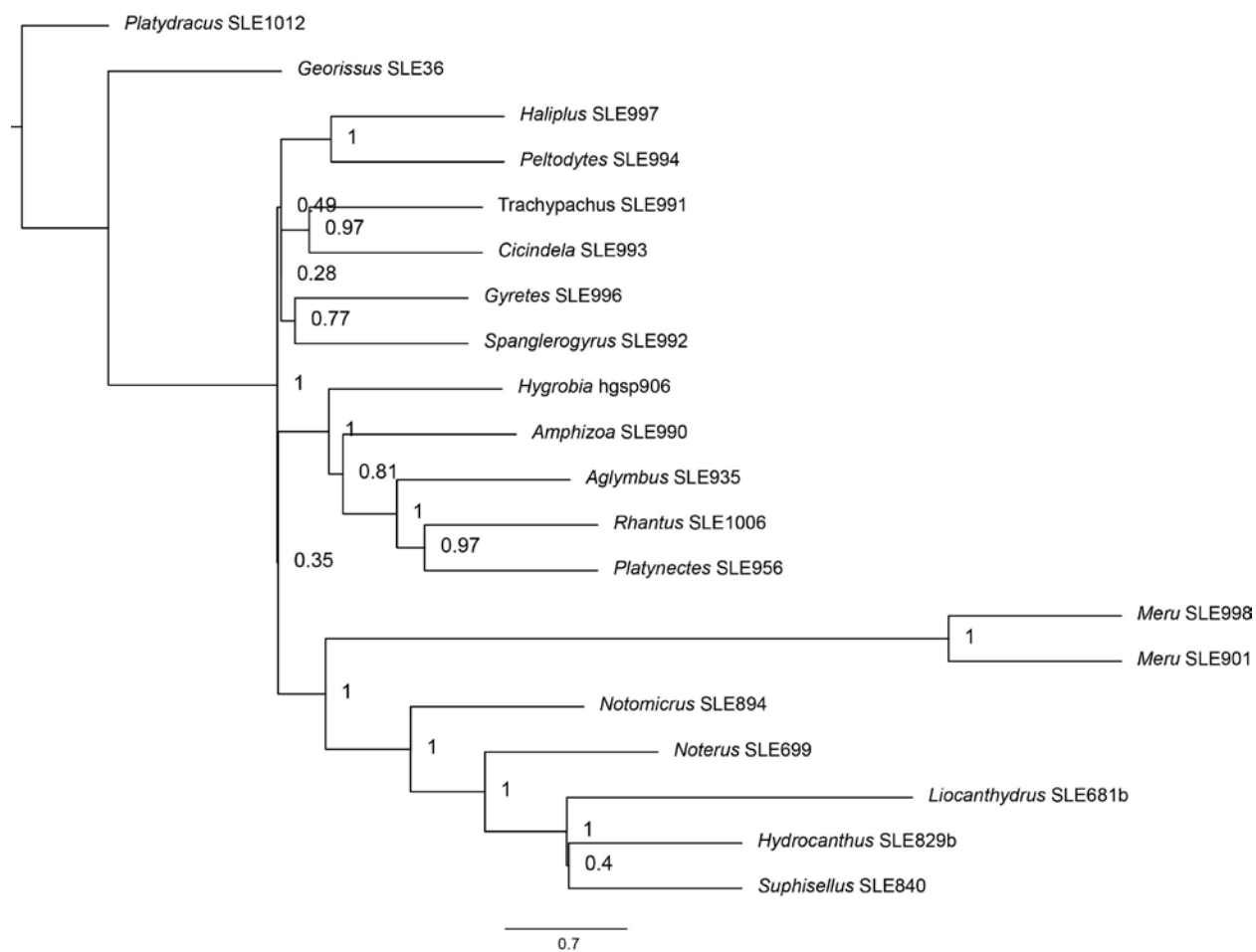
Figure S3. Species tree recovered by ASTRAL II analysis of gene trees generated by CloudForest. Values at nodes indicate posterior probability.
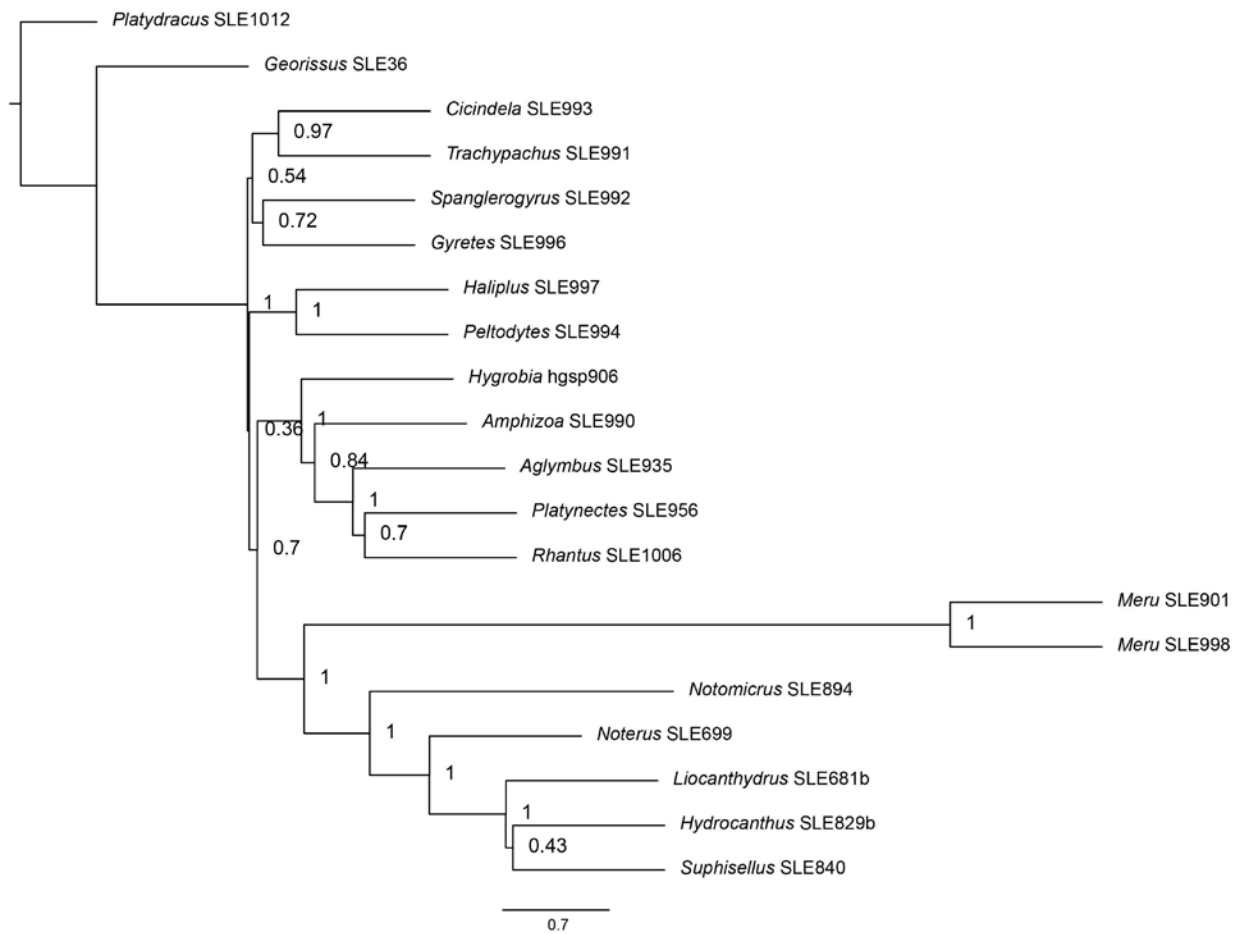
Figure S4. Species tree recovered by ASTRAL II analysis of gene trees generated by RAxML. Values at nodes indicate posterior probability.
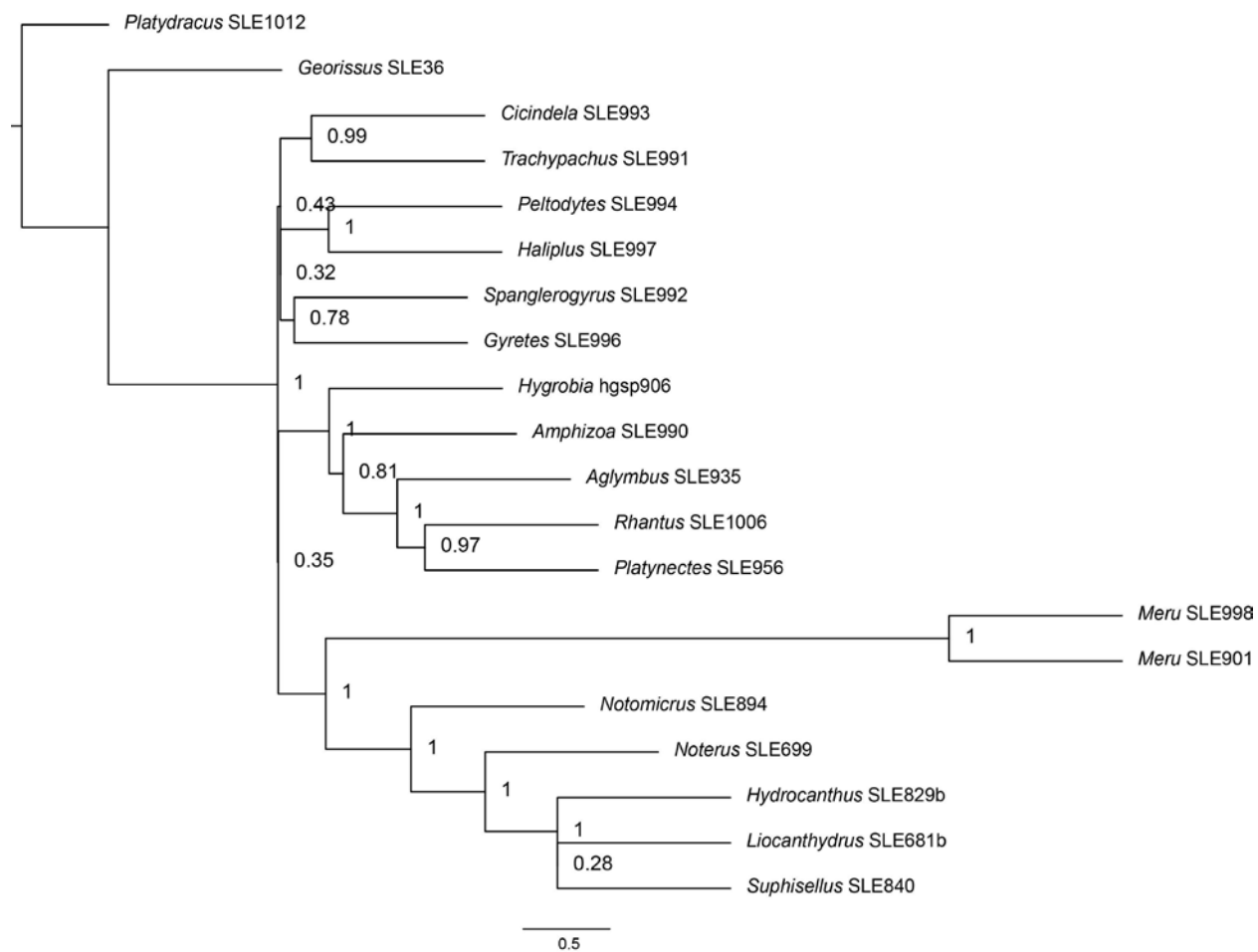
Figure S5. Species tree recovered by ASTRID analysis of gene trees generated by CloudForest. Values at nodes indicate posterior probability.
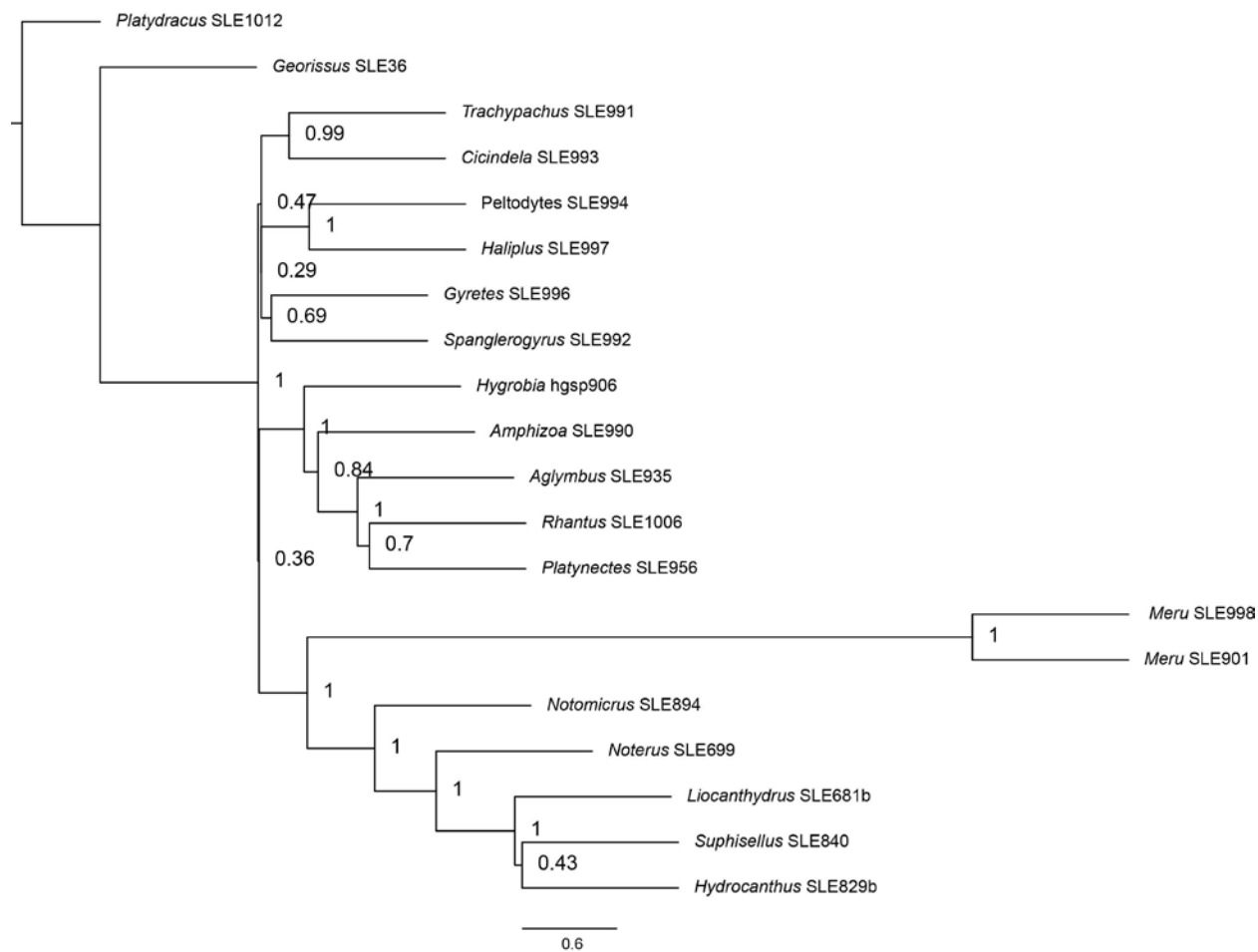
Figure S6. Species tree recovered by ASTRID analysis of gene trees generated by RAxML. Values at nodes indicate posterior probability
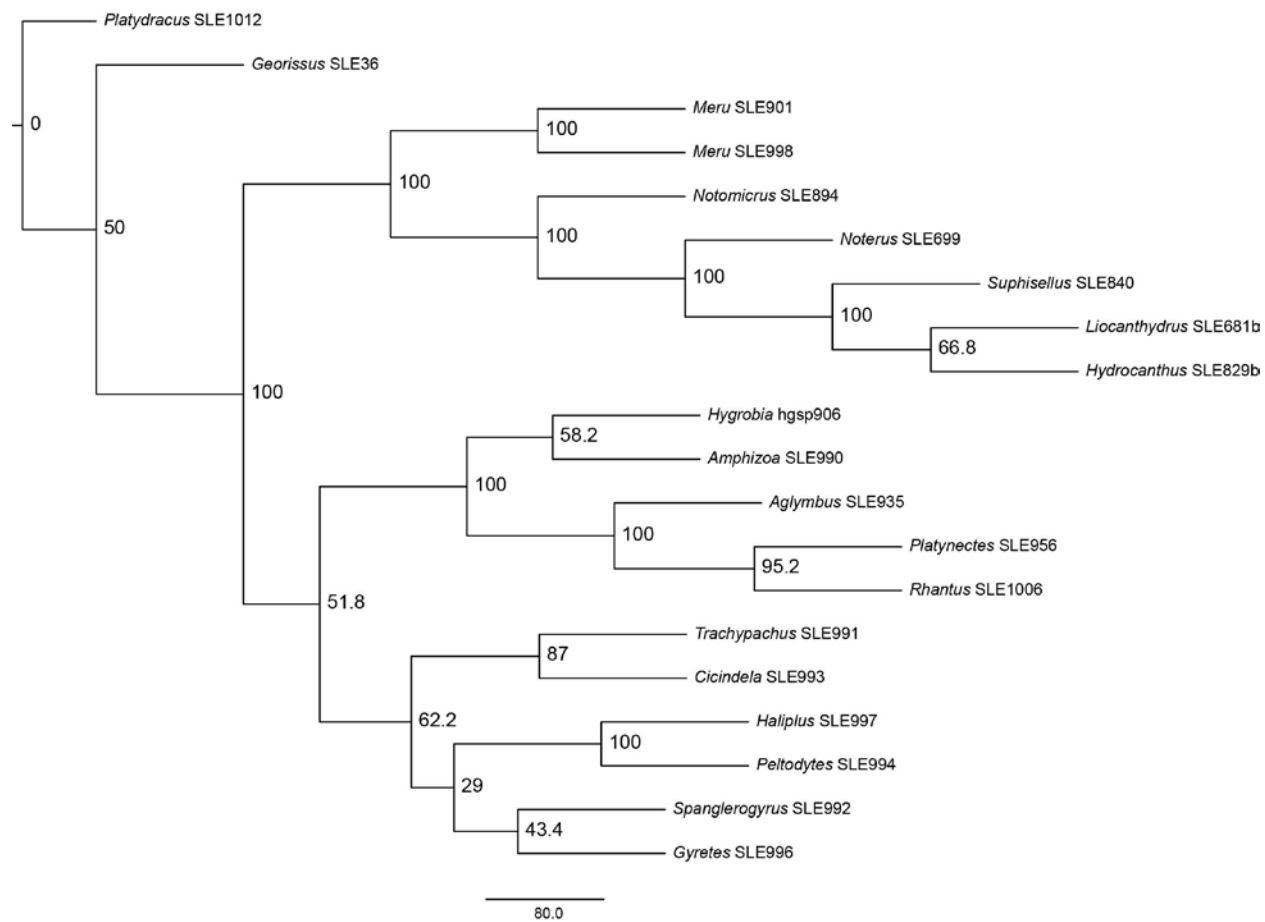
Figure S7. Species tree recovered by SVDquartets. Values at nodes indicate bootstrap support.

Supplementary Tables

Table S1.1. Comparison of summary metrics for different levels of data matrix completeness. *% complete matrix*: levels of data filtering for matrix constructions, i.e. the required minimum percent of taxa present for each UCE locus for inclusion in data matrix; *Alignment length*: total length (bp) of concatenated alignments for respective levels of data completeness; *UCE loci*: number of UCEs present in respective matrices; *Mean UCE length*: mean length of UCE loci for respective matrix; *% missing data per sample*: Missing data (% bp) present for each sample for respective matrix. Missing data metrics calculated using custom Rscript following methods at https://github.com/laninsky/UCE_processing_steps.

% complete matrix	50%	60%	70%	80%	90%	100%
Alignment length (bp)	83,547	59,786	40,790	19,661	5,376	541
UCE loci	305	215	144	70	19	1
Mean UCE length (bp)	347	278	283	281	283	541
% missing data (bp) per sample						
SLE36	48.32	43.44	37.97	29.63	28.62	15.15
SLE1012	40.06	34.87	33.79	25.36	17.26	13.68
SLE993	39.80	34.53	30.9	21.84	11.48	0
SLE991	31.54	23.40	19.15	11.45	3.89	9.42
SLE990	34.80	27.70	25.26	19.04	9.36	9.06
SLE956	34.93	24.88	16.29	11.71	1.08	9.24
SLE935	44.48	38.36	33.68	26.73	3.2	17.38
SLE1006	25.43	17.64	15.36	11.27	1.23	0
SLE992	31.84	25.52	21.6	14.21	0.15	0
SLE996	37.60	30.16	25.12	14.87	13.04	8.68
SLE994	34.10	26.74	20.36	10	14.47	0
SLE997	25.50	20.41	21.1	13.48	9.9	0
Hgsp906	26.63	20.70	18.29	10.89	3.59	0
SLE901	31.01	28.59	21.44	19.52	15.49	0
SLE998	46.85	44.50	38.78	38.88	29.99	0
SLE829	26.62	25.43	19.45	13.93	19.44	2.96
SLE699	38.60	30.66	22.12	17.52	8.98	0
SLE681	52.12	47.73	40.58	34.65	33.2	12.38
SLE894	31.40	26.14	23.12	14.17	4.87	0
SLE840	65.94	60.93	52.87	43.7	31.19	13.12

Table S1.2. Number of sites with missing data for respective number of samples, for different levels of data filtering. Left column: Number of samples missing per site. Right columns: number of sites missing corresponding number of samples (left column) for different levels of data filtering. Missing data metrics calculated using custom Rscript following methods at https://github.com/laninsky/UCE_processing_steps.

# missing samples	# sites with missing data for different levels of data completeness					
	50%	**60%**	**70%**	**80%**	**90%**	**100%**
0	403	403	403	403	403	403
1	764	764	764	764	764	25
2	2340	2340	2340	2340	2340	14
3	3785	3785	3785	3785	814	9
4	7285	7285	7285	7285	299	24
5	7040	7040	7040	1895	111	1
6	8962	8962	8962	1084	317	35
7	8869	8869	4045	740	162	30
8	9920	9920	2307	605	166	0
9	10389	4405	1773	332	0	0
10	13357	3238	1442	428	0	0
11	5251	1826	644	0	0	0
12	3962	949	0	0	0	0
13	1220	0	0	0	0	0
14	0	0	0	0	0	0

Chapter 2.

Supplementary Figures

Supp. Fig. S1. ML, Full

0.2
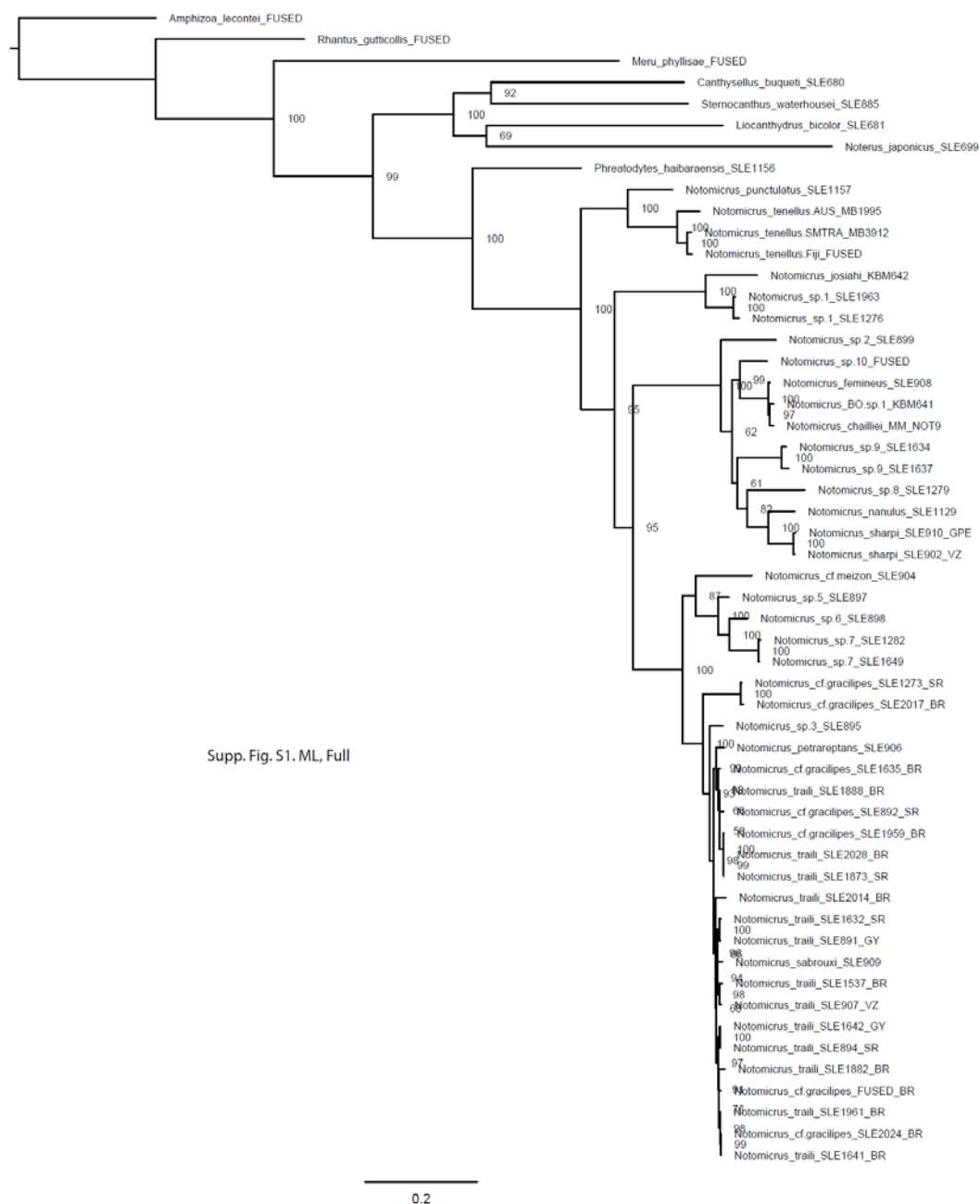
Figure S2.1. Phylogeny of Notomicrinae recovered by Maximum Likelihood analysis of Full dataset in IQ-Tree 2. Numbers at nodes represent UFBoot bootstrap support values.
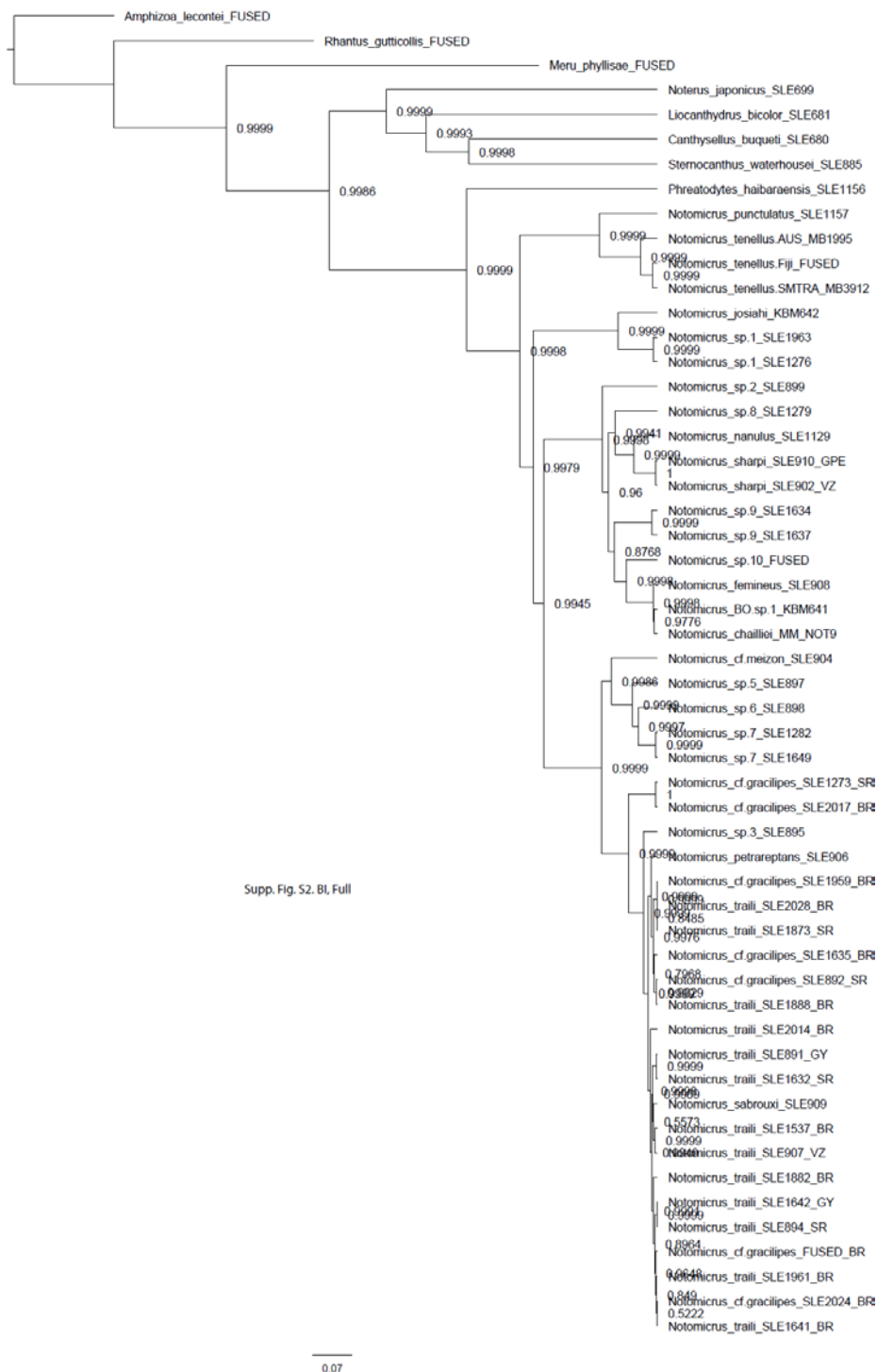
Figure S2.2. Phylogeny of Notomicrinae recovered by Bayesian inference of Full dataset in BEAST2. Maximum clade credibility tree recovered using strict clock with no calibrations. Numbers at nodes represent posterior probability values.

Supp Fig. S3. ML, Trimmed, All models

0.6
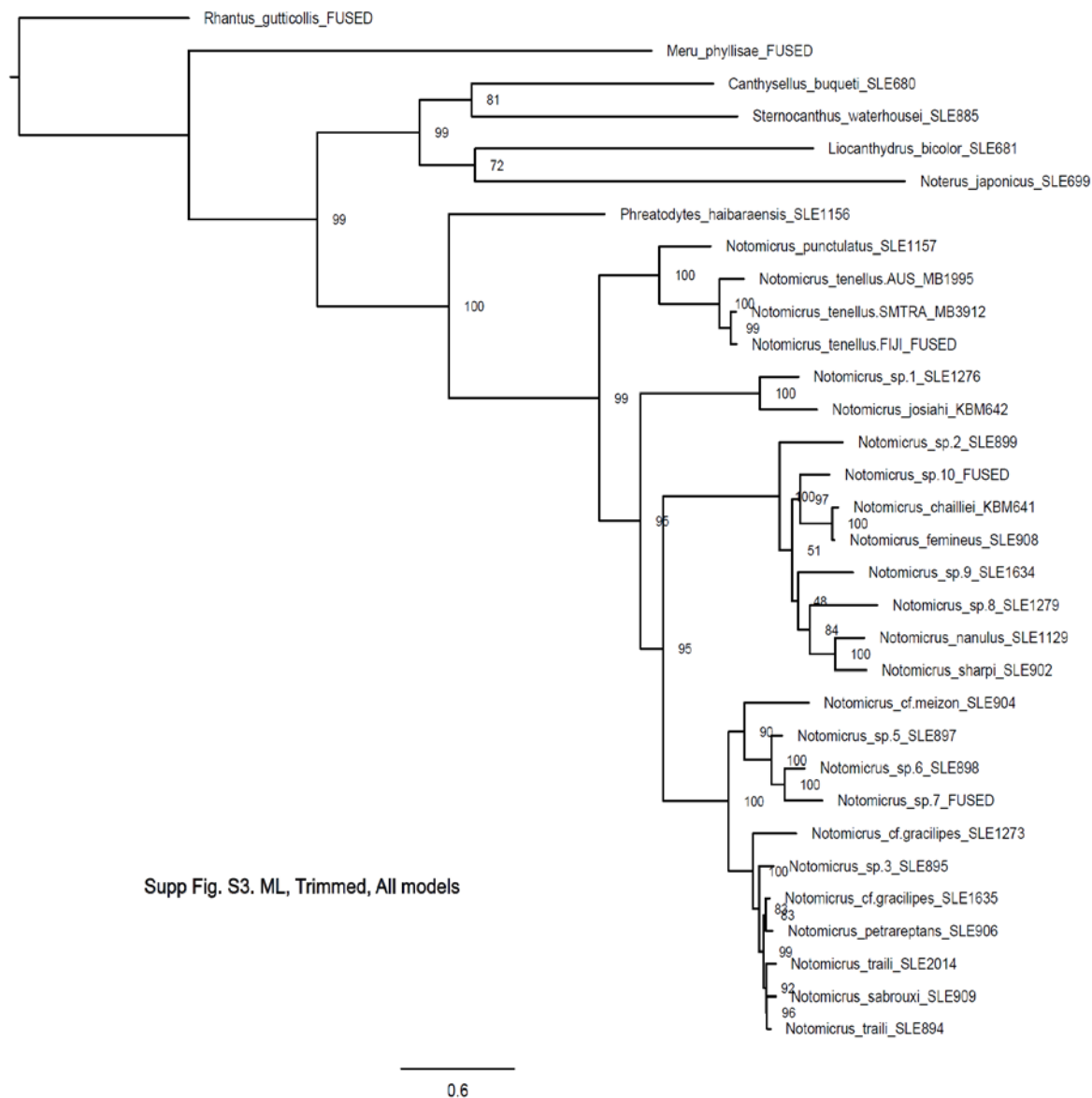
Figure S2.3. Phylogeny of Notomicrinae recovered by Maximum Likelihood analysis of Trimmed dataset in IQ-Tree 2; models search included all available models. Numbers at nodes represent UFBoot bootstrap support values.

Supp. Fig. S4. ML, Trimmed, BEAST Models
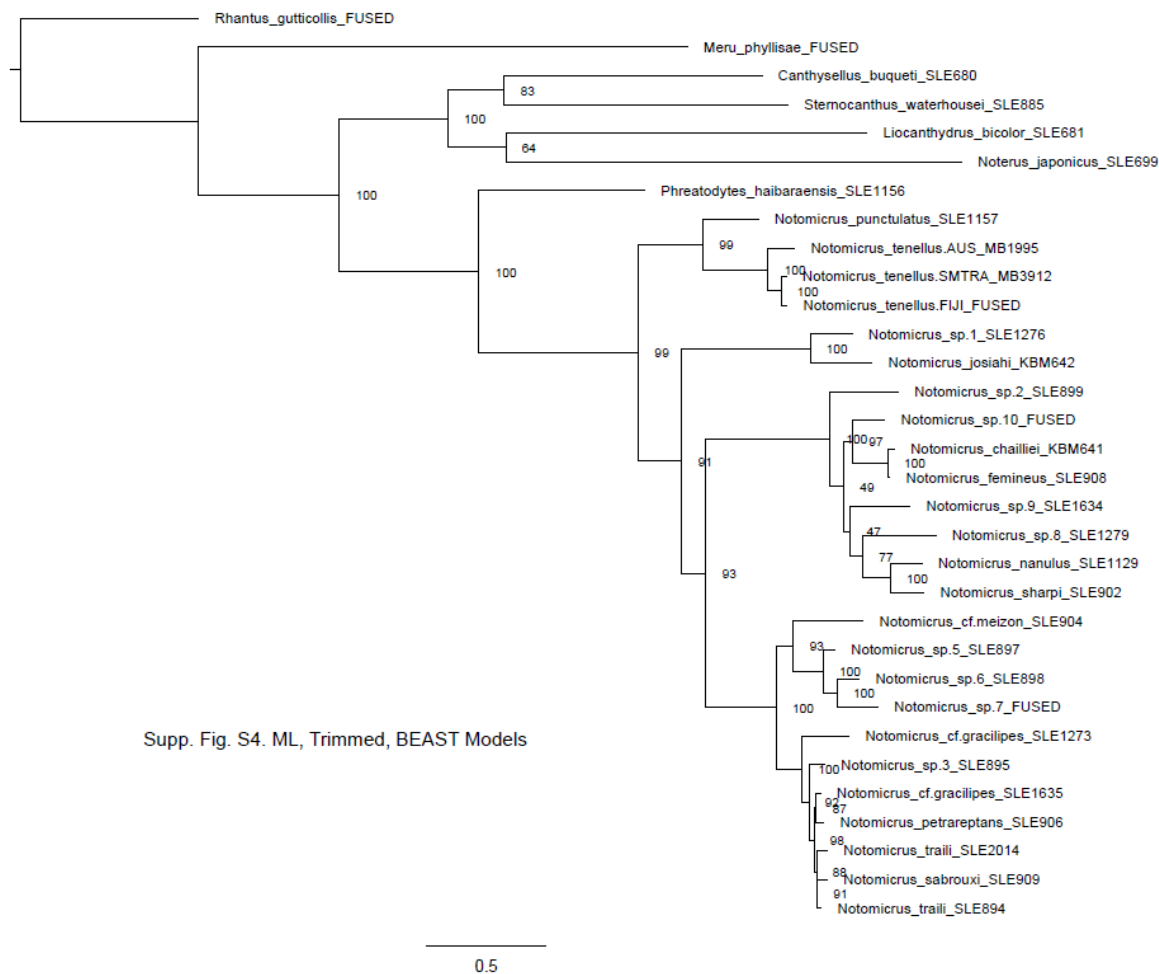
Figure S2.4. Phylogeny of Notomicrinae recovered by Maximum Likelihood analysis of Trimmed dataset in IQ-Tree 2; model search included only models available in BEAST2. Numbers at nodes represent UFBoot bootstrap support values support values.

Supp. Fig. S5. BI, Trimmed, Strict, No calibration
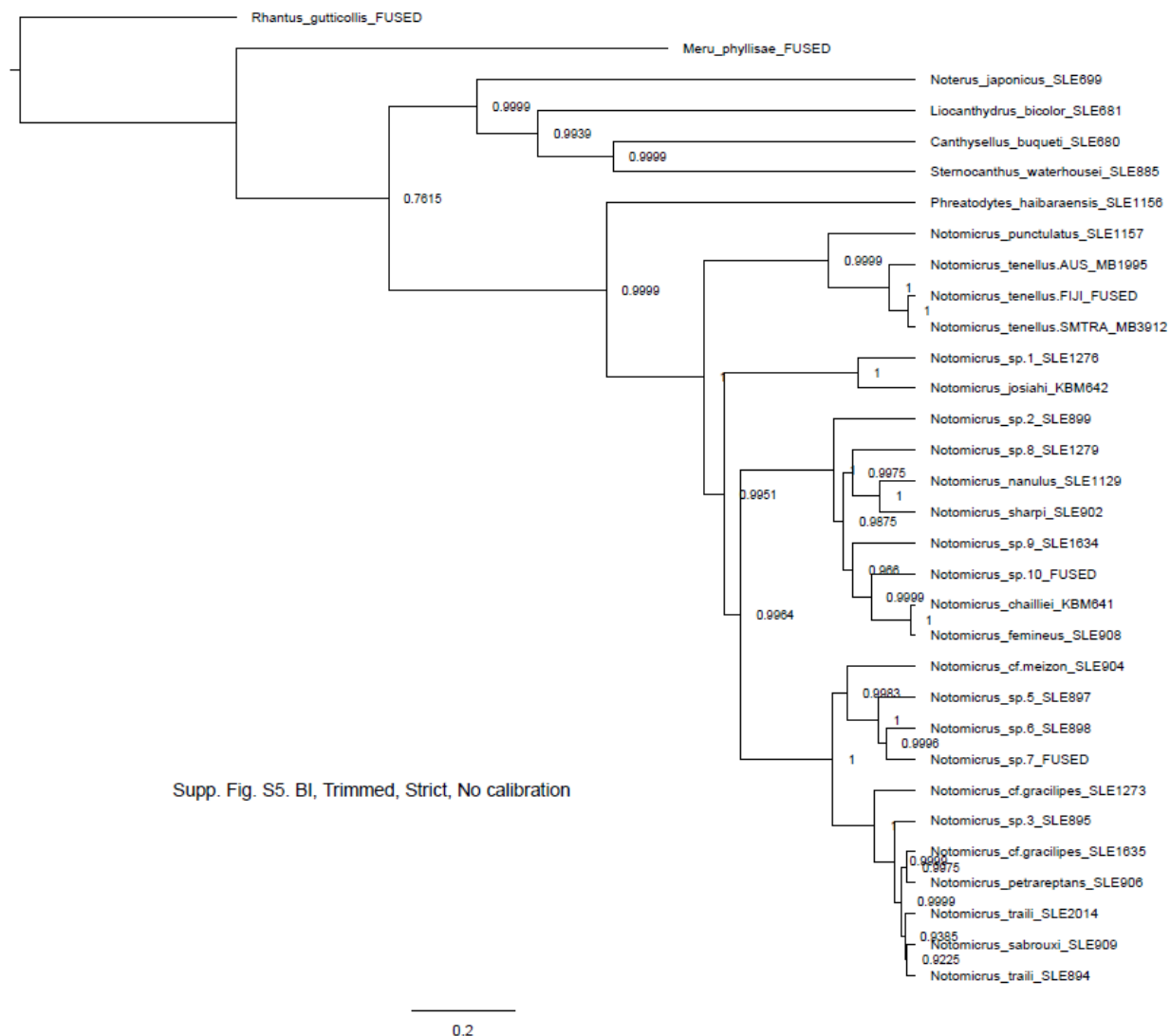
Figure S2.5. Phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using strict clock with no calibrations. Numbers at nodes represent posterior probability values.

Supp. Fig. S6. BI, Trimmed, RCY

30.0
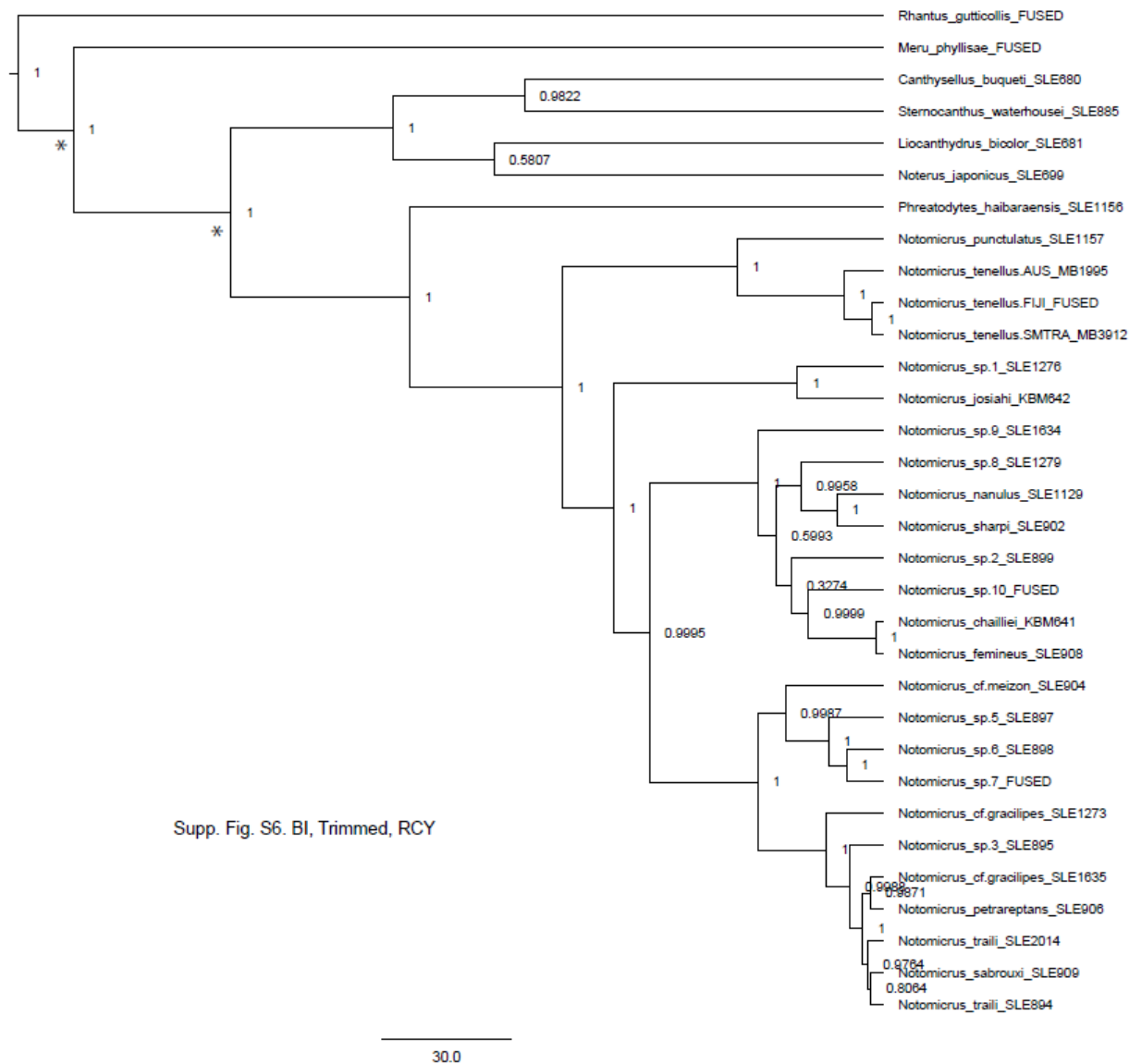
Figure S2.6. Phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using relaxed clock and Yule tree prior (RCY), with secondary node calibrations. Numbers at nodes represent posterior probability values, asterisks indicate calibration points.
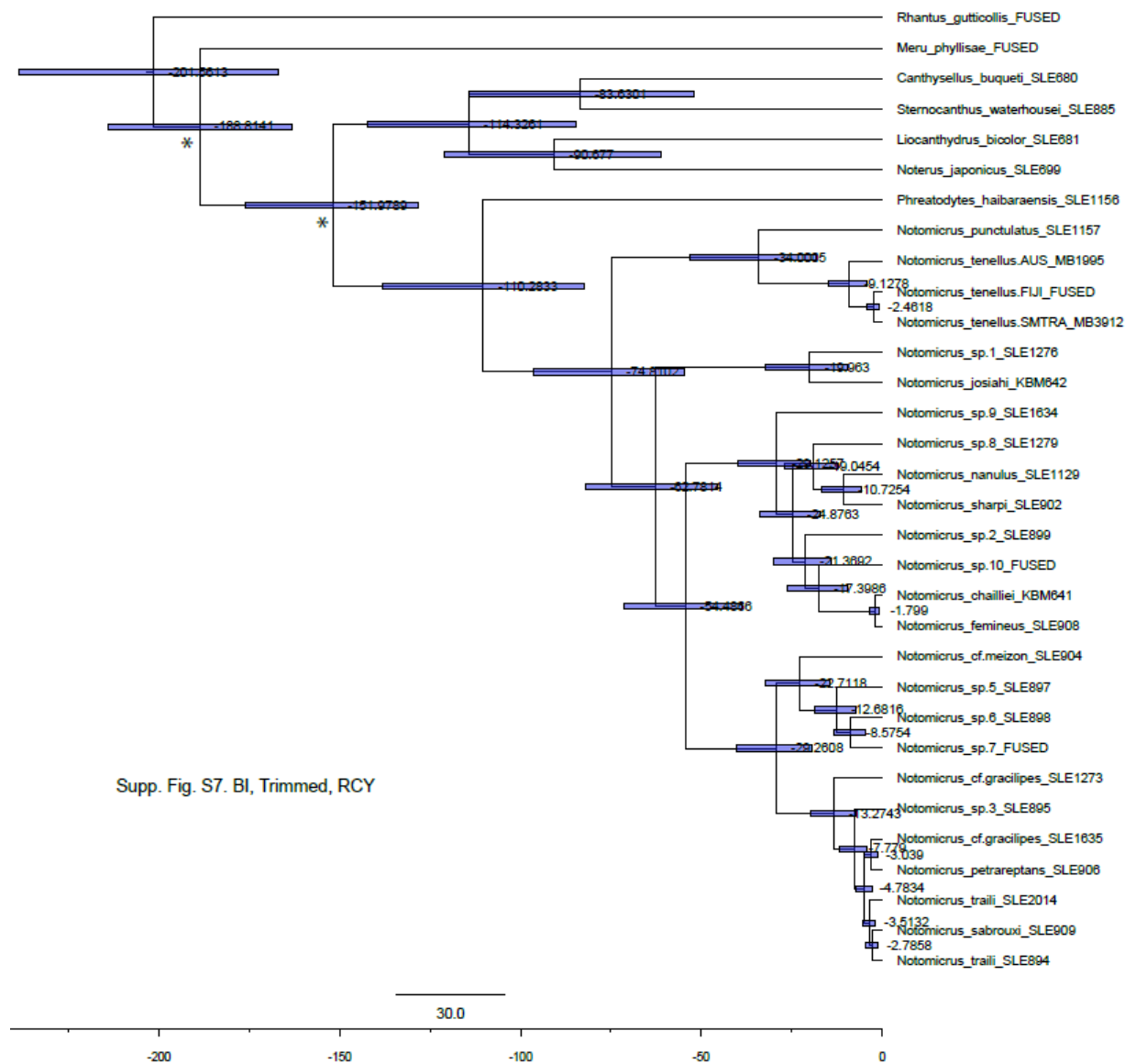
Figure S2.7. Dated phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using relaxed clock and Yule tree prior (RCY), with secondary node calibrations. Numbers at nodes represent recovered clade ages, node bars indicate 95% HPD, asterisks indicate calibration points.

Rhantus_gutticollis_FUSED
Meru_phyllisae_FUSED
Canthysellus_buqueti_SLE680
0.9805
Sternocanthus_waterhousei_SLE885
1
Liocanthydrus_bicolor_SLE681
0.559
Noterus_japonicus_SLE699
1
Phreatodytes_haibaraensis_SLE1156
Notomicrus_punctulatus_SLE1157
1
Notomicrus_tenellus.AUS_MB1995
1
Notomicrus_tenellus.FIJI_FUSED
1
Notomicrus_tenellus.SMTRA_MB3912
Notomicrus_sp.1_SLE1276
1
Notomicrus_josiahi_KBM642
Notomicrus_sp.9_SLE1634
Notomicrus_sp.8_SLE1279
0.9951
Notomicrus_nanulus_SLE1129
1
Notomicrus_sharpi_SLE902
0.5994
Notomicrus_sp.2_SLE899
0.3329
Notomicrus_sp.10_FUSED
0.9999
Notomicrus_chailliei_KBM641
1
Notomicrus_femineus_SLE908
Notomicrus_cf.meizon_SLE904
0.9985
Notomicrus_sp.5_SLE897
1
Notomicrus_sp.6_SLE898
0.9999
Notomicrus_sp.7_FUSED
Notomicrus_cf.gracilipes_SLE1273
Notomicrus_sp.3_SLE895
Notomicrus_cf.gracilipes_SLE1635
0.9982
Notomicrus_petrareptans_SLE906
Notomicrus_traili_SLE2014
0.9736
Notomicrus_sabrouxi_SLE909
0.8108
Notomicrus_traili_SLE894

0.9999
0.9993

Supp. Fig. S8. BI, Trimmed, RCBD

30.0

Figure S2.8. Phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using relaxed clock and birth-death tree prior (RCBD), with secondary node calibrations. Numbers at nodes represent posterior probability, asterisks indicate calibration points.

Figure S2.9. Dated phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using relaxed clock and birth-death tree prior (RCBD), with secondary node calibrations. Numbers at nodes represent recovered clade ages, node bars indicate 95% HPD, asterisks indicate calibration points.

Rhantus_gutticollis_FUSED

Meru_phyllisae_FUSED

Noterus_japonicus_SLE699

Liocanthydrus_bicolor_SLE681

Canthysellus_buqueti_SLE680

Sternocanthus_waterhousei_SLE885

Phreatodytes_haibaraensis_SLE1156

Notomicrus_punctulatus_SLE1157

Notomicrus_tenellus.AUS_MB1995

Notomicrus_tenellus.FIJI_FUSED

Notomicrus_tenellus.SMTRA_MB3912

Notomicrus_sp.1_SLE1276

Notomicrus_josiahi_KBM642

Notomicrus_sp.2_SLE899

Notomicrus_sp.8_SLE1279

Notomicrus_nanulus_SLE1129

Notomicrus_sharpi_SLE902

Notomicrus_sp.9_SLE1634

Notomicrus_sp.10_FUSED

Notomicrus_chailliei_KBM641

Notomicrus_femineus_SLE908

Notomicrus_cf.meizon_SLE904

Notomicrus_sp.5_SLE897

Notomicrus_sp.6_SLE898

Notomicrus_sp.7_FUSED

Notomicrus_cf.gracilipes_SLE1273

Notomicrus_sp.3_SLE895

Notomicrus_cf.gracilipes_SLE1635

Notomicrus_petrareptans_SLE906

Notomicrus_traili_SLE2014

Notomicrus_sabrouxi_SLE909

Notomicrus_traili_SLE894

Supp. Fig. S10. BI, Trimmed, STY

20.0

Figure S2.10. Phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using strict clock and Yule tree prior (STY), with secondary node calibrations. Numbers at nodes represent posterior probability, asterisks indicate calibration points.

Rhantus_gutticollis_FUSED

Meru_phyllisae_FUSED

Noterus_japonicus_SLE699

Liocanthydrus_bicolor_SLE681

Canthysellus_buqueti_SLE680

Sternocanthus_waterhousei_SLE885

Phreatodytes_haibaraensis_SLE1156

Notomicrus_punctulatus_SLE1157

Notomicrus_tenellus.AUS_MB1995

Notomicrus_tenellus.FIJI_FUSED

Notomicrus_tenellus.SMTRA_MB3912

Notomicrus_sp.1_SLE1276

Notomicrus_josiahi_KBM642

Notomicrus_sp.2_SLE899

Notomicrus_sp.8_SLE1279

Notomicrus_nanulus_SLE1129

Notomicrus_sharpi_SLE902

Notomicrus_sp.9_SLE1634

Notomicrus_sp.10_FUSED

Notomicrus_chailliei_KBM641

Notomicrus_femineus_SLE908

Notomicrus_cf.meizon_SLE904

Notomicrus_sp.5_SLE897

Notomicrus_sp.6_SLE898

Notomicrus_sp.7_FUSED

Notomicrus_cf.gracilipes_SLE1273

Notomicrus_sp.3_SLE895

Notomicrus_cf.gracilipes_SLE1635

Notomicrus_petrareptans_SLE906

Notomicrus_traili_SLE2014

Notomicrus_sabrouxi_SLE909

Notomicrus_traili_SLE894

Supp. Fig. S11. BI, Trimmed, STY

20.0

Figure S2.11. Dated phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using strict clock and Yule tree prior (STY), with secondary node calibrations. Numbers at nodes represent recovered clade ages, node bars indicate 95% HPD, asterisks indicate calibration points.

Supp. Fig. S12. BI, Trimmed, STBD
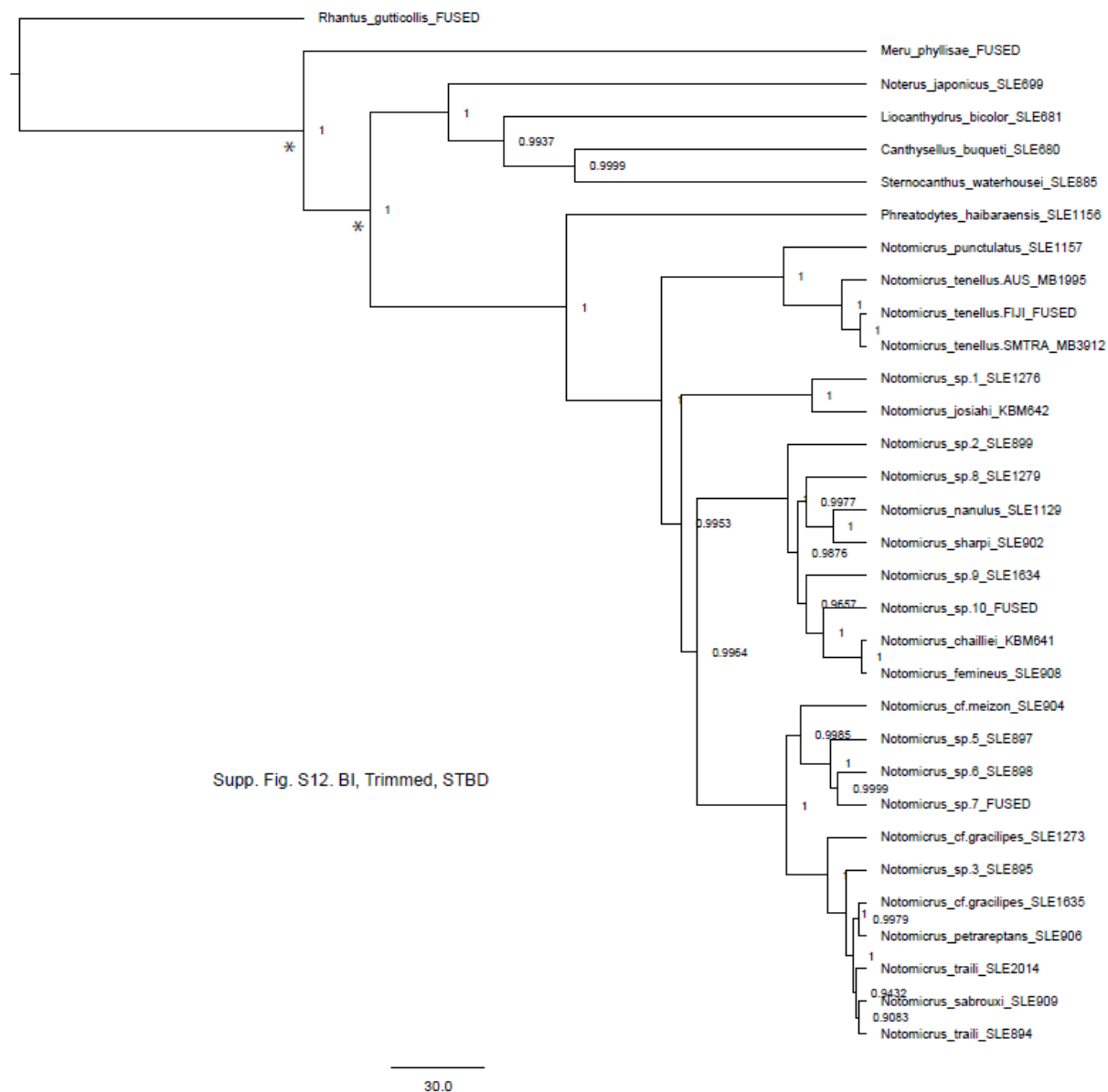
Figure S2.12. Phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using strict clock and birth-death tree prior (STBD), with secondary node calibrations. Numbers at nodes represent posterior probability, asterisks indicate calibration points.

Supp. Fig. S13. BI, Trimmed, STBD
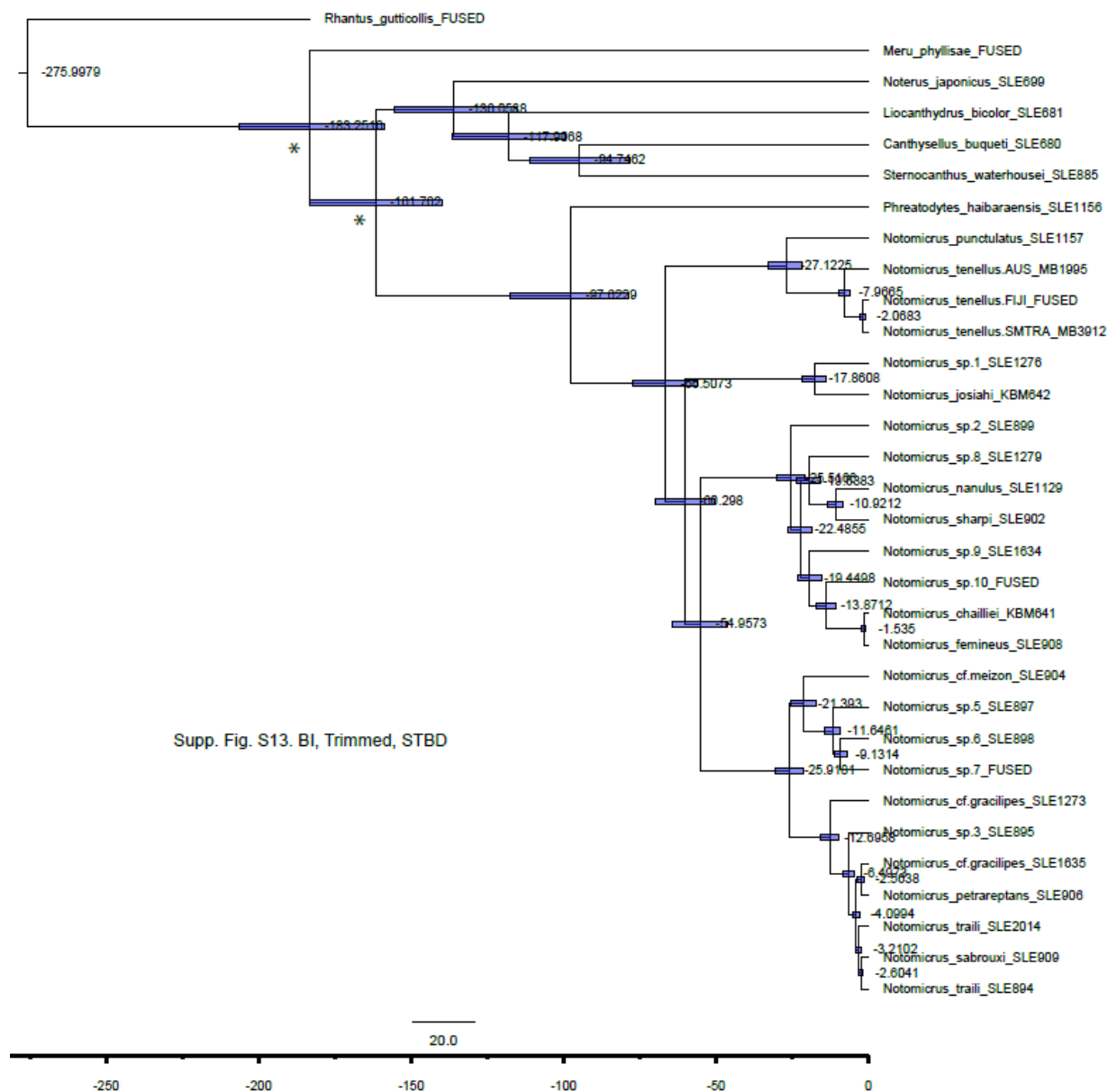
Figure S2.13. Dated phylogeny of Notomicrinae recovered by Bayesian Inference of Trimmed data set in BEAST2. Maximum clade credibility tree recovered using strict clock and birth-death tree prior (STBD), with secondary node calibrations. Numbers at nodes represent recovered clade ages, node bars indicate 95% HPD, asterisks indicate calibration points.
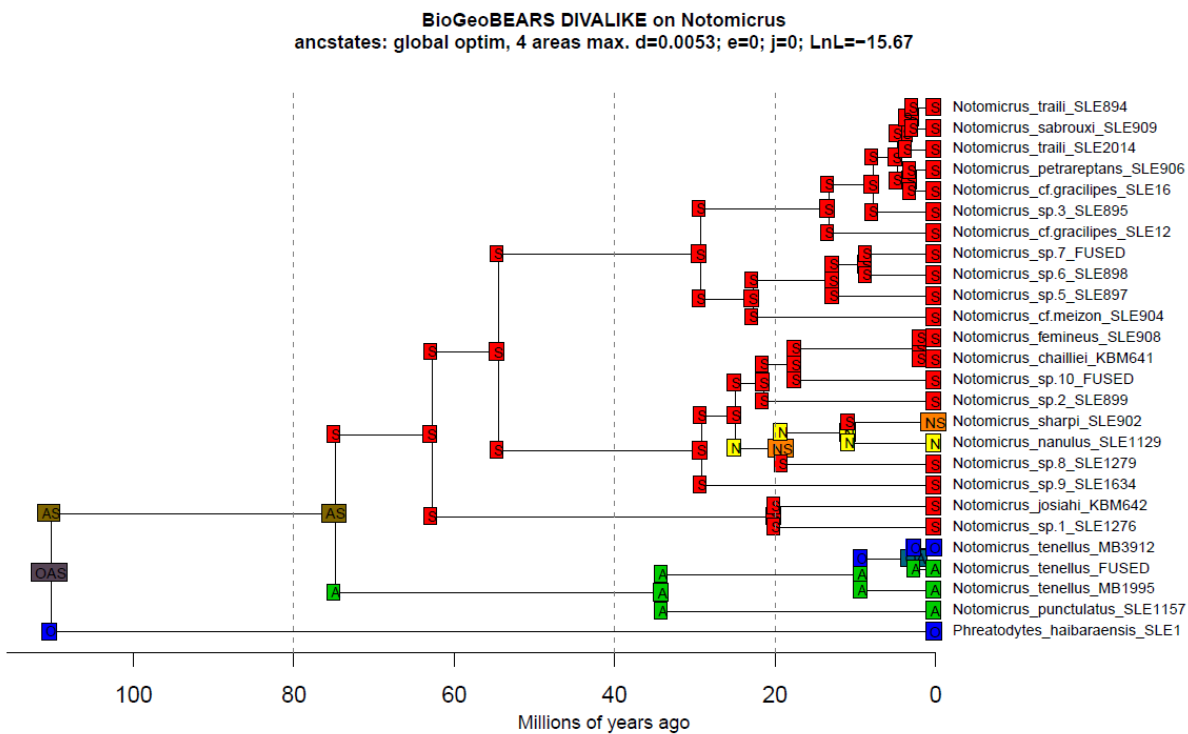
Figure S2.14. Ancestral range reconstruction of Notomicrinae. Reconstructed using DEC model in BioGeoBears. Boxes indicate recovered most probable ancestral ranges/states.
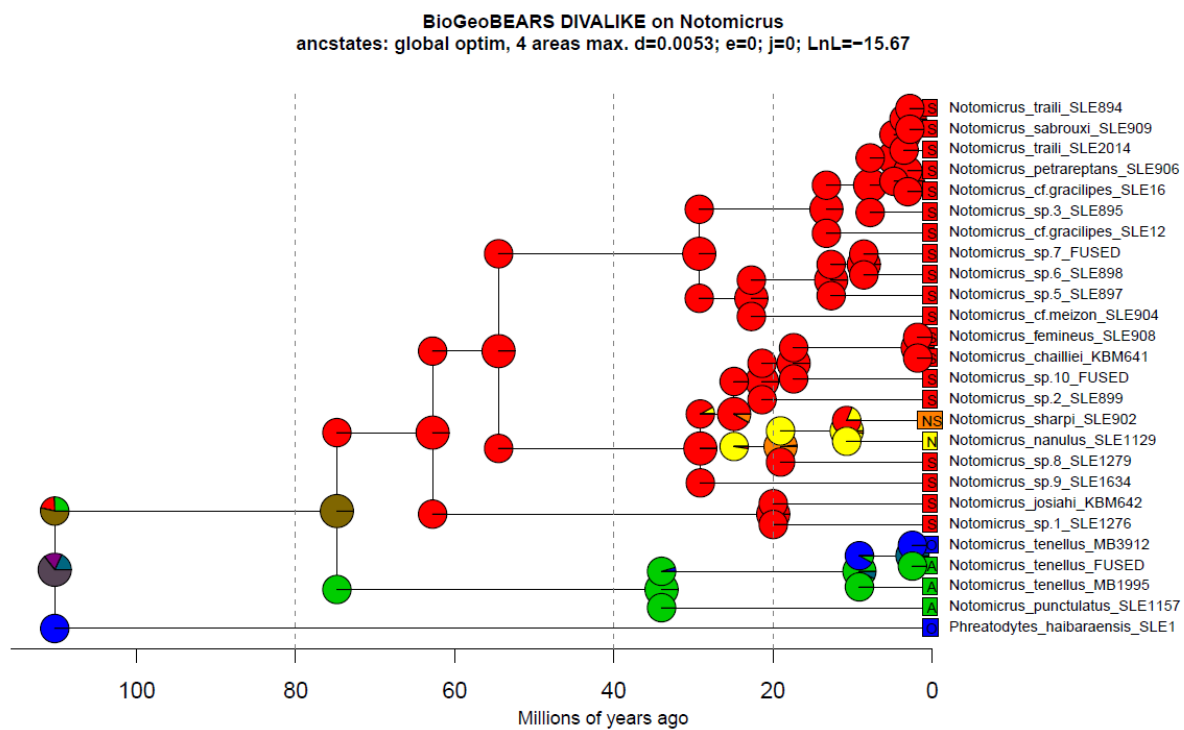
Figure S2.15. Ancestral range reconstruction of Notomicrinae. Reconstructed using DEC model in BioGeoBears. Pie charts indicate relative probability of ancestral ranges/states.
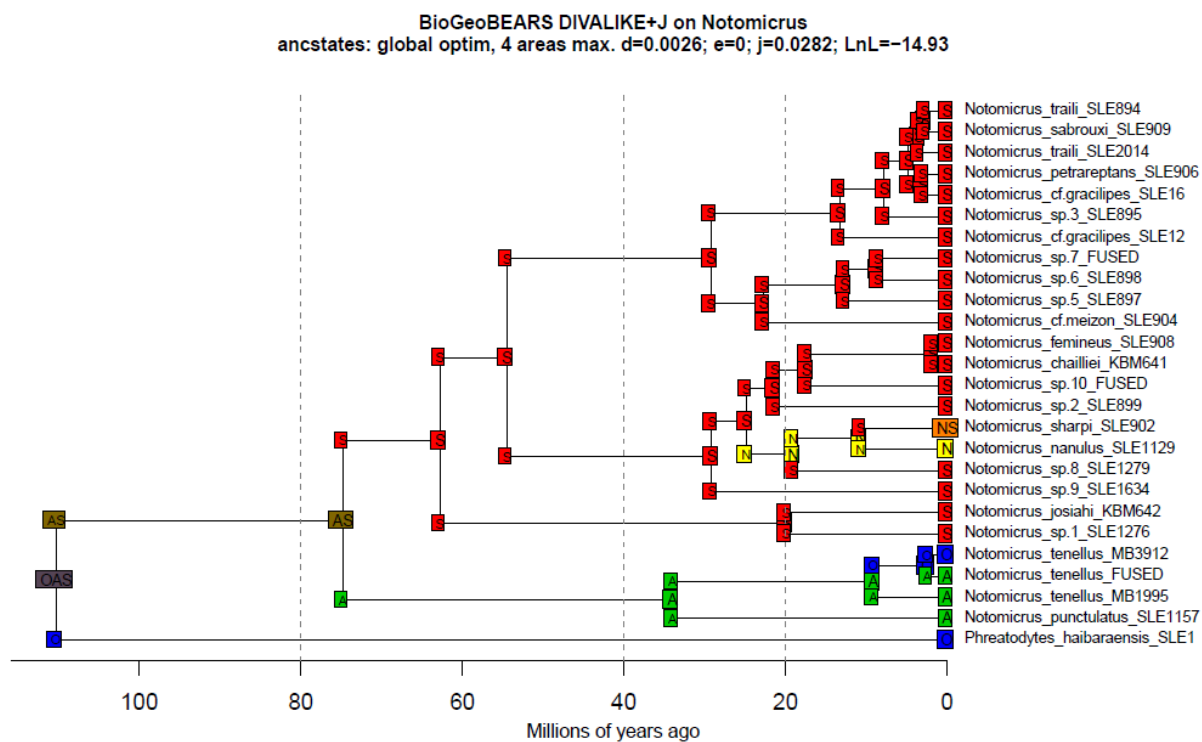
Figure S2.16. Ancestral range reconstruction of Notomicrinae. Reconstructed using DEC+j model in BioGeoBears. Boxes indicate recovered most probable ancestral ranges/states.
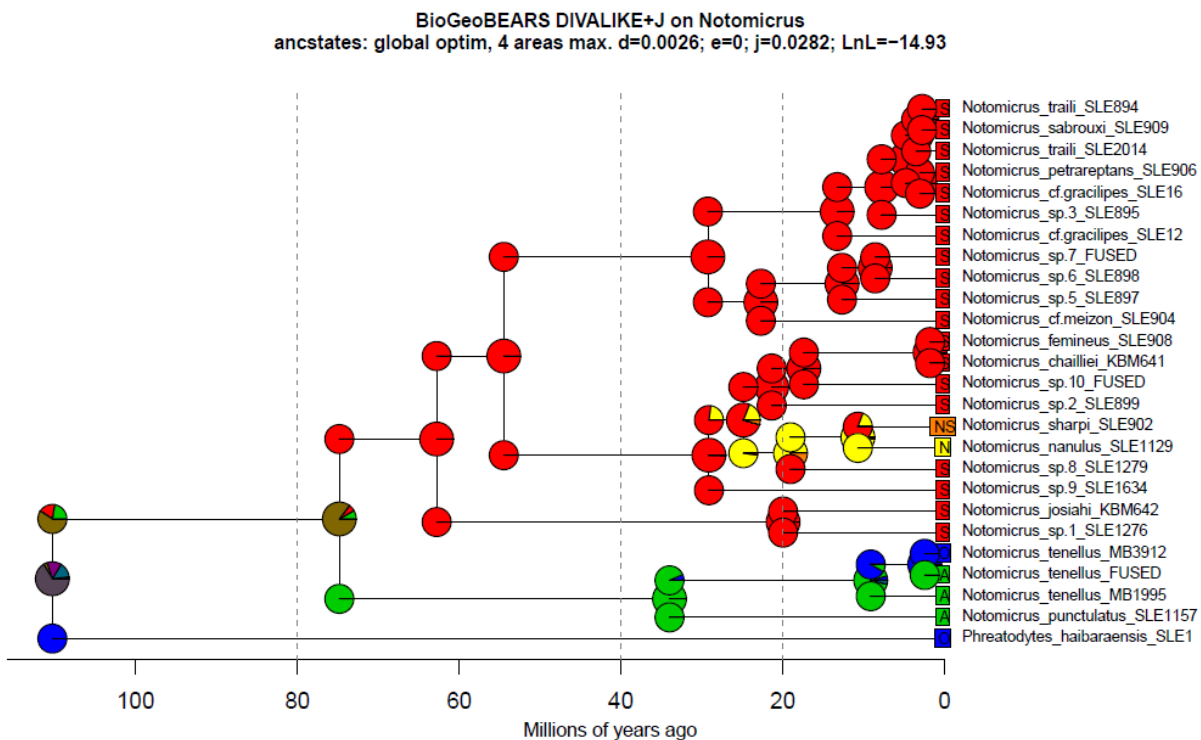
Figure S2.17. Ancestral range reconstruction of Notomicrinae. Reconstructed using DEC+j model in BioGeoBears. Pie charts indicate relative probability of ancestral ranges/states.
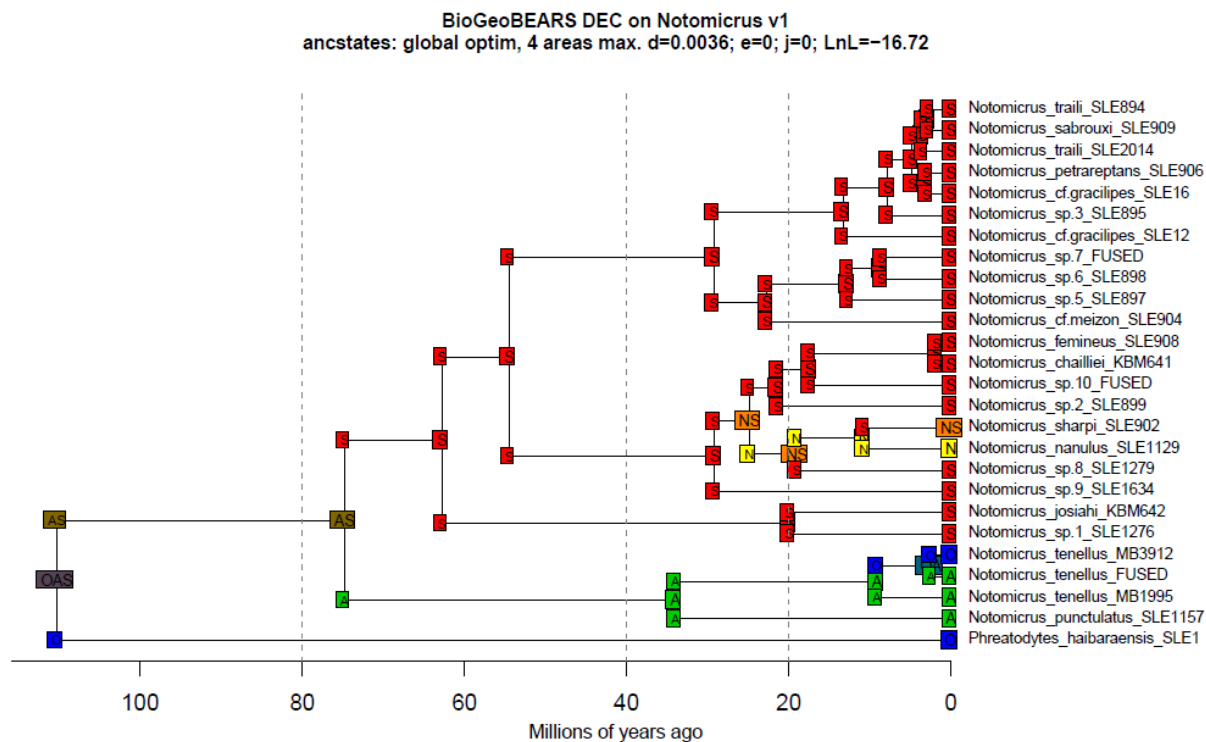
Figure S2.18. Ancestral range reconstruction of Notomicrinae. Reconstructed using DIVALIKE model in BioGeoBears. Boxes indicate recovered most probable ancestral ranges/states.

BioGeoBEARS DEC on Notomicrus v1
ancstates: global optim, 4 areas max. d=0.0036; e=0; j=0; LnL=−16.72
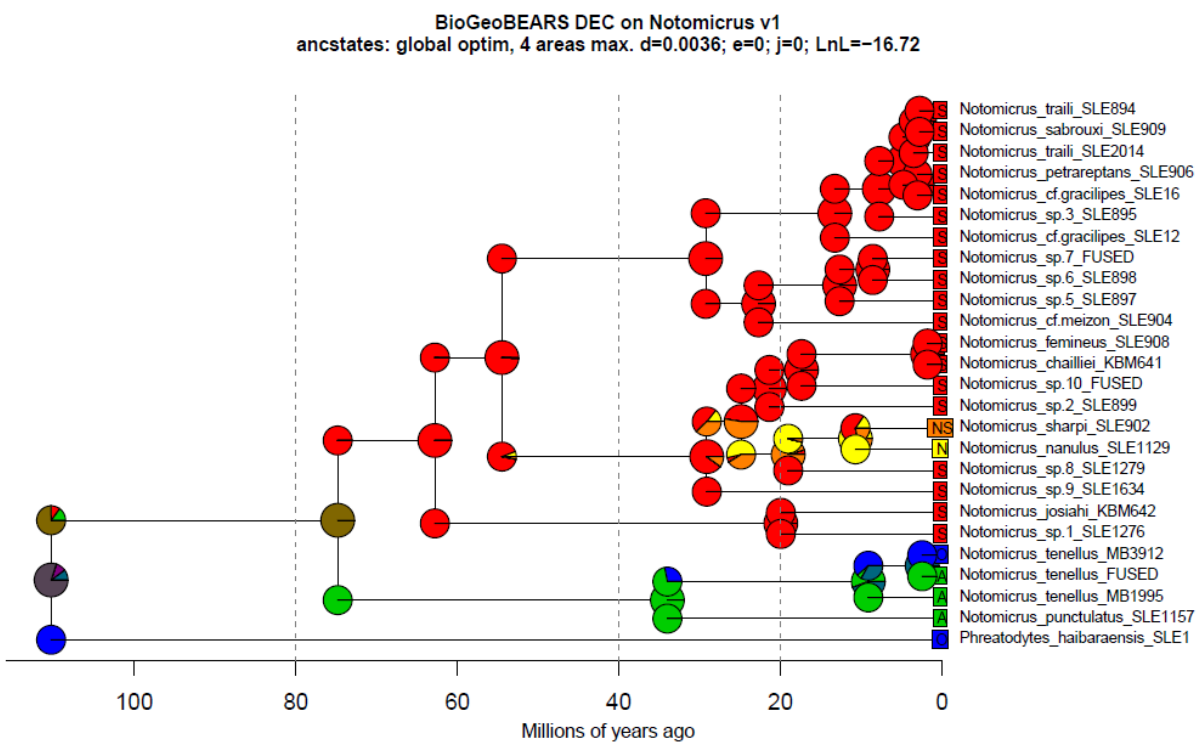
Figure S2.19. Ancestral range reconstruction of Notomicrinae. Reconstructed using DIVALIKE model in BioGeoBears. Pie charts indicate relative probability of ancestral ranges/states.
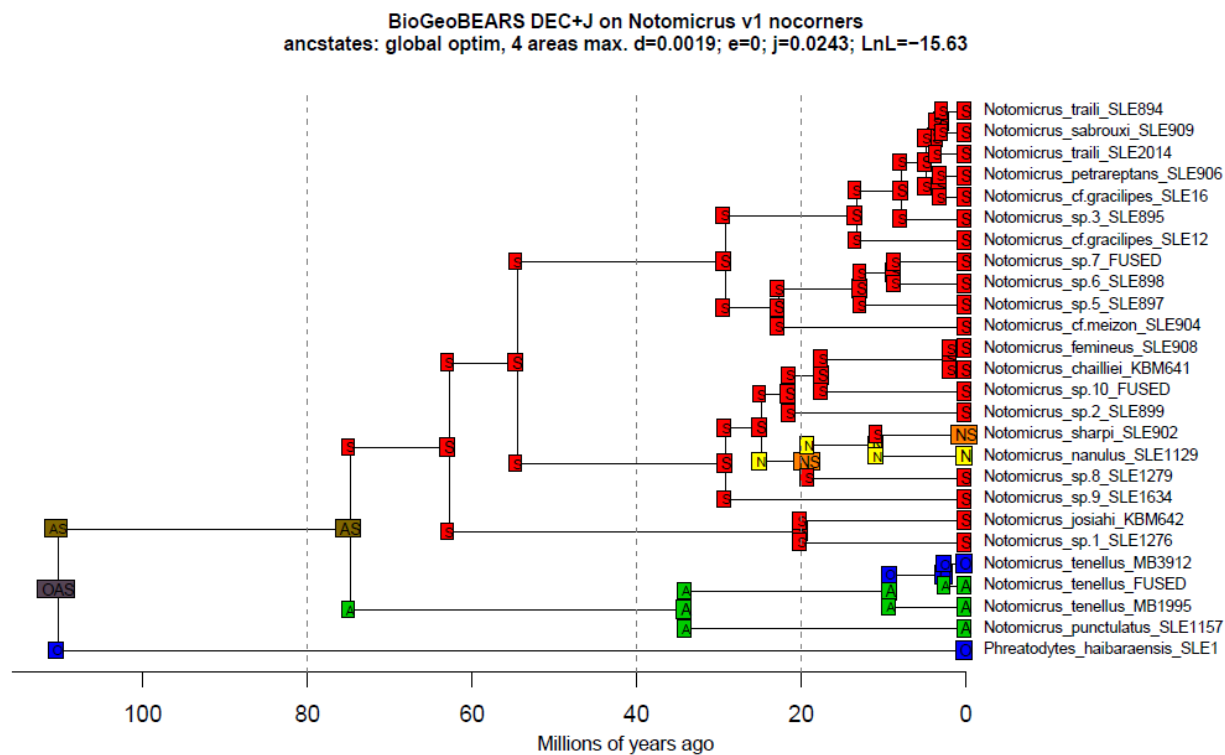
Figure S2.20. Ancestral range reconstruction of Notomicrinae. Reconstructed using DIVALIKE+j model in BioGeoBears. Boxes indicate recovered most probable ancestral ranges/states.
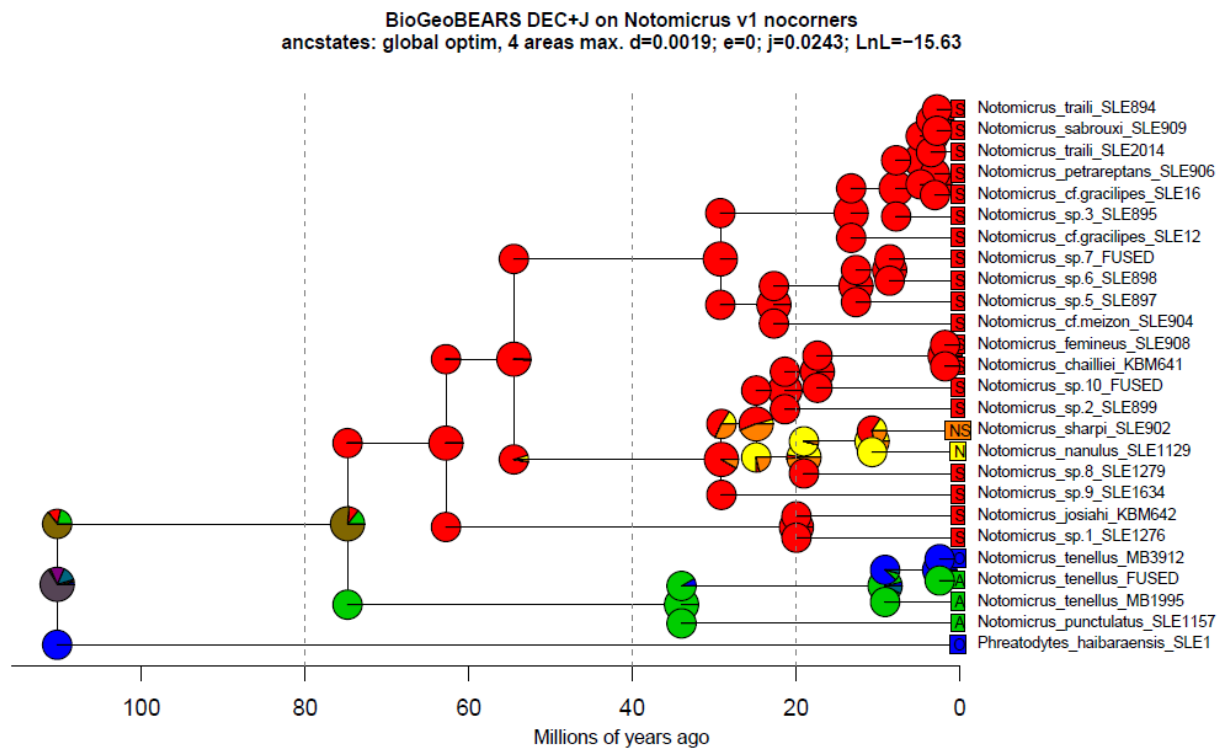
Figure S2.21. Ancestral range reconstruction of Notomicrinae. Reconstructed using DIVALIKE+j model in BioGeoBears. Pie charts indicate relative probability of ancestral ranges/states.

Supplementary Tables

Table S1. List of specimens used in this study. GenBank accession codes for each successfully sequenced or downloaded gene fragments are provided in the right-hand columns.

Family	Subfamily	Genus	species	Code	Locality	CO1	CAD	H3	16S	28S
Amphizoidae		*Amphizoa*	*lecontei*	FUSED	N/A	KJ548509	EU677532	KJ548745	AY071771	EU677680
Dytiscidae		***Rhantus***	***gutticollis***	**FUSED**	N/A	KJ637973	KJ638036	KJ637998	KJ637884	EU797382
Meruidae		***Meru***	***phyllisae***	**FUSED**	Venezuela	AY071809	MW043832	MW043864	FM163591	-
Noteridae	Noterinae	*Canthysellus*	*buqueti*	SLE680	Suriname	KY055888	MW043797	KY055941	KY055999	KY056117
Noteridae	Noterinae	*Sternocanthus*	*waterhousei*	SLE885	Australia	KY055880	MW043829	KY055930	KY055988	KY056105
Noteridae	Noterinae	*Liocanthydrus*	*bicolor*	SLE681	Suriname	KY055879	MW043830	KY055929	KY055987	KY056104
Noteridae	Noterinae	*Noterus*	*japonicus*	SLE699	Mongolia	KY055868	MW043831	-	KY055977	KY056093
Noteridae	**Notomicrinae**	***Phreatodytes***	***haibaraensis***	**SLE1156**	Japan	MW041967	-	MW043863	MW035912	-
Noteridae	**Notomicrinae**	***Notomicrus***	***punctulatus***	**SLE1157**	New Caledonia	MW041964	MW043826	MW043859	MW035909	MW035872
Noteridae	**Notomicrinae**	***Notomicrus***	***tenellus***	**MB1995**	Australia	MW041966	MW043828	MW043862	MW035911	MW035875
Noteridae	**Notomicrinae**	***Notomicrus***	***tenellus***	**MB3912**	Sumatra	KT607942	KT607974	MW043860	KT607922	MW035873
Noteridae	**Notomicrinae**	***Notomicrus***	***tenellus***	**FUSED**	Fiji	MW041965	MW043827	MW043861	MW035910	MW035874
Noteridae	**Notomicrinae**	***Notomicrus***	***josiahi***	**KBM642**	Venezuela	KY055867	MW043825	KJ548802	KJ548385	KY056092
Noteridae	Notomicrinae	*Notomicrus*	*sp.1*	SLE1963	Brazil: Amapá	MW041962	-	-	-	-
Noteridae	**Notomicrinae**	***Notomicrus***	***sp.1***	**SLE1276**	Brazil: Amazonas	MW041963	MW043824	MW043858	MW035908	MW035871
Noteridae	**Notomicrinae**	***Notomicrus***	***sp.2***	**SLE899**	Suriname	MW041961	MW043823	MW043857	MW035907	MW035870
Noteridae	**Notomicrinae**	***Notomicrus***	***sp.10***	**FUSED**	Peru: Madre de Dios	MW041926	MW043799	MW043834	MW035877	MW035849
Noteridae	**Notomicrinae**	***Notomicrus***	***femineus***	**SLE908**	Guadeloupe	MW041925	MW043798	MW043833	MW035876	MW035848
Noteridae	**Notomicrinae**	***Notomicrus***	***chailliei***	**KBM641**	Bolivia	KY055866	-	KY055920	KY055976	KY056091
Noteridae	Notomicrinae	*Notomicrus*	*chailliei*	MMNOT9	Guadeloupe	KT006271	-	-	-	-
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 9***	**SLE1634**	Brazil: Minas Gerais	MW041956	MW043819	MW043852	MW035902	MW035865
Noteridae	Notomicrinae	*Notomicrus*	*sp. 9*	SLE1637	Brazil: Roraima	MW041957	-	MW043853	MW035903	MW035866
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 8***	**SLE1279**	Brazil: Amazonas	MW041955	MW043818	MW043851	MW035901	MW035864
Noteridae	**Notomicrinae**	***Notomicrus***	***nanulus***	**SLE1129**	USA: Alabama	MW041958	MW043820	MW043854	MW035904	MW035867
Noteridae	Notomicrinae	*Notomicrus*	*sharpi*	SLE910	Guadeloupe	MW041959	MW043821	MW043855	MW035905	MW035868
Noteridae	**Notomicrinae**	***Notomicrus***	***sharpi***	**SLE902**	Venezuela	MW041960	MW043822	MW043856	MW035906	MW035869
Noteridae	**Notomicrinae**	***Notomicrus***	***cf. meizon***	**SLE904**	Venezuela	MW041954	MW043817	MW043850	MW035900	MW035863
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 5***	**SLE897**	Suriname	MW041950	MW043814	MW043846	MW035897	MW035859
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 6***	**SLE898**	Suriname	MW041951	MW043815	MW043847	MW035898	MW035860
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 7***	**SLE1282**	Brazil: Amazonas	MW041952	-	MW043848	MW035899	MW035861
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 7***	**SLE1649**	Suriname	MW041953	MW043816	MW043849	-	MW035862
Noteridae	**Notomicrinae**	***Notomicrus***	***cf. gracilipes***	**SLE1273**	Suriname	MW041948	MW043813	MW043845	MW035895	MW035858
Noteridae	Notomicrinae	*Notomicrus*	*cf. gracilipes*	SLE2017	Brazil: Amazonas	MW041949	-	-	MW035896	-
Noteridae	**Notomicrinae**	***Notomicrus***	***sp. 3***	**SLE895**	Suriname	MW041947	MW043812	MW043844	MW035894	MW035857
Noteridae	**Notomicrinae**	***Notomicrus***	***petrareptans***	**SLE906**	Suriname	MW041931	MW043801	MW043836	MW035881	-
Noteridae	**Notomicrinae**	***Notomicrus***	***cf. gracilipes***	**SLE1635**	Brazil: Minas Gerais	MW041930	MW043800	MW043835	MW035880	MW035850
Noteridae	Notomicrinae	*Notomicrus*	*cf. gracilipes*	SLE1888	Brazil: Para	MW041933	-	-	MW035882	-
Noteridae	Notomicrinae	*Notomicrus*	*cf. gracilipes*	SLE892	Suriname	MW041932	MW043802	MW043837	-	MW035851
Noteridae	Notomicrinae	*Notomicrus*	*cf. gracilipes*	SLE1959	Brazil: Amazonas	MW041927	-	-	MW035878	-

Table S2.2. List of gene fragments and primers used in this study.

Gene	Location	Primer	Direction	Sequence	Reference
COI	mitochondrial	Jerry	Forward	CAACAYTTATTTTGATTTTTTGG	Simon et al. 1994
COI	mitochondrial	Pat	Reverse	ATCCATTACATATAATCTGCCATA	Simon et al. 1994
CAD	nuclear	CD439F	Forward	TTCAGTGTACARTTYCAYCCHGARCAYAC	Wild & Maddison (2008)
CAD	nuclear	CD688R	Reverse	TGTATACCTAGAGGATCDACRTTYTCCATRTTRCA	Wild & Maddison (2008)
H3	nuclear	H3aF	Forward	ATGGCTCGTACCAAGCAGACGGC	Colgan et al. 1998
H3	nuclear	H3aR	Reverse	ATATCCTTGGGCATGATGGTGAC	Colgan et al. 1998
16S	mitochondrial	LR-N-13398	Forward	CGCCTGTTTAACAAAAACA	Simon et al. 1994
16S	mitochondrial	LR-J-12887	Reverse	CCGGTCTGAACTCAGATCACGT	Simon et al. 1994
28S	nuclear	NLF184-21	Forward	ACCCGCTGAAYTTAAGCATAT	Van der Auwera et al.1994
28S	nuclear	LS1041R	Reverse	TACGGACRTCCATCAGGGTTTCCCCTGACTTC	Maddison 2008

Table S2.3. Partitioning scheme with input and output subsets for 'Full' and 'Trimmed' datasets

Input subset number	Input subsets	Subset size 'Full' (bp)	Subset size 'Trimmed' (bp)	'Full' subset assignment	'Trimmed' subset assignment
1	COI codon pos. 1	244	243	1	1
2	COI codon pos. 2	244	243	2	2
3	COI codon pos. 3	244	243	3	3
4	CAD codon pos. 1	251	251	4	4
5	CAD codon pos. 2	251	251	2	2
6	CAD codon pos. 3	251	251	6	6
7	H3 codon pos. 1	111	111	4	2
8	H3 codon pos. 2	111	111	2	2
9	H3 codon pos. 3	111	111	9	9
10	16S	527	527	10	10
11	18S	1,077	1,040	11	4

Table S2.4. Type specimens and depositories used for sample identification. Depositories and abbreviations: Snow Entomological Collection, University of Kansas Natural History Museum, Lawrence KS, USA **(SEMC)**; British Museum of Natural History, London **(NHM);** Florida State Collection of Arthropods, Gainesville, FL, USA **(FSCA)**; US National Museum of Natural History, Washington DC, USA **(USNM)** Asterisk(*) = extracted, sequenced; Double asterisk(**) = extracted, sequencing failed.

Species	Authority	Types viewed	Depository
Notomicrus brevicornis	Sharp, 1882	Syntypes	NHM
Notomicrus femineus	Manuel, 2015	Paratypes	SEMC
Notomicrus gracilipes	Sharp, 1882	Syntypes	NHM
Notomicrus huttoni	Young, 1978	Paratypes	FSCA, USNM
Notomicrus josiahi	Miller, 2013	Paratypes	SEMC
Notomicrus malkini	Young, 1978	Paratypes	FSCA
Notomicrus meizon	Guimaraes and Ferreira-Jr, 2019	Paratypes	1.00
Notomicrus petrareptans	Baca and Short, 2018	Holotype, Paratypes	SEMC
Notomicrus sabrouxi	Manuel, 2015	Paratypes	SEMC
Notomicrus sharpi	Balfour-Browne, 1939	Holotype	NHM
Notomicrus teramnus	Guimaraes and Ferreira-Jr, 2019	Holotype, Patatype	0.81
Notomicrus traili	Sharp, 1882	Syntype	NHM

Supp Table references. Ch.2

Baca, S. M., and A.E. Short, A. E. Z. (2018). *Notomicrus petrareptans* sp. n., a new seep-dwelling species of Noteridae from Suriname (Coleoptera: Adephaga). Zootaxa. 4388(2): 182-190.

Balfour-Browne J. 1939. A contribution to the study of the Dytiscidae. - I. (Coleoptera, Adephaga). The Annals and Magazine of Natural History (11) 3: 97–114.

Colgan, D. J., A. McLauchlan, G. D. F. Wilson, S. P. Livingston, G. D. Edgecombe, J. Macaranas, ... and M. R. Gray. (1998). Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. Australian Journal of Zoology. 46(5): 419–437.

Guimarães, B. A. and N. Ferreira-Jr. (2019). Two new species and new records of Notomicrus Sharp, 1882 (Coleoptera: Noteridae) from Brazil. Zootaxa, 4629(2): 263–270.

Maddison, D. R. (2008). Systematics of the North American beetle subgenus Pseudoperyphus (Coleoptera: Carabidae: Bembidion) based upon morphological, chromosomal, and molecular data. Annals of Carnegie Museum. 77(1): 147–193.

Manuel, M. (2015). The genus Notomicrus in Guadeloupe, with description of three new species (Coleoptera: Noteridae). Zootaxa. 4018 (4): 506–534.

Miller, K. B. (2013). Notomicrus josiahi, a new species of Noteridae (Coleoptera) from Venezuela. Zootaxa, 3609(2): 243–247.

Sharp, D., (1882). On aquatic carnivorous Coleoptera or Dytiscidae. Scientific Transactions of the Royal Dublin Society. 2: 179–1003.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Ento. Soc. Am. 87(6): 651–701.

Van der Auwera, G., S. Chapelle, and R. De Wächter. (1994). Structure of the large ribosomal subunit RNA of Phytophthora megasperma, and phylogeny of the oomycetes. FEBS letters. 338(2): 133-136.

Wild, A. L., and D. R. Maddison. (2008). Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. Molecular Phylogenetics and Evolution, 48(3), 877–891.

Young F. N. (1978). The New World species of the water-beetle genus Notomicrus (Noteridae). Systematic Entomology. 3(3): 285–293.

Supplementary Materials and Methods.

Supplementary Materials and Methods
    Taxon Sampling.

We were able to include a single species of *Phraetodytes* (*P. haibaraensis* Kato et al. 2010). Within

Notomicrus, we sampled 13 of the 17 described species, in addition to 11 putative new species, for

a total of 24 ingroup species (Supp Table S1). Our initial phylogeny investigating the relationships

within *Notomicrus* included 47 ingroup samples (including *Phreatodytes*) and seven outgroups for

a total of 54 samples (Supp. Table S2). We used reduced sampling for divergence time estimation

and biogeographic analysis. Following initial ML and phylogenetic analyses, both ML and BI, we

reduced sampling to individual terminals for each tentative species, for a total of 54 samples. In

the case of *Notomicrus tenellus*, we included terminals across its wide distribution, as it appears

multiple species are subsumed under this name. In the case of the *traili* complex, we included a

representative for each larger clade within the group, including *N. sabrouxi* Manuel, 2015. We

also trimmed out some of the more distant outgroup Amphizoidae. This resulted in a dataset of

We used type material of 13 known species to accurately identify described versus undescribed

species where needed.

    DNA extraction and data collection

DNA was extracted from whole beetles, stored in ≥95% ethanol, using a Qiagen DNeasy kit

(Qiagen, Hilden, Germany), following the manufacturers protocols. As notomicrines are very

small, the following extraction modifications were added to maximize DNA yield, following

several of the recommendations of Craaude et al. (2019): (1) beetle abdomens were separated from

the thorax to allow lysis buffer access to thoracic tissue; (2) each sample was lysed overnight and

vortexed several times during lysing; (3) DNA was eluted in two subsequent elutions of 100ul and

50-75ul, respectively, with heated elution buffer AE (56° C), and allowing each a 15 minute incubation prior to centrifuging.

We used the primers listed in Table 1 in polymerase chain reactions to recover sequence data of the following mitochondrial and nuclear gene fragments: cytochrome oxidase subunit 1 mtDNA (CO1), carbamoylphosphate synthetase nDNA(CAD), histone 3 nDNA(H3), 16S mtrDNA and 28S rDNA. Polymerase chain reactions followed Baca et al. (2017b; H3, 16S, 28S), Simoes et al. (2019; CAD) and the Beetle DNA lab (https://zsm-entomology.de/wiki/The_Beetle_D_N_A_Lab; accessed 2019; COI), see Table 1. Amplicons were diluted where necessary and sent to GENEWIZ-Boston, MA (Brooks Automation Inc., Chelmsford, MA, USA) for Sanger sequencing).

Alignment and phylogenetic analyses.

Alignment and partitioning. Sequence data were assembled, aligned and concatenated with Geneious 10.2.2 (Biomatters, http://www.geneious.com). Trace files were visually inspected and refined before generating contigs. Data for each fragment were aligned using MAFFT v7.450 (Katoh and Standley, 2013) with default settings (Algorithm = Auto; Scoring matrix = 200 PAM, k = 2; Gap open penalty = 1.53; Offset value = 0.123). Alignments were visually inspected and protein coding alignments were translated to amino acid reading frames to verify absence of stop codons. Individual fragment alignments were then trimmed and concatenated. The initial alignment of 54 samples was 3422bp in length; the down-sampled alignment of 32 samples was 3,382bp in length.

To search for optimal models for phylogenetic inference, alignments were partitioned *a priori* by alignment features, following Baca et al. (2017b); respective alignments were divided into subsets of individual gene fragments with protein coding genes (COI, CAD, H3) each further divided by

codon position, for a total of 11 partitions for each alignment (Supp. Table S3). Models were searched using IQ-Tree 2 (v.2.0-rc1; Minh et al. 2020) using the -m TESTMERGE or -m TESTMERGEONLY functions which implement the *greedy* search algorithm of PartitionFinder (Lanfear et al. 2012) and allows the merging of input partitions to increase model fit and reduce the risk of over-parameterization, similar to PartitionFinder. We used an edge-linked model search which assumes the same underlying branch lengths for all partitions, but allows for different rates. For Bayesian inference, models were restricted to those available in the inference software using the -mset option, however this was done only to recover optimal merging of partitions (see below). For Maximum Likelihood inference (hereby ML) we used IQ-Tree 2 using the -m TESTMERGE option (model search followed by phylogenetic inference), with edge-linked partitions (-p or -spp). Support was assessed using UFBoot 2 Ultrafast bootstrapping (Hoang et al., 2017) with 1000 replicates (-B or -bb 1000). We considered a UFBoot bootstrap support of 95% or higher to be strong support for a recovered clade.

Bayesian inference analyses were conducted on both the initial and reduced alignments in BEAST2 2.6.1 (Bouckaert et al. 2019) via the CIPRES Science Gateway (Miller et al. 2010), with the .xml file prepared in BEAUti 2 (Bouckaert et al. 2019). We used the merged partitioning scheme recovered by the TESTMERGEONLY option in IQ-Tree 2. We unlinked site models across partitions, but linked clock models and trees. We did not use the models recovered by IQ-Tree for site models. Instead we implemented bModelTest (Bouckaert et al. 2017), which uses a Bayesian reversible jump to sample among different substitution models (including site heterogeneity and invariant site selection), with mutation rate estimated. This eliminates potential bias in using models recovered by ML methods (Bouckaert et al. 2017). Analyses were ran with an uncalibrated strict clock, under a Yule tree model with an exponential birthrate prior (initial =

1.0, +/- infinite bounds), with a random initial tree. Analyses consisted of two independent runs of 50 million generations each, with trace and trees logged every 2500 generations. Log files were together viewed in Tracer 1.6 (Rambaut et al. 2017; https://beast.community/tracer) to assess convergence and ensure sufficient sampling of parameters, with an effective sample size (ESS) value of 200 considered as the threshold for good convergence. Tree files were combined and annotated with LogCombiner 2.6.2 and TreeAnnotator 2.6.2 (https://beast.community/). Posterior values of 0.95 or higher was considered strong support for a given clade.

Divergence time estimation.

Divergence times were estimated using the reduced sampling alignment (with sampling trimmed to species' representatives; see above) in BEAST2, with .xml infiles generated using BEAUti 2. As above, we used the partitioning scheme recovered by IQ-tree, with unlinked site models, and linked clock and tree models. For all partition site models, we used bModelTest with mutation rate estimated. In lieu of fixing the tree topology, we implemented the recovered tree from the uncalibrated analyses above as the starting tree. We implemented four different combinations of clock and tree models to test the effect on divergence times: strict vs. uncorrelated lognormal relaxed clock models and Yule vs Birth Death tree models (Table 1). We did not test the effect of unlinked clock models as initial analyses resulted in very poor MCMC convergence. Exponential birthrate (Yule and Birth Death) and deathrate (Birth Death) tree prior distributions were respectively used across all analyses. For both Strict clock rate and uncorrelated lognormal relax clock mean, we used a uniform prior distribution, with an initial value of 0.01, and bounded from 0.00001 to 1.0 to prevent numerical instability in the analysis.

There are currently no described fossil Noteridae; consequently, we used secondary node calibrations.

There are currently no described fossil Noteridae; consequently, we used secondary node calibrations. Using the ages recovered by Toussaint et al. (2017b), we implemented MRCA node calibrations on Meruidae+Noteridae (191.004 Mya; 95% HPD = 167.2, 216.0) and Noteridae (153.896 Mya; 95% HPD = 124.6, 183.3). Nodes were calibrated using a normal distribution, with the mean set to the respective mean ages of Toussaint et al. (2017b), with Sigma values set so that the 95% prior distribution density quantiles matched as near as possible the respective height posterior densities intervals (95% HPD) of Toussaint et al. (2017b) (Meruidae+Noteridae: Mean = 191.004, Sigma = 14.7, Offset = 0.4; Noteridae: Mean = 153.896, Sigma = 18.0, Offset = 0.0).

All analyses were run on CIPRES and consisted of two independent runs of 100 million generations with traces logged and trees sampled every 5000 generations. Logs were imported into Tracer 1.6 to assess analysis convergence, with an effective sample size (ESS) value of 200 considered the threshold for good convergence. Recovered tree files were then combined in LogCombiner. Respective Maximum Clade Credibility (MCC) trees were generated with TreeAnnotator, with mean node heights, 95% HPD intervals and 20% burn-ins.

The fit of the four model combinations implemented were compared using marginal likelihoods estimated with Nested Sampling (Skilling 2006; Maturana Russel et al. 2018a). Nested Sampling (NS) differs from "power posterior" methods like Steppingstones Sampling (SS) or Path Sampling (PS) by exploring parameter space from the prior which becomes increasingly restricted to areas of high likelihood, rather than relying on power posterior densities (Maturana Russel 2018). The method is accurate and uses less prior tuning and is more generally applied than SS or PS (Maturana Russel et al 2018; and additionally calculates the standard deviation of the estimated

marginal likelihoods (see Maturana Russel et al., 2018). We set up the NS runs by manually editing the BEAUti generated .xml infiles. Our NS runs consisted of the following parameters: chain length = 20,000; particle count = 12; subchain length = 10,000. To select the best fit model combination, we calculated Bayes Factors (BF; Kass and Raftery, 1995) using the recovered marginal likelihood estimates (MLEs), calculated as a ratio of the marginal likelihoods: $2\log(BF)=2x(MLE_1 - MLE_2)$. A BF $\geq$ 10 was used as the threshold for strong support of a given model combination Kass and Raftery, 1995.

Ancestral Range Reconstruction.

Biogeographic inference was conducted with the BioGeoBEARS (Matzke, 2018) package in R v. 3.5.0 (R Core Team, 2018). This package reconstructs ancestral ranges in Bayesian and likelihood frameworks, offering models (and model testing) of various combinations of anagenetic and cladogenetic evolutionary processes, including, dispersal, extinction, vicariance, sympatry and others. Analyses were conducted under the Dispersal Extinction Cladogenesis (DEC) and Dispersal Vicariance Analysis (DIVALIKE) models; both include parameters for dispersal, extinction, and vicariance, but differ in including sympatric cladogenesis (included in DEC), and widespread vicariance (included in DIVALIKE) parameters. We also included iterations of these analyses with the jump dispersal (j) parameter. This models founder-event cladogenesis, allowing a daughter lineage to inhabit a new range while its sister inherits the ancestral range. Inclusion of the +j parameter has shown to increase model likelihood (Matzke, 2012; 2014), but has been suggested to be flawed by Ree and Sanmartín (2018). As the imposed models of range evolution do not account for time, the probability of time-dependent anagenetic processes of range expansion via dispersal and extinction are underestimated, and the contribution of cladogenetic events to the likelihood artificially increases. The inclusion of the j parameter further exacerbates this, first by

adding an additional cladogenetic process of range evolution, and second by j being a free cladogenetic parameter, that, at high values, can over-power any probability of non-jump events (Ree and Sanmartín, 2018). Because of the potential effect on inference, and the likely inappropriateness of comparing models with j to those without (Ree and Sanmartín, 2018), we are excluding the +j parameter from our biogeographic models.

For the analyses we used the MCC time-calibrated tree from the above BI analysis preferred by BF of marginal likelihoods. Outgroups were manually pruned to only include notomicrine taxa (*Phreatodytes + Notomicrus*). We included the following regions in our analyses: Oriental, O; Oceania, A; Nearctic, N; and South America/Neotropical, S. These encompass all regions inhabited by known Notomicrinae. For disambiguation, Oceania includes Australia and Polynesia, roughly separated from the Oriental region by the Wallace line; the oriental region here includes Japan in our coding; Nearctic is restricted to North America, excluding south and eastern Mexico; Neotropical includes South and Central America and the Antilles. We followed the time-slice, dispersal rate, and adjacency and areas allowed scheme of Toussaint et al. (2017a) (excluding the African and Palearctic regions) as our focal clade age and range were very similar. In brief, we used four time slices, each with its own dispersal rate multiplier matrix following Toussaint et al. (2017a). Baseline dispersal rates (*d*) between areas were penalized by scalars (*s*) according to distance and geographic barriers along the shortest path between areas: adjacent areas with no barrier were not penalized, $s = 1.0$ ($d_{adj} = 1.0d$); areas separated by a small water barrier were scaled by 0.75 ($d_{swb} = 0.75*d$); areas separated by another area were scaled by $s = 0.5$ ($d_a = 0.5*d$); areas separated by a larger water barrier were scaled by $s = 0.25$ ($d_{lwb} = 0.25*d$). The final dispersal rate matrices were calculated for each time slice following this scheme, with consideration to

multiple barriers, with each time slice allowed a different adjacency matrix. See Toussaint et al. (2017a) for complete details.

References

Baca, S.M., E. F. A. Toussaint, K.B Miller, and A. E. Z. Short. (2017b). Molecular phylogeny of the aquatic beetle family Noteridae (Coleoptera: Adephaga) with an emphasis on partitioning strategies. Molecular Phylogenetics and Evolution. 107: 282–292.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, ... and A. J. Drummond. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLOS Computational. Biology. 10 (4): e1003537.

Bouckaert, R., and A. J. Drummond. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evolutionary Biology 17.1: 42.

Bouckaert, R., T. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, ... and M. Matschiner. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 15(4): e1006650.

Cruaud, A., S. Nidelet, P. Arnal, A. Weber, L. Fusu, A. Gumovsky, ... and J. Y. Rasplus. (2019). Optimized DNA extraction and library preparation for minute arthropods: application to target enrichment in chalcid wasps used for biocontrol. Mol. Ecol. Resour. 19(3): 702–710.

Hoang, D. T., O. Chernomor, A. Von Haeseler, B. Q. Minh, and L. S. Vinh. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35(2): 518–522.

Kass, R. E., and A.E. Raftery. (1995). Bayes factors. J Am Stat Assoc. 90(430): 773–795.

Kato, M., A. Kawakita, T. Kato. (2010). Colonization to aquifers and adaptations to subterranean interstitial life by a water beetle clade (Noteridae) with description of a new Phreatodytes species. Zool. Sci. 27: 717–722.

Katoh, K., and D. M. Standley, (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Bio. Evol. 30(4): 772–780.

Lanfear, R., B. Calcott, S. Y. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Bio. Evol. 29(6): 1695–1701.

Manuel, M. (2015) The genus Notomicrus in Guadeloupe, with description of three new species (Coleoptera: Noteridae). Zootaxa. 4018 (4): 506–534.

Maturana Russel, P., B. J. Brewer, S. Klaere, and R. R. Bouckaert (2019). Model selection and parameter inference in phylogenetics using Nested Sampling. Syst. Biol. 68(2): 219–233.

Matzke, N. J. (2013). Probabilistic Historical Biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Ph. D. Dissertation, Department of Integrative Biology, University of California, Berkeley. Order No. 3616487, ProQuest Dissertations and Theses, pp. 1–240. Open-access at: http://search.proquest.com/docview/1526024556.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63(6): 951–970.

Matzke, N.J. (2015). http://phylo.wikidot.com/biogeobears; accessed 24.vii.2020).

Matzke, N. J. (2018). BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts. Version 1.1.1, published on GitHub. https://github.com/nmatzke/BioGeoBEARS.

Minh, B. Q., H.A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler, and R. Lanfear, (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Bio. Evol. 37(5): 1530–1534.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele, and M. A. Suchard (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Bio. 67(5): 901.

Ree, R. H., and I. Sanmartín. 2018. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. 45(4). 741–749.

Simões, M. V., Baca, S. M., Toussaint, E. F., Windsor, D. M., & Short, A. E. (2019). Solving a thorny situation: DNA and morphology illuminate the evolution of the leaf beetle tribe Dorynotini (Coleoptera: Chrysomelidae: Cassidinae). Zoological Journal of the Linnean Society, 185(4), 1123–1136.

Toussaint, E. F. A, D. Bloom, and A. E. Z. Short. (2017a). Cretaceous West Gondwana vicariance shaped giant water scavenger beetle biogeography. J. Biogeogr. 44(9): 1952–1965.

Toussaint, E. F.A., M. Seidel, E. Arriaga-Varela, J. Hájek, D. Kral, L. Sekerka, and M. Fikáček. 2017b. The peril of dating beetles. Syst. Entomol. 42(1): 1–10.

Chapter 4.

Supplementary Figures.

Figure S4.1. 70%CM-ET maximum likelihood cladogram inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.

N. josiahi SLE2072 [Venezuela]

N. nanulus SLE1127 [USA]

N. sp. 6 SLE1649 [Suriname]

N. gracilipes SLE2063 [Guatemala]
100/100
N. cf.gracilipes SLE1273 [Suriname]
100/100
N. traili SLE2017 [Brazil_Amazonas]
100/100

N. sp.3 SLE895 [Suriname]

N. petrareptans SLE2055 [Suriname]
100/100
N. petrareptans SLE2059 [Suriname]
100/100

100/100 N.gracilipes SLE1635 [Brazil_M_Gerais]

N. cf.gracilipes SLE1959 [Brazil_Amazonas]
1000/1000
N. cf.gracilipes SLE2074 [Brazil_Amapa]
100/100
100/100 gracilipes SLE2049 [Suriname]

N. traili SLE1885 [Brazil_Para]
100/100
N. traili SLE1888 [Brazil_Para]

100/90 traili SLE1873 [Suriname]
100/100
N. cf.gracilipes SLE1879 [Suriname]
11.5/65
100/100 traili SLE2020 [Brazil_Amapa]
100/89
N. traili SLE2028 [Brazil_Amapa]
99.9/93
N. traili SLE2034 [Brazil_Amapa]

N. cf.gracilipes SLE2024 [Brazil_Rondonia]
100/100
N. traili SLE1876 [Brazil_MG_do_Sul]
100/100
N. traili SLE2014 [Brazil_Rondonia]

N. traili SLE1639 [Brazil_Goias]

N. traili SLE1882 [Brazil_MG_do_Sul]
100/100
100/100 traili SLE2078 [Brazil_MG_do_Sul]

100/100 sobrouxi SLE2053 [Guadeloupe]
100/100
N. sobrouxi SLE909 [Guadeloupe]
100/100
N. traili SLE907 [Venezuela_Aragua]

100/100 traili SLE1636 [Venezuela_Bolivar]
5.8/85
100/100 traili SLE1534 [Brazil_Roraima]
100/100
N. traili SLE1537 [Brazil_Roraima]

N. traili SLE1961 [Brazil_Bahia]

N. traili SLE1641 [Brazil_Roraima]
100/100
N. traili SLE1870 [Suriname]
31.8/87
N. traili SLE1891 [Suriname]
100/100

N. traili SLE2065 [Brazil_Para]

100/100 traili SLE1631 [Guyana]
76/95
N. traili SLE2057 [Guyana]

N. traili SLE1644 [Suriname]
99.6/71
N. traili SLE2036 [Guyana]

N. traili SLE1632 [Suriname]
98/77
N. traili SLE1638 [Suriname]

88.8 traili SLE1633 [Suriname]
100/53
N. traili SLE1642 [Guyana]
90.2/44
N. traili SLE1640 [Guyana]
100/72
N. traili SLE2061 [Suriname]

70CM-ET Maximum likelihood

Node values: UFboot
ML tree/ML consensus tree

0.04

Figure S4.2. 70%CM-ET maximum likelihood tree inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.

Figure S4.3. 70%CM-GB maximum likelihood cladogram inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.

Figure S4.4. 70%CM-GB maximum likelihood tree inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.
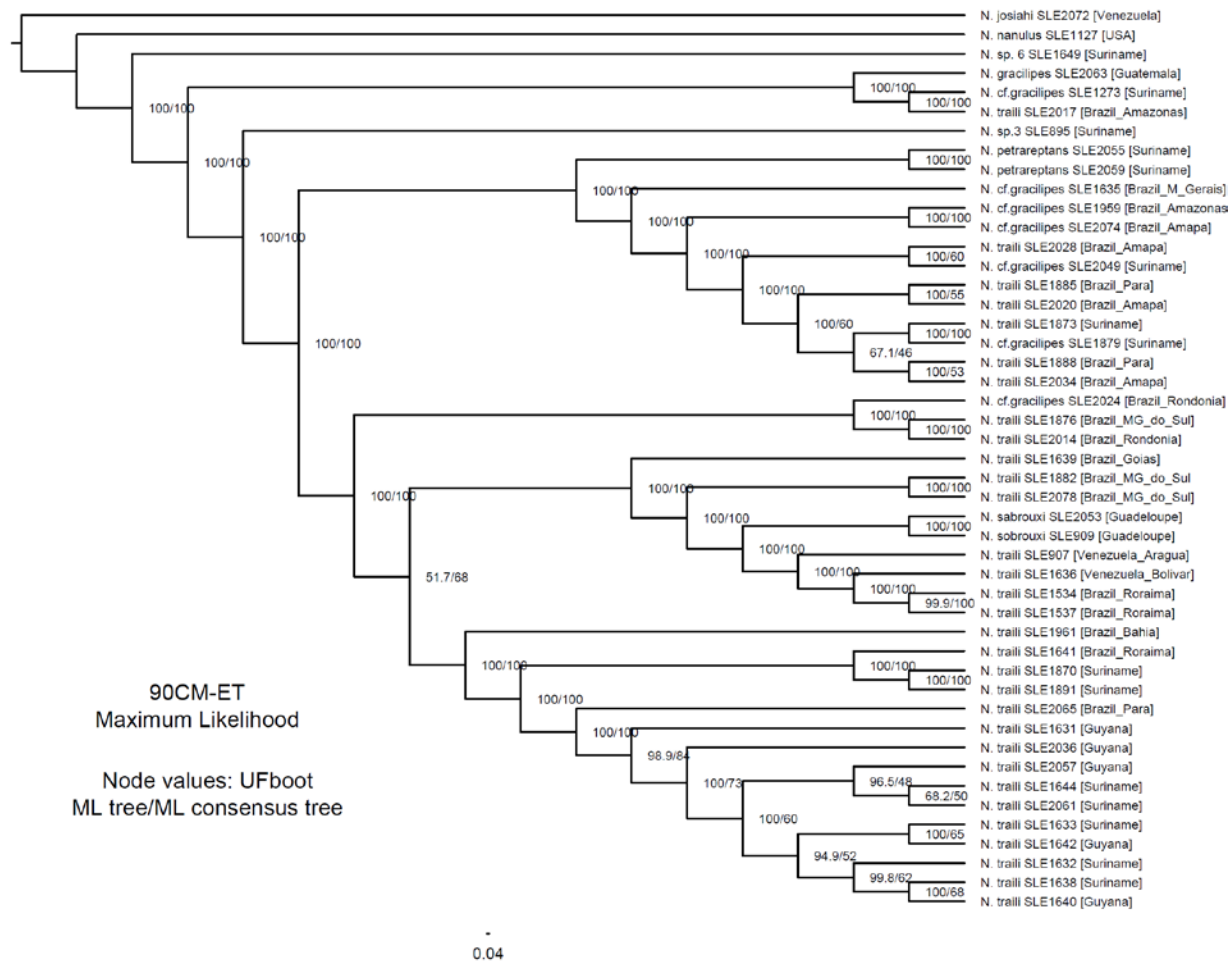
Figure S4.5. 90%CM-ET maximum likelihood cladogram inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.
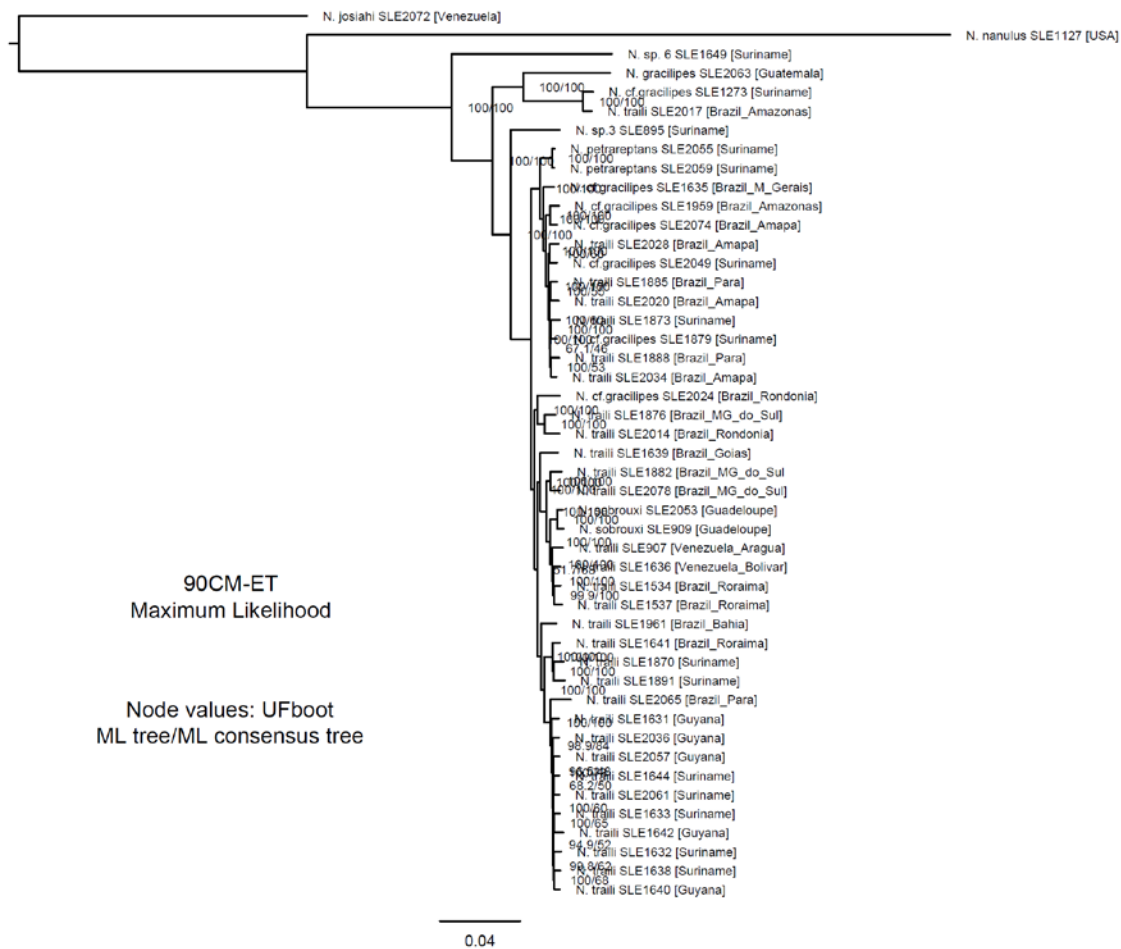
Figure S4.6. 90%CM-ET maximum likelihood tree inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.
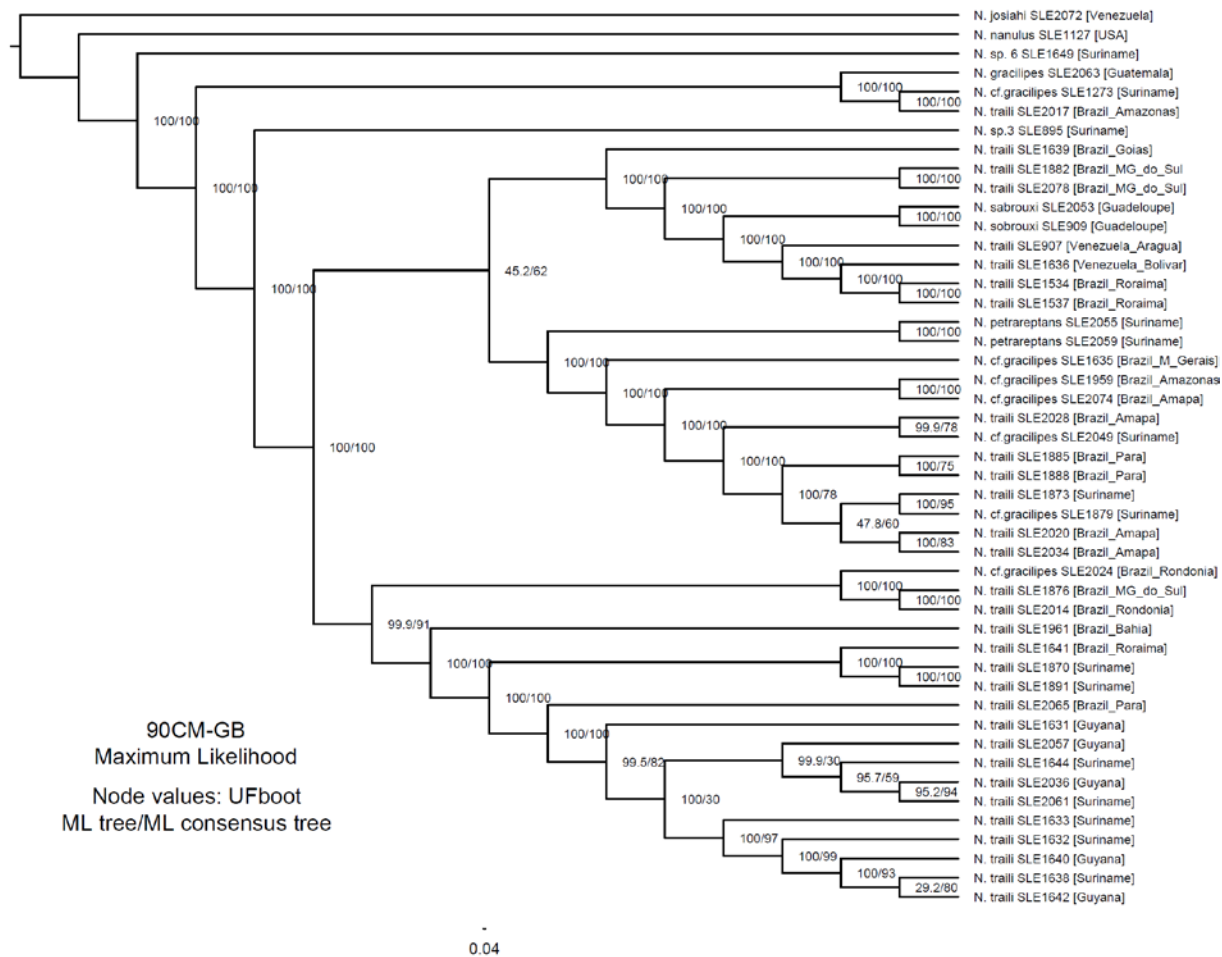
Figure S4.7. 90%CM-GB maximum likelihood cladogram inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.
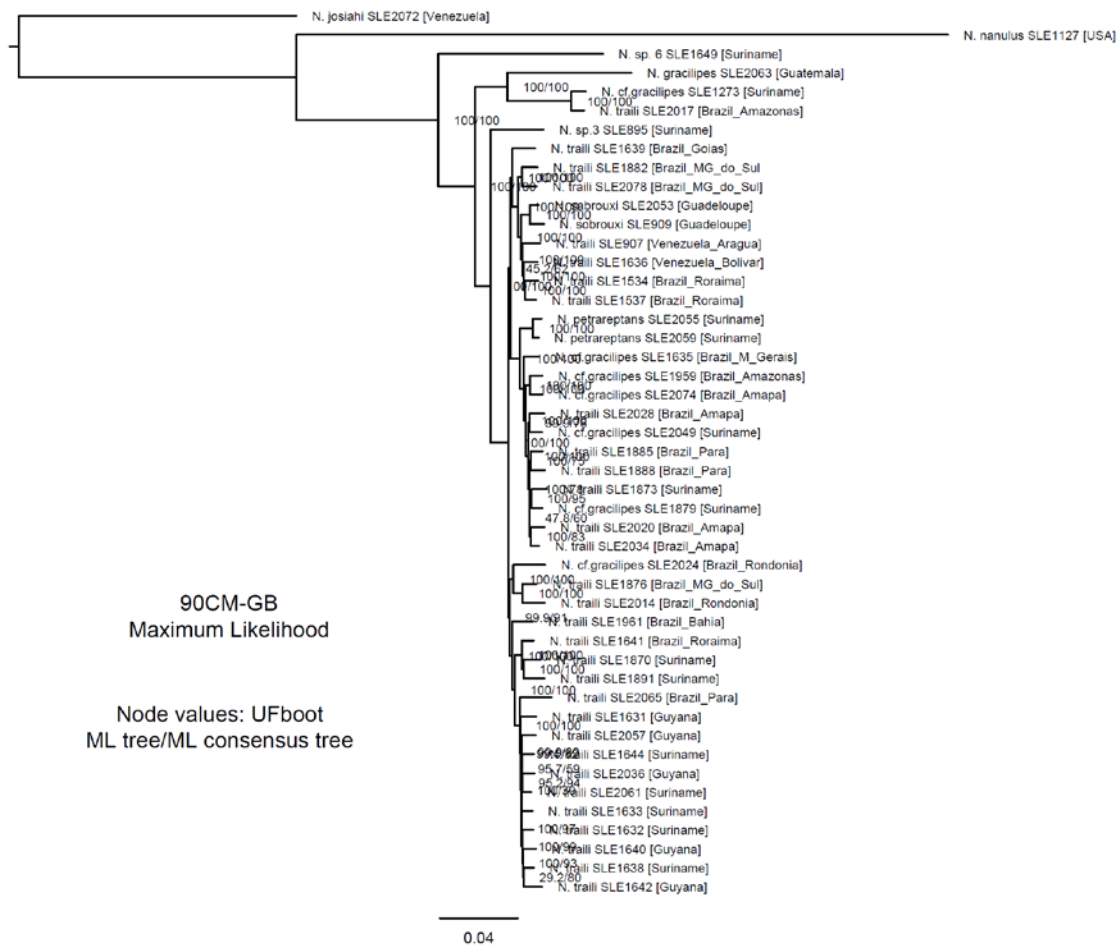
Figure S4.8. 90%CM-GB maximum likelihood tree inferred by IQ-Tree 2 under the GTR+G model. Values indicate UFboot support: Maximum Likelihood tree / Consensus tree.
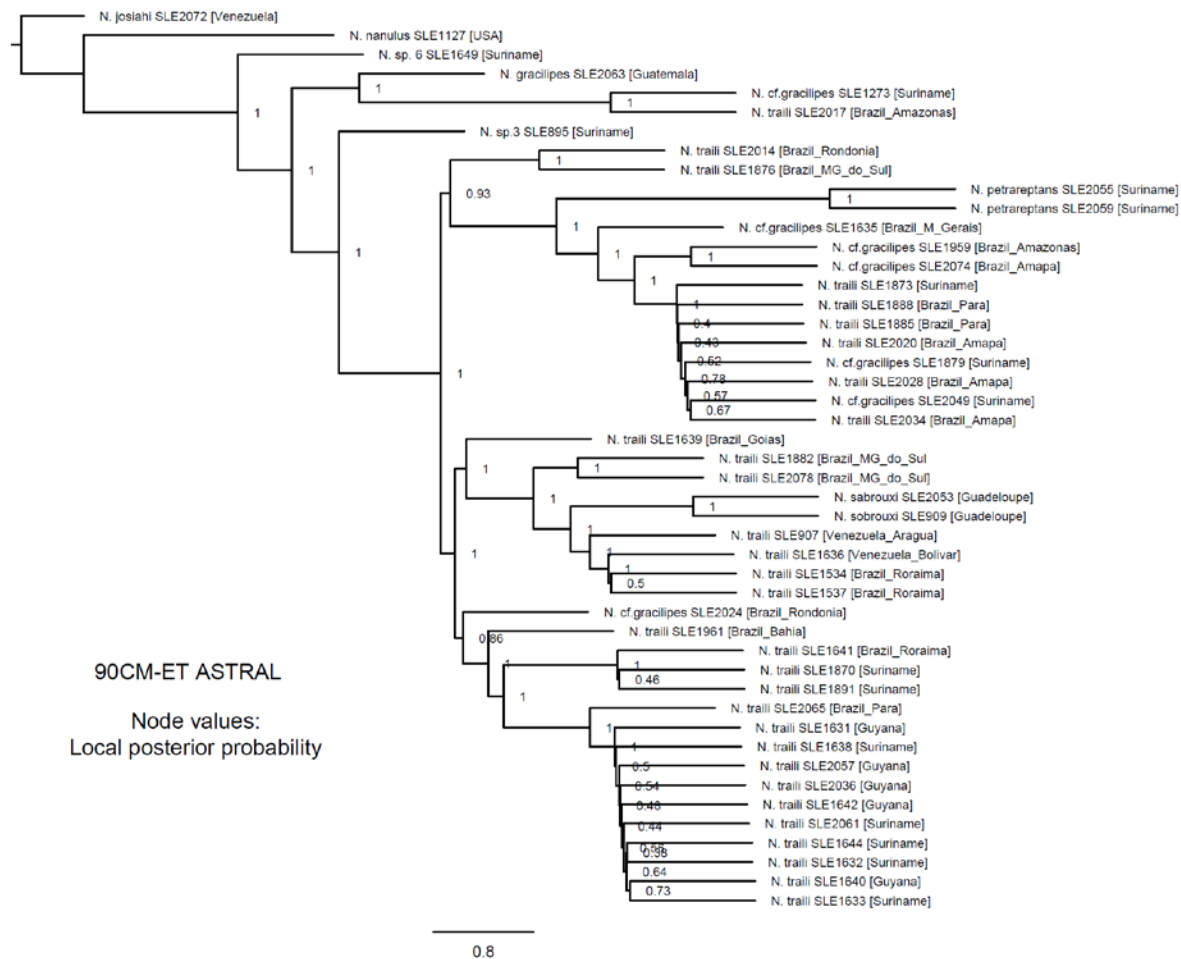
Figure S4.9. 90%CM-ET ASTRAL tree inferred by ASTRAL III. Values indicate local posterior probability.
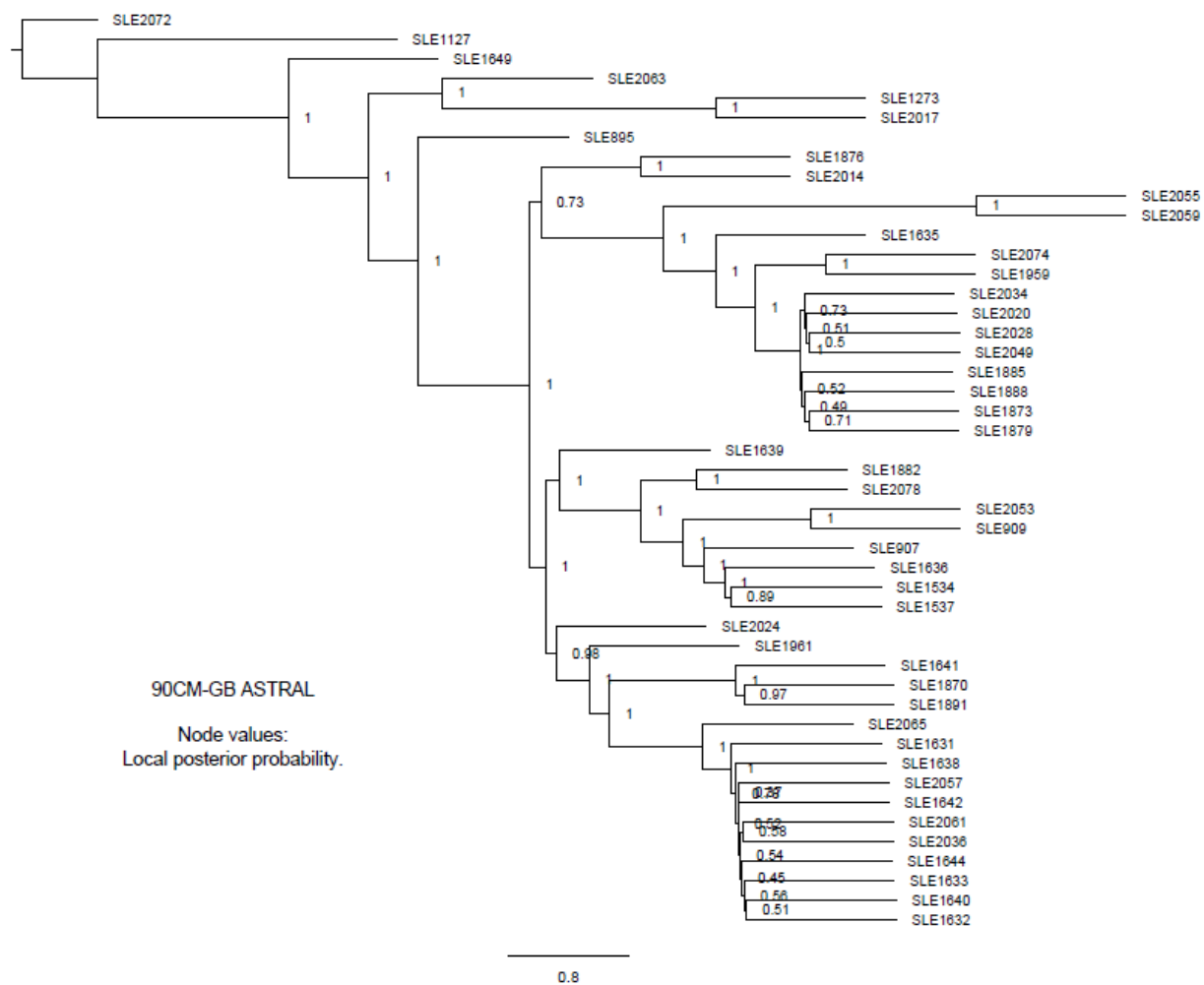
Figure S4.10. 90%CM-GB ASTRAL tree inferred by ASTRAL III. Values indicate local posterior probability.
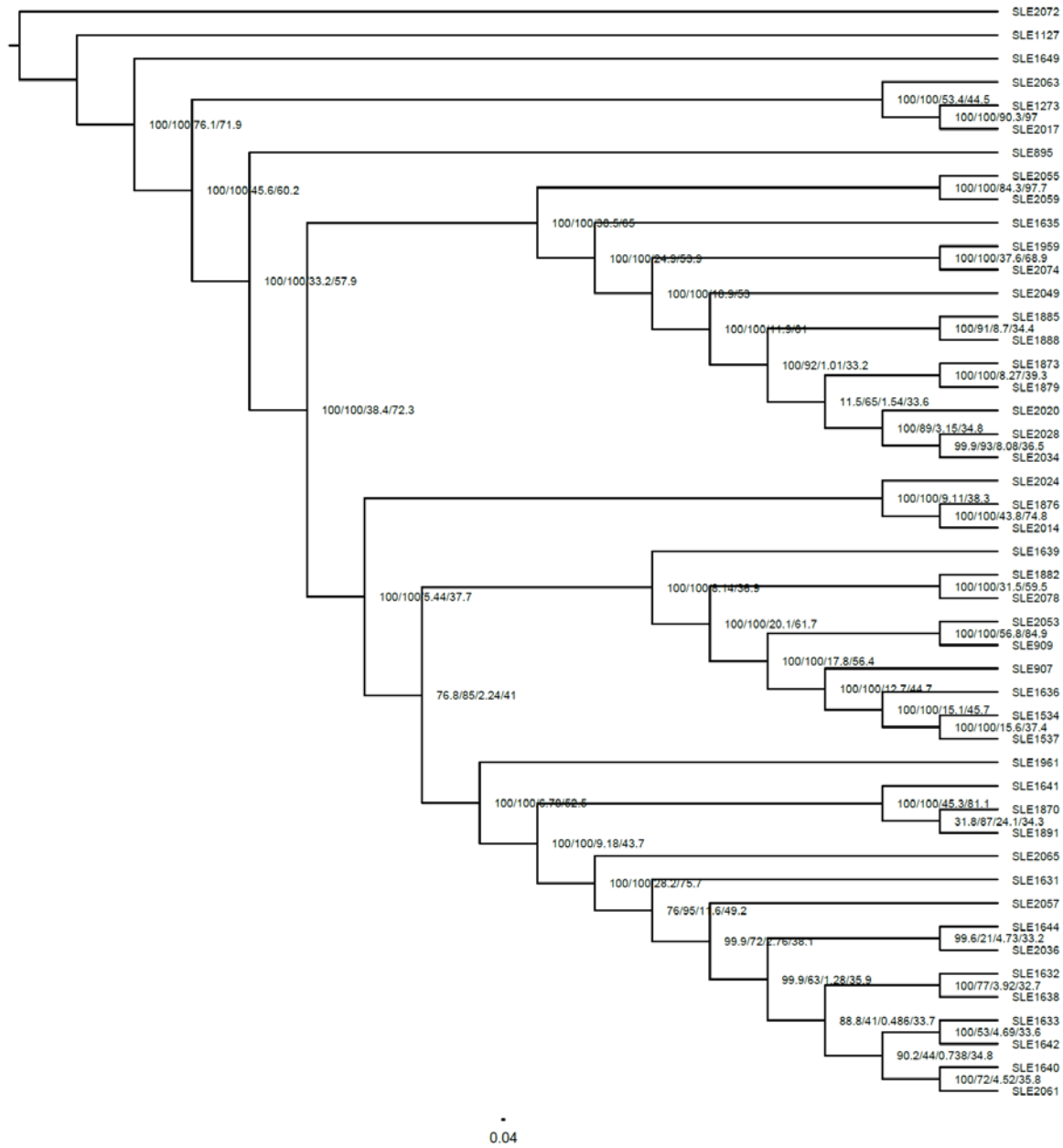
Figure S4.11. 70%CM-ET concordance factors. Values: UFboot Maximum likelihood/UFboot consensus/gCF/sCF
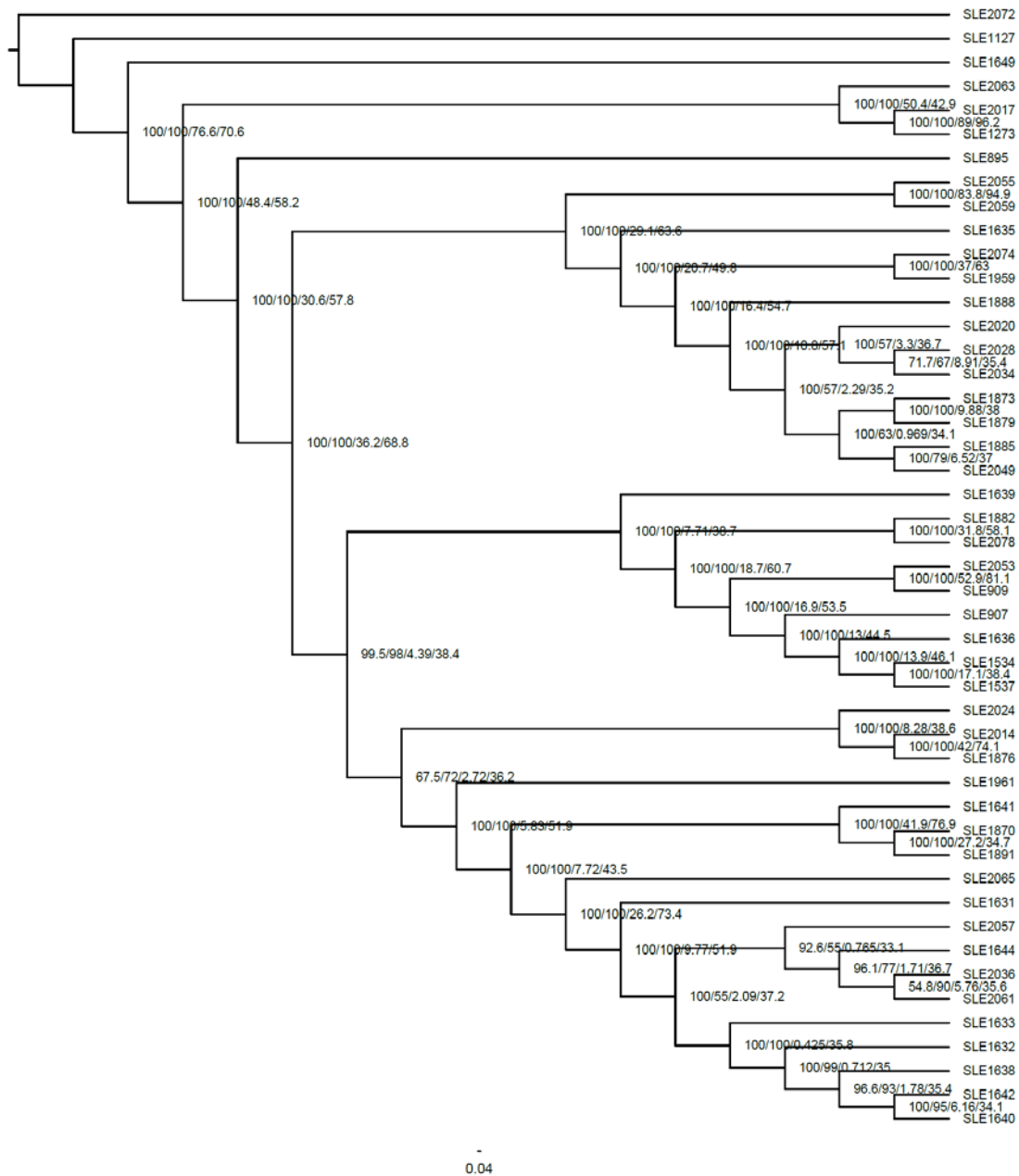
Figure S4.12. 70%CM-GB concordance factors. Values: UFboot Maximum likelihood/UFboot consensus/gCF/sCF

Figure S4.13. 90%CM-ET concordance factors. Values: UFboot Maximum likelihood/UFboot consensus/gCF/sCF

Figure S4.14. 90%CM-GB concordance factors. Values: UFboot Maximum likelihood/UFboot consensus/gCF/sCF
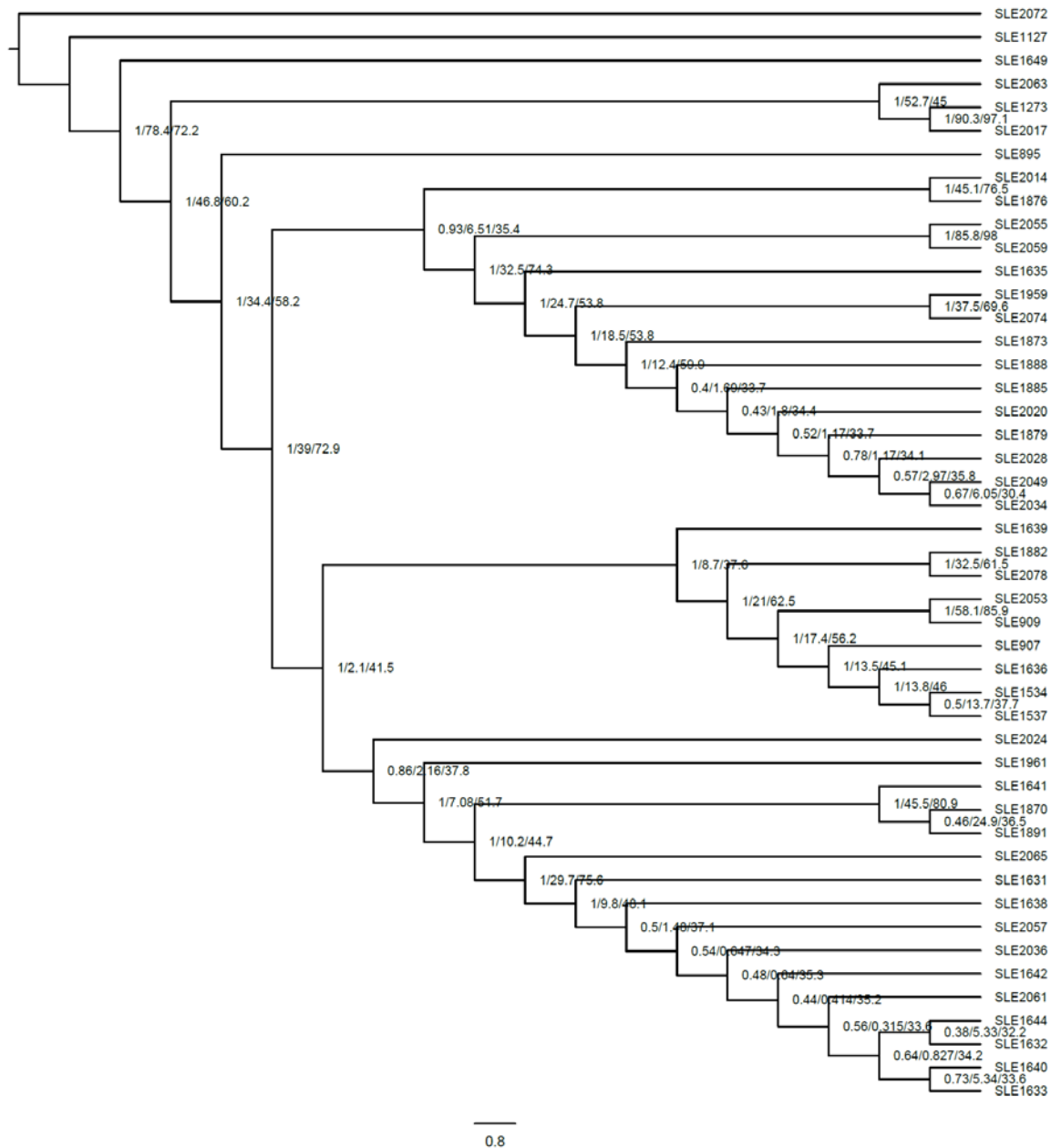
Figure S4.15. 90%CM-ET-ASTRAL concordance factors. Values: local posterior probability/gCF/sCF
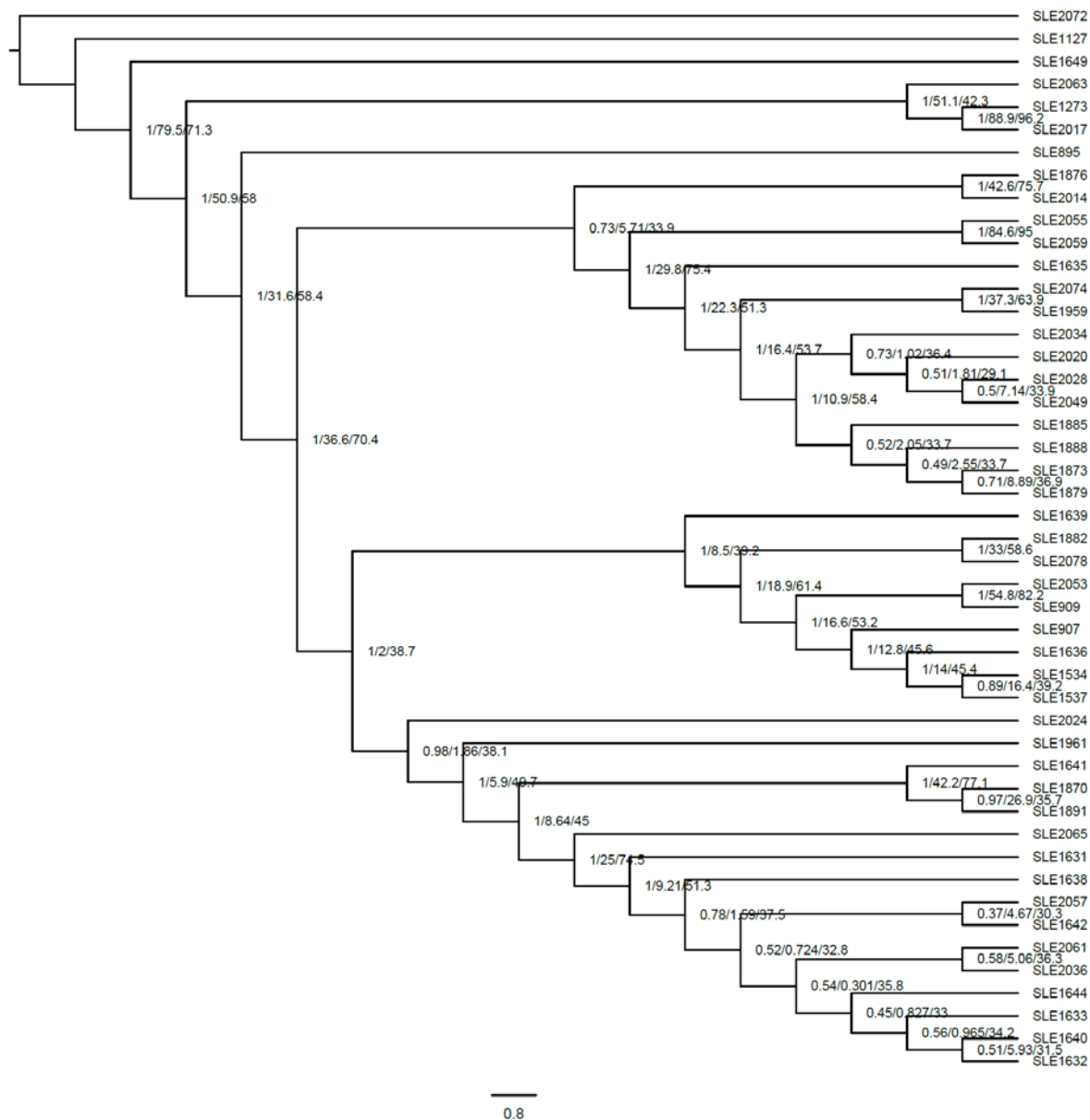
Figure S4.16. 90%CM-GB-ASTRAL concordance factors. Values: local posterior probability/gCF/sCF
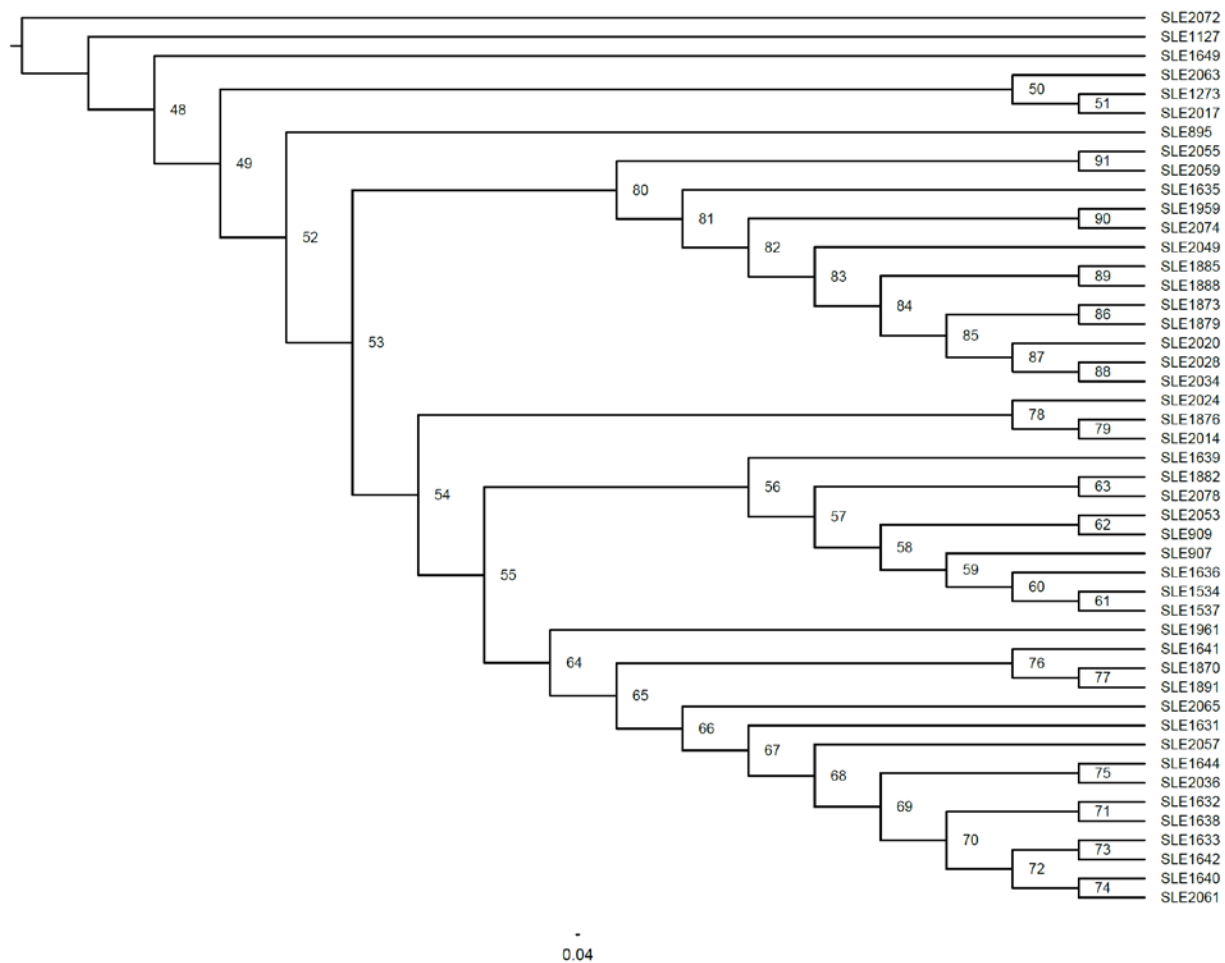
Figure S4.17. 70%CM-ET concordance factors ID. Branch labels correspond to 'ID' Table S4.4
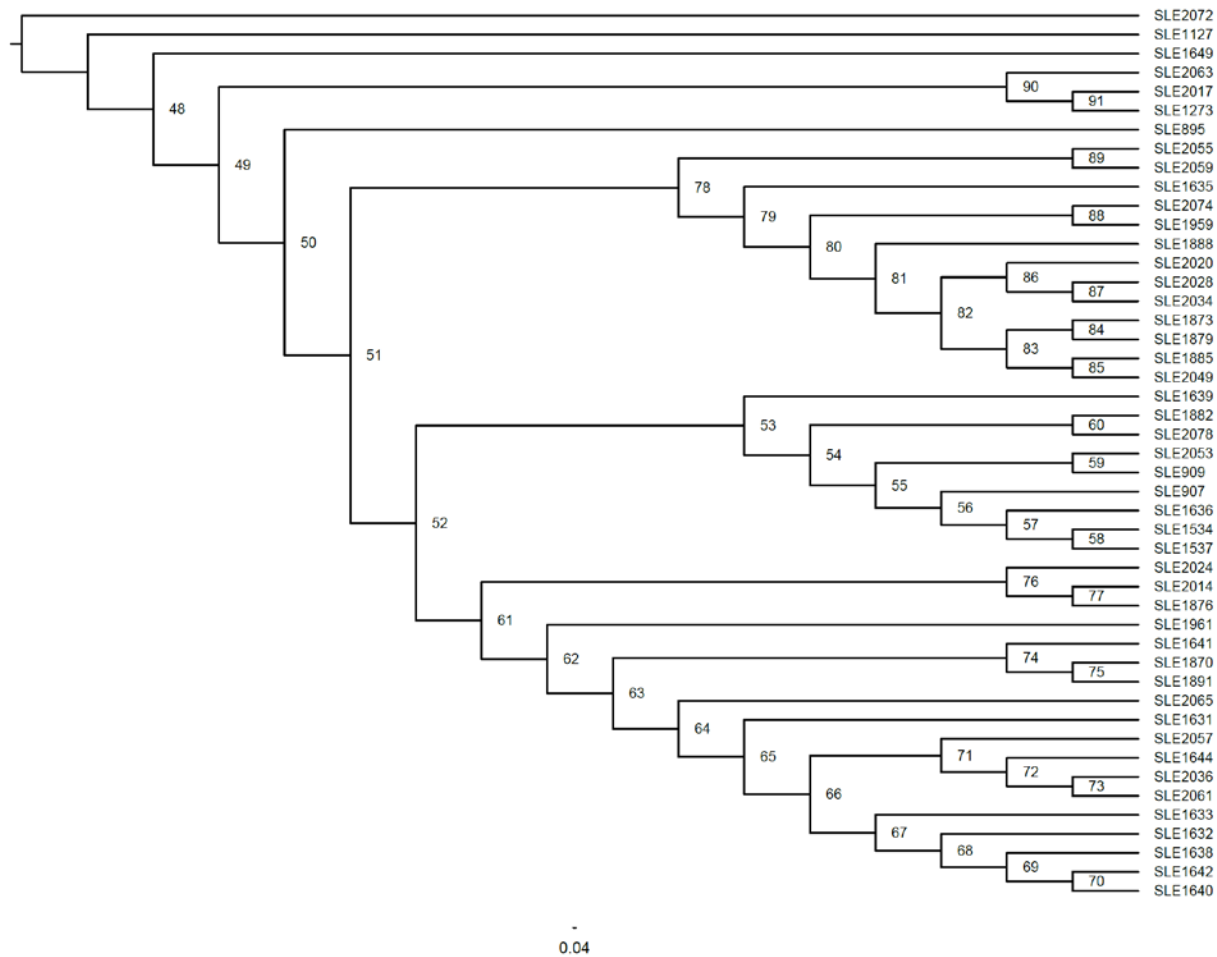
Figure S4.18. 70%CM-GB concordance factors ID. Branch labels correspond to 'ID' Table S4.5
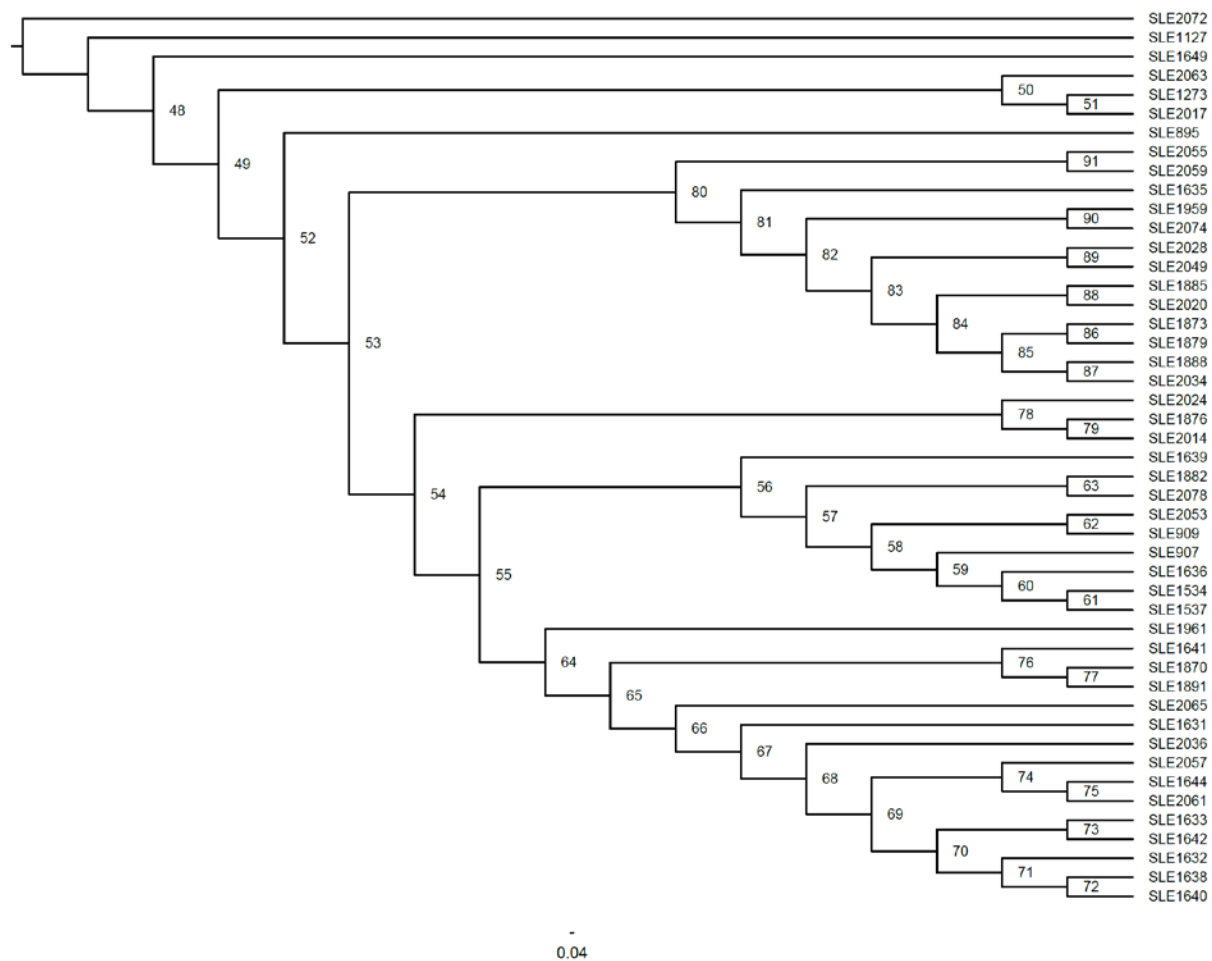
Figure S4.19. 90%CM-ET concordance factors ID. Branch labels correspond to 'ID' Table S4.6
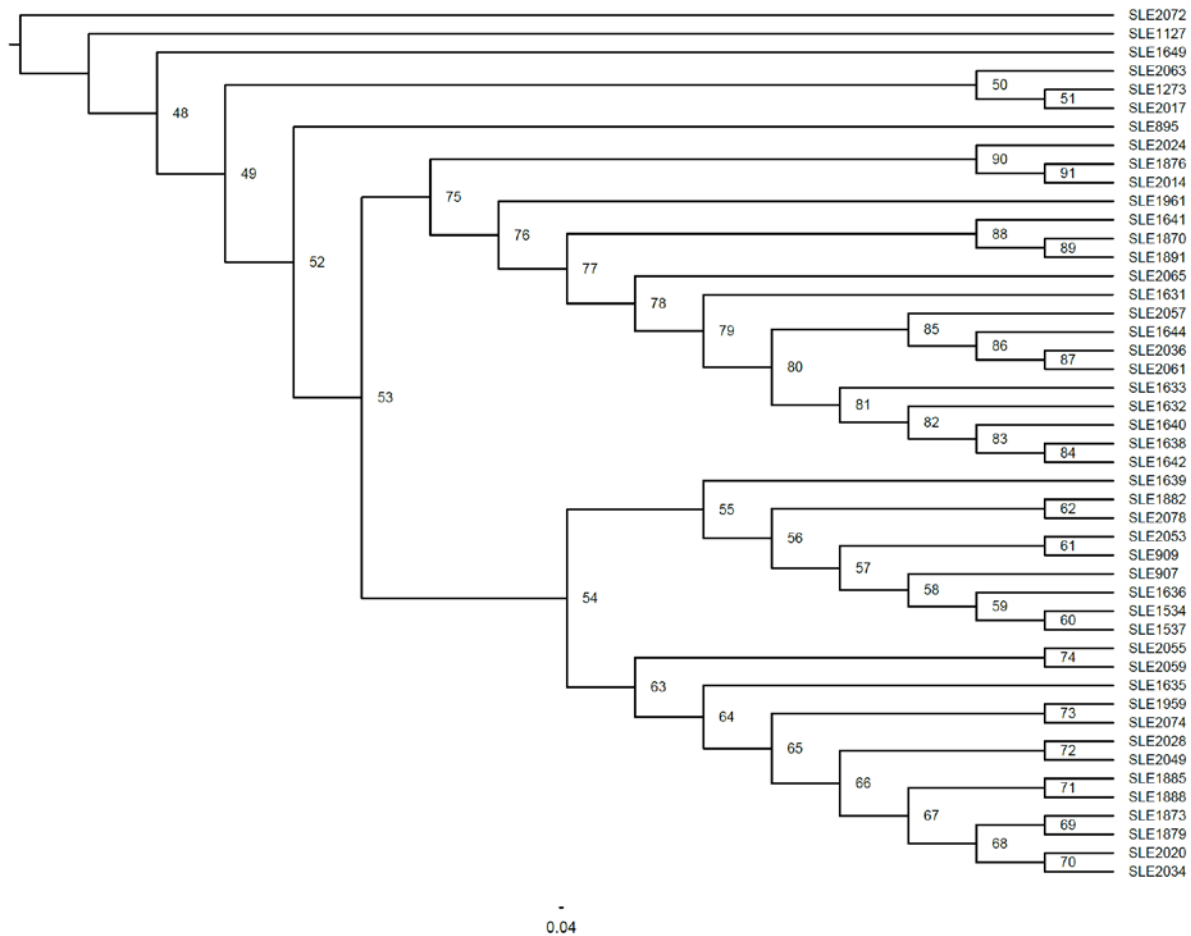
Figure S4.20. 90%CM-GB concordance factors ID. Branch labels correspond to 'ID' Table S4.7
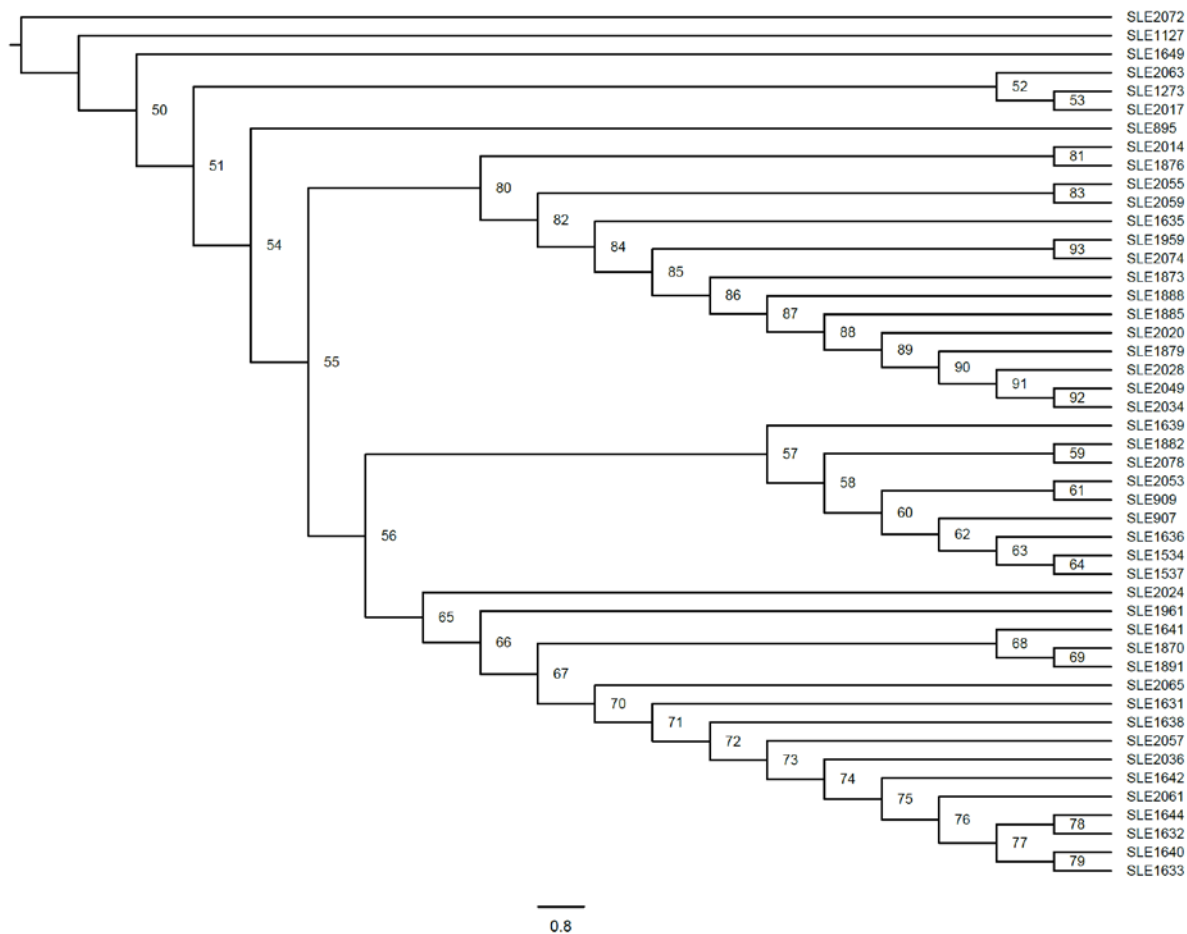
Figure S4.21. 90%CM-ET-ASTRAL concordance factors ID. Branch labels correspond to 'ID' Table S4.8
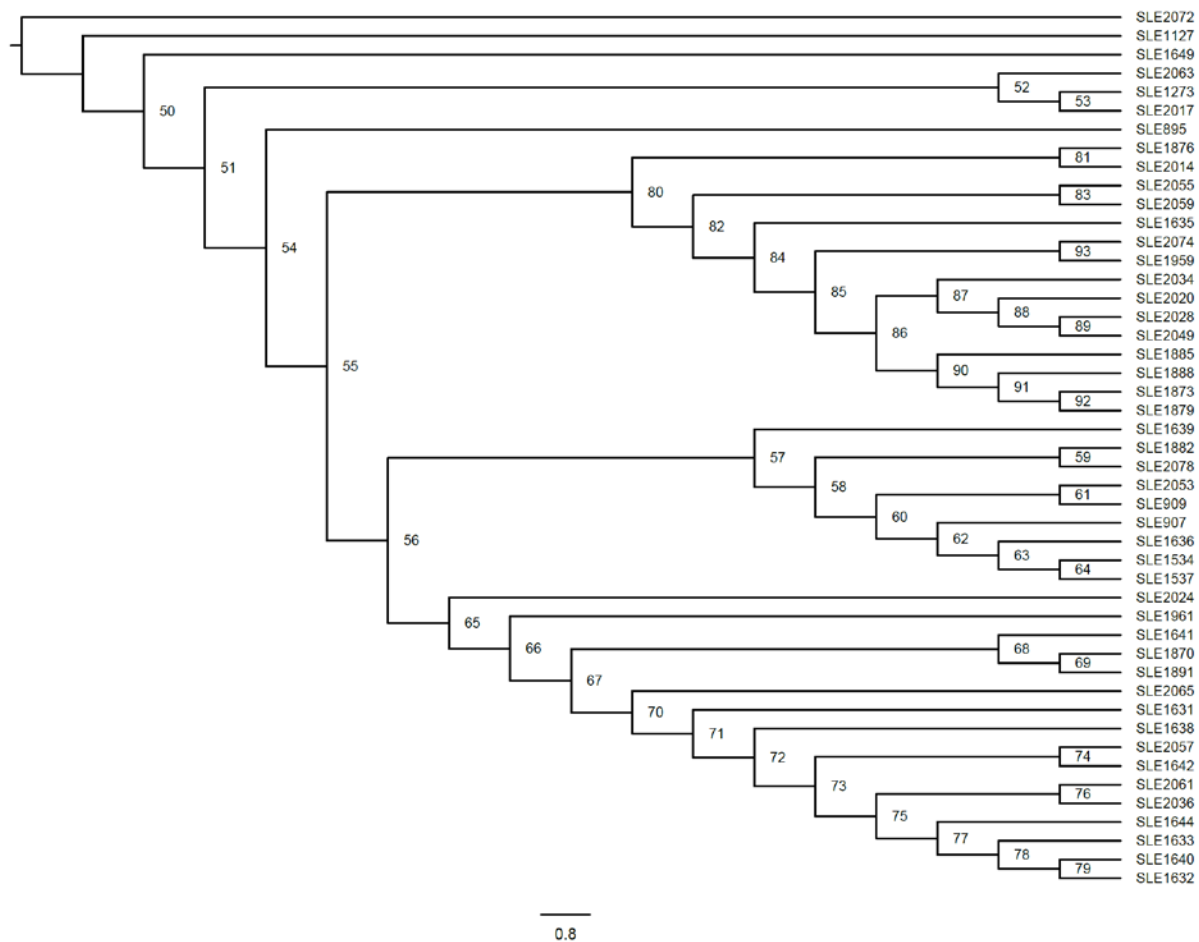
Figure S4.22. 90%CM-GB-ASTRAL concordance factors ID. Branch labels correspond to 'ID' Table S4.9.

Supplementary Tables

Table S4.1. Taxon sampling. Samples in descending order of tree tips in Fig. 4.2 of main text. Single Asterisk (*) indicates outgroup; double asterisk (**) indicates sample was dried museum specimen.

Taxon	Sample code	Country	State/Province	Year collected	Field code	Lat.	Long.
N. josiahi *	SLE2072	Venezuela	Amazonas	2009	VZ09-0116-01X	4.9808°	-67.7391°
N. nanulus *	SLE1127	USA	Florida	2016	SMB060816-A	30.0402°	84.9836°
N. sp. 6*	SLE1649	Suriname	Sipiliwini	2016	SR16-0316-01B	4.6739°	-56.1847°
N. gracilipes **	SLE2063	Guatemala	Retalhuleu		N/A	N/A	N/A
N. gracilipes	SLE1273	Suriname	Brokopondo	2017	SR17-0322-02A	4.9489°	-55.1804°
N. gracilipes	SLE2017	Brazil	Amazonas	2018	BR18-0710-02B	-10.9176°	-62.3770°
N. sp. 3	SLE895	Suriname	Sipiliwini	2010	SR10-0819-01A	2.1753°	-56.7874°
N. petrareptans	SLE2055	Suriname	Sipiliwini	2010	SR10-0901-01A	2.1829°	-56.7873°
N. petrareptans	SLE2059	Suriname	Sipiliwini	2010	SR10-0901-01A	2.1829°	-56.7873°
N. c.f.*gracilipes*	SLE1635	Brazil	Minas Gerais	2018	BR18-0228-02A	-15.1707°	-42.8035°
N. c.f.*gracilipes*	SLE1959	Brazil	Amazonas	2018	BR18-0706-04A	-8.4410°	-61.6603°
N. c.f.*gracilipes*	SLE2074	Brazil	Amapá	2018	BR18-0705-03E	-8.4376°	-61.7506°
N. c.f.*gracilipes*	SLE2049	Suriname	Sipiliwini	2013	SR13-0824-03A	3.7913°	-56.1495°
N. traili	SLE1885	Brazil	Para	2018	BR18-0203-01E	-1.4929°	-54.5157°
N. traili	SLE1888	Brazil	Para	2018	BR18-0203-01H	-1.4929°	-54.5157°
N. traili	SLE1873	Suriname	Sipiliwini	2017	SR17-0330-04A	2.0040°	-55.9710°
N. c.f.*gracilipes*	SLE1879	Suriname	Sipiliwini	2017	SR17-0401-01A	2.0109°	-55.9845°
N. traili	SLE2020	Brazil	Amapá	2018	BR18-0720-04A	3.8116°	-51.7837°
N. traili	SLE2028	Brazil	Amapá	2018	BR18-0724-01A	0.8090°	-51.9797°
N. traili	SLE2034	Brazil	Amapá	2018	BR18-0722-03A	1.9314°	-50.8610°
N. c.f.*gracilipes*	SLE2024	Brazil	Rondonia	2018	BR18-0708-01A	-8.9081°	-62.1792°
N. traili	SLE1876	Brazil	Mato Grosso do Sul	2018	BR18-0622-03A	-20.4509°	-55.6218°
N. traili	SLE2014	Brazil	Rondonia	2018	BR18-0710-02B	-10.9176°	-62.3770°
N. traili	SLE1639	Brazil	Goias	2018	BR18-0222-01A	-14.2774°	-47.7508°
N. traili	SLE1882	Brazil	Mato Grosso do Sul	2018	BR18-0625-02A	-19.3062°	-57.6038°
N. traili	SLE2078	Brazil	Mato Grosso do Sul	2018	BR18-0625-01A	-18.9518°	-57.6664°
N. sabrouxi	SLE2053	Guadeloupe	Basse-Terre	2013	MM-190813-2	16.1799°	-61.6729°

N. sabrouxi	SLE909	Guadeloupe	Basse-Terre	2013	MM-190813-2	16.1799°	-61.6729°
N. traili	SLE907	Venezuela	Aragua	2009	VZ09-0104-02B	10.3937°	-67.7960°
N. traili	SLE1636	Venezuela	Bolivar	2010	VZ10-0713-03A	7.3848°	-61.3253°
N. traili	SLE1534	Brazil	Roraima	2018	BR18-0111-03B	2.3837°	-60.5227°
N. traili	SLE1537	Brazil	Roraima	2018	BR18-0111-03A	2.3837°	-60.5227°
N. traili	SLE1961	Brazil	Bahia	2018	BR18-0224-03A	-11.6749°	-40.9964°
N. traili	SLE1641	Brazil	Roraima	2018	BR18-0117-04A	0.9131°	-59.5733°
N. traili	SLE1870	Suriname	Sipiliwini	2017	SR17-0402-02B	2.0040°	-55.9710°
N. traili	SLE1891	Suriname	Sipiliwini	2019	SR19-0313-01B	4.4060°	-57.2207°
*N. traili**	SLE2065	Brazil	Para		N/A	N/A	N/A
N. traili	SLE1631	Guyana	Region 8	2014	GY14-0319-02A	5.3044°	-59.8376°
N. traili	SLE2057	Guyana	Region 9	2013	GY13-1031-01A	2.1818°	-59.3384°
N. traili	SLE1644	Suriname	Sipiliwini	2010	SR10-0904-01A	2.36293°	-56.6977°
N. traili	SLE2036	Suriname	Sipiliwini	2010	SR10-0819-01A	2.1753°	-56.7874°
N. traili	SLE1632	Suriname	Sipiliwini	2012	SR12-0310-02A	2.4770°	-55.6290°
N. traili	SLE1638	Suriname	Sipiliwini	2012	SR12-0727-03A	4.7080°	-56.2193°
N. traili	SLE1633	Suriname	Sipiliwini	2013	SR13-0813-04A	3.7913°	-56.1494°
N. traili	SLE1642	Guyana	Region 6	2014	GY14-0925-01D	4.1540°	-58.1771°
N. traili	SLE1640	Guyana	Region 6	2104	GY14-0928-01A	4.1540°	-58.1771°
N. traili	SLE2061	Suriname	Sipiliwini	2010	SR10-0819-01A	2.1754°	-56.7874°

Table S4.2. Proportions of missing data and gaps in concatenated alignments by sample. Values indicate missing data ('?'), gaps ('-'), and combined missing data and gaps ('?' + '-') as percent of total respective alignment length. Samples are ordered as in tips in Figure 4.2 and Table S4.1.

Taxon	Sample	70% CM-ET '?'	'-'	'?' + '-'	70% CM-GB '?'	'-'	'?' + '-'	90% CM-ET '?'	'-'	'?' + '-'	90% CM-GB '?'	'-'	'?' + '-'
N. josiahi	SLE2072	36.85	18.75	55.59	27.47	12.89	40.37	33.08	19.63	52.71	22.36	13.74	36.10
N. nanulus	SLE1127	45.34	18.20	63.54	26.63	26.52	53.15	42.38	19.12	61.49	22.27	28.11	50.38
N. sp. 6	SLE1649	19.64	22.24	41.88	13.26	9.86	23.12	16.95	23.28	40.23	10.48	9.93	20.41
N. gracilipes	SLE1273	13.55	24.81	38.36	9.58	6.04	15.62	11.12	25.83	36.95	6.99	5.88	12.87
N. gracilipes	SLE2063	74.43	7.39	81.82	31.88	44.19	76.07	73.16	7.94	81.10	27.81	46.99	74.80
N. gracilipes	SLE2017	14.28	24.74	39.02	7.43	11.26	18.69	11.46	26.06	37.52	4.40	11.53	15.93
N. sp. 3	SLE895	8.92	27.86	36.78	5.34	7.61	12.95	7.29	28.69	35.98	3.59	7.40	10.99
N. petrareptans	SLE2055	7.85	27.92	35.78	5.68	4.10	9.78	5.44	29.29	34.73	3.51	3.77	7.28
N. petrareptans	SLE2059	8.36	28.28	36.63	6.54	4.56	11.10	5.84	29.39	35.23	3.91	4.22	8.14
N. c.f.*gracilipes*	SLE1635	7.66	27.86	35.52	5.30	4.18	9.47	6.10	28.70	34.81	3.60	3.88	7.48
N. c.f.*gracilipes*	SLE1959	10.29	27.34	37.63	5.26	9.15	14.40	8.49	28.34	36.83	3.26	9.13	12.39
N. c.f.*gracilipes*	SLE2074	9.85	27.71	37.57	7.57	5.13	12.70	7.95	28.62	36.57	5.45	4.94	10.39
N. c.f.*gracilipes*	SLE2049	5.85	28.73	34.58	3.96	4.13	8.09	4.47	29.55	34.02	2.43	3.85	6.28
N. traili	SLE1885	8.18	28.14	36.32	4.85	6.77	11.61	6.15	29.11	35.26	2.50	6.81	9.30
N. traili	SLE1888	13.23	26.38	39.61	5.24	12.95	18.19	11.89	27.00	38.89	3.30	13.07	16.37
N. traili	SLE1873	10.27	27.43	37.71	4.33	10.47	14.80	8.10	28.31	36.41	1.60	10.58	12.18
N. c.f.*gracilipes*	SLE1879	9.86	27.33	37.19	4.60	9.70	14.30	8.02	28.37	36.39	2.61	9.67	12.28
N. traili	SLE2020	8.23	28.17	36.41	5.20	6.92	12.12	7.00	28.89	35.89	3.90	6.64	10.54
N. traili	SLE2028	7.60	28.14	35.74	4.98	6.13	11.11	5.82	29.01	34.83	3.01	6.00	9.00
N. traili	SLE2034	5.62	28.70	34.31	4.06	3.06	7.12	4.27	29.39	33.66	2.54	2.80	5.35
N. c.f.*gracilipes*	SLE2024	18.54	23.42	41.96	3.79	18.27	22.06	18.09	24.11	42.20	2.77	18.97	21.73
N. traili	SLE1876	9.96	27.33	37.29	4.17	9.96	14.14	7.86	28.56	36.42	2.01	9.99	11.99
N. traili	SLE2014	31.60	18.61	50.20	5.25	28.91	34.17	30.57	19.29	49.86	3.20	29.85	33.05
N. traili	SLE1639	7.03	28.35	35.38	5.16	4.19	9.36	5.02	29.49	34.51	3.29	3.79	7.07
N. traili	SLE1882	7.36	28.42	35.78	4.81	6.11	10.92	5.48	29.38	34.86	2.60	6.13	8.73
N. traili	SLE2078	11.71	27.17	38.88	10.52	2.97	13.49	8.57	28.64	37.22	7.40	2.72	10.12
N. sabrouxi	SLE2053	6.53	28.20	34.73	4.90	3.20	8.10	4.77	29.33	34.10	3.17	3.03	6.20
N. sabrouxi	SLE909	11.86	26.58	38.44	7.93	7.22	15.15	8.59	28.10	36.69	4.69	7.09	11.78

N. traili	SLE907	8.67	27.26	35.93	6.64	2.77	9.41	6.72	28.18	34.89	4.42	2.43	6.85
N. traili	SLE1636	9.33	27.13	36.46	7.16	3.45	10.61	6.20	28.66	34.86	4.19	3.17	7.36
N. traili	SLE1534	5.77	28.78	34.55	4.28	3.18	7.45	4.14	29.75	33.89	2.68	2.87	5.55
N. traili	SLE1537	4.04	29.10	33.13	2.90	2.45	5.36	2.59	30.10	32.69	1.59	2.20	3.79
N. traili	SLE1961	7.75	28.08	35.84	6.78	2.42	9.20	5.10	29.26	34.35	3.88	2.22	6.10
N. traili	SLE1641	5.42	28.44	33.87	3.93	2.48	6.41	3.79	29.45	33.25	2.51	2.12	4.63
N. traili	SLE1870	7.01	28.06	35.06	2.84	7.79	10.64	5.27	29.12	34.39	1.22	7.66	8.89
N. traili	SLE1891	10.95	26.41	37.35	3.82	10.79	14.61	9.99	27.28	37.27	2.78	10.99	13.78
N. traili	SLE2065	51.49	12.65	64.14	14.03	38.96	52.99	50.45	13.17	63.61	11.01	40.67	51.69
N. traili	SLE1631	6.50	28.17	34.67	4.26	4.12	8.38	4.87	29.10	33.97	2.57	3.85	6.42
N. traili	SLE2057	8.76	27.53	36.29	6.32	4.95	11.27	7.24	28.42	35.66	4.55	4.91	9.46
N. traili	SLE1644	6.59	28.06	34.65	4.92	3.26	8.18	4.42	29.50	33.92	3.15	2.99	6.13
N. traili	SLE2036	7.63	28.03	35.66	3.93	7.36	11.29	6.17	28.91	35.08	2.25	7.36	9.60
N. traili	SLE1632	5.53	28.87	34.40	4.56	2.65	7.21	4.02	29.71	33.73	2.86	2.45	5.30
N. traili	SLE1638	5.96	28.87	34.83	5.25	2.63	7.87	3.96	29.83	33.79	3.07	2.38	5.45
N. traili	SLE1633	6.53	28.39	34.92	5.72	2.32	8.04	4.69	29.37	34.06	3.74	2.03	5.76
N. traili	SLE1640	5.95	28.13	34.08	4.40	1.88	6.28	4.52	28.92	33.44	2.73	1.69	4.42
N. traili	SLE1642	10.23	26.62	36.85	3.74	10.55	14.29	8.07	27.72	35.79	1.44	10.47	11.91
N. traili	SLE2061	6.79	28.09	34.88	5.24	2.88	8.12	4.84	29.26	34.10	3.41	2.74	6.15

Table S4.3. Concordance Factor column legend. Tables S4.4–S4.9**.** Concordance factor output column terminology**.** As listed in IQ-Tree concordance factor analysis output. Branch length omitted. NOTE: (gCF+gDF1+gDF2+gDFP) = 100% and (gCF_N+gDF1_N+gDF2_N+gDFP_N) = gN

Column Label	Description
ID	Branch ID
gCF	Gene concordance factor (=gCF_N/gN %)
gCF_N	Number of trees concordant with the branch
gDF1	Gene discordance factor for NNI-1 branch (=gDF1_N/gN %)
gDF1_N	Number of trees concordant with NNI-1 branch
gDF2	Gene discordance factor for NNI-2 branch (=gDF2_N/gN %)
gDF2_N	Number of trees concordant with NNI-2 branch
gDFP	Gene discordance factor due to polyphyly (=gDFP_N/gN %)
gDFP_N	Number of trees decisive but discordant due to polyphyly
gN	Number of trees decisive for the branch
sCF	Site concordance factor averaged over 200 quartets (=sCF_N/sN %)
sCF_N	sCF in absolute number of sites
sDF1	Site discordance factor for alternative quartet 1 (=sDF1_N/sN %)
sDF1_N	sDF1 in absolute number of sites
sDF2	Site discordance factor for alternative quartet 2 (=sDF2_N/sN %)
sDF2_N	sDF2 in absolute number of sites
sN	Number of informative sites averaged over 200 quartets
Label	Recovered support for respective branch. UFboot (IQ-Tree) or local posterior Probability (ASTRAL). UFboot value is given for both Maximum Likelihood tree and consensus Tree, separated by slash (/).

Table S4.4. 70%CM-ET concordance factors. See Table S4.3 for column label and value descriptions. IDs correlate to branch labels in Fig. S4.17.

ID	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label
48	76.08	477	7.18	45	7.02	44	9.73	61	627	71.91	8647.44	13.62	1637.06	14.46	1739.22	12023.72	100/100
49	45.57	442	15.26	148	8.14	79	31.03	301	970	60.23	3779.81	16.84	1052.97	22.93	1359.41	6192.19	100/100
50	53.43	397	7.13	53	10.63	79	28.8	214	743	44.51	825.07	31.37	579.87	24.12	446.19	1851.115	100/100
51	90.32	597	3.48	23	4.69	31	1.51	10	661	96.95	2431.73	1.33	32.64	1.72	42.43	2506.79	100/100
52	33.16	372	7.31	82	12.75	143	46.79	525	1122	57.93	2421.79	22.23	962.94	19.84	880.24	4264.96	100/100
53	38.38	451	1.11	13	1.36	16	59.15	695	1175	72.3	2713.47	14.57	533.82	13.13	483.38	3730.67	100/100
54	5.44	68	1.04	13	4.32	54	89.2	1115	1250	37.68	716.58	27.6	524.21	34.72	666.69	1907.47	100/100
55	2.24	28	0.56	7	3.6	45	93.61	1171	1251	40.95	831.9	23.5	466.51	35.55	718.13	2016.535	76.8/85
56	8.14	96	0.76	9	1.19	14	89.92	1061	1180	36.85	740.75	32.72	671.43	30.43	622.71	2034.88	100/100
57	20.07	234	1.11	13	2.4	28	76.42	891	1166	61.69	1337.42	18.84	400.9	19.46	416.17	2154.48	100/100
58	17.8	218	3.67	45	4.24	52	74.29	910	1225	56.37	998.8	22.39	393.03	21.24	372.74	1764.565	100/100
59	12.69	147	6.48	75	8.64	100	72.19	836	1158	44.69	661.26	28.59	423.19	26.72	396.43	1480.865	100/100
60	15.1	166	4.82	53	8.1	89	71.97	791	1099	45.68	591.99	26.04	335.77	28.28	364.01	1291.755	100/100
61	15.61	170	14.14	154	14.05	153	56.2	612	1089	37.39	482.62	32.9	423.09	29.7	382.79	1288.495	100/100
62	56.82	621	2.65	29	2.2	24	38.33	419	1093	84.94	1620.5	7.31	136.01	7.75	146.19	1902.69	100/100
63	31.48	334	3.86	41	5.37	57	59.28	629	1061	59.52	1040.4	20.12	344.11	20.36	351.79	1736.295	100/100
64	6.78	79	0.6	7	1.8	21	90.82	1059	1166	52.46	1062.77	22.11	434.63	25.42	497.41	1994.805	100/100
65	9.18	107	5.58	65	4.12	48	81.12	945	1165	43.73	772.55	31.93	559.01	24.34	418.29	1749.84	100/100
66	28.18	290	1.46	15	1.46	15	68.9	709	1029	75.74	776.17	15.44	156.33	8.83	90.1	1022.59	100/100
67	11.6	114	5.19	51	2.85	28	80.37	790	983	49.21	248.73	23.68	117.03	27.11	133.76	499.51	76/95
68	2.76	31	2.32	26	4.63	52	90.29	1014	1123	38.09	331.79	31.75	276.61	30.16	262.46	870.865	99.9/72
69	1.28	15	0.17	2	1.45	17	97.1	1138	1172	35.9	287.16	31.38	250.84	32.72	261.63	799.625	99.9/63
70	0.49	6	0.57	7	0.4	5	98.54	1217	1235	33.71	275.51	31.2	253.65	35.09	284.75	813.905	88.8/41
71	3.92	44	0.09	1	0.45	5	95.55	1073	1123	32.72	290.23	32.23	284.7	35.05	307.76	882.685	100/77
72	0.74	9	0.98	12	0.49	6	97.79	1192	1219	34.75	270.99	32.36	249.36	32.89	256.77	777.115	90.2/44
73	4.69	53	1.06	12	1.15	13	93.1	1053	1131	33.65	262.87	33.31	261.9	33.04	259.18	783.95	100/53
74	4.52	51	1.77	20	0.97	11	92.74	1047	1129	35.82	287.25	36.37	287.73	27.8	220.46	795.43	100/72
75	4.73	54	0.7	8	0.44	5	94.13	1075	1142	33.2	260.87	34.2	269.44	32.6	256.68	786.98	99.6/21
76	45.34	540	0.84	10	0.92	11	52.9	630	1191	81.11	1845.67	9.4	206.76	9.49	212.39	2264.815	100/100
77	24.08	269	24.53	274	22.2	248	29.19	326	1117	34.28	388.3	34.99	395.76	30.73	349.46	1133.51	31.8/87
78	9.11	109	1.34	16	2.42	29	87.12	1042	1196	38.28	722.73	26.51	500.94	35.21	667.25	1890.915	99.9/72
79	43.78	479	5.94	65	3.75	41	46.53	509	1094	74.78	1427.89	12.77	235.03	12.44	231.41	1894.325	100/100
80	30.53	377	1.05	13	1.13	14	67.29	831	1235	64.97	1449.15	17.44	377.89	17.59	382.52	2209.555	100/100
81	24.87	289	6.63	77	7.23	84	61.27	712	1162	53.92	956.75	20.71	363.38	25.37	448.96	1769.085	100/100
82	18.86	219	3.62	42	5.08	59	72.44	841	1161	53.05	795.96	24.05	355.08	22.9	340.96	1491.99	100/100
83	11.94	141	1.1	13	1.1	13	85.86	1014	1181	61.01	839.91	18.54	253.15	20.45	280.35	1373.395	100/100
84	1.01	12	1.52	18	1.94	23	95.53	1134	1187	33.19	392.42	33.81	400.44	33.01	392.17	1185.025	100/92
85	1.54	19	1.79	22	0.81	10	95.86	1180	1231	33.59	390.65	34.25	398.06	32.16	373.84	1162.54	11.5/65
86	8.27	93	1.07	12	0.53	6	90.13	1014	1125	39.33	484.64	28.97	356.37	31.7	390.84	1231.84	100/100
87	3.15	37	2.47	29	1.45	17	92.94	1092	1175	34.78	426.58	31.57	390.54	33.65	412.22	1229.33	100/89
88	8.08	88	7.62	83	7.16	78	77.13	840	1089	36.55	480.37	29.11	382.25	34.34	450.6	1313.205	99.9/93
89	8.7	98	0.8	9	0.8	9	89.71	1011	1127	34.36	389.37	34.31	389.09	31.33	354.26	1132.71	100/91
90	37.6	408	3.59	39	4.33	47	54.47	591	1085	68.93	1066.17	14.99	228.06	16.08	243.84	1538.065	100/100
91	84.35	927	0.55	6	0.82	9	14.29	157	1099	97.71	2855.36	1.11	28.83	1.18	30.59	2914.775	100/100

Table S4.5. 70%CM-GB concordance factors. See Table S4.3 for column label and value descriptions. IDs correlate to branch labels in Fig. S4.18.

ID	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label
48	76.59	481	7.32	46	6.85	43	9.24	58	628	70.63	10123.82	14.43	2061.86	14.94	2144.49	14330.17	100/100
49	48.4	470	6.49	63	14.42	140	30.69	298	971	58.24	4589.09	24.31	1768.62	17.44	1378.75	7736.455	100/100
50	30.63	344	7.48	84	13.45	151	48.44	544	1123	57.83	3139.31	21.8	1260.91	20.37	1208.68	5608.89	100/100
51	36.22	426	1.28	15	1.11	13	61.39	722	1176	68.85	3175.06	16.34	761.07	14.81	685.92	4622.035	100/100
52	4.39	55	1.36	17	1.44	18	92.81	1162	1252	38.35	1076.54	31.96	904.9	29.69	813.73	2795.16	99.5/98
53	7.71	91	1.1	13	0.76	9	90.43	1068	1181	38.66	1076.92	31.88	888.53	29.46	827.34	2792.78	100/100
54	18.68	218	1.11	13	2.31	27	77.89	909	1167	60.72	1675.86	18.95	515.1	20.33	556.31	2747.26	100/100
55	16.88	207	3.92	48	3.92	48	75.29	923	1226	53.52	1197.93	24.36	543.84	22.12	493.94	2235.71	100/100
56	12.95	150	6.3	73	7.86	91	72.88	844	1158	44.48	891.29	30.02	601.6	25.5	512.02	2004.905	100/100
57	13.92	153	5.55	61	9.55	105	70.97	780	1099	46.06	773.44	26.99	454.07	26.94	450.53	1678.025	100/100
58	17.06	186	14.5	158	15.69	171	52.75	575	1090	38.35	647.98	31.19	522.38	30.46	510.62	1680.97	100/100
59	52.93	579	2.93	32	2.19	24	41.96	459	1094	81.1	2098.02	10.12	251.48	8.78	222.75	2572.235	100/100
60	31.83	338	3.48	37	4.71	50	59.98	637	1062	58.15	1265	19.96	428.05	21.88	476.18	2169.225	100/100
61	2.72	34	1.6	20	0.56	7	95.13	1191	1252	36.18	874.64	40.04	967.75	23.78	569.36	2411.745	67.5/72
62	5.83	68	0.77	9	1.46	17	91.94	1072	1166	51.92	1075.66	25.56	520.12	22.52	458.98	2054.755	100/100
63	7.72	90	4.97	58	4.46	52	82.85	966	1166	43.55	975.51	32.27	722.24	24.18	535.99	2233.735	100/100
64	26.24	270	0.58	6	1.46	15	71.72	738	1029	73.44	839.59	16.46	189.67	10.1	116.32	1145.58	100/100
65	9.77	96	4.37	43	1.93	19	83.93	825	983	51.94	291.33	21.71	118.11	26.35	144.26	553.7	100/100
66	2.09	25	0.42	5	0	0	97.49	1165	1195	37.2	456.7	31.37	384.31	31.43	388.66	1229.66	100/55
67	0.42	5	0.25	3	0.51	6	98.81	1163	1177	35.75	398.51	32.51	365.29	31.74	357.09	1120.89	100/100
68	0.71	8	0.62	7	4.72	53	93.94	1055	1123	35	431.48	32.12	387.86	32.88	403.09	1222.415	100/99
69	1.78	20	5.26	59	1.34	15	91.62	1028	1122	35.44	405.26	34.01	391.38	30.56	350.64	1147.275	96.6/93
70	6.16	67	5.06	55	5.33	58	83.46	908	1088	34.14	373.37	28.77	310.08	37.09	406.81	1090.25	100/95
71	0.77	9	0.34	4	0.34	4	98.55	1159	1176	33.07	365.28	34.89	384.11	32.04	353.63	1103.02	92.6/55
72	1.71	19	6.41	71	1.62	18	90.25	1000	1108	36.72	383.75	30.92	322.93	32.36	339.81	1046.48	96.1/77
73	5.76	63	5.4	59	4.03	44	84.81	927	1093	35.57	367.69	33	337.4	31.43	320.9	1025.99	54.8/90
74	41.95	500	1.34	16	0.59	7	56.12	669	1192	76.9	2071.63	11.3	304.1	11.8	314.51	2690.235	100/100
75	27.19	304	24.15	270	21.56	241	27.1	303	1118	34.69	502.68	34.69	502.82	30.62	443.24	1448.73	100/100
76	8.28	99	1.67	20	2.01	24	88.04	1053	1196	38.61	802.72	26.51	558.89	34.87	731.07	2092.675	100/100
77	42.05	460	5.21	57	4.94	54	47.81	523	1094	74.12	1487.95	12.97	256.42	12.9	252.31	1996.675	100/100
78	29.13	360	1.13	14	0.57	7	69.17	855	1236	63.57	1898.02	18.27	535.93	18.16	535.24	2969.18	100/100
79	20.74	241	6.2	72	6.02	70	67.04	779	1162	49.82	1183.29	28.9	689.09	21.28	501.81	2374.185	100/100
80	16.37	190	3.19	37	4.48	52	75.97	882	1161	54.74	1108.27	22.94	467.45	22.33	449.32	2025.025	100/100
81	10.82	126	0.94	11	2.75	32	85.49	996	1165	57.06	963.64	22.57	380.74	20.37	344.1	1688.465	100/100
82	2.29	27	0.93	11	1.19	14	95.58	1125	1177	35.25	518.81	33.38	493.59	31.37	460.9	1473.3	100/57
83	0.97	12	1.29	16	1.13	14	96.61	1197	1239	34.09	533.24	32.86	510.29	33.05	515.99	1559.51	100/63
84	9.88	111	2.14	24	2.49	28	85.49	960	1123	37.97	555.4	30.47	449.7	31.57	462.12	1467.22	100/100
85	6.52	74	2.91	33	2.91	33	87.67	995	1135	36.95	579.11	31.79	504.61	31.26	492.75	1576.465	100/79
86	3.3	39	0.93	11	0.85	10	94.92	1121	1181	36.72	590.89	31.06	502.6	32.22	517.73	1611.215	100/57
87	8.91	97	8.82	96	9.37	102	72.91	794	1089	35.39	577.71	27.63	452.93	36.98	606.32	1636.955	71.7/67
88	37.02	402	4.05	44	2.12	23	56.81	617	1086	63.02	1216.26	17.93	341.77	19.05	368.68	1926.7	100/100
89	83.82	922	0.55	6	0.45	5	15.18	167	1100	94.94	3471.58	2.55	90.7	2.51	87.39	3649.665	100/100
90	50.4	375	7.39	55	12.5	93	29.7	221	744	42.89	881.34	33.71	692.64	23.4	480.75	2054.725	100/100
91	88.96	588	5.45	36	4.39	29	1.21	8	661	96.25	2619.43	2.09	56.55	1.66	44.93	2720.905	100/100

Table S4.6. 90%CM-ET concordance factors. See Table S4.3 for column label and value descriptions. IDs correlate to branch labels in Fig. S4.19.

ID	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label
48	78.45	444	6.71	38	6.71	38	8.13	46	566	72.25	7970.26	13.32	1468.08	14.42	1590.81	11029.15	100/100
49	46.83	391	15.33	128	7.07	59	30.78	257	835	60.34	3385.49	16.76	939.71	22.9	1219.91	5545.1	100/100
50	52.74	346	7.01	46	12.5	82	27.74	182	656	45.09	757.91	30.58	512.58	24.33	407.79	1678.265	100/100
51	90.25	537	3.53	21	5.04	30	1.18	7	595	97.06	2192.08	1.33	29.61	1.61	35.48	2257.165	100/100
52	34.36	325	7.08	67	11.73	111	46.83	443	946	58.63	2060.46	22.11	802.72	19.25	729.59	3592.77	100/100
53	39.04	376	1.04	10	1.14	11	58.77	566	963	72.52	2371.01	14.47	460.95	13.01	417.63	3249.58	100/100
54	5.61	56	1.3	13	4.5	45	88.59	885	999	38.46	657.99	27.29	461.99	34.26	589.85	1709.82	100/100
55	2	20	0.8	8	3.3	33	93.89	938	999	41.24	716.81	22.85	386.52	35.91	617.68	1721.015	51.7/68
56	8.7	84	0.93	9	1.14	11	89.22	861	965	37.2	640.23	32.26	572.37	30.53	539.13	1751.72	100/100
57	21.02	202	1.14	11	2.71	26	75.13	722	961	63.04	1242.33	17.88	346.59	19.07	370.74	1959.645	100/100
58	17.36	172	3.23	32	4.04	40	75.38	747	991	55.13	836.26	23	345.67	21.86	329.16	1511.085	100/100
59	13.52	129	6.71	64	7.44	71	72.33	690	954	45.08	593.88	28.97	381.17	25.95	341.18	1316.225	100/100
60	13.82	128	5.4	50	8.64	80	72.14	668	926	45.97	521.87	26.39	298.5	27.64	311.09	1131.45	100/100
61	13.68	126	15.2	140	13.14	121	57.98	534	921	37.9	395.28	32.84	340.21	29.26	303.17	1038.65	99.9/100
62	58.05	537	2.7	25	2.16	20	37.08	343	925	85.6	1499.68	6.33	106.92	8.07	139.19	1745.785	100/100
63	32.52	294	4.2	38	5.2	47	58.08	525	904	61.36	943.76	19.04	289.49	19.6	298.97	1532.22	100/100
64	7.08	68	1.04	10	1.46	14	90.43	869	961	53.1	937.42	22.19	383.88	24.7	426.41	1747.71	100/100
65	10.2	98	6.24	60	4.06	39	79.5	764	961	44.71	691.96	31.61	486.91	23.68	360.89	1539.755	100/100
66	29.7	259	1.26	11	1.03	9	68	593	872	75.45	650.38	15.54	133.14	9.01	77.61	861.125	100/100
67	9.8	83	4.13	35	2.48	21	83.59	708	847	48.59	202.15	24	97.94	27.41	111.63	411.705	98.9/84
68	1.47	14	1.05	10	4.62	44	92.86	885	953	39.08	270.91	29.21	202.69	31.71	218.07	691.66	100/73
69	0.41	4	0.1	1	0.1	1	99.39	973	979	33.05	219.85	33.09	219.71	33.86	225.15	664.705	100/60
70	0.4	4	0	0	0.2	2	99.4	993	999	35.47	236.1	31.59	211.13	32.94	219.65	666.87	94.9/52
71	1.55	15	1.13	11	0.93	9	96.39	935	970	32.21	225.84	31.83	222.56	35.96	250.85	699.24	99.8/62
72	4.37	40	3.5	32	3.93	36	88.2	807	915	40.12	288.3	26.46	189.72	33.41	242.72	720.74	100/68
73	3.8	36	0.32	3	0.53	5	95.35	903	947	34.65	237.15	32.25	219.44	33.1	225.07	681.65	100/65
74	0.63	6	0.73	7	0.1	1	98.53	940	954	34	218.03	35.19	226.13	30.8	196.58	640.73	96.5/48
75	4.35	39	2.57	23	4.46	40	88.62	794	896	36.42	236.16	31.57	204.72	32.01	207.94	648.81	68.2/50
76	45.48	443	0.51	5	0.92	9	53.08	517	974	81.53	1504.35	9.01	158.6	9.47	171.47	1834.42	100/100
77	24.95	233	24.41	228	22.16	207	28.48	266	934	36.65	335.33	31.49	289.57	31.86	296.47	921.365	100/100
78	9.38	91	1.75	17	2.16	21	86.7	841	970	38.49	615.9	26.83	428.96	34.69	556.79	1601.65	100/100
79	44.95	414	5.65	52	4.13	38	45.28	417	921	76.21	1270.58	11.92	195.14	11.87	194.95	1660.66	100/100
80	32.53	324	0.9	9	1	10	65.56	653	996	65.9	1321.74	17	328.99	17.11	334.64	1985.36	100/100
81	24.71	237	7.09	68	7.92	76	60.27	578	959	53.32	797.39	21.27	315.02	25.41	379.71	1492.11	100/100
82	18.52	178	3.33	32	4.68	45	73.47	706	961	53.51	682.67	23.37	294.67	23.12	293.1	1270.43	100/100
83	12.44	124	0.3	3	0.6	6	86.66	864	997	59.84	692.36	19	216.95	21.17	243.96	1153.265	100/100
84	1.1	11	0.9	9	1.5	15	96.49	963	998	34.41	366.08	32.85	351.51	32.74	348.36	1065.945	100/60
85	0.7	7	0.8	8	0.5	5	97.99	975	995	34.4	335.66	32.57	318.49	33.03	321	975.135	67.1/46
86	8.19	78	1.79	17	1.16	11	88.87	846	952	38.4	392.23	28.37	289.77	33.23	341.23	1023.23	100/100
87	5.4	51	2.12	20	1.9	18	90.58	856	945	34.05	336.43	31.98	316.57	33.98	334.72	987.72	100/53
88	6.29	59	0.32	3	0.64	6	92.75	870	938	33.77	366.17	34.34	371.77	31.89	345.58	1083.505	100/55
89	7.72	73	0.63	6	1.59	15	90.06	852	946	35.25	378.89	32.2	345.86	32.55	349.99	1074.725	100/60
90	37.5	342	3.4	31	4.28	39	54.82	500	912	69.57	927.8	14.29	187.67	16.15	212.83	1328.295	100/100
91	85.76	795	0.54	5	0.65	6	13.05	121	927	97.81	2458.47	1.13	24.75	1.06	23.95	2507.16	100/100

Table S4.7. 90%CM-GB concordance factors. See Table S4.3 for column label and value descriptions. IDs correlate to branch labels in Fig. S4.20

ID	gCF	gCF N	gDF1	gDF1 N	gDF2	gDF2 N	gDFP	gDFP N	gN	sCF	sCF N	sDF1	sDF1 N	sDF2	sDF2 N	sN	Label
48	79.51	450	6.36	36	6.71	38	7.42	42	566	71.34	9381.75	13.98	1838.38	14.68	1932.26	13152.38	100/100
49	50.9	425	14.37	120	5.75	48	28.98	242	835	58.28	4167.27	17.29	1235.67	24.43	1611.06	7013.99	100/100
50	51.07	335	7.93	52	12.35	81	28.66	188	656	42.82	791.58	33.65	622.33	23.52	434.81	1848.715	100/100
51	88.91	529	4.37	26	6.22	37	0.5	3	595	96.28	2324.32	1.74	41.8	1.98	47.67	2413.783	100/100
52	31.61	299	8.35	79	13.32	126	46.72	442	946	58.51	2680.14	21.92	1071.18	19.57	984.81	4736.125	100/100
53	36.55	352	0.1	1	0.93	9	62.41	601	963	70.82	2695	13.72	530.61	15.45	600.27	3825.865	100/100
54	1.6	16	4.6	46	1.7	17	92.09	920	999	31.91	733.75	37.71	854.46	30.38	675.69	2263.89	45.2/62
55	8.5	82	1.14	11	2.9	28	87.46	844	965	40.59	925.61	26.95	627.04	32.46	757.33	2309.98	100/100
56	18.94	182	0.83	8	2.19	21	78.04	750	961	61.73	1500.64	18.06	438.85	20.21	489.7	2429.185	100/100
57	16.65	165	3.33	33	3.63	36	76.39	757	991	53.37	996.26	24.34	457.38	22.29	419.04	1872.68	100/100
58	12.79	122	7.44	71	7.55	72	72.22	689	954	45.81	781.15	29.23	497.1	24.96	424.33	1702.58	100/100
59	14.04	130	6.37	59	9.5	88	70.09	649	926	45.42	656.86	27.08	392.28	27.5	397.29	1446.425	100/100
60	16.4	151	14.88	137	14.33	132	54.4	501	921	39.17	536.35	30.84	419.15	29.99	405.56	1361.055	100/100
61	54.81	507	2.81	26	2.92	27	39.46	365	925	82.29	1805.25	8.93	190.74	8.78	189.94	2185.93	100/100
62	32.96	298	3.54	32	4.87	44	58.63	530	904	59.35	1092.73	19.14	348.33	21.51	395.82	1836.875	100/100
63	29.82	297	0	0	0.1	1	70.08	698	996	74.36	2206.19	12.63	358.59	13.01	370.3	2935.075	100/100
64	22.31	214	6.26	60	6.57	63	64.86	622	959	51.05	974.96	21.49	409.2	27.46	528.03	1912.185	100/100
65	16.44	158	3.12	30	4.58	44	75.86	729	961	52.9	871.23	22.75	373.09	24.35	400.9	1645.205	100/100
66	10.93	109	0.1	1	0.5	5	88.47	882	997	58.28	875.45	19.96	297.49	21.76	324.14	1497.07	100/100
67	1.2	12	1.51	15	0.6	6	96.69	963	996	33.99	439.2	33.74	436.5	32.27	418.45	1294.14	100/78
68	0.9	9	2.11	21	0.7	7	96.29	959	996	33.21	403.55	33.98	414.08	32.81	397.75	1215.37	47.8/60
69	8.91	85	1.99	19	1.99	19	87.11	831	954	37.58	476.04	30.24	384.61	32.18	406.03	1266.665	100/95
70	7.46	70	1.81	17	1.49	14	89.23	837	938	38.01	489.13	29.45	382.62	32.54	415.02	1286.76	100/83
71	9.11	86	0.64	6	0.85	8	89.41	844	944	33.08	385.66	33.92	399.73	33.01	386.47	1171.845	100/75
72	7.29	69	1.16	11	1.27	12	90.27	854	946	33.47	446.08	31.57	421.95	34.96	471.22	1339.245	99.9/78
73	37.28	340	2.08	19	4.5	41	56.14	512	912	63.54	1018.57	18.6	299.68	17.86	285.79	1604.04	100/100
74	84.57	784	0.43	4	0.43	4	14.56	135	927	94.88	2871.16	2.76	81.22	2.36	67.11	3019.49	100/100
75	3.1	31	0.7	7	1.3	13	94.89	948	999	37.77	692.6	31.61	586.01	30.62	574.87	1853.47	99.9/91
76	5.83	56	0.73	7	1.46	14	91.99	884	961	51.8	944.19	25.65	459.18	22.54	409.39	1812.76	100/100
77	8.64	83	4.58	44	4.27	41	82.52	793	961	44.74	839.89	30.7	577.99	24.56	457.28	1875.15	100/100
78	25	218	0.92	8	0.92	8	73.17	638	872	73.67	686.99	16.68	157.17	9.65	90.45	934.6	100/100
79	9.21	78	4.49	38	1.89	16	84.42	715	847	50.91	243.46	21.89	100.67	27.2	125.4	469.525	99.5/82
80	1.65	16	0	0	0.1	1	98.25	955	972	36.28	352.67	31.35	301.27	32.37	316.15	970.08	100/30
81	0.31	3	0.31	3	0	0	99.38	956	962	35.23	329.43	31.67	299.96	33.09	311.92	941.31	100/97
82	0.43	4	4.6	43	0.86	8	94.12	880	935	35.12	360.72	32.94	334.87	31.94	318.33	1013.905	100/99
83	2.54	24	0.95	9	5.93	56	90.57	855	944	38.02	375.7	29.28	288.9	32.7	326.29	990.89	100/93
84	5.06	47	5.49	51	5.92	55	83.53	776	929	31.22	270.24	35.56	313.88	33.22	293.12	877.24	29.2/80
85	0.52	5	0.31	3	0.1	1	99.06	947	956	33.89	312.12	35.1	321.44	31.01	283.6	917.155	99.9/30
86	1.3	12	5.09	47	1.19	11	92.42	854	924	37.2	323.77	30.34	262.64	32.46	283.25	869.645	95.7/59
87	5.01	46	5.01	46	3.92	36	86.06	790	918	36.24	306.36	30.78	257.62	32.98	275.7	839.67	95.2/94
88	42.2	411	1.13	11	1.03	10	55.65	542	974	76.96	1678.91	11.76	249.76	11.29	242.99	2171.66	100/100
89	26.87	251	24.09	225	21.09	197	27.94	261	934	35.63	381.79	33.16	358.01	31.22	338.03	1077.825	100/100
90	9.18	89	1.34	13	1.86	18	87.63	850	970	38.71	668.43	27.34	478.88	33.95	585.47	1732.775	100/100
91	42.67	393	4.99	46	5.54	51	46.8	431	921	75.05	1258.76	12.29	201.91	12.66	209.13	1669.79	100/100

Table S4.8. 90%CM-ET-ASTRAL concordance factors. See Table S4.3 for column label and value descriptions. IDs correlate to branch labels in Fig. S4.21.

ID	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label
49	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
50	78.45	444	6.71	38	6.71	38	8.13	46	566	72.24	7945.31	13.33	1463.74	14.44	1587.43	10996.48	1
51	46.83	391	15.33	128	7.07	59	30.78	257	835	60.18	3339.08	16.93	934.26	22.9	1205.79	5479.125	1
52	52.74	346	12.5	82	7.01	46	27.74	182	656	44.95	758.01	24.29	408.51	30.76	517.5	1684.01	1
53	90.25	537	3.53	21	5.04	30	1.18	7	595	97.11	2199.25	1.31	29.25	1.58	35.37	2263.86	1
54	34.36	325	11.73	111	7.08	67	46.83	443	946	58.17	2191.26	19.56	780.29	22.27	860.14	3831.69	1
55	39.04	376	0.21	2	0.1	1	60.64	584	963	72.88	2350.49	14.38	454.19	12.74	409.67	3214.34	1
56	2.1	21	0.9	9	1.2	12	95.8	957	999	41.52	710.79	30.12	519.84	28.36	490.49	1721.11	0.5
57	8.7	84	0.52	5	1.55	15	89.22	861	965	37.57	662.69	29.94	536	32.49	587.13	1785.805	1
58	21.02	202	2.71	26	1.14	11	75.13	722	961	62.45	1208.21	19.49	370.06	18.07	341.6	1919.86	1
59	32.52	294	4.2	38	5.2	47	61.48	525	904	61.48	972.39	19.1	297.79	19.42	304.21	1574.38	1
60	17.36	172	4.04	40	3.23	32	75.38	747	991	56.17	877.11	21.65	335.45	22.18	343.69	1556.24	1
61	58.05	537	2.7	25	2.16	20	37.08	343	925	85.94	1481.41	6.23	104.16	7.83	132.88	1718.435	1
62	13.52	129	7.44	71	6.71	64	72.33	690	954	45.15	575.56	26.01	331.56	28.84	368.91	1276.03	1
63	13.82	128	8.64	80	5.4	50	72.14	668	926	46.01	511.08	27.71	304.5	26.28	290.96	1106.535	1
64	13.68	126	15.2	140	13.14	121	57.98	534	921	37.73	397.87	32.92	344.83	29.35	307.4	1050.095	0.5
65	2.16	21	0.72	7	2.06	20	95.05	922	970	37.79	644.03	23.39	392.35	38.82	657.38	1693.76	0.86
66	7.08	66	4.94	46	1.82	17	86.16	803	932	51.66	827.9	24.59	387.8	23.75	374.01	1589.705	1
67	10.2	98	4.06	39	6.24	60	79.5	764	961	44.67	671.07	23.64	347.2	31.69	471.55	1489.815	1
68	45.48	443	0.51	5	0.92	9	53.08	517	974	80.92	1502.71	9.34	164.89	9.74	175.81	1843.395	1
69	24.95	233	24.41	228	22.16	207	28.48	266	934	36.55	338.32	31.59	293.48	31.86	297.85	929.635	0.46
70	29.7	259	1.03	9	1.26	11	68	593	872	75.59	666.45	8.76	77.17	15.65	138.29	881.915	0.44
71	9.8	83	4.13	35	2.48	21	83.59	708	847	48.1	201.17	24.28	99.33	27.63	113.07	413.56	1
72	1.48	14	3.92	37	2.01	19	92.58	873	943	37.08	272.11	33.23	243.11	29.69	218.91	734.125	0.5
73	0.65	6	3.56	33	0.97	9	94.82	879	927	34.32	246.07	29.85	213.2	35.83	256.38	715.64	0.54
74	0.64	6	5.44	51	0.11	1	93.81	879	937	35.34	230.3	32.75	212.54	31.91	206.85	649.685	0.48
75	0.41	4	3.73	36	0.62	6	95.24	920	966	35.21	235.02	31.83	210.76	32.96	220.33	666.1	0.48
76	0.31	3	4.93	47	0.42	4	94.33	899	953	33.63	220.4	31.03	202.72	35.34	228.32	651.44	0.56
77	0.83	8	1.65	16	0.83	8	96.69	935	967	33.63	225.36	33.31	218.91	32.44	215.72	659.98	0.64
78	5.33	50	1.39	13	1.28	12	92	863	938	34.25	208.19	35.76	228.8	28.36	206.22	643.205	0.38
79	5.34	50	2.03	19	0.96	9	85.29	858	936	33.61	224.01	32.64	216.82	33.75	222.51	663.335	0.73
80	6.51	65	5.61	56	2.6	26	91.67	852	999	35.4	597.4	36.63	611.33	27.97	465.79	1674.51	0.93
81	45.15	428	2.22	21	1.79	17	50.84	482	948	76.53	1357.47	11.7	207.1	11.77	206.65	1771.22	1
82	32.53	324	1.81	18	1.51	15	64.16	639	996	74.3	1561.34	11.84	234.77	13.86	274.43	2070.525	1
83	85.76	795	0.54	5	0.65	6	13.05	121	927	98.01	2489.7	1.02	23.93	0.97	23.34	2556.96	1
84	24.71	237	7.09	68	7.92	76	60.27	578	959	53.8	796.47	20.8	305.73	25.39	376.97	1479.16	1
85	18.52	178	3.33	32	4.68	45	73.47	706	961	53.8	683.52	23.47	294.01	22.73	286.77	1264.395	1
86	12.38	121	1.54	15	1.13	11	84.95	830	977	59.92	675.31	18.36	205.39	21.72	245.01	1125.71	0.4
87	1.69	16	7.82	74	2.01	19	88.48	837	946	33.67	335.89	32.44	321.76	33.89	338.54	996.185	0.43
88	1.8	17	8.9	84	0.64	6	88.67	837	944	34.37	337.38	34.37	336.25	34.37	306.73	980.355	0.52
89	1.17	11	6.29	59	0.96	9	91.58	859	938	33.72	379.27	33.28	374.06	33.28	368.18	1121.505	0.78
90	1.17	11	6.5	61	1.49	14	90.84	853	939	34.06	358.57	33.22	353.08	32.72	348.3	1059.945	1
91	2.97	28	7.22	68	1.91	18	87.9	828	942	35.76	380.32	29.36	312.9	34.88	369.09	1062.295	0.57
92	6.05	56	7.56	70	7.13	66	79.27	734	926	30.37	342.34	32.7	371.06	36.93	417.48	1130.875	0.67
93	37.5	342	3.4	31	4.28	39	54.82	500	912	69.62	929.42	14.23	186.74	16.15	212.97	1329.125	1

Table S4.9. 90%CM-GB-ASTRAL concordance factors. See Table S4.3 for column label and value descriptions. IDs correlate to branch labels in Fig. S4.22.

ID	gCF	gCF_N	gDF1	gDF1_N	gDF2	gDF2_N	gDFP	gDFP_N	gN	sCF	sCF_N	sDF1	sDF1_N	sDF2	sDF2_N	sN	Label
49	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
50	79.51	450	6.36	36	6.71	38	7.42	42	566	71.28	9347.52	14.04	1839.04	14.68	1926.24	13112.79	1
51	50.9	425	14.37	120	5.75	48	28.98	242	835	58.03	3926.34	17.4	1180.02	24.57	1528.39	6634.74	1
52	51.07	335	12.35	81	7.93	52	28.66	188	656	42.29	783.9	23.62	437.74	34.08	631.91	1853.54	1
53	88.91	529	4.37	26	6.22	37	0.5	3	595	96.25	2314.84	1.75	41.74	2.01	47.89	2404.475	1
54	31.61	299	13.32	126	8.35	79	46.72	442	946	58.36	2689.28	19.65	990.71	21.99	1074.54	4754.525	1
55	36.55	352	0.21	2	0.31	3	62.93	606	963	70.36	2825.38	15.79	635.82	13.85	561.57	4022.76	1
56	2	20	0.5	5	1.7	17	95.8	957	999	38.7	872.98	31.19	715.45	30.11	667.83	2256.25	1
57	8.5	82	0.41	4	0.41	4	90.67	875	965	39.24	902.51	29.26	678.04	29.23	411.27	2317.245	0.89
58	18.94	182	2.19	21	0.83	8	78.04	750	961	61.38	1532.63	20.49	509.06	18.13	449.88	2491.56	0.98
59	32.96	298	3.54	32	4.87	44	58.63	530	904	58.59	1054.58	19.73	346.88	21.68	389.27	1790.72	1
60	16.65	165	3.63	36	3.33	33	76.39	757	991	53.17	1022.2	22.33	430.77	24.51	473.43	1926.395	1
61	54.81	507	2.81	26	2.92	27	39.46	365	925	82.22	1887.32	9.08	202.41	8.7	195.68	2285.41	1
62	12.79	122	7.55	72	7.44	71	72.22	689	954	45.63	774.91	25.06	424.51	29.31	496.29	1695.71	1
63	14.04	130	9.5	88	6.37	59	70.09	649	926	45.37	634.37	27.54	384.06	27.09	380.86	1399.28	1
64	16.4	151	14.88	137	14.33	132	54.4	501	921	39.22	543.54	30.51	418.9	30.27	411.27	1373.705	1
65	1.86	18	0.93	9	1.65	16	95.57	927	970	38.13	756.68	22.88	455.08	38.99	778.85	1990.6	0.98
66	5.9	55	4.83	45	1.61	15	87.66	817	932	49.67	886.5	24.23	425.09	26.1	454.18	1765.76	1
67	8.64	83	4.27	41	4.58	44	82.52	793	961	44.96	843.38	24.25	454.7	30.79	577.22	1875.29	1
68	42.2	411	1.13	11	1.03	10	55.65	542	974	77.06	1682.32	11.65	249.72	11.29	245.64	2177.68	1
69	26.87	251	24.09	225	21.09	197	27.94	261	934	35.71	388.74	32.96	360.72	31.33	343.39	1092.835	0.97
70	25	218	0.92	8	0.92	8	73.17	638	872	74.52	690.3	9.6	89.54	15.87	147.48	927.31	1
71	9.21	78	4.49	38	1.89	16	84.42	715	847	51.31	244.28	21.96	101.5	26.73	123.48	469.255	1
72	1.59	15	3.71	35	0.85	8	93.85	885	943	37.49	399.78	32.12	339.39	30.39	325.4	1064.56	0.78
73	0.72	7	1.03	10	0.41	4	97.83	946	967	32.84	320.15	33.95	325.46	33.2	319.67	965.275	1
74	4.67	44	0.11	1	0.21	2	95.02	896	943	30.3	269.99	34.13	307.87	35.56	319.05	896.905	0.37
75	0.3	3	0.7	7	0.1	1	98.9	986	997	35.78	320.26	31.37	279.92	32.84	293.12	893.29	0.54
76	5.06	48	0.53	5	0.42	4	93.99	891	948	36.26	316.42	33.45	293.71	30.28	260.46	870.58	0.58
77	0.83	8	1.24	12	0.62	6	97.31	941	967	33	295.83	34.52	311.44	32.47	295.58	902.855	0.45
78	0.96	9	4.82	45	1.82	17	92.39	862	933	34.16	329.65	32.96	315.94	32.88	318.08	963.66	0.56
79	5.93	54	5.71	52	4.62	42	83.74	762	910	31.46	300.76	34.27	326.31	34.27	329.75	956.82	0.51
80	5.71	57	4.6	46	2.2	22	87.49	874	999	33.9	696.56	35.56	721.96	30.54	609.67	2028.18	0.73
81	42.62	404	1.69	16	1.79	17	53.9	511	948	75.66	1452.11	12.41	235.27	11.92	228.4	1915.77	1
82	29.82	297	1.31	13	0.9	9	67.97	677	996	75.4	1864.79	11.51	278.22	13.09	319.99	2462.995	1
83	84.57	784	0.43	4	0.43	4	14.56	135	927	95.02	3027.57	2.67	82.58	2.31	71.02	3181.165	1
84	22.31	214	6.26	60	6.57	63	64.86	622	959	51.27	969.87	21.56	406.15	27.17	519.83	1895.835	1
85	16.44	158	3.12	30	4.58	44	75.86	729	961	53.72	880.87	22.48	367.89	23.8	389.99	1638.75	1
86	10.92	109	0.3	3	0.4	4	88.38	882	998	58.38	859.31	20.96	304.89	20.66	304.17	1468.365	1
87	1.02	10	0.82	8	2.05	20	96.11	938	976	36.41	464.22	34.1	431.04	29.48	379.51	1274.755	0.73
88	1.81	17	7.46	70	2.56	24	88.17	827	938	29.07	405.94	37.29	517.7	33.64	466.09	1389.72	0.51
89	7.14	65	9.45	86	6.59	60	76.81	699	910	33.92	463.45	32.95	451.77	33.12	459.69	1374.905	0.5
90	2.05	20	0.92	9	2.15	21	94.87	925	975	33.66	429.35	33.96	431.54	32.37	410.99	1271.875	0.52
91	2.55	24	9.12	86	2.44	23	85.9	810	943	33.66	393.82	32.87	384.48	33.46	393.43	1171.725	0.49
92	8.89	82	7.7	71	7.16	66	76.25	703	922	36.91	462.86	28.89	363.79	34.2	432.04	1258.675	0.71
93	37.28	340	4.5	41	2.08	19	56.14	512	912	63.94	1023.96	17.62	281.99	18.44	296.46	1602.415	1