Syst. Biol. XX(XX):XX–XX, 2025

https://doi.org/10.1093/sysbio/syaf001
Advance Access Publication January 8, 2025

Evaluating UCE Data Adequacy and Integrating Uncertainty in a Comprehensive Phylogeny of Ants

Marek L. Borowiec[1,*], Y. Miles Zhang[2,6], Karen Neves[3], Manuela O. Ramalho[4], Brian L. Fisher[5], Andrea Lucky[6], and Corrie S. Moreau[7,8,*]

[1]*Department of Agricultural Biology and C. P. Gillette Museum, Colorado State University, 1177H Campus Delivery, Fort Collins, CO 80523, USA*
[2]*Institute of Ecology and Evolution, Ashworth Laboratories, University of Edinburgh, King's Buildings, Charlotte Auerbach Rd, Edinburgh, EH9 3FL, UK*
[3]*Instituto de Biologia, Universidade Federal de Uberlândia, Avenida Pará 1720, Campus Umuarama, CEP 38 405-320, Uberlândia, Minas Gerais, Brazil*
[4]*Department of Biology, West Chester University, 750 South Church St., West Chester, PA 19383, USA*
[5]*California Academy of Sciences, 55 Music Concourse Drive, San Francisco, CA 94118, USA*
[6]*Entomology and Nematology Department, University of Florida, 1881 Natural Area Dr., Gainesville, FL 32611, USA*
[7]*Department of Entomology, Cornell University, 129 Garden Ave., Ithaca, NY 14853, USA*
[8]*Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Rd., Ithaca, NY 14853, USA*
[*]*Correspondence to be sent to: E-mail: marek.borowiec@colostate.edu; E-mail: corrie.moreau@cornell.edu.*

Received 8 July 2024; reviews returned 11 December 2024; accepted 6 January 2025
Associate Editor: Brant Faircloth

Abstract.—While some relationships in phylogenomic studies have remained stable since the Sanger sequencing era, many challenging nodes remain, even with genome-scale data. Incongruence or lack of resolution in the phylogenomic era is frequently attributed to inadequate data modeling and analytical issues that lead to systematic biases. However, few studies investigate the potential for random error or establish expectations for the level of resolution achievable with a given empirical data set and integrate uncertainties across methods when faced with conflicting results. Ants are the most species-rich lineage of social insects and one of the most ecologically important terrestrial animals. Consequently, ants have garnered significant research attention, including their systematics. Despite this, there has been no comprehensive genus-level phylogeny of the ants inferred using genomic data that thoroughly evaluates both signal strength and incongruence. In this study, we provide insight into and quantify uncertainty across the ant tree of life by utilizing the most taxonomically comprehensive ultraconserved elements data set of ants to date, including 277 (81%) of recognized ant genera from all 16 extant subfamilies, and representing over 98% of described species. We use simulations to establish expectations for resolution, identify branches with less-than-expected concordance, and dissect the effects of data and model selection on recalcitrant nodes. Simulations show that hundreds of loci are needed to resolve recalcitrant nodes on our genus-level ant phylogeny. This demonstrates the continued role of random error in phylogenomic studies. Our analyses provide a comprehensive picture of support and incongruence across the ant phylogeny, while offering a more nuanced depiction of uncertainty and significantly expanding generic sampling. We use a consensus approach to integrate uncertainty across different analyses and find that assumptions about root age exert substantial influence on divergence dating. Our results suggest that advancing the understanding of ant phylogeny will require not only more data but also more refined phylogenetic models. We also provide a workflow for identifying under-supported nodes in concatenation analyses, outline a pragmatic way to reconcile conflicting results in phylogenomics, and introduce a user-friendly locus selection tool for divergence dating. [keyword: Formicidae; incongruence; phylogenomics; random error; recalcitrant nodes; systematic bias; ultraconserved elements.]

Sanger sequencing revolutionized our understanding of evolutionary relationships among organisms. Early molecular phylogenetic studies of ants (order Hymenoptera, family Formicidae), the world's most species-rich lineage of social organisms, confirmed some phylogenetic predictions of morphology-based systematics, put forth a slew of new hypotheses, and identified areas of uncertainty (Ward 2014; Borowiec et al. 2020). Since then, studies featuring increased taxon and locus sampling, including genome-wide data, confirmed many of the relationships uncovered by these early molecular studies, but also revealed many recalcitrant nodes (Blaimer et al. 2015, 2018; Branstetter et al. 2017c; Borowiec 2019b; Romiguier et al. 2022). These include relationships for which the topology is inconsistent across phylogenetic methods, either accompanied by low nodal support or showing highly supported but conflicting results. This discordance is a familiar pattern across the tree of life, with many examples from shallow (e.g., Pease et al. 2016; Edelman et al. 2019; Zhao et al. 2023) and deep relationships (e.g., McCormack et al. 2012; Li et al. 2021a; Morales-Briones et al. 2021).

These phylogenetic uncertainties are not surprising given well-documented biological processes that violate common assumptions of phylogenetic inference methods on bifurcating phylogenies, including incomplete lineage sorting (Maddison 1997; Edwards 2009a), reticulate evolution (McDade 1990), convergence (Edwards

2009b; Li et al. 2010; Zou and Zhang 2020), and recombination (Schierup and Hein 2000). This is compounded by other potential sources of error and bias that can propagate across complicated phylogenomic workflows (Guang et al. 2016; Philippe et al. 2017). Increased awareness of these issues spurred development of tools to better account for, describe, and visualize incongruence in phylogenomic data (Salichos and Rokas 2013; Smith et al. 2015; Arcila et al. 2017; Shen et al. 2017; Sayyari et al. 2018; Bravo et al. 2019; Minh et al. 2020a; Steenwyk et al. 2023; Lanfear and Hahn 2024). Although large phylogenomic data sets are susceptible to systematic bias, few studies explore the potential of random error to impact the result or use simulations to establish an expectation of what resolution is feasible with the data at hand (but see e.g., Giarla and Esselstyn 2015; Cloutier et al. 2019). Furthermore, different approaches, such as concatenation versus coalescent methods, perform better in different contexts (e.g., Bryant and Hahn 2020), and these contexts are expected to vary within large phylogenies of ancient clades. Despite these limitations, there is currently no universally accepted framework for integrating uncertainties across different phylogenetic inference methods.

Recent years brought considerable progress in resolving phylogenetic relationships of ants, but there has been no comprehensive quantification of support and incongruence across the ant phylogeny using genomic data, and no studies asking what level of resolution should be expected for the ants. Most studies thus far have focused on individual subfamilies or on resolving select, contentious relationships. The considerable significance of the study system, increasing accessibility of genomic resources, and recently refined generic taxonomy merit a new, global look at the ant phylogeny.

In this study, we use a newly generated, comprehensive genus-level data set to answer 3 questions about phylogenomic inference of the ant tree of life: 1) How much sequence data are needed to reconstruct a robust phylogeny? 2) What are the impacts of data type and model selection at recalcitrant nodes? and 3) How can uncertainty be integrated across inference methods?

Current State of the Ant Phylogeny

With over 14,000 described species and counting (Bolton 2024), ants are one of the most ecologically important animal lineages (Risch and Carroll 1982; Folgarait 1998; Parker and Kronauer 2021). Despite accounting for only 1.5% of insect species, they are estimated to comprise 15–20% of all terrestrial animal biomass, exceeding 25% in tropical ecosystems where they are most abundant (Hölldobler and Wilson 1990; McGlynn 1999; Tuma et al. 2020). Because of their diversity, ubiquity, and sociality, ants are well-studied in comparison to other arthropod clades of comparable size (Andersen and Majer 2004; Hoffmann 2010; Lach 2010). This is reflected by the abundance of studies attempting to resolve the global ant phylogeny (Fig. 1). First attempts at resolving higher-level ant relationships used exclusively morphological data, including early studies preceding statistical phylogenetics as well as later ones using parsimony (Brown 1954; Wilson et al. 1967; Wilson 1971; Taylor 1978; Dlussky and Fedoseeva 1988; Baroni Urbani 1989; Hölldobler and Wilson 1990; Baroni Urbani et al. 1992; Ward 1994; Grimaldi et al. 1997). The late 1990s and early 2000s saw the first efforts to reconstruct the broad-scale ant phylogeny using molecular data (Sullender and Johnson 1998; Ohnishi et al. 2003; Ward and Brady 2003; Saux et al. 2004; Ouellette et al. 2006). Many of these early studies relied on ribosomal genes, occasionally combined with morphology (but see Astruc et al. 2004; Ward and Downie 2005). In 2006, 2 landmark studies included multiple nuclear gene fragments from representatives of all subfamilies and 50–60% of genera recognized at the time (Brady et al. 2006; Moreau et al. 2006). These trees laid a foundation for subsequent comprehensive phylogenies within subfamilies (Schultz and Brady 2008; Ward et al. 2010; Schmidt 2013; Brady et al. 2014; Chomicki et al. 2015; Ward et al. 2015; Ward and Fisher 2016; Borowiec et al. 2019). Moreau and Bell (2013) and Blanchard and Moreau (2017) combined previously published Sanger data for phylogenies representing up to 271 ant genera. Nelsen et al. (2018) also consolidated previously published sequences to infer the largest molecular phylogeny of the ants to date based on molecular data for over 1,400 species. Economo et al. (2018) combined published molecular data for 687 species with phylogenetic grafting of all remaining species based on taxonomy to obtain a species-level tree of ants.

These phylogenetic studies overlapped with the introduction of genomic data, first used to investigate generic relationships within subfamilies (Blaimer et al. 2015, 2018; Branstetter et al. 2017a, 2022; Borowiec 2019b; Griebenow 2020; Camacho et al. 2022b), and more recently broadened to relationships among subfamilies (Branstetter et al. 2017c; Romiguier et al. 2022). Branstetter et al. (2017c) presented a phylogeny based on ultraconserved elements (UCEs) (Faircloth et al. 2012; McCormack et al. 2012) for 15 out of 16 subfamilies and 80 ant genera, as well as a tree based on 12 nuclear loci that included 299 ant genera valid at the time. Most recently, Romiguier et al. (2022) generated genomic information from over 4,000 protein-coding loci and inferred a tree for representatives of all subfamilies and 63 ant genera. Collectively, these studies led to a dramatic refinement of generic boundaries across ants, resulting in as few as ca. 5% of the 342 currently recognized genera remaining nonmonophyletic (e.g., Bolton and Fisher 2014; Schmidt and Shattuck 2014; Ward et al. 2015, 2016; Borowiec 2016b; Ward and Fisher 2016; Camacho et al. 2022b; Griebenow 2024).

The current understanding of ant phylogeny (Fig. 1) recognizes that all 16 extant ant subfamilies are reciprocally monophyletic and grouped into 3 clades: leptanilloids (= leptanillomorphs sensu

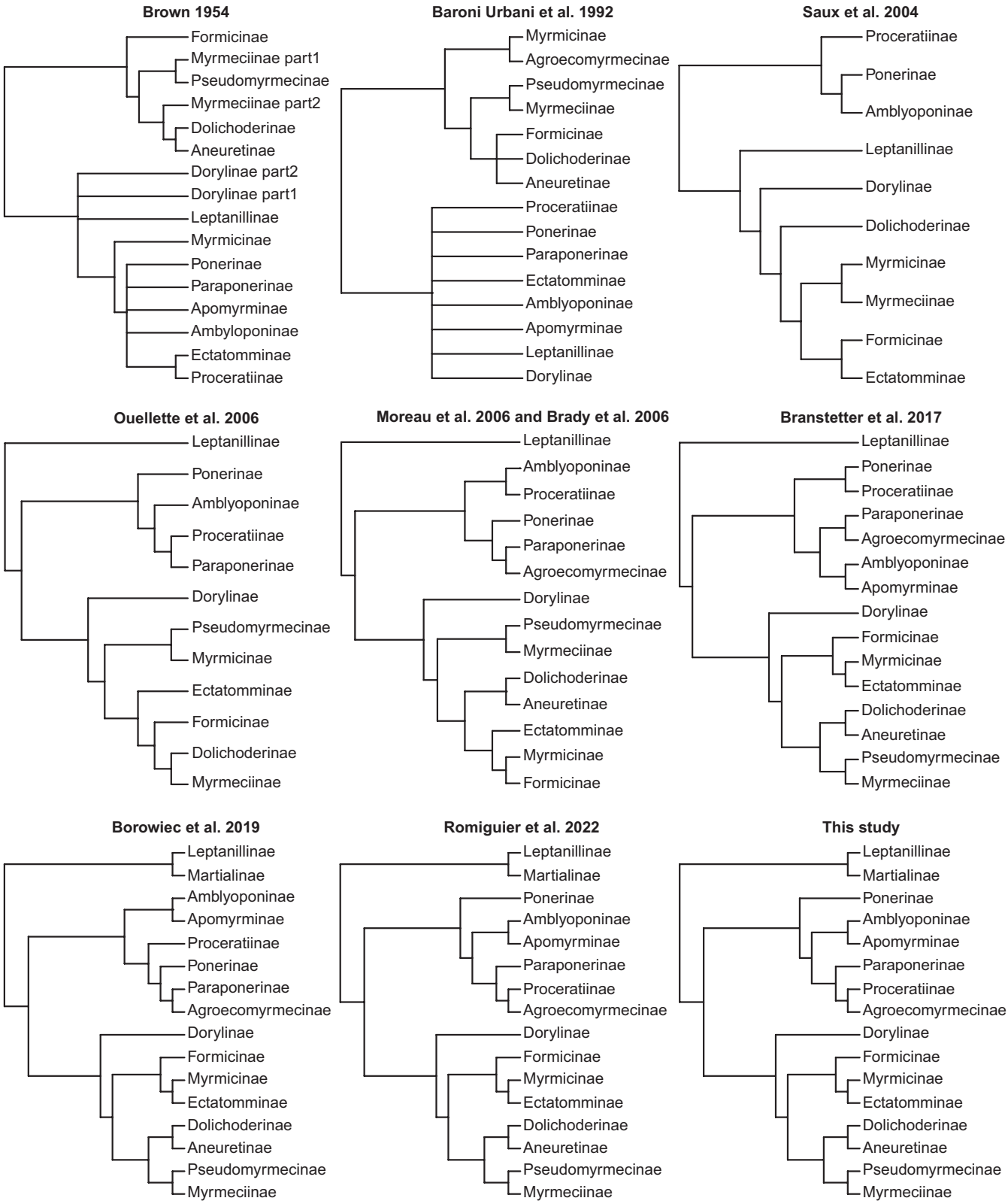

FIGURE 1. History of phylogenetic relationships among ant subfamilies from select studies, adapted to match current subfamily classification. Consensus subfamily-level relationships recovered in this study are identical to those of Romiguier et al. 2022.

Boudinot et al. (2022)), poneroids, and formicoids (Moreau et al. 2006; Borowiec et al. 2020). The leptanilloids are a small group of approximately 80 species of subterranean ants belonging to 2 subfamilies, Leptanillinae and Martialinae. The position of the monotypic Martialinae has been a subject of great interest and considerable disagreement (Rabeling et al. 2008; Kück et al. 2011; Moreau and Bell 2013; Ward and Fisher 2016; Branstetter et al. 2017c; Borowiec et al. 2019; Cai 2024). Upon its initial discovery, *Martialis heureka*, the sole member of the subfamily, was recovered as the sister lineage to all other extant ants (Rabeling et al. 2008). This finding ignited some controversy, and a data reanalysis suggested that Martialinae may, in fact, be sister to the clade containing poneroids and formicoids, to the exclusion of Leptanillinae (Kück et al. 2011). Recent analyses using Sanger and genomic data present increasing evidence that Martialinae is, in fact, sister to Leptanillinae (Borowiec et al. 2019; Boudinot et al. 2022; Romiguier et al. 2022), although one re-analysis of published Sanger and protein coding genomic data continued to argue that *Martialis* is sister to ants minus Leptanillinae (Cai 2024). The poneroids contain about 1,500 species classified in 6 small to medium-sized subfamilies, namely, Agroecomyrmecinae, Amblyoponinae, Apomyrminae, Paraponerinae, Ponerinae, and Proceratiinae. The relationships among poneroid subfamilies are contentious and genomic studies thus far have shown poor support and conflicting topologies (Branstetter et al. 2017c; Romiguier et al. 2022). Finally, the most diverse clade, the formicoids, are a grouping of the remaining 8 subfamilies Aneuretinae, Dolichoderinae, Dorylinae, Ectatomminae, Formicinae, Myrmeciinae, Myrmicinae, and Pseudomyrmecinae (Brady et al. 2006; Moreau et al. 2006). The formicoids contain most ant species, or about 88% of described diversity (Borowiec et al. 2020). The relationships among formicoid subfamilies are well resolved and uncontroversial (Ward 2014; Borowiec et al. 2020); however, many relationships among formicoid genera remain contentious or unresolved.

Study Objectives

Evolutionary relationships eluding resolution despite application of genomic data are not unique to ants. Large-scale phylogenomic studies across the tree of life, including plants (e.g., Drew et al. 2014), animal phyla (e.g., Li et al. 2021a), and other clades of insects (e.g., Tihelka et al. 2021) offer many examples of unresolved or controversial relationships. This is not surprising given analytical limitations of phylogenetic methods, the deep timescales under study, and the complexity of the evolutionary process. Furthermore, studies showing pervasive topological incongruence across whole genomes show that a bifurcating tree may not always be the best depiction of evolutionary relationships (Pease et al. 2016; Suh 2016; Edelman et al. 2019).

These analytical limitations are likely at play across the backbone ant phylogeny, but we currently lack understanding of how much difficulty one should expect in inferring the ant tree of life. To examine relationships at subfamily and generic levels in ants, we assembled a new data set including newly sequenced UCEs for representatives of 176 ant genera. We combined these with publicly available sequences for a total of 280 ant taxa in 277 genera, representing 81% of the 342 ant genera valid at the time of this writing (Bolton 2024). These genera represent 98% of described ant species, offering unprecedented coverage of ant diversity with genomic data.

To establish a baseline expectation for the level of difficulty in resolving ant relationships, we simulated alignment data resembling our empirical sequences. We explored the influence of analysis type, model, and locus selection on inference using empirical data. To highlight areas on the tree potentially affected by systematic bias, we compared simulated and empirical results.

No individual phylogenetic analysis can address all sources of potential bias and negative impact, and different approaches will be advantageous in different situations (e.g., Bryant and Hahn 2020). In recognition of this observation, we integrated uncertainty across and within methods by constructing a consensus topology from bootstrap trees generated in multiple analyses. Consensus tree methods are universally used to summarize support within an analysis, uncertainty in Bayesian phylogenetics, or information in gene trees (Holder et al. 2008; Degnan et al. 2009) but are rarely used to express uncertainty across methods in phylogenomics (but see Oliveros et al. 2019 for a similar topology reconciliation approach). We used this consensus tree to produce a time-calibrated phylogeny to serve as a new resource for researchers studying ants in a comparative evolutionary framework.

Materials and Methods

For the overview of data processing and analysis see workflow visualization in Fig. 2.

Data Availability

Assembled contigs, alignments, trees, analysis configuration files and logs, and custom scripts can be found on Zenodo at https://doi.org/10.5281/zenodo.14338504 and Dryad at https://doi.org/10.5061/dryad.547d7wmhb. Trimmed reads generated for this study have been deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA1195116 (http://www.ncbi.nlm.nih.gov/bioproject/1195116).

Taxon Sampling

Our taxon sampling consists of 280 ant taxa (277 genera) and 12 outgroups in 10 families of aculeate

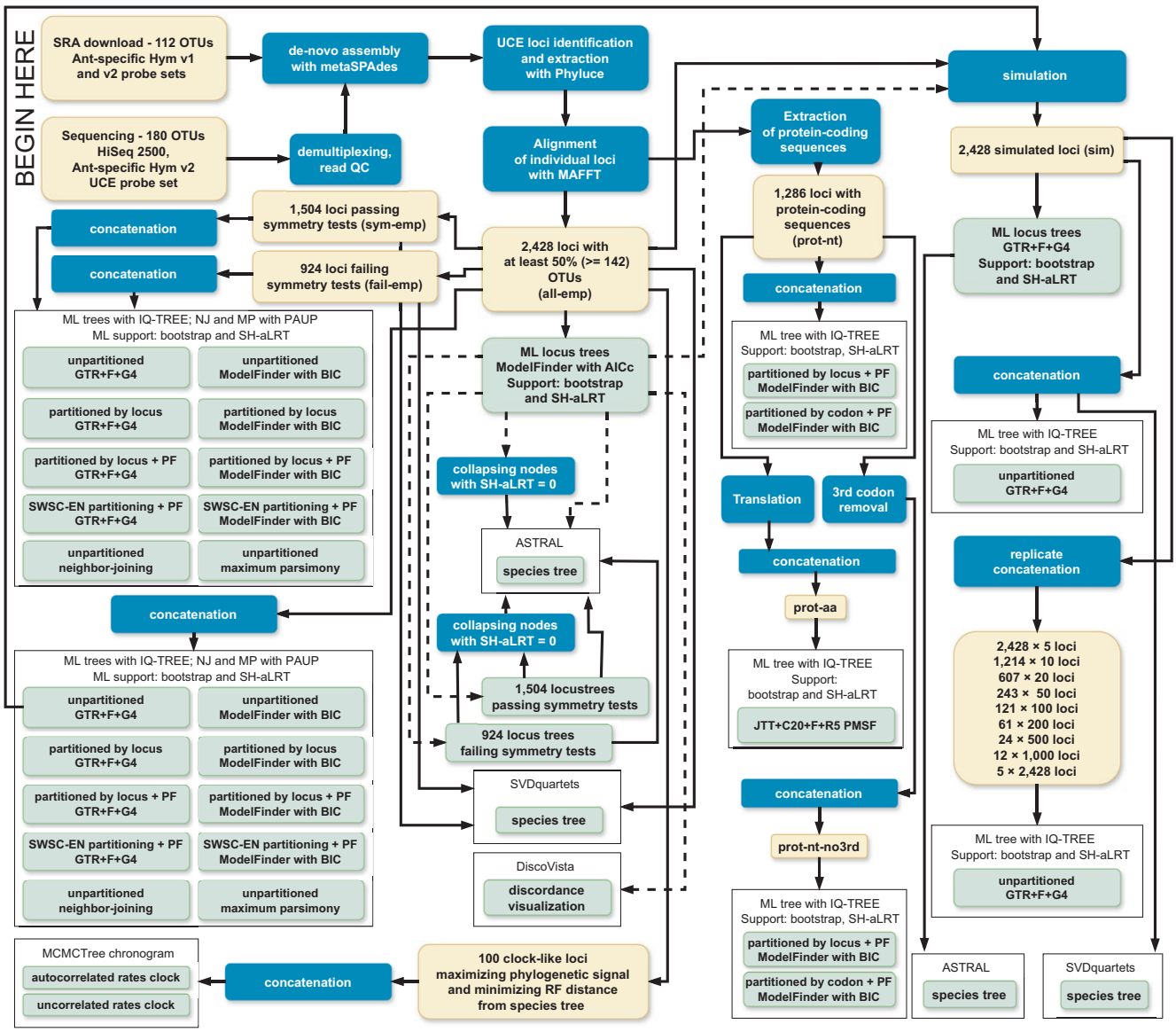

FIGURE 2. Visual overview of analysis workflow including data sets, data transformations, and analyses. BIC = Bayesian information criterion; GTR = general time reversible; OTU = operational taxonomic unit; PF = Partition Finder; QC = quality control; SH-aLRT = approximate likelihood-ratio test; SRA = Sequence Read Archive; SWSC-EN = entropy-based sliding-window site characteristics. See Table 2 and Methods text for data set name abbreviations. "ModelFinder" analyses used the best inferred model, other models were set a priori. The many lines originating at the full set of gene trees are dashed for clarity. For better readability, please refer to the PDF version of this figure, as the printed version may have font size limitations.

Hymenoptera. We focused on obtaining a single representative of as many ant genera as possible. Further refinement of generic boundaries was not the focus of this study, and we did not attempt to include multiple lineages of the several known nonmonophyletic genera such as *Monomorium* (Ward et al. 2015) or *Stigmatomma* (Ward and Fisher 2016).

We generated new Ultraconserved Element data for representatives of 176 ant genera and 4 outgroups. The taxonomy, accessions, vouchers and references for all newly sequenced samples are in Supplementary Table S1.

We identified 104 ant genera missing from our study and 8 additional outgroups for which Ultraconserved Element or whole genome data were publicly available, and downloaded their sequence reads from NCBI Sequence Read Archive. Since the inception of this project, 3 of the 104 genera have become synonyms: *Anomalomyrma* is now included in *Protanilla*, and *Noonilla* and *Yavnella* are both synonyms of *Leptanilla* (Griebenow 2024). Downloaded sequences included 111 SRA read samples from previous ant phylogenies based on UCEs, including sequences from nine previous studies (Table 1), using both the original Hymenoptera UCE

Table 1. Studies from which published UCE sequences were downloaded to supplement newly generated data

Study	Study DOI	SRA BioProject number	Bait set
Blaimer et al. 2015	10.1186/s12862-015-0552-5	PRJNA293213	v1
Branstetter et al. 2017a	10.1098/rspb.2017.0095	PRJNA379607	v1
Branstetter et al. 2017b	10.1016/j.cub.2017.03.027	PRJNA379583	v1
Ješovnik et al. 2017	10.1111/syen.12228	PRJNA354064	v1
Branstetter et al. 2017c	10.1111/2041-210X.12742	PRJNA360290	v2
Blaimer et al. 2018	10.1093/isd/ixy013	PRJNA473845	v2
Borowiec 2019b	10.1093/sysbio/syy088	PRJNA504894	v2
Griebenow 2020	10.25849/myrmecol.news_030:229	PRJNA629360	v2
Longino and Branstetter (2020)	10.1093/isd/ixaa004	PRJNA563172	v2

bait set Hymenoptera 1.5Kv1 (Faircloth et al. 2015), and the updated ant-specific probes Hymenoptera 2.5Kv2A (Branstetter et al. 2017c). The full list of downloaded sequences and their corresponding studies is available in Supplementary Table S2. In addition to downloading previously published UCE reads, we also harvested UCEs from the published genome of *Linepithema humile* (Smith et al. 2011).

Molecular Data Collection and Sequencing

We extracted DNA from all newly sequenced specimens at the Cornell Arthropod Biosystematics and Biodiversity lab. The samples were subjected to destructive DNA extraction following the recommendations of the Qiagen DNEasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA) and Moreau (2014) to avoid contamination. Each extracted specimen came from a collecting event containing multiple conspecifics, and one remaining specimen was designated as the voucher for each extraction and deposited in a museum collection (Supplementary Table S1). We followed the protocol described in more detail by Branstetter et al. (2017c) targeting 2,524 ultraconserved element (UCE) loci. The DNA was quantified using a Qubit fluorometer (HS Assay Kit, Life Technologies Inc., Carlsbad, CA) and sheared through Covaris M220 (Covaris, Woburn, MA) and Qsonica Q800R sonicators, targeting 400–600 bp average fragment size. Sheared DNA was subjected to the modified genomic DNA library protocol using the Kapa Hyper Prep Library Kit (Kapa Biosystems, Roche Inc.) described by Faircloth et al. (2015). Purification was performed with SPRI bead cleanup and AMPure (Rohland and Reich 2012). Then, custom dual-indexing barcodes (Faircloth and Glenn 2012) were incorporated in the library amplification. We quantified the libraries again using Qubit and qPCR (Kapa qPCR reagents; ViiA 7 Real-Time PCR System, Thermo Fisher Scientific Inc.) and grouped the libraries into pools (4–10 libraries/pool) according to the similarity of the Qubit values. We subjected these library pools to a vacuum centrifuge to adjust the concentrations to 147 ng/μL^{-1}.

The hybridization reactions (Hybs, Blocks and baits reagents) were conducted following recommendations of the myBaits Expert UCE kit (Arbor Biosciences), and the libraries were enriched with the Hym2.5Kv2A bait set (Branstetter et al. 2017c), which targets 2,524 UCE loci common across Hymenoptera and Formicidae. These new enriched libraries were purified, resuspended and amplified following recommendations from Faircloth et al. (2015). We confirmed the success of enrichment, quantification and fragment size (fragments ranging from 300 to 900 bp) using Bioanalyzer (Agilent Genomics, Santa Clara, CA). For select samples in one enrichment pool, a BluePippin instrument (Sage Science, Beverly, MA) was used to size-select to a range of 300–800 bp. We again pooled samples into 3 final pools which were submitted to the University of Oregon GC3F iLab and Novogene Corporation, Sacramento, CA for Illumina sequencing on an Illumina HiSeq 4000 (150 bp paired-end; Illumina Inc., San Diego, CA).

Processing of UCE Data

We demultiplexed newly generated sequences using the BBMap (Bushnell 2014) script demuxbyname2.sh. We trimmed the FASTQ files using Illumiprocessor (Faircloth 2011), a wrapper around Trimmomatic (Bolger et al. 2014) with default settings (LEADING:5, TRAILING:15, SLIDINGWINDOW:4:15, MINLEN:40). After trimming and quality control, we counted reads for each sample. We omitted from further analyses 4 samples for which the read count was very low at fewer than 100,000 reads (average number of reads per sample was 5.5 million).

Following the recent findings suggesting SPAdes is the optimal assembler for UCEs (Allio et al. 2020; Elst et al. 2021; Bossert et al. 2024), we assembled all reads, both newly generated and those downloaded from NCBI SRA, with SPAdes v3.14.0 (Bankevich et al. 2012) using the metagenomic mode (MetaSPAdes) with default settings.

We downloaded the ant-specific UCE bait set described by Branstetter et al. 2017c (Hym2.5Kv2A) and used Phyluce v1.6.7 (Faircloth 2015) to obtain alignments of UCE loci from our assembled contigs. Orthology assessment in Phyluce is performed by matching the assembled contigs to enrichment bait sequences with phyluce_assembly_match_contigs_to_probes (we used default settings: min_coverage = 50, min_identity = 80). This step generates a SQLite database, which is then used to build individual FASTA files for the 2,524 orthologous loci with phyluce_assembly_get_match_counts, phyluce_assembly_get_fastas_from_match_counts, and phyluce_assembly_explode_get_fastas_file.

Read, assembled contigs, and individual locus statistics for all samples can be found in Supplementary Table S3.

Alignment and Trimming

For downstream analyses we considered only loci with at least 50% taxa present. We aligned all locus files using MAFFT v7.407 (Katoh and Standley 2013) under default settings. Once aligned, we trimmed each locus with Gblocks v0.91b (Castresana 2000) using settings relaxed from default (−b3 = 12 −b4 = 7 −b5 = h −p = n). We then trimmed our data with Spruceup v2020.2.19 (Borowiec 2019a). Because our data show a large disparity of within-locus nucleotide variation, we split our data into 2 equal parts by arranging loci according to proportion of variable sites. The proportion of variable sites and taxon occupancy was computed using AMAS v0.98 (Borowiec 2016a). We concatenated loci from the top and bottom halves of these proportion of variable sites separately, which we call fast- and slow-evolving loci, respectively. We ran Spruceup on each matrix without a guide tree, under an uncorrected p-distance setting with a window size of 20 and overlap of 10. We inspected plots resulting from Spruceup trimming and decided on 0.9 lognorm cutoffs for both slow-evolving and fast-evolving loci, which removed 0.36% or 332,760 and 0.74% or 775,620 of sites, respectively. We did not manually trim any sequences to a different cutoff value. We split the 2 trimmed data sets into 2,428 individual locus alignments and concatenated the alignments for the final matrix (abbreviated name "all-emp") with AMAS. See Table 2 for summaries of concatenated matrices and Supplementary Table S4 for statistics of individual locus alignments.

Symmetry Tests

To examine the potential impact of model violations on the inferred concatenated phylogeny, we used symmetry tests to identify loci that are likely to violate common assumptions of stationarity, homogeneity, and reversibility, as implemented in IQ-TREE under default settings (Naser-Khdour et al. 2019). The status of each locus can be found in Supplementary Table S4.

Simulations

To establish a baseline expectation for assessing the impacts of factors decreasing support and concordance, we simulated alignments closely resembling our empirical data. First, we inferred a concatenated unpartitioned tree using a matrix of all loci and individual gene trees under GTR + F + G4 model in IQ-TREE v2.1.2 (Minh et al. 2020b). Using a custom Python script, from each locus IQ-TREE log file we extracted alignment length, base frequencies, estimated rate parameters, and alpha parameter of the rate variation gamma distribution and used these to simulate 2428 loci mimicking empirical UCE sequence data. We also used the formula (locus tree length / no. taxa present in locus) / (concatenated tree length / no. of all taxa) to scale the trees accounting for the fact that some alignments were missing taxa. All loci were simulated on a full tree and were then masked with missing data for individual loci after simulation. Missing data patterns, i.e., coordinates of all gap character cells, were extracted from empirical alignments to mask the corresponding cells in simulated alignments using custom Python code (Supplementary Data script simulate_alignments.py). A limitation of our simulations is the assumption of a random distribution of variation along loci. Since variation in UCE loci is concentrated in the flanking regions, masking with empirical missing data patterns may produce unrealistic numbers of variable sites in the simulated alignments. This effect is slight in our data, with 42.9% proportion of parsimony-informative sites in empirical alignments compared to 50.2% in simulated alignments (Supplementary Data directory "8-simulations"). Following simulations, we generated replicate alignments ("sim-" prefix in Supplementary Data) for various numbers of simulated loci randomly selected without replacement. We used AMAS to create matrices corresponding to approximately 5 replicates of the entire data set irrespective of the number of loci included: 2,428 replicates of 5 loci, 1,214 replicates of 10 loci, 607 replicates of 20 loci, 243 replicates of 50 loci, 121 replicates of 100 loci, 61 replicates of 200 loci, 24 replicates of 500 loci, 12 replicates of 1,000 loci, and 5 replicates containing all 2,428 loci. We repeated this procedure for empirical loci. We inferred trees on simulated data replicates using the model and parameters

Table 2. Properties of empirical matrices assembled for this study. All matrices contain 292 taxa. All lengths represent trimmed alignments.

Data set	No. loci	Length nt or aa	**Avg. taxa per locus**	Missing %	Avg. locus length	Parsimony informative %
All UCE loci—nucleotides (all-emp)	2,428	680,924	247	20.4%	280	42.9%
UCE loci passing symmetry tests—nucleotides (sym-emp)	1,504	349,722	254	17.6%	233	41.0%
UCE loci failing symmetry tests—nucleotides (fail-emp)	924	331,202	236	23.3%	358	44.9%
Protein-coding—amino acids (prot-aa)	1,286	118,458	212	36.0%	92	20.4%
Protein-coding—nucleotides (prot-nt)	1,286	383,764	212	31.2%	298	42.7%
Protein-coding—nucleotides, 3rd codon positions removed (prot-nt-no3rd)	1,286	255,883	212	31.9%	199	13.6%
Clocklike loci for divergence analyses—nucleotides (clocklike)	100	47,263	230	24.6%	468	47.2%

used for simulations, thus setting up a best-case inference scenario. The topology under which the data was simulated can be found in Supplementary Figure S46.

Simulations to Empirical Comparison

To compare the empirical data set with simulations, we analyzed support in simulated and empirical data using gene concordance factors (Lanfear and Hahn 2024) as implemented in IQ-TREE v2.1.2 (Minh et al. 2020b) (--gcf and --scf arguments). Using a custom R script, we computed concordance factors at each node in the unpartitioned GTR + F + G4 tree divided by subtending branch length. We scaled this support measure by branch length so that values follow an approximately linear distribution, as opposed to the exponential distribution seen for unscaled gene concordance factors (Supplementary Data). We then plotted this index for each node based on trees from individual locus alignments of empirical alignments against simulated alignments. The resulting scatter plot can be used to compare the performance of empirical and simulated data: Points at the diagonal would show no decrease in concordance relative to simulated data. The further below the diagonal a node is, the lower the concordance is over simulated data. Supplementary Figure S46 shows node numbers corresponding to those in the analyses and in Supplementary Tables S9–10.

Extraction of Protein-Coding Sequences

In addition to nucleotide matrices, we also extracted protein-coding sequences from UCE loci using the workflow presented in Borowiec (2019b). We downloaded and combined protein FASTA files from assembled genomes of *Camponotus floridanus* (v7.5, RefSeq GCF_003227725.1) (Bonasio et al. 2010), and *Harpegnathos saltator* (v8.5, RefSeq GCF_003227715.1) (Bonasio et al. 2010). We then used uce_to_protein.py script relying on Python 3.6.7, BioPython 1.72 (Cock et al. 2009), SQLite 3.7.17, and BLASTX v2.9.0 (Camacho et al. 2009) to match unaligned UCE contigs to protein sequences from the 2 genomes. This workflow produces 3 matrices: an amino acid matrix (abbreviated "prot-aa"), a matrix with nucleotides corresponding to the protein-coding amino acids ("prot-nt"), and a version of "prot-nt" with 3rd codon positions removed ("prot-nt-no3rd"). We trimmed the resulting alignments individually with Gblocks using the same settings as described for UCE loci above and with Spruceup as follows. "Prot-aa" was trimmed under the uncorrected distance method without a guide tree, windows of 10 amino acids with overlap of 5, full taxon sample, and lognorm criterion with cutoffs of 0.95, 0.97, 0.98, and 0.99. "Prot-nt" and "prot-nt-no3rd" were trimmed using an identical scheme except window size and overlap 20 and 10, respectively, and cutoffs set to 0.9, 0.925, 0.95, 0.97, 0.98, and 0.99. Following examination of Spruceup plots for all taxa, we retained the 0.97 cutoff matrix for "prot-aa," resulting in 0.77% or 265,140 individual characters removed. We used the 0.9 cutoff matrices for "prot-nt" and "prot-nt-no3rd" for downstream analyses. These cutoffs resulted in removal of 0.8% or 897,990 individual characters in "prot-nt" and 5.73% or 4,279,815 individual characters in "prot-nt-no3rd" data sets. We did not manually trim any sequences to different cutoff values. See Table 2 for summaries of concatenated matrices and Supplementary Table S4 for statistics of individual locus alignments.

Individual Locus Analyses

To obtain gene trees, we inferred maximum likelihood trees using IQ-TREE with the best-fitting model determined by ModelFinder under the AICc criterion (Kalyaanamoorthy et al. 2017). To assess support, we used 2,000 ultrafast bootstrap replicates (Hoang et al. 2017) and 1,000 replicates of the Shimodaira-Hasegawa approximate likelihood-ratio test (SH-like aLRT; Shimodaira and Hasegawa 1999; Guindon et al. 2010; Minh et al. 2020b). We also increased the number of unsuccessful tree search iterations before stopping from 100 to 200 and decreased the strength of nearest neighbor interchange from the default of 0.5 to 0.2, per suggestion for short alignments from IQ-TREE documentation. Locus statistics can be found in Supplementary Table S4.

Concatenated Analyses

Trimming, symmetry tests, and extraction of protein-coding sequences described above resulted in six6 empirical matrices for concatenated analyses: 1) all loci—nucleotides (abbreviated as "all-emp"), 2) loci passing symmetry tests—nucleotides ("sym-emp"), 3) loci failing symmetry tests—nucleotides ("fail-emp"), 4) protein-coding sequences—amino acids ("prot-aa"), 5) protein-coding sequences—nucleotides ("prot-nt"), and 6) protein-coding sequences—nucleotides with 3rd codon positions removed ("prot-nt-no3rd"). The properties of these matrices are in Table 2.

We analyzed the empirical ("all-emp"), loci passing symmetry tests ("sym-emp"), and loci failing symmetry tests ("fail-emp") data sets under the GTR + F + G4 model and under the best models selected using ModelFinder, using 4 different partitioning schemes: 1) unpartitioned, 2) partitioned by locus, 3) partitioned by locus with merging, and 4) partitioned by sliding window side characteristics (SWSC, Tagliacollo and Lanfear 2018).

The protein-coding matrices prot-nt and prot-nt-no3rd were analyzed partitioned by locus. The prot-aa matrix was analyzed under the posterior mean site frequency (PMSF) method with 20-component mixture (Wang et al. 2018).

Additionally, we performed 500 bootstrap replicates for neighbor-joining (uncorrected distance) and maximum parsimony analyses on all-emp, sym-emp, and sym-fail matrices with PAUP* v4.0a (Swofford 2003). Details of the resulting concatenated analyses

can be found in Supplementary Table 5 and trees in Supplementary Figures 1–35.

Species Tree Analyses

We performed coalescence species tree analyses in ASTRAL v5.7.7 (Zhang et al. 2018). We performed analyses: 1) using all 2,428 gene trees from individual locus analyses, 2) using 1,504 gene trees that passed symmetry tests, and 3) using 924 gene trees that failed symmetry tests. Each analysis was conducted using 1) unaltered individual gene trees, and 2) gene trees with nodes with SH-like aLRT support of zero collapsed, following the recommendation of Simmons and Gatesy (2021). We used Newick utilities (Junier and Zdobnov 2010) to collapse nodes.

We also performed coalescent analysis in ASTRAL-MP v5.15.5 (Yin et al. 2019) based on bootstrap replicates inferred for each locus tree. Because of computational constraints, we used 100 randomly selected bootstrap replicates from each locus tree for the concatenated ASTRAL analysis. We then copied the 100 resulting ASTRAL bootstrap trees 20 times for 2,000 trees for the consensus analysis (see below). In addition to ASTRAL, we used SVDquartets as implemented in PAUP* v4.0a (Swofford 2003) on all-emp, sym-emp, and fail-emp matrices with 500 bootstrap replicates. Details on the resulting species tree analyses can be found in Supplementary Table 5 and trees in Supplementary Figures 36–44.

Consensus Tree Building

To create a consensus tree with support we included bootstrap replicates from 4 analyses we deemed representative of their respective categories:

- Full data set under ModelFinder with SWSC-EN partitioning. We included this data because it represents the full data set with model and partitioning scheme achieving the best likelihood.
- Symmetry tests-passing loci under ModelFinder partitioned by locus. We included this analysis to represent data that is less likely to violate common model assumptions in concatenated analyses, represented by analysis with the best likelihood for this alignment.
- Protein-coding amino acid sequences under PMSF model. We included these results to represent analysis of a qualitatively different data set that provides less information but is likely less biased in certain aspects than nucleotide alignments, under a complex protein mixture model.
- Coalescent analysis in ASTRAL based on bootstrap replicates inferred for all gene trees. We included this analysis to represent a coalescent model approach to ant phylogeny, accounting for incomplete lineage sorting.

The concatenated analyses each included 2,000 trees from ultrafast bootstrap (Hoang et al. 2017). We then constructed an extended majority-rule consensus tree using IQ-TREE from all bootstrap replicates. We used this consensus topology to infer branch lengths under ModelFinder with SWSC-EN partitioning using all loci. We did not include any trees from concatenated analyses comprising only loci failing symmetry tests. The resulting consensus tree can be found in Figure 5 and Supplementary Figure S45.

Visualizing Support

We used DiscoVista (Sayyari et al. 2018) to visualize how support for the monophyly of 27 groupings changes across all 44 analyses. We used 95% support threshold to indicate strong support or strong rejection in these visualizations. Clade definitions are available in Supplementary Table S6 and analyses are outlined in Figure 2. We also used DiscoVista to compare support for the same groupings across all 2,428 empirical gene trees, across 1,504 gene trees passing symmetry tests, and across all simulated loci (Figure 3).

Ancient Introgression

We tested whether introgression in ancient lineages could be responsible for the conflict between the phylogenies using the test statistic Δ (Huson et al. 2005; Vanderpool et al. 2020). We followed Cunha et al. (2022) by identifying nodes for which more than 5% of gene trees were discordant with the species tree (376 internal branches) and calculated the *P*-value of each observed *Z*-score. For this one-tailed test, evidence of introgression would be indicated by *Z*-scores of at least 1.65 at threshold *P*-value of 0.05.

Incomplete Lineage Sorting

To measure gene and site concordance across the ant phylogeny, we performed concordance factor analysis on consensus, empirical, and simulated data. This was done with functions (--gcf and --scf 500 arguments) implemented in IQ-TREE v2.1.2 (Minh et al. 2020b) on input of concatenated matrices, species trees, and individual locus trees. We also investigated which branches may be impacted by incomplete lineage sorting (ILS). To this end, we performed chi-squared tests on gene and site concordance factors on each branch to see whether they violate the assumption of equal discordance frequencies, which is indicative of ILS (Lanfear and Hahn 2024). Gene and site concordance and discordance factors for the consensus tree can be found in Supplementary Table S9 and those for empirical data inferred under the unpartitioned GTR + F + G4 model and simulated data in Supplementary Table S10.

Divergence Dating

Following Smith et al. (2018), we used the "gene shopping" approach to divergence dating. These authors used simulated and empirical data and found that using a subset of loci least likely to violate the molecular

clock resulted in reliable and more precise divergence estimates than uncorrelated clocks. We identified a subset of loci with desirable properties of clock-likeness, information content, and low topological conflict with the species tree for our divergence analyses under strict molecular clock. We minimized root-to-tips variance and Robinson–Foulds (RF) distances to a reference tree (Robinson and Foulds 1981; Penny and Hendy 1985) and maximized average branch lengths as loci selection criteria. We computed each metric and ranked each of our loci accordingly. We then used the sum of weighted ranks to identify the most desirable loci. Our weights were 0.5 for clock-likeness, 0.3 for average branch lengths, and 0.2 for RF distances. Visual inspection of gene trees for highest and lowest-ranking loci confirms that this approach successfully identifies loci with low branch length variance, reasonable topology, and significant information content. Custom code written for this purpose is freely available as a well-documented and user-friendly R script "kinda_date.R" (https://github.com/marekborowiec/kinda_date). Its functionality is similar to SortaDate (Smith et al. 2018) but it can be used with unrooted gene trees, correctly sorts loci according to multiple criteria, allows for custom weights of these criteria, and requires only R (R Core Team 2021) libraries ape (Paradis and Schliep 2019) (tested with v5.7-1) and phytools (Revell 2024) (tested with v2.1-1). Locus statistics and ranking can be found in Supplementary Table S7.

We then selected the 100 best-ranking loci for unpartitioned downstream analyses in MCMCTree (Yang 2007) using the fixed, consensus tree topology and likelihood approximation (Reis and Yang 2011). Using published literature, we identified fossils that can be confidently assigned to crown groups of clades in our taxon sampling scheme. When multiple fossils of the same age potentially informed a set of nested nodes, we generally chose to calibrate only the shallowest node. This resulted in a set of 37 fossil calibrations. We conservatively assumed that each fossil could belong to either the crown or stem group of a clade in which it has been placed in the literature, except for those that had been explicitly placed using total-evidence tip dating (Ronquist et al. 2012). In other words, a fossil always calibrates the node preceding the most recent common ancestor of the least inclusive clade in which the fossil has been placed. On each such node we placed a minimum age constraint corresponding to the youngest accepted age of the deposit from which the fossil was reported. In MCMCTree this means a Cauchy distribution truncated at minimum age with 0.1 offset and scale parameter equal to 1. The distribution has a soft bound, meaning that 2.5% of prior density mass is smaller than the minimum age (Yang 2020). See Supplementary Table S8 for details on fossil calibrations used and corresponding citations. MCMCTree analyses require a root age constraint, and we explored 2 alternative approaches. The first approach used a uniform prior between 129 and 158 Ma with soft bounds. These ages correspond to the minimum and maximum of 95% highest posterior density (HPD) estimates obtained for Aculeata by a recent comprehensive phylogeny of Hymenoptera (Blaimer et al. 2023). The other approach used a less informative prior. It assumed a uniform distribution from the present to a soft bound at 224 Ma, the maximum of 95% HPD from a study that recovered the oldest recently inferred age for Aculeata (Peters et al. 2017).

Following Mongiardino Koch et al. (2022), who demonstrated that clock model choice is potentially the most impactful parameter in divergence dating, we replicated each root constraint analysis with 6 runs under each independent and autocorrelated molecular clock models. We did not compare these results to strict clock inference, however, as it is not recommended under the likelihood approximation in the MCMCTree official tutorials (Reis et al. 2017), and our attempt at inference without approximation proved to be too computationally expensive. We ran each chain for 5,000,000 generations, sampling every 500 generations with a burnin of 10%, or 500,000 generations. We combined all runs and visually examined the posterior using Tracer (Rambaut et al. 2018), verifying that it was successful and that ESS values for all parameters exceeded 300. Trees resulting from divergence dating analyses are visualized in Supplementary Figures S49–52.

Results and Discussion

After excluding samples with fewer than 100,000 reads, sequencing yielded an average of 5.76 million reads (range 308,532–43,105,119, $n = 180$) per specimen. Assembly resulted in an average of 120,645 contigs (range 1,346–699,932) across all samples. Average UCE contig length was 1,074bp. There were 2,428 loci with at least 50% of taxa present and average locus length of 280bp following trimming. See Supplementary Tables S3 and 4 for details on reads, assemblies, and UCE loci.

We acknowledge the shorter average UCE locus length in our data set compared to recent UCE ant studies (e.g., Branstetter et al. 2017c) despite using similar trimming strategies. We suspect this discrepancy stems from the inclusion of ~15% of rare, critical taxa with low DNA quality or enriched using the older v1 probe set. This introduced missing data and led to more aggressive trimming. This disproportionately impacts coalescent analyses, which will rely on locus trees with relatively low resolution.

The concatenated matrix of all loci contained 292 sequences with 680,924 columns, 489,709 variable sites and 292,009 parsimony-informative sites. Symmetry tests were passed by 1,504 of the 2,428 loci. Following alignment and trimming there were a total of 1,753 loci with protein-coding sequences, 1,286 of which had at least 50% of taxa, and 30 amino acid positions. These 1,286 loci were selected for downstream analyses. They had on average 212 taxa (73% of all taxa).

Simulations: How Much Data Are Enough?

Assuming the topology and branch lengths of a concatenated tree inferred under unpartitioned GTR + F + G4 model ("true tree") are reasonable approximations of the true topology and branch lengths in the Formicidae tree, simulations show that very many loci are necessary to recover the correct topology using data similar to ours. When analyzing RF distances of simulated replicates to the true tree (Fig. 3a), at least 100 loci are necessary to recover the true tree, while only replicates of 1,000 loci and the full data set (2,428 loci) recover the true tree with 100% consistency. Average RF distance to the true tree for alignments composed of 5 loci is 78, 43.2 for 10 loci, 17.6 for 20 loci, 4.7 for 50 loci, 1.7 for 100 loci, and 0.1 for 500 loci (Supplementary Data).

Comparison between expected gene concordance in a naive empirical analysis and simulations reveals the branches that most severely underperform concordance expectations (Fig. 3b). The topology inferred under the unpartitioned GTR + F + G4 model reveals that some of the most affected branches are in the clades Dorylinae and Crematogastrini. This suggests these branches are particularly impacted by model violations in a simple concatenated analysis and thus unlikely to represent true relationships.

Side by side comparisons of DiscoVista (Sayyari et al. 2018) plots show that the majority of empirical loci lack signal to resolve most of the contentious relationships in ants, and the same is true for simulated loci (Fig. 3c). These results show that hundreds or even thousands of short markers, such as UCE loci, are needed to resolve certain backbone relationships. Furthermore, many of these relationships are likely impacted by systematic biases, such as insufficient modeling of the heterogeneity of the evolutionary process across sites and taxa, and phenomena such as incomplete lineage sorting.

This underscores the fact that caution is needed in interpreting results of phylogenomic studies of ants and likely other taxa, for as the amount of data used increases, systematic bias may result in increasing confidence in the wrong topology (Jermiin et al. 2004; Philippe et al. 2005; Sullivan and Joyce 2005). Moreover, random error still plays a role in phylogenomics, not only at the level of individual gene trees, but also in the trees generated from the concatenated alignments of hundreds of loci containing extremely short branches.

Phylogeny of the Formicidae

Disagreement among analyses and general trends.—Few clades on the ant phylogeny receive equivocal support, while most vary from being moderately or strongly supported to moderately to strongly rejected, depending on analysis. Many of the most difficult to place taxa, such as *Paraponera*, *Tatuidris*, *Santschiella*, *Oecophylla*, and others, are placed on long terminal branches connected to very short internal edges (Fig. 4, Supplementary Figs. S1–44). Visual examination of support in analysis blocks shows that data selection, for example all loci versus loci failing symmetry tests, exerts stronger influence than model or partitioning scheme selection, although maximum parsimony and neighbor joining trees tend to be different from maximum likelihood trees in individual data treatments. Furthermore, the model used, GTR versus best-fitting choice, has more impact than partitioning scheme on support for several clades.

Consensus tree relationships, support, and concordance.—The consensus tree provides a single phylogenetic hypothesis while accounting for and highlighting uncertainty across analyses (Fig. 5). All subfamilies are unambiguously monophyletic and relationships among the subfamilies are identical to those recovered by Romiguier et al. (2022). This includes formicoid subfamilies, which have been stable in previous studies (Brady et al. 2006; Moreau et al. 2006; Moreau and Bell 2013; Branstetter et al. 2017c). Contentious relationships within poneroids and modest uncertainty over the rooting of the ant tree are consistent with disagreements observed in prior work. Other areas of contention identified in previous studies include the poorly-resolved backbone of the Dorylinae (Borowiec 2019b), the position of Bothriomyrmecini within Dolichoderinae (Ward et al. 2010), the position of the genera *Gesomyrmex*, *Gigantiops*, *Oecophylla*, and *Santschiella* within Formicinae (Blaimer et al. 2015), and a number of relationships within the largest subfamily, Myrmicinae, including positions of several genera in the Attini (Ward et al. 2015; Branstetter et al. 2017a), and relationships among genus groups of Crematogastrini (Blaimer et al. 2018).

Concordance factors on the consensus tree (Fig. 5) show a wide range of values. An average branch is concordant with 16.8% or 318 loci (range 0–1,607), but many internal branches are concordant with fewer loci, and site concordance factor is on average 43.8% (27.3–92.5%) (Supplementary Table S9). Testing for incomplete lineage sorting (ILS) reveals there are 187 branches for which the pattern cannot be explained by ILS alone, based on discordance patterns for either loci or sites. However, this test is only valid if we assume random sampling of unlinked loci or sites across the genome (Lanfear and Hahn 2024). We compared ILS violation patterns between empirical and simulated alignments and trees generated for this study and found that similar numbers of branches violate ILS assumptions for both data sets. There are 106 branches for which site patterns violate ILS assumptions in simulated data versus 97 branches in empirical trees inferred under GTR model, 29 branches violating ILS based on loci in both data sets, and 35 versus 55 branches violating ILS for both loci and sites in simulated and empirical data, respectively. There are 119 out of 289 internal branches conforming to ILS assumptions in data simulated without ILS. This suggests that much of the ILS signal in our empirical data may be obscured by stochastic patterns unrelated to biological phenomena. This is consistent with overall low signal in UCE loci (Figs. 3a, 3c) and perhaps explains why Δ test statistics did not detect

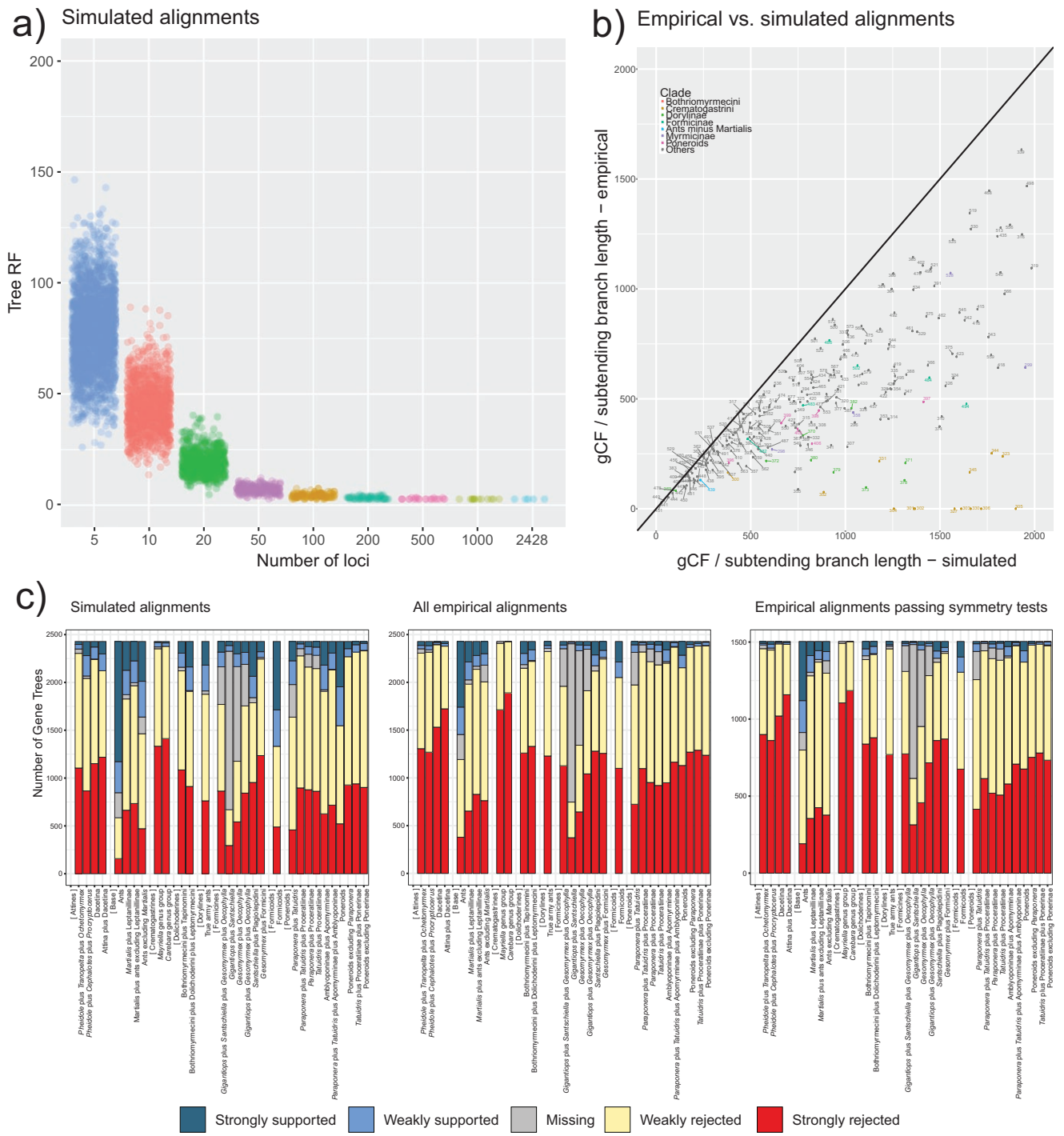

FIGURE 3. (previous page). Comparisons of empirical and simulated data. a) Robinson–Foulds distances of simulated loci alignments to the true tree, demonstrating the difficulty of correctly resolving all the nodes on the ant phylogeny: none of the concatenated alignments of 50 or fewer loci recovered the tree under which they were simulated. Alternative visualizations of these distances, including boxplots with means are available in Supplementary Data directory "7-emp-vs-sim-comparison." b) Comparison of gene concordance factors in simulated versus empirical loci. Numbers represent nodes in Supplementary Figure S46 and distances from the diagonal are in Supplementary Table S9. If our analyses were able to perfectly capture the phylogenetic process, there would be little difference between nodal support in empirical and simulated alignments and all nodes would lie along the diagonal. The plot is limited to a nodal support index of 2,000. The further a node is below the diagonal, the more it underperforms expectations under the correct inference model. Gene and site concordance factors on the empirical tree are in Supplementary Figure S47 and on the simulated tree in Supplementary Figure S48. c) DiscoVista plots showing support for select clades in individual gene trees. Alignments were simulated on the tree in Supplementary Figure S46, inferred under the unpartitioned GTR + F + G4 model. Strong support and rejection are defined by 95% threshold. These results highlight that few individual UCE loci decisively support most clades, even under simulations. Note that more loci strongly support the leptanilloid clade (*Martialis* plus Leptanillinae) in empirical data than the alternative arrangement of either *Martialis* being sister to all ants or sister to ants excluding Leptanillinae. In simulated alignments *Martialis* is sister to all other ants and that is the most commonly recovered topology there. For better readability, please refer to the PDF version of this figure, as the printed version may have font size limitations.

FIGURE 4. Heatmap showing support for select clades across 44 analyses. For clarity, analyses are grouped into colored blocks on the x-axis and clades are grouped into y-axis blocks affecting particular parts of the ant phylogeny, for example, "[Attines]," [Base]," etc. Concat = concatenation; SpTree = species tree; ASTRAL = ASTRAL analysis; SVDQ = SVDquarters analysis; all vs symtest_pass vs symtest_fail refers to subset of loci analyzed: all loci vs passing symmetry tests loci vs failing symmetry tests loci; GTR = GTR+F+G4 model; MFP = ModelFinder-selected model for each alignment/partition; PF = partition merging; unpart = unpartitioned analysis; by locus = partitioned by locus; SWSC = partitioned using sliding-window site characteristics; by codon = partitioned by locus and codon position; collapsed vs uncollapsed: all nodes with SH-aLRT support of 0 collapsed vs nodes with SH-aLRT support of 0 not collapsed; MP = maximum parsimony; NJ = neighbor-joining; AA = amino acids; NT = nucleotides. Refer to Supplementary Table S5 and the methods section for additional details of the analyses and Supplementary Table S6 for clade definitions. For better readability, please refer to the PDF version of this figure, as the printed version may have font size limitations.

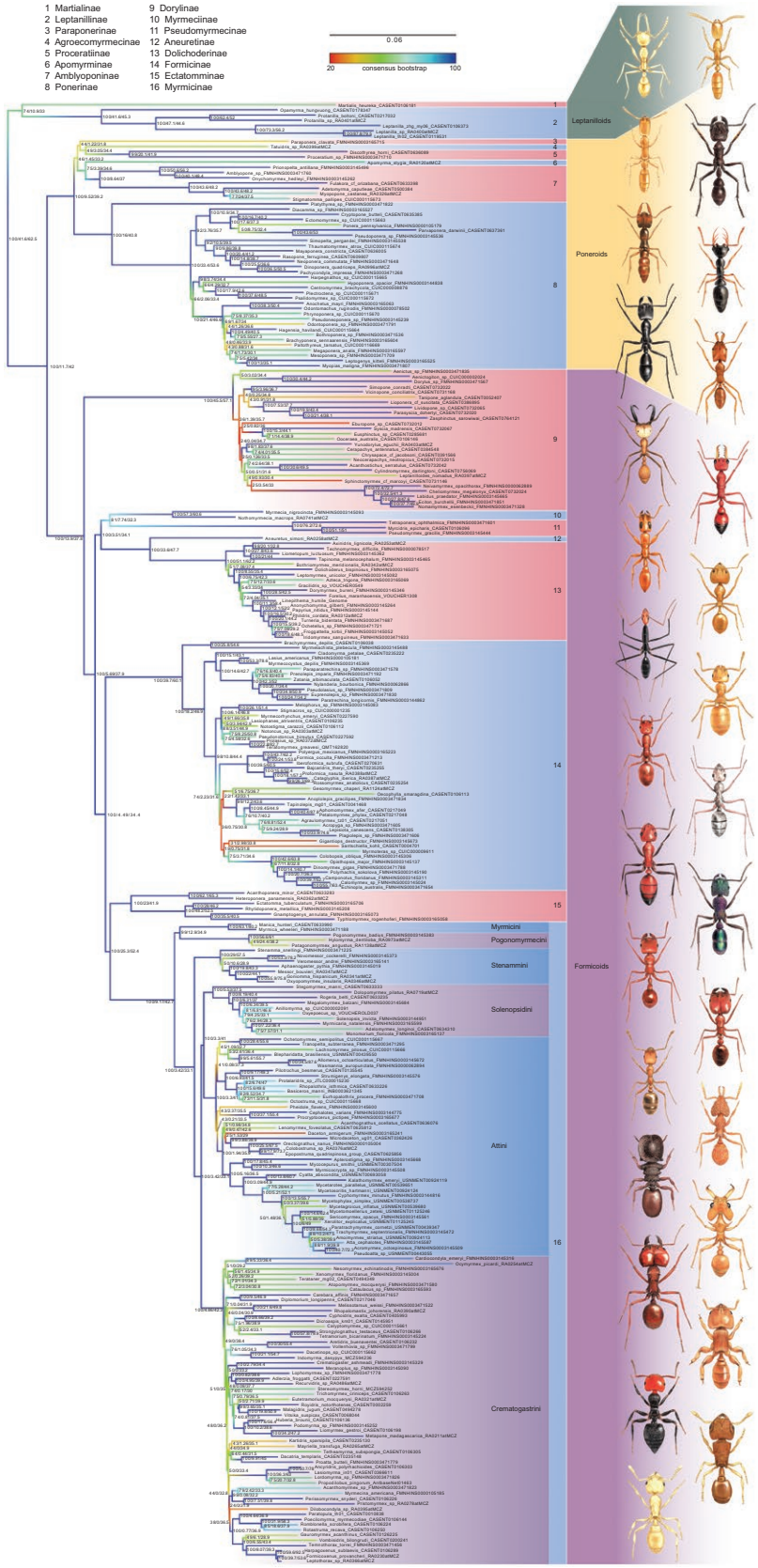

FIGURE 5. (previous page). Phylogram of the ants based on consensus topology with branch lengths inferred under SWSC-EN partitioning, partition merging, and ModelFinder model selection in IQ-TREE. Major clades, subfamilies, and tribes of the largest subfamily Myrmicinae are partitioned into colored taxon blocks. Branches are colored according to bootstrap support across the 4 analyses included in the consensus tree

any hybridization in our data (Huson et al. 2005). These results suggest caution is warranted in extracting biological interpretations of concordance factors and Δ test statistics from UCE data.

Rooting of the ant tree.—Our results support the "leptanilloid clade" (Borowiec et al. 2019, 2020; Romiguier et al. 2022), alternatively called "leptanillomorph clade" (Boudinot et al. 2022), as the sister group to all other ants in analyses of the amino acid matrix under finite mixture models, all maximum likelihood analyses of loci passing symmetry tests, and all ASTRAL species trees, as well as in the neighbor-joining tree of loci failing symmetry tests (Fig. 4). Analysis of individual gene trees (Fig. 3c, Supplementary Table S9) shows that there are more loci strongly supporting Martialinae + Leptanillinae than there are loci strongly supporting Martialinae + ants excluding Leptanillinae, which explains the strong support in ASTRAL analyses. Analyses of loci that fail symmetry tests, on the other hand, show maximum support for Martialinae as sister to all other ants. Maximum likelihood topologies from analyses of all loci are intermediate: they place Martialinae as sister to all ants with moderate to strong support, which is lowered when more complex and better-fitting (Supplementary Table S5) models are applied. The hypothesis that *Martialis* is sister to ants excluding Leptanillinae, which was originally recovered by Kück et al. (2011) and more recently promoted by Cai (2024), is moderately supported by 2 maximum likelihood trees of protein-coding loci with the 3rd codon position removed and strongly or moderately rejected by all other analyses. Previous findings show that the position of *Martialis* can be affected by compositional bias (Borowiec et al. 2019) and loci failing symmetry tests are likely to exacerbate this bias (Borowiec 2019b; Naser-Khdour et al. 2019). Concordance factors (Fig. 5, Supplementary Table S9) also show the leptanilloid clade has the most individual gene trees supporting it. We interpret the directional trend in which support for Leptanillinae + Martialinae increases as better-fitting models are applied and loci prone to compositional bias are removed as evidence supporting the leptanilloid clade.

Relationships among poneroid subfamilies.—We find near-universal support for the poneroid clade of subfamilies but the topology within it is contentious (Fig. 4). Our consensus topology of poneroids shows an initial split between the Ponerinae and the remaining subfamilies, with the latter clade splitting into Apomyrminae and Amblyoponinae on one side and (Paraponerinae, (Agroecomyrmecinae,Proceratiinae)) on the other. This poneroid topology is different from some of the recently proposed trees (Branstetter et al. 2017c; Borowiec et al. 2019) but consistent with the genome-wide amino acid-based phylogeny of Romiguier et al. (2022). Alternative arrangements, particularly Paraponerinae + Agroecomyrmecinae monophyly, are supported to some extent by most analyses, moderately so by loci passing symmetry tests, although only in analyses with inferior log likelihood values (Fig. 4). Inclusion of the extremely rare *Ankylomyrma*, the only other extant genus of Agroecomyrmecinae absent from our study, would break up the long branch leading to Agroecomyrmecinae and could potentially improve resolution (Ward et al. 2015).

Other relationships.—Early molecular studies (e.g., Brady et al. 2006; Moreau et al. 2006) showed strong support for the clade informally called formicoids, and we recover it with strong support in all analyses (Fig. 4). The universal support for this clade in molecular data stands in remarkable contrast with lack of known morphological synapomorphies (Boudinot et al. 2022).

Borowiec (2019b) showed that the topology within the subfamily Dorylinae is sensitive to model violations due to saturation and compositional heterogeneity, particularly regarding monophyly of the so-called "true army ants." Our analyses show a similar pattern to that at the root of the ant tree (Fig. 3), where concatenated and ASTRAL coalescent analyses of all loci strongly support monophyletic true army ants. This support is even stronger in concatenated analyses of loci failing symmetry tests. When only loci passing symmetry tests are included, there is strong rejection of true army ant monophyly. Our analyses also show many backbone relationships of the doryline radiation with low support.

The most recent comprehensive genus-level phylogeny of Dolichoderinae (Ward et al. 2010) found the position of the tribe Bothriomyrmecini to be either sister to Tapinomini or a clade of Dolichoderini + Leptomyrmecini. The former arrangement is supported by the majority of analyses using all loci and only loci passing symmetry tests. Bothriomyrmecini + (Dolichoderini + Leptomyrmecini) is universally supported by analyses of protein coding sequences, including nucleotide and amino acid-based phylogenies, as well as loci failing symmetry tests. This is an unusual pattern, as in most cases the amino acid matrix does not support the same relationship as loci failing symmetry tests (Fig. 4).

As in Blaimer et al. (2015), the positions of *Gesomyrmex*, *Gigantiops*, *Oecophylla*, and *Santschiella*, 4 morphologically derived and long-branched genera of Formicinae,

building for visual identification of discordant areas of the tree. Values at nodes are: Bootstrap proportion in consensus / gene concordance factor/site concordance factor. Scale bar in substitutions per site. Outgroups are omitted. Ant illustrations not to scale, top to bottom, left column: *Martialis*, *Apomyrma*, *Stigmatomma*, *Dinoponera*, *Eciton*, *Tetraponera*, *Leptomyrmex*, *Polyergus*, *Colobopsis*, *Pogonomyrmex*, *Monomorium*, *Cephalotes*, *Atta*, *Crematogaster*, *Temnothorax*; right column: *Leptanilla*, *Paraponera*, *Thaumatomyrmex*, *Anochetus*, *Myrmecia*, *Aneuretus*, *Lasius*, *Cataglyphis*, *Rhytidoponera*, *Solenopsis*, *Strumigenys*, *Daceton*, *Melissotarsus*, *Mayriella*. Original artwork by Carim Nahaboo. For better readability, please refer to the PDF version of this figure, as the printed version may have font size limitations.

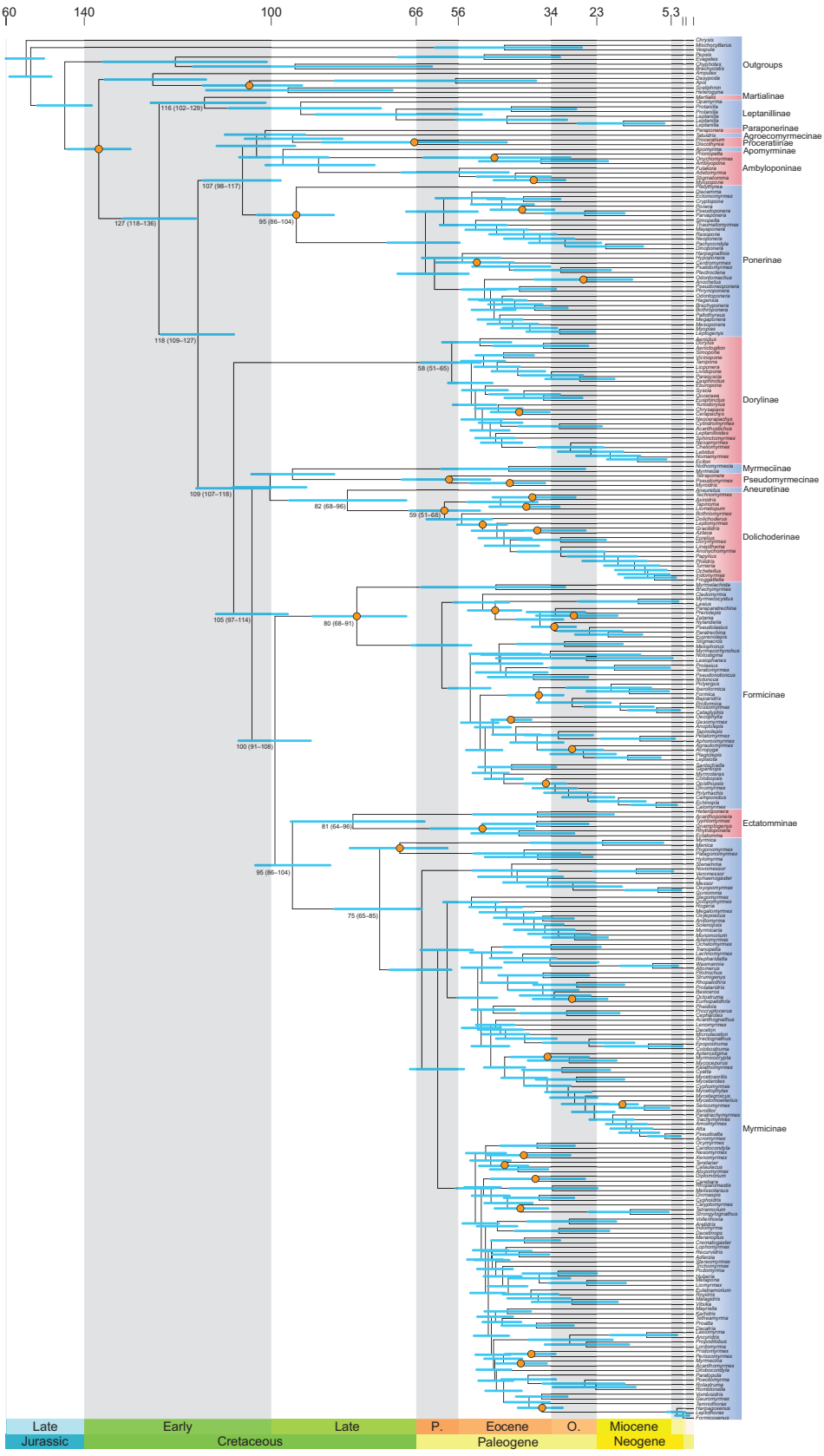

FIGURE 6. (previous page). Chronogram of ant evolution inferred in MCMCTree under likelihood approximation, root age uniform prior with soft bounds at 129 and 158 Ma, and independent rates clock using fixed consensus topology and 100 loci selected based on clock-likeness,

are difficult to ascertain. Few loci are available to inform these relationships in our data (Fig. 3c). Our consensus phylogeny shows *Gesomyrmex* and *Oecophylla* to be sister lineages with 51% bootstrap support and *Gigantiops* and *Santschiella* as sister lineages with even lower (31%) support. Both sister pairs are supported by most analyses of loci passing symmetry tests, but otherwise the pattern of supporting analyses is divergent and complex.

In the myrmicine tribe Attini (Fig. 5), the monophyly of subtribes Attina and Dacetina (Branstetter et al. 2017a), as well as their sister relationship are strongly supported by most analyses, apart from neighbor-joining SVDquartet trees (Fig. 4). The sister group of one of the most species-rich ant genera, *Pheidole*, has been mostly accepted to be the clade uniting *Procryptocerus* and *Cephalotes* (Ward et al. 2015), although alternatives have been proposed (Moreau 2008). We did not find universal support for this relationship (Fig. 5), with the most common alternative sister group for *Pheidole* consisting of a clade uniting *Tranopelta* and *Ochetomyrmex*. This alternative is supported mostly by analyses of loci failing symmetry tests and the majority of ASTRAL analyses. In the analysis of amino acid data, *Pheidole* is sister (with low support) to a clade consisting of the genera *Ochetomyrmex*, *Tranopelta*, *Blepharidatta*, *Wasmannia*, and *Allomerus* (Fig. 4; Supplementary Fig. S15).

The *Carebara* genus-group and the *Mayriella* genus-group in the myrmicine tribe Crematogastrini (Fig. 5) were identified as monophyletic in concatenation and nonmonophyletic in coalescence analyses of Blaimer et al. (2018). Our study shows the same pattern, although support is weaker for both in concatenated analyses of loci failing symmetry tests. Several branches at the backbone of Crematogastrini are notable for having gene concordance factors equal to zero, meaning they are not concordant with any individual gene trees in the data. This is significantly fewer than in simulated data, where at least 10 loci are concordant with any individual branch.

Divergence Dating

Recent studies have recovered a wide range of ages for crown group ants, with confidence intervals ranging from 103 to 220 Ma and an average of just under 150 Ma (Borowiec et al. 2020). Here we find that the choice of root node prior, which is required by MCMCTree, common software used for Bayesian divergence dating, has enormous impact on the inferred ages. Using a 129 to 158 Ma soft-bound uniform prior on the age of Aculeata, derived from a recent study considering all of Hymenoptera (Blaimer et al. 2023), we recovered the mean age of the crown ants to be 127 Ma (118–136 Ma 95% highest posterior density or HPD) under independent clock inference (Fig. 6; Supplementary Figs. S49–52; Supplementary Data). Correlated clock results were slightly younger at 121 Ma mean age and 115 and 127 Ma range.

In contrast, using a less-informative prior of uniform distribution with a soft bound at maximum of 224 Ma produced much older trees, with independent clock estimate putting the age of crown ants at 188 Ma (161–215 Ma 95% HPD) and correlated clock at 164 Ma (146–182 Ma 95% HPD).

The younger divergences inferred here are consistent with those obtained in some recent studies (Borowiec et al. 2019; Boudinot et al. 2022), while the older ages are similar to other published results (e.g., Moreau et al. 2006; Blanchard and Moreau 2017; Economo et al. 2018). Comprehensive investigation of factors underlying the inference of timing of ant evolution is beyond the scope of the study. Substantial progress in this area awaits future explicit incorporation of fossils and their morphology in the fossilized birth-death process and tip-dating frameworks (Ronquist et al. 2012; Heath et al. 2014; Boudinot et al. 2022). In the meantime, we hope that our calibrated backbone ant phylogenies will be useful to comparative biologists.

Conclusion

Our simulations show that hundreds of loci are needed to confidently resolve the most recalcitrant nodes of the ant phylogeny under the best-case scenario of inference using the generating model and with no systematic bias. It is worth noting here that our trimmed UCE loci were quite short compared to recent UCE ant phylogenies (280bp compared to ~600bp in Branstetter et al. 2017c, for example), and data sets with longer UCE alignments would likely provide better resolution and a more optimistic best-case simulation scenario.

The implications of this for phylogenomics in general are 3-fold. First, collecting more genetic data may be beneficial for some phylogenetic questions. Second, random error is still at play in phylogenomics when extremely short branches are encountered. Third, this result underscores that the small amount of information may be present for certain relationships could easily be overwhelmed by systematic error accumulated across large alignments, leading to inflated support for incorrect groupings. Potential violations leading to systematic bias include the ubiquitous assumptions of stationarity, reversibility, and homogeneity (SRH)

similarity to the reference tree, and phylogenetic signal. Select nodes contain mean age as well as minimum and maximum ages in 95% highest posterior density in parentheses. Circles at nodes signify fossil calibrations. See Supplementary Table S8 for calibration details and Supplementary Data for tree files. Choice of alternative root prior led to much older divergence time estimates (Supplementary Figures S51–52). For better readability, please refer to the PDF version of this figure, as the printed version may have font size limitations.

(Naser-Khdour et al. 2019), other analytical issues (Philippe et al. 2017), and biological processes leading to incongruence (Bravo et al. 2019; Schrempf and Szöllősi 2020; Steenwyk et al. 2023).

Model violations resulting in systematic bias, such as compositional heterogeneity across taxa, are often correlated with phylogenetic distance among taxa (Foster 2004; Cruaud et al. 2021; Prebus and Rabeling 2024). At the same time, reduced representation phylogenomic workflows, such as the UCE pipeline, recover more data when more closely related taxa are included. Therefore, studies assembling supermatrices of ancient clades, such as this one, are likely to be inferior to focused inference on smaller clades, which can maximize taxon and character sampling while simultaneously minimizing the potential for systematic bias.

Currently, there are no models relaxing SRH assumptions and simultaneously accounting for biological phenomena such as incomplete lineage sorting as well as analytical errors propagating across phylogenomic workflows (Bryant and Hahn 2020; Simion et al. 2020). Such models would require an enormous amount of information for reliable inference of the many parameters being estimated. However, very little information is available along the short branches often subtending the most difficult nodes of the ant (or any other) phylogeny. This means that very complex models may still not be able to resolve those relationships but could provide a more accurate picture of uncertainty. It is also clear that phylogenetic histories vary substantially across genomes (e.g., Pease et al. 2016; Edelman et al. 2019) and systematists need to consider situations where a bifurcating tree is not the best approximation of the phylogenetic process.

An alternative to using complex models is eliminating loci or sites that contribute to bias (Philippe et al. 2017). Measuring contribution and susceptibility to bias is difficult, however, as we lack user-friendly tools for investigating impacts of multiple sources of bias in a unified framework. The potential fragility of results at recalcitrant nodes illustrates one difficulty with data selection: How much should investigators weigh phylogenetic signal, which may be correlated with systematic bias (cf. Borowiec 2019b), when searching for optimal markers (Mongiardino Koch 2021)? The answer will depend on a quantitative understanding of potentially biasing influences present across the tree. Our approach used here to compare inference under empirical and simulated scenarios can highlight underperforming nodes, but it says nothing about underlying causes. Bayesian posterior predictive approaches (Bollback 2002; Brown 2014; Brown and Thomson 2018) can also be applied, but are currently prohibitively computationally expensive and will require even more molecular data. In a study of a recently diverged ant clade (Prebus 2021), posterior prediction resulted in rejection of UCE loci containing 70.7% of parsimony-informative sites (Prebus pers. comm.). Given our finding that a large amount of data is needed to resolve contentious nodes even absent model violations, such aggressive trimming is not viable for a global ant phylogeny inferred from ultraconserved elements.

While our approach of choosing only certain analyses is subjective and the simple consensus tree is imperfect, we view it as an improvement over the common paradigm of presenting one tree from a preferred analysis or simply contrasting results from phylogenies obtained using different approaches. Conceptually, our approach resembles model ensembling, a common machine learning technique shown to improve predictive inference in certain scenarios (Dietterich 2000). In the absence of a comprehensive framework to assess the impact of diverse model violations across the tree and alignment, we hope our consensus hypothesis will be useful for comparative analyses across the ants.

Incremental progress will continue in ant phylogenetics as additional genomic resources for ants become available (Boomsma et al. 2017) and taxon sampling further increases. For reasons explained above, focused phylogenetic studies of individual clades are more likely to contribute to resolving difficult relationships than new global ant phylogenies. Discoveries of rare, phylogenetically isolated lineages (e.g., Rabeling et al. 2008; Yamane et al. 2008) will likely continue in undersampled biodiversity hotspots (Guénard et al. 2012). It is also probable that certain nodes will remain intractable, as reticulate phylogenetic histories are real and result from well-documented biological phenomena (Maddison 1997; Huson et al. 2010; Mallet et al. 2016; Steenwyk et al. 2023). However, genomic resources will enable a more complete picture of incongruence, as has been the case in other organisms (Edelman et al. 2019; Knyshov et al. 2023; Mirarab et al. 2024). Renewed interest in insect morphology enabled by microcomputed tomography is producing large amounts of new information for the ants (Richter et al. 2022, 2023). Future work will refine our understanding of the timeline of ant evolution by explicitly incorporating morphology and the rich and constantly expanding fossil record into phylogenomic inference (Ronquist et al. 2012; Zhang et al. 2016). Future studies of ant phylogeny should pay attention to bias and incongruence, as ultimately both more data and better models will contribute to a better understanding of ant phylogeny.

Supplementary Material

Data available from the Dryad Digital Repository: https://dx.doi.org/10.5061/dryad.547d7wmhb.

Funding

We acknowledge the University of Florida Research Computing and the Smithsonian Institution High Performance Cluster for providing computational resources and support that have contributed to the

research results reported in this publication. AL gratefully acknowledges the support of the National Science Foundation (DEB 2026772). CSM thanks the National Science Foundation for support (DBI 2210800, DEB 1900357). KN thanks the LAB facilities of the National Museum of Natural History, the Smithsonian Institution OCIO team, and National Science Foundation (DEB 1654829 to Ted Schultz) for support.

Permitting

Ant samples used in this study comply with the regulations for export and exchange of research samples outlined in the Convention of Biology Diversity and the Convention on International Trade in Endangered Species of Wild Fauna and Flora. Collecting permits in Costa Rica were granted by the "ComisiónInstitucional de Biodiversidad" (Institutional Biodiversity Committee, University of Costa Rica; Resolution Number 028); and authorized by the owners of each property. In Madagascar, permits to research, collect, and export arthropods were obtained from the Ministry of the Environment and Forest as part of an ongoing collaboration between the California Academy of Sciences and the Ministry of Environment and Forest, Madagascar National Parks and ParcBotanique et Zoologique de Tsimbazaza. Samples collected as part of Ant Course with permits issued to BLF were issued by: Sabah, Malaysia: Danum Valley Management Committee, Forest Research Centre (FRC) of the Sabah Forest Department and the Research Promotion and Co-Ordination Committee of the Economic Planning Unit, and Sabah Biodiversity Council; Uganda: Uganda National Council of Science and Technology, Uganda Wildlife Authority Makerere University Biological Station; Peru: Gestion Forestal y de Fauna Silvestre; Mozambique: Ministerio Da Terra, Ambiente E Desenvolvimento Rual, Administracoa National Das Areas De Conservacao; Venezuela: Rancho Grande Field Station of the Universidad Central de Venezuela, Maracay, Venezuela.

Acknowledgments

MLB thanks Dominik Schneider for discussions on combining uncertainty from separate analyses. YMZ was supported in part by the Oak Ridge Institute for Science and Education (ORISE) fellowship and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 101024056. Many thanks to Zachary Griebenow, Phil Ward, three anonymous reviewers, and the editors for insightful feedback that improved this manuscript. Thanks to Vilas Brown for helping with voucher data curation.

References

Allio R., Schomaker-Bastos A., Romiguier J., Prosdocimi F., Nabholz B., Delsuc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol. Ecol. Resour. 20(4):892–905. doi:10.1111/1755-0998.13160

Andersen A.N., Majer J.D. 2004. Ants show the way down under: invertebrates as bioindicators in land management. Front. Ecol. Environ. 2(6):291–298. doi:10.1890/1540-9295(2004)002[0292:ASTWDU]2.0.CO;2

Arcila D., Ortí G., Vari R., Armbruster J.W., Stiassny M.L.J., Ko K.D., Sabaj M.H., Lundberg J., Revell L.J., Betancur-R, R. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. Nat. Ecol. Evol. 1(2):1–10. doi:10.1038/s41559-016-0020

Astruc C., Julien J.F., Errard C., Lenoir A. 2004. Phylogeny of ants (Formicidae) based on morphology and DNA sequence data. Mol. Phylogenet. Evol. 31(3):880–893. doi:10.1016/j.ympev.2003.10.024

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. J. Comput. Biol. 19(5):455–477. doi:10.1089/cmb.2012.0021

Baroni Urbani C. 1989. Phylogeny and behavioural evolution in ants, with a discussion of the role of behaviour in evolutionary processes. Ethol. Ecol. Evol. 1:137–168. doi:10.1080/08927014.1989.9525520

Baroni Urbani C., Bolton B., Ward P.S. 1992. The internal phylogeny of ants (Hymenoptera: Formicidae). Syst. Entomol. 17(4):301–329. doi:10.1111/j.1365-3113.1992.tb00553.x

Blaimer B.B., Brady S.G., Schultz T.R., Lloyd M.W., Fisher B.L., Ward P.S. 2015. Phylogenomic methods outperform traditional multilocus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evol. Biol. 15:e271. doi:10.1186/s12862-015-0552-5

Blaimer B.B., Santos B.F., Cruaud A., Gates M.W., Kula R.R., Mikó I., Rasplus J.-Y., Smith D.R., Talamas E.J., Brady S.G., Buffington M.L. 2023. Key innovations and the diversification of Hymenoptera. Nat. Commun. 14(1):1212. doi:10.1038/s41467-023-36868-4

Blaimer B.B., Ward P.S., Schultz T.R., Fisher B.L., Brady S.G. 2018. Paleotropical diversification dominates the evolution of the hyperdiverse ant tribe Crematogastrini (Hymenoptera: Formicidae). Insect Syst. Divers. 2(5):1–14. doi:10.1093/isd/ixy013

Blanchard B.D., Moreau C.S. 2017. Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. Evolution 71(2):315–328. doi:10.1111/evo.13117

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120. doi:10.1093/bioinformatics/btu170

Bollback J.P. 2002. Bayesian model adequacy and choice in phylogenetics. Mol. Biol. Evol. 19(7):1171–1180. doi:10.1093/oxfordjournals.molbev.a004175

Bolton B. 2024. An online catalog of the ants of the world. Available from: https://antcat.org. (accessed April 2024)

Bolton B., Fisher B.L. 2014. The Madagascan endemic myrmicine ants related to *Eutetramorium* (Hymenoptera: Formicidae): Taxonomy of the genera *Eutetramorium* Emery, *Malagidris* nom. N., *Myrmisaraka* gen. N., *Royidris* gen. N., and *Vitsika* gen. N. Zootaxa 3791(1):1–99. doi:10.11646/Zootaxa.3791.1.1

Bonasio R., Zhang G., Ye C., Mutti N.S., Fang X., Qin N., Donahue G., Yang P., Li Q., Li C., Zhang P., Huang Z., Berger S.L., Reinberg D., Wang J., Liebig J. 2010. Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science 329(5995):1068–1071. doi:10.1126/science.1192428

Boomsma J.J., Brady S.G., Dunn R.R., Gadau J., Heinze J., Keller L., Moreau C.S., Sanders N.J., Schrader L., Schultz T.R.; others. 2017. The Global Ant Genomics Alliance (GAGA). Myrmecol. News 25:61–66.

Borowiec M.L. 2016a. AMAS: A fast tool for alignment manipulation and computing of summary statistics. PeerJ 4:e1660. doi:10.7717/PeerJ.1660

Borowiec M.L. 2016b. Generic revision of the ant subfamily Dorylinae (Hymenoptera, Formicidae). ZooKeys 608:1–280. doi:10.3897/ZooKeys.608.9427

Borowiec M.L. 2019b. Convergent evolution of the army ant syndrome and congruence in big-data phylogenetics. Syst. Biol. 68(4):642–656. doi:10.1093/sysbio/syy088
Borowiec M.L. 2019a. Spruceup: Fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. J. Open Source Softw. 4(42):1635. doi:10.21105/joss.01635
Borowiec M.L., Moreau C.S., Rabeling C. 2020. Ants: phylogeny and classification. In: Starr C., editor. Encyclopedia of social insects. Cham, Switzerland: Springer. p. 52–69. doi:10.1007/978-3-319-90306-4_155-1
Borowiec M.L., Rabeling C., Brady S.G., Fisher B.L., Schultz T.R., Ward P.S. 2019. Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. Mol. Phylogenet. Evol. 134:111–121. doi:10.1016/j.ympev.2019.01.024
Bossert S., Pauly A., Danforth B.N., Orr M.C., Murray E.A. 2024. Lessons from assembling UCEs: a comparison of common methods and the case of *Clavinomia* (Halictidae). Mol. Ecol. Resour. 24:e13925. doi:10.1111/1755-0998.13925
Boudinot B.E., Khouri Z., Richter A., Griebenow Z.H., Kamp T. van de, Perrichot V., Barden P. 2022. Evolution and systematics of the Aculeata and kin (Hymenoptera), with emphasis on the ants (Formicoidea: †@@@idae fam. nov., Formicidae). bioRxiv. doi:10.1101/2022.02.20.480183
Brady S.G., Fisher B.L., Schultz T.R., Ward P.S. 2014. The rise of army ants and their relatives: Diversification of specialized predatory doryline ants. BMC Evol. Biol. 14(1):93. doi:10.1186/1471-2148-14-93
Brady S.G., Schultz T.R., Fisher B.L., Ward P.S. 2006. Evaluating alternative hypotheses for the early evolution and diversification of ants. Proc. Natl. Acad. Sci. USA 103:18172–18177. doi:10.1073/pnas.0605858103
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. 2017b. Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. Curr. Biol. 27(7):1019–1025. doi:10.1016/j.cub.2017.03.027
Branstetter M.G., Ješovnik A., Sosa-Calvo J., Lloyd M.W., Faircloth B.C., Brady S.G., Schultz T.R. 2017a. Dry habitats were crucibles of domestication in the evolution of agriculture in ants. Proc. Biol. Sci. 284(1852):20170095. doi:10.1098/rspb.2017.0095
Branstetter M.G., Longino J.T., Reyes-López J.L., Brady S.G., Schultz T.R. 2022. Out of the temperate zone: a phylogenomic test of the biogeographical conservatism hypothesis in a contrarian clade of ants. J. Biogeogr. 49(9):1640–1653. doi:10.1111/jbi.14462
Branstetter M.G., Longino J.T., Ward P.S., Faircloth B.C. 2017c. Enriching the ant tree of life: Enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods Ecol. Evol. 8(6):768–776. doi:10.1111/2041-210x.12742
Bravo G.A., Antonelli A., Bacon C.D., Bartoszek K., Blom M.P.K., Huynh S., Jones G., Knowles L.L., Lamichhaney S., Marcussen T., Morlon H., Nakhleh L.K., Oxelman B., Pfeil B., Schliep A., Wahlberg N., Werneck F.P., Wiedenhoeft J., Willows-Munro S., Edwards S.V. 2019. Embracing heterogeneity: coalescing the tree of life and the future of phylogenomics. PeerJ 7:e6399. doi:10.7717/PeerJ.6399
Brown J.M. 2014. Predictive approaches to assessing the fit of evolutionary models. Syst. Biol. 63(3):289–292. doi:10.1093/sysbio/syu009
Brown J.M., Thomson R.C. 2018. Evaluating model performance in evolutionary biology. Annu. Rev. Ecol. Evol. Syst. 49(1):95–114. doi:10.1146/annurev-ecolsys-110617-062249
Brown W.L. 1954. Remarks on the internal phylogeny and subfamily classification of the family Formicidae. Insectes Soc. 1(1):21–31. doi:10.1007/BF02223148
Bryant D., Hahn M.W. 2020. The concatenation question. In: Scornavacca C., Delsuc F., Galtier N., editors. Phylogenetics in the genomic era. No commercial publisher | Authors open access book. p. 3.4:1–3.4:23. Available from: https://hal.science/hal-02535651
Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner (LBNL-7065E). Berkeley, CA (United States): Lawrence Berkeley National Lab. (LBNL). https://www.osti.gov/biblio/1241166
Cai C. 2024. Ant backbone phylogeny resolved by modelling compositional heterogeneity among sites in genomic data. Commun. Biol. 7(1):1. doi:10.1038/s42003-024-05793-7
Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. 2009. BLAST+: architecture and applications. BMC Bioinf. 10(1):421. doi:10.1186/1471-2105-10-421
Camacho G.P., Franco W., Branstetter M.G., Pie M.R., Longino J.T., Schultz T.R., Feitosa R.M. 2022b. UCE phylogenomics resolves major relationships among ectaheteromorph ants (Hymenoptera: Formicidae: Ectatomminae, Heteroponerinae): a new classification for the subfamilies and the description of a new genus. Insect Syst. Divers. 6(1):5. doi:10.1093/isd/ixab026
Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17(4):540–552. doi:10.1093/oxfordjournals.molbev.a026334
Chomicki G., Ward P.S., Renner S.S. 2015. Macroevolutionary assembly of ant/plant symbioses: *Pseudomyrmex* ants and their ant-housing plants in the Neotropics. Proc. Biol. Sci. 282(1819):20152200. doi:10.1098/rspb.2015.2200
Cloutier A., Sackton T.B., Grayson P., Clamp M., Baker A.J., Edwards S.V. 2019. Whole-genome analyses resolve the phylogeny of flightless birds (Palaeognathae) in the presence of an empirical anomaly zone. Syst. Biol. 68(6):937–955. doi:10.1093/sysbio/syz019
Cock P.J.A., Antao T., Chang J.T., Chapman B.A., Cox C.J., Dalke A., Friedberg I., Hamelryck T., Kauff F., Wilczynski B., Hoon M. J. L. de. 2009. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11):1422–1423. doi:10.1093/bioinformatics/btp163
Cruaud A., Delvare G., Nidelet S., Sauné L., Ratnasingham S., Chartois M., Blaimer B.B., Gates M., Brady S.G., Faure S., van Noort S., Rossi J.-P., Rasplus J.-Y. 2021. Ultra-conserved elements and morphology reciprocally illuminate conflicting phylogenetic hypotheses in Chalcididae (Hymenoptera, Chalcidoidea). Cladistics 37(1):1–35. doi:10.1111/cla.12416
Cunha T.J., Reimer J.D., Giribet G. 2022. Investigating sources of conflict in deep phylogenomics of vetigastropod snails. Syst. Biol. 71(4):1009–1022. doi:10.1093/sysbio/syab071
Degnan J.H., DeGiorgio M., Bryant D., Rosenberg N.A. 2009. Properties of consensus methods for inferring species trees from gene trees. Syst. Biol. 58(1):35–54. doi:10.1093/sysbio/syp008
Dietterich T.G. 2000. Ensemble methods in machine learning. In: Multiple classifier systems. MCS 2000. Lecture notes in computer science, 1857. Berlin, Heidelberg, Germany: Springer. doi:10.1007/3-540-45014-9_1
Dlussky G.M., Fedoseeva E.B. 1988. Origin and early stages of evolution in ants. [In Russian.]. In: Cretaceous biocenotic crisis and insect evolution. [In Russian.]. Moscow, Russia: Nauka. p. 70–144.
Drew B.T., Ruhfel B.R., Smith S.A., Moore M.J., Briggs B.G., Gitzendanner M.A., Soltis P.S., Soltis D.E. 2014. Another look at the root of the angiosperms reveals a familiar tale. Syst. Biol. 63(3):368–382. doi:10.1093/sysbio/syt108
Economo E.P., Narula N., Friedman N.R., Weiser M.D., Guénard B. 2018. Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nat. Commun. 9(1):1778. doi:10.1038/s41467-018-04218-4
Edelman N.B., Frandsen P.B., Miyagi M., Clavijo B., Davey J., Dikow R.B., García-Accinelli G., Van Belleghem S.M., Patterson N., Neafsey D.E., Challis R., Kumar S., Moreira G.R.P., Salazar C., Chouteau M., Counterman B.A., Papa R., Blaxter M., Reed R.D., Dasmahapatra K.K., Kronforst M., Joron M., Jiggins C.D., McMillan W.O., Di Palma F., Blumberg A.J., Wakeley J., Jaffe D., Mallet J. 2019. Genomic architecture and introgression shape a butterfly radiation. Science 366(6465):594–599. doi:10.1126/science.aaw2090
Edwards S.V. 2009a. Is a new and general theory of molecular systematics emerging? Evolution 63(1):1–19. doi:10.1111/j.1558-5646.2008.00549.x
Edwards S.V. 2009b. Natural selection and phylogenetic analysis. Proc. Natl. Acad. Sci. USA 106(22):8799–8800. doi:10.1073/pnas.0904103106
Elst T. van, Eriksson T.H., Gadau J., Johnson R.A., Rabeling C., Taylor J.E., Borowiec M.L. 2021. Comprehensive phylogeny of

Myrmecocystus honey ants highlights cryptic diversity and infers evolution during aridification of the American Southwest. Mol. Phylogenet. Evol. 155:107036. doi:10.1016/j.ympev.2020.107036
Faircloth B.C. 2011. Illumiprocessor—software for Illumina read quality filtering. https://github.com/faircloth-lab/illumiprocessor. (accessed April 2024). doi:10.6079/J9ILL
Faircloth B.C. 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32(5):786–788. doi: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 Hymenoptera. Mol. Ecol. Resour. 15(3):489–501. doi:10.1111/1755-0998.12328
Faircloth B.C., Glenn T.C. 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. PLoS One 7(8):e42543. doi:10.1371/journal.pone.0042543
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. Syst. Biol. 61(5):717–726. doi:10.1093/sysbio/sys004
Folgarait P.J. 1998. Ant biodiversity and its relationship to ecosystem functioning: a review. Biodivers. Conserv. 7:1221–1244. doi:10.1023/A:1008891901953
Foster P.G. 2004. Modeling compositional heterogeneity. Syst. Biol. 53(3):485–495. doi:10.1080/10635150490445779
Giarla T.C., Esselstyn J.A. 2015. The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomics of Philippine shrews. Syst. Biol. 64(5):727–740. doi:10.1093/sysbio/syv029
Griebenow Z. 2020. Delimitation of tribes in the subfamily Leptanillinae (Hymenoptera: Formicidae), with a description of the male of *Protanilla lini* Terayama, 2009. Myrmecol. News 30:229–250. doi:10.25849/MYRMECOL.NEWS_030:229
Griebenow Z. 2024. Systematic revision of the ant subfamily Leptanillinae (Hymenoptera, Formicidae). ZooKeys 1189:83–184. doi:10.3897/ZooKeys.1189.107506
Grimaldi D., Agosti D., Carpenter J.M. 1997. New and rediscovered primitive ants (Hymenoptera: Formicidae) in Cretaceous amber from New Jersey, and their phylogenetic relationships. Am. Mus. Novit. 3208:1–43. Available from: http://hdl.handle.net/2246/3615
Guang A., Zapata F., Howison M., Lawrence C.E., Dunn C.W. 2016. An integrated perspective on phylogenetic workflows. Trends Ecol. Evol. 31(2):116–126. doi:10.1016/j.tree.2015.12.007
Guénard B., Weiser M.D., Dunn R.R. 2012. Global models of ant diversity suggest regions where new discoveries are most likely are under disproportionate deforestation threat. Proc. Natl. Acad. Sci. USA 109(19):7368–7373. doi:10.1073/pnas.1113867109
Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59(3):307–321. doi:10.1093/sysbio/syq010
Heath T.A., Huelsenbeck J.P., Stadler T. 2014. The fossilized birth-death process for coherent calibration of divergence-time estimates. Proc. Natl. Acad. Sci. USA 111(29):E2957–E2966. doi:10.1073/pnas.1319091111
Hoang D.T., Chernomor O., Haeseler A. von, Minh B.Q., Vinh L.S. 2017. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35(2):518–522. doi:10.1093/molbev/msx281
Hoffmann B.D. 2010. Using ants for rangeland monitoring: global patterns in the responses of ant communities to grazing. Ecol. Indic. 10(2):105–111. doi:10.1016/j.ecolind.2009.04.016
Holder M.T., Sukumaran J., Lewis P.O. 2008. A justification for reporting the majority-rule consensus tree in Bayesian phylogenetics. Syst. Biol. 57(5):814–821. doi:10.1080/10635150802422308
Hölldobler B., Wilson E.O. 1990. The ants. Cambridge (MA): Belknap Press of Harvard University Press.
Huson D.H., Klöpper T., Lockhart P.J., Steel M.A. 2005. Reconstruction of reticulate networks from gene trees. In: Miyano S., Mesirov J., Kasif S., Istrail S., Pevzner P. A., Waterman M., editors. Research in computational molecular biology. Berlin, Heidelberg, Germany: Springer. p. 233–249. doi:10.1007/11415770_18
Huson D.H., Rupp R., Scornavacca C. 2010. Phylogenetic networks: concepts, algorithms and applications. Cambridge (MA): Cambridge University Press. doi:10.1017/CBO9780511974076
Jermiin L., Ho S.Y., Ababneh F., Robinson J., Larkum A.W. 2004. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. Syst. Biol. 53(4):638–643. doi:10.1080/10635150490468648
Ješovnik A., Sosa-Calvo J., Lloyd M.W., Branstetter M.G., Fernandez F., Schultz T.R. 2017. Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (Hymenoptera: Formicidae): ultraconserved elements (UCEs) resolve a recent radiation. Syst. Entomol. 42(3):523–542. doi:10.1111/syen.12228
Junier T., Zdobnov E.M. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. Bioinformatics 26(13):1669–1670. doi:10.1093/bioinformatics/btq243
Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A., Jermiin L.S. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14(6):587–589. doi:10.1038/nmeth.4285
Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30(4):772–780. doi:10.1093/molbev/mst010
Knyshov A., Gordon E.R.L., Masonick P.K., Castillo S., Forero D., Hoey-Chamberlain R., Hwang W.S., Johnson K.P., Lemmon A.R., Moriarty Lemmon E., Standring S., Zhang J., Weirauch C. 2023. Chromosome-aware phylogenomics of assassin bugs (Hemiptera: Reduvioidea) elucidates ancient gene conflict. Mol. Biol. Evol. 40(8):msad168. doi:10.1093/molbev/msad168
Kück P., Garcia F.H., Misof B., Meusemann K. 2011. Improved phylogenetic analyses corroborate a plausible position of *Martialis heureka* in the ant tree of life. PLoS One 6(6):e21031. doi:10.1371/journal.pone.0021031
Lach L., editor. 2010. Ant ecology. New York: Oxford University Press.
Lanfear R., Hahn M. 2024. The meaning and measure of concordance factors in phylogenomics. EcoEvoRxiv. doi:10.32942/X27617
Li Y., Liu Z., Shi P., Zhang J. 2010. The hearing gene Prestin unites echolocating bats and whales. Curr. Biol. 20(2):R55–R56. doi:10.1016/j.cub.2009.11.042
Li Y., Shen X.-X., Evans B., Dunn C.W., Rokas A. 2021a. Rooting the animal tree of life. Mol. Biol. Evol. 38(10):4322–4333. doi:10.1093/molbev/msab170
Longino J.T., Branstetter M.G. (2020). Phylogenomic species delimitation, taxonomy, and 'bird guide' Identification for the Neotropical ant genus *Rasopone* (Hymenoptera: Formicidae). Insect Systematics and Diversity 4(2):1–33. doi:10.1093/isd/ixaa004
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 46(3):523–536. doi:10.1093/sysbio/46.3.523
Mallet J., Besansky N., Hahn M.W. 2016. How reticulated are species? Bioessays 38(2):140–149. doi:10.1002/bies.201500149
McCormack J.E., Faircloth B.C., Crawford N.G., Gowaty P.A., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22(4):746–754. doi:10.1101/gr.125864.111
McDade L. 1990. Hybrids and phylogenetic systematics I. Patterns of character expression in hybrids and their implications for cladistic analysis. Evolution 44(6):1685–1700. doi:10.1111/j.1558-5646.1990.tb03856.x
McGlynn T.P. 1999. The worldwide transfer of ants: geographical distribution and ecological invasions. J. Biogeogr. 26(3):535–548. doi:10.1046/j.1365-2699.1999.00310.x
Minh B.Q., Hahn M.W., Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37(9):2727–2733. doi:10.1093/molbev/msaa106
Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020b. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37(5):1530–1534. doi:10.1093/molbev/msaa015
Mirarab S., Rivas-González I., Feng S., Stiller J., Fang Q., Mai U., Hickey G., Chen G., Brajuka N., Fedrigo O., Formenti G., Wolf J.B.W., Howe K., Antunes A., Schierup M.H., Paten B., Jarvis E.D.,

Zhang G., Braun E.L. 2024. A region of suppressed recombination misleads neoavian phylogenomics. Proc. Natl. Acad. Sci. USA 121(15):e2319506121. doi:10.1073/pnas.2319506121
Mongiardino Koch N. 2021. Phylogenomic subsampling and the search for phylogenetically reliable loci. Mol. Biol. Evol. 38(9):4025–4038. doi:10.1093/molbev/msab151
Mongiardino Koch N., Thompson J.R., Hiley A.S., McCowin M.F., Armstrong A.F., Coppard S.E., Aguilera F., Bronstein O., Kroh A., Mooi R., Rouse G.W. 2022. Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. eLife 11:e72460. doi:10.7554/eLife.72460
Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A., Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. Syst. Biol. 70(2):219–235. doi:10.1093/sysbio/syaa066
Moreau C.S. 2008. Unraveling the evolutionary history of the hyperdiverse ant genus *Pheidole* (Hymenoptera: Formicidae). Mol. Phylogenet. Evol. 48(1):224–239. doi:10.1016/j.ympev.2008.02.020
Moreau C.S. 2014. A practical guide to DNA extraction, PCR, and gene-based DNA sequencing in insects. Halteres 5:32–42. Available from: https://www.biodiversitylibrary.org/page/49900760
Moreau C.S., Bell C.D. 2013. Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. Evolution 67(8):2240–2257. doi:10.1111/evo.12105
Moreau C.S., Bell C.D., Vila R., Archibald S.B., Pierce N.E. 2006. Phylogeny of the ants: Diversification in the age of angiosperms. Science 312:101–104. doi:10.1126/science.1124891
Naser-Khdour S., Minh B.Q., Zhang W., Stone E.A., Lanfear R. 2019. The prevalence and impact of model violations in phylogenetic analysis. Genome Biol. Evol. 11(12):3341–3352. doi:10.1093/gbe/evz193
Nelsen M.P., Ree R.H., Moreau C.S. 2018. Ant-plant interactions evolved through increasing interdependence. Proc. Natl. Acad. Sci. USA 115(48):12253–12258. doi:10.1073/pnas.1719794115
Ohnishi H., Imai H.T., Yamamoto M.-T. 2003. Molecular phylogenetic analysis of ant subfamily relationship inferred from rDNA sequences. Genes Genet. Syst. 78(6):419–425. doi:10.1266/ggs.78.419
Oliveros C.H., Field D.J., Ksepka D.T., Barker F.K., Aleixo A., Andersen M.J., Alström P., Benz B.W., Braun E.L., Braun M.J., Bravo G.A., Brumfield R.T., Chesser R.T., Claramunt S., Cracraft J., Cuervo A.M., Derryberry E.P., Glenn T.C., Harvey M.G., Hosner P.A., Joseph L., Kimball R.T., Mack A.L., Miskelly C.M., Peterson A.T., Robbins M.B., Sheldon F.H., Silveira L.F., Smith B.T., White N.D., Moyle R.G., Faircloth B.C. 2019. Earth history and the passerine superradiation. Proc. Natl. Acad. Sci. USA 116(16):7916–7925. doi:10.1073/pnas.1813206116
Ouellette G.D., Fisher B.L., Girman D.J. 2006. Molecular systematics of basal subfamilies of ants using 28S rRNA (Hymenoptera: Formicidae). Mol. Phylogenet. Evol. 40(2):359–369. doi:10.1016/j.ympev.2006.03.017
Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35(3):526–528. doi:10.1093/bioinformatics/bty633
Parker J., Kronauer D.J.C. 2021. How ants shape biodiversity. Curr. Biol. 31(19):R1208–R1214. doi:10.1016/j.cub.2021.08.015
Pease J.B., Haak D.C., Hahn M.W., Moyle L.C. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 14(2):e1002379. doi:10.1371/journal.pbio.1002379
Penny D., Hendy M.D. 1985. Testing methods of evolutionary tree construction. Cladistics 1(3):266–278. doi:10.1111/j.1096-0031.1985.tb00427.x
Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., Kozlov A., Podsiadlowski L., Petersen M., Lanfear R., Diez P.A., Heraty J., Kjer K.M., Klopfstein S., Meier R., Polidori C., Schmitt T., Liu S., Zhou X., Wappler T., Rust J., Misof B., Niehuis O. 2017. Evolutionary history of the Hymenoptera. Curr. Biol. 27(7):1013–1018. doi:10.1016/j.cub.2017.01.027
Philippe H., Delsuc F., Brinkmann H., Lartillot N. 2005. Phylogenomics. Annu. Rev. Ecol. Evol. Syst. 36:541–562. doi:10.1146/annurev.ecolsys.35.112202.130205
Philippe H., Vienne D. M. de, Ranwez V., Roure B., Baurain D., Delsuc F. 2017. Pitfalls in supermatrix phylogenomics. Eur. J. Taxon. 283:1–25. doi:10.5852/ejt.2017.283
Prebus M., Rabeling C. 2024. Phylogenomics resolve the systematics and biogeography of the ant tribe Myrmicini and tribal relationships within the hyperdiverse ant subfamily Myrmicinae. bioRxiv. doi:10.1101/2024.08.01.606207
Prebus M.M. 2021. Phylogenomic species delimitation in the ants of the *Temnothorax salvini* group (Hymenoptera: Formicidae): an integrative approach. Syst. Entomol. 46(2):307–326. doi:10.1111/syen.12463
R Core Team. 2021. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing [Computer Software]. Available from: https://www.R-project.org/
Rabeling C., Brown J.M., Verhaagh M. 2008. Newly discovered sister lineage sheds light on early ant evolution. Proc. Natl. Acad. Sci. USA 105(39):14913–14917. doi:10.1073/pnas.0806187105
Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67(5):901–904. doi:10.1093/sysbio/syy032
Reis M., Yang Z. 2011. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Mol. Biol. Evol. 28(7):2161–2172. doi:10.1093/molbev/msr045
Reis M.dos, Álvarez-Carretero S., Yang Z. 2017. MCMCTree Tutorials [Software Documentation]. Available from: http://abacus.gene.ucl.ac.uk/software/MCMCtree.Tutorials.pdf
Revell L.J. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ 12:e16505. doi:10.7717/PeerJ.16505
Richter A., Boudinot B.E., Garcia F.H., Billen J., Economo E.P., Beutel R.G. 2023. Wonderfully weird: The head anatomy of the armadillo ant, *Tatuidris tatusia* (Hymenoptera: Formicidae: Agroecomyrmecinae), with evolutionary implications. Myrmecol. News 33:35–75. doi:10.25849/myrmecol.news_033:035
Richter A., Boudinot B., Yamamoto S., Katzke J., Beutel R.G. 2022. The first reconstruction of the head anatomy of a Cretaceous insect, †*Gerontoformica gracilis* (Hymenoptera: Formicidae), and the early evolution of ants. Insect Syst Divers 6(5):4. doi:10.1093/isd/ixac013
Risch S.J., Carroll C.R. 1982. The ecological role of ants in two Mexican agroecosystems. Oecologia 55(1):114–119. doi:10.1007/BF00386726
Robinson D.F., Foulds L.R. 1981. Comparison of phylogenetic trees. Math. Biosci. 53(1):131–147. doi:10.1016/0025-5564(81)90043-2
Rohland N., Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22(5):939–946. doi:10.1101/gr.128124.111
Romiguier J., Borowiec M.L., Weyna A., Helleu Q., Loire E., La Mendola C., Rabeling C., Fisher B.L., Ward P.S., Keller L. 2022. Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality. Curr. Biol. 32(3):2942–2947.e4. doi:10.1016/j.cub.2022.05.001
Ronquist F., Klopfstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst. Biol. 61(6):973–999. doi:10.1093/sysbio/sys058
Salichos L., Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497(7449):327–331. doi:10.1038/Nature12130
Saux C., Fisher B.L., Spicer G.S. 2004. Dracula ant phylogeny as inferred by nuclear 28S rDNA sequences and implications for ant systematics (Hymenoptera: Formicidae: Amblyoponinae). Mol. Phylogenet. Evol. 33(2):457–468. doi:10.1016/j.ympev.2004.06.017
Sayyari E., Whitfield J.B., Mirarab S. 2018. DiscoVista: Interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. 122:110–115. doi:10.1016/j.ympev.2018.01.019
Schierup M.H., Hein J. 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156(2):879–891. doi:10.1093/genetics/156.2.879
Schmidt C. 2013. Molecular phylogenetics of ponerine ants (Hymenoptera: Formicidae: Ponerinae). Zootaxa 3647:201–250. doi:10.11646/Zootaxa.3647.2.1
Schmidt C.A., Shattuck S.O. 2014. The higher classification of the ant subfamily Ponerinae (Hymenoptera: Formicidae), with a

review of ponerine ecology and behavior. Zootaxa 3817(1):1–242. doi:10.11646/Zootaxa.3817.1.1
Schrempf D., Szöllősi G. 2020. The sources of phylogenetic conflicts. In: Scornavacca C., Delsuc F., Galtier N., editors. Phylogenetics in the genomic era. No commercial publisher | Authors open access book. p. 3.1:1–3.1:23. Available from: https://hal.science/hal-02535482
Schultz T.R., Brady S.G. 2008. Major evolutionary transitions in ant agriculture. Proc. Natl. Acad. Sci. USA 105(14):5435–5440. doi:10.1073/pnas.0711024105
Shen X.-X., Hittinger C.T., Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1(5):0126. doi:10.1038/s41559-017-0126
Shimodaira H., Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16(8):1114–1116. doi:10.1093/oxfordjournals.molbev.a026201
Simion P., Delsuc F., Philippe H. 2020. To what extent current limits of phylogenomics can be overcome? In: Scornavacca C., Delsuc F., Galtier N., editors. Phylogenetics in the genomic era. No commercial publisher | Authors open access book. p. 2.1:1–2.1:34. Available from: https://hal.science/hal-02535366
Simmons M.P., Gatesy J. 2021. Collapsing dubiously resolved gene-tree branches in phylogenomic coalescent analyses. Mol. Phylogenet. Evol. 158:107092. doi:10.1016/j.ympev.2021.107092
Smith C.D., Zimin A., Holt C., Abouheif E., Benton R., Cash E., Croset V., Currie C.R., Elhaik E., Elsik C.G., Fave M.-J., Fernandes V., Gadau J., Gibson J.D., Graur D., Grubbs K.J., Hagen D.E., Helmkampf M., Holley J.-A., Hu H., Ibarraran Viniegra A.S., Johnson B.R., Johnson R.M., Khila A., Kim J.W., Laird J., Mathis K.A., Moeller J.A., Muñoz-Torres M.C., Murphy M.C., Nakamura R., Nigam S., Overson R.P., Placek J.E., Rajakumar R., Reese J.T., Robertson H.M., Smith C.R., Suarez A.V., Suen G., Suhr E.L., Tao S., Torres C.W., van Wilgenburg E., Viljakainen L., Walden K.K.O., Wild A.L., Yandell M., Yorke J.A., Tsutsui N.D. 2011. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proc. Natl. Acad. Sci. USA 108(14):5673–5678. doi:10.1073/pnas.1008617108
Smith S.A., Brown J.W., Walker J.F. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One 13(5):e0197433. doi:10.1371/journal.pone.0197433
Smith S.A., Moore M.J., Brown J.W., Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15(1):150. doi:10.1186/s12862-015-0423-0
Steenwyk J.L., Li Y., Zhou X., Shen X.-X., Rokas A. 2023. Incongruence in the phylogenomics era. Nat. Rev. Genet. 24(12):834–850. doi:10.1038/s41576-023-00620-x
Suh A. 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. Zool. Scr. 45((S1):50–62. doi:10.1111/zsc.12213
Sullender B.W., Johnson M.J. 1998. A preliminary molecular phylogeny for the Formicidae. Social Insects at the Turn of the Millennium Proceedings of XIII Congress of IUSSI; Adelaide, Australia, 29. 460 p.
Sullivan J., Joyce P. 2005. Model selection in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 36:445–466. doi:10.1146/annurev.ecolsys.36.102003.152633
Swofford D.L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and Other Methods) (4.0) [Computer software]. Sunderland (MA): Sinauer Associates.
Tagliacollo V.A., Lanfear R. 2018. Estimating improved partitioning schemes for ultraconserved elements. Mol. Biol. Evol. 35(7):1798–1811. doi:10.1093/molbev/msy069
Taylor R.W. 1978. *Nothomyrmecia macrops*: a living-fossil ant rediscovered. Science 201:979–985. doi:10.1126/science.201.4360.979
Tihelka E., Cai C., Giacomelli M., Lozano-Fernandez J., Rota-Stabelli O., Huang D., Engel M.S., Donoghue P.C.J., Pisani D. 2021. The evolution of insect biodiversity. Curr. Biol. 31(19):R1299–R1311. doi:10.1016/j.cub.2021.08.057
Tuma J., Eggleton P., Fayle T.M. 2020. Ant–termite interactions: an important but under-explored ecological linkage. Biol. Rev. Camb. Philos. Soc. 95(3):555–572. doi:10.1111/brv.12577
Vanderpool D., Minh B.Q., Lanfear R., Hughes D., Murali S., Harris R.A., Raveendran M., Muzny D.M., Hibbins M.S., Williamson R.J., Gibbs R.A., Worley K.C., Rogers J., Hahn M.W. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. PLoS Biol. 18(12):e3000954. doi:10.1371/journal.pbio.3000954
Wang H.-C., Minh B.Q., Susko E., Roger A.J. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst. Biol. 67(2):216–235. doi:10.1093/sysbio/syx068
Ward P.S. 1994. *Adetomyrma*, an enigmatic new ant genus from Madagascar (Hymenoptera: Formicidae), and its implications for ant phylogeny. Syst. Entomol. 19:159–175. doi:10.1111/j.1365-3113.1994.tb00585.x
Ward P.S. 2014. The phylogeny and evolution of ants. Annu. Rev. Ecol. Evol. Syst. 45:23–43. doi:10.1146/annurev-ecolsys-120213-091824
Ward P.S., Blaimer B.B., Fisher B.L. 2016. A revised phylogenetic classification of the ant subfamily Formicinae (Hymenoptera: Formicidae), with resurrection of the genera *Colobopsis* and *Dinomyrmex*. Zootaxa 4072(3):343–357. doi:10.11646/zootaxa.4072.3.4
Ward P.S., Brady S.G. 2003. Phylogeny and biogeography of the ant subfamily Myrmeciinae (Hymenoptera: Formicidae). Invertebr. Syst. 17(3):361–386. doi:10.1071/is02046
Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. 2010. Phylogeny and biogeography of dolichoderine ants: effects of data partitioning and relict taxa on historical inference. Syst. Biol. 59(3):342–362. doi:10.1093/sysbio/syq012
Ward P.S., Brady S.G., Fisher B.L., Schultz T.R. 2015. The evolution of myrmicine ants: phylogeny and biogeography of a hyperdiverse ant clade (Hymenoptera: Formicidae). Syst. Entomol. 40(1):61–81. doi:10.1111/syen.12090
Ward P.S., Downie D.A. 2005. The ant subfamily Pseudomyrmecinae (Hymenoptera: Formicidae): phylogeny and evolution of big-eyed arboreal ants. Syst. Entomol. 30(2):310–335. doi:10.1111/j.1365-3113.2004.00281.x
Ward P.S., Fisher B.L. 2016. Tales of dracula ants: the evolutionary history of the ant subfamily Amblyoponinae (Hymenoptera: Formicidae). Syst. Entomol. 41(3):683–693. doi:10.1111/syen.12186
Wilson E.O. 1971. The insect societies. Cambridge (MA): Harvard University Press.
Wilson E.O., Carpenter F.M., Brown W.L. 1967. The first Mesozoic ants. Science 157:1038–1040. doi:10.1126/science.157.3792.1038
Yamane S., Bui T.V., Eguchi K. 2008. *Opamyrma hungvuong*, a new genus and species of ant related to Apomyrma (Hymenoptera: Formicidae: Amblyoponinae). Zootaxa 1767(1):55–63. doi:10.11646/zootaxa.1767.1.3
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24(8):1586–1591. doi:10.1093/molbev/msm088
Yang Z. 2020. PAML: Phylogenetic analysis by maximum likelihood [Software Documentation]. Available from: http://abacus.gene.ucl.ac.uk/software/pamlDOC.pdf
Yin J., Zhang C., Mirarab S. 2019. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. Bioinformatics 35(20):3961–3969. doi:10.1093/bioinformatics/btz211
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. 19(6):153. doi:10.1186/s12859-018-2129-y
Zhang C., Stadler T., Klopfstein S., Heath T.A., Ronquist F. 2016. Total-evidence dating under the fossilized birth–death process. Syst. Biol. 65(2):228–249. doi:10.1093/sysbio/syv080
Zhao M., Kurtis S.M., White N.D., Moncrieff A.E., Leite R.N., Brumfield R.T., Braun E.L., Kimball R.T. 2023. Exploring conflicts in whole genome phylogenetics: a case study within manakins (Aves: Pipridae). Syst. Biol. 72(1):161–178. doi:10.1093/sysbio/syac062
Zou Z., Zhang J. (2020). The nature and phylogenomic impact of sequence convergence. In: Scornavacca C., Delsuc F., Galtier N., editors. Phylogenetics in the genomic era. No commercial publisher | Authors open access book. p. 4.6:1–4.6:17. Available from: https://hal.science/hal-02536347