Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees

Highlights

- UCE phylogenomics provides a highly resolved phylogeny of the stinging wasps

- Ants are the sister group to bees and apoid wasps

- Bees are nested inside crabronid wasps and sister to Pemphredoninae+Philanthinae

- Outgroup choice and taxon sampling can strongly impact phylogenomic inference

Authors

Michael G. Branstetter,
Bryan N. Danforth, James P. Pitts, ...,
Michael W. Gates, Robert R. Kula,
Seán G. Brady

Correspondence

mgbranstetter@gmail.com

In Brief

Branstetter et al. present a densely sampled phylogeny of the stinging wasps, inferred using UCE phylogenomic data. They confirm that ants are sister to bees and apoid wasps and that bees are specialized crabronid wasps. They also demonstrate that taxon sampling can have a strong impact on phylogenetic results even when using genome-scale data.

CrossMark

CellPress

Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees

Michael G. Branstetter,[1,2,8,9,*] Bryan N. Danforth,[3] James P. Pitts,[4] Brant C. Faircloth,[5] Philip S. Ward,[6] Matthew L. Buffington,[7] Michael W. Gates,[7] Robert R. Kula,[7] and Seán G. Brady[2]

[1]Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112, USA
[2]Department of Entomology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, 10th Street and Constitution Avenue NW, Washington, DC 20560, USA
[3]Department of Entomology, 3119 Comstock Hall, Cornell University, Ithaca, NY 14853, USA
[4]Department of Biology, Utah State University, 5305 Old Main Hill, Logan, UT 84322-5305, USA
[5]Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA
[6]Department of Entomology and Nematology, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA
[7]Systematic Entomology Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, US Department of Agriculture, c/o Department of Entomology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, 10th Street and Constitution Avenue NW, Washington, DC 20560, USA
[8]Lead Contact
[9]Twitter: @bramic21
*Correspondence: mgbranstetter@gmail.com
http://dx.doi.org/10.1016/j.cub.2017.03.027

SUMMARY

The stinging wasps (Hymenoptera: Aculeata) are an extremely diverse lineage of hymenopteran insects, encompassing over 70,000 described species and a diversity of life history traits, including ectoparasitism, cleptoparasitism, predation, pollen feeding (bees [Anthophila] and Masarinae), and eusociality (social vespid wasps, ants, and some bees) [1]. The most well-studied lineages of Aculeata are the ants, which are ecologically dominant in most terrestrial ecosystems [2], and the bees, the most important lineage of angiosperm-pollinating insects [3]. Establishing the phylogenetic affinities of ants and bees helps us understand and reconstruct patterns of social evolution as well as fully appreciate the biological implications of the switch from carnivory to pollen feeding (pollenivory). Despite recent advancements in aculeate phylogeny [4–11], considerable uncertainty remains regarding higher-level relationships within Aculeata, including the phylogenetic affinities of ants and bees [5–7]. We used ultraconserved element (UCE) phylogenomics [7, 12] to resolve relationships among stinging-wasp families, gathering sequence data from >800 UCE loci and 187 samples, including 30 out of 31 aculeate families. We analyzed the 187-taxon dataset using multiple analytical approaches, and we evaluated several alternative taxon sets. We also tested alternative hypotheses for the phylogenetic positions of ants and bees. Our results present a highly supported phylogeny of the stinging wasps. Most importantly, we find unequivocal evidence that ants are the sister group to bees+apoid wasps (Apoidea) and that bees are nested within a paraphyletic Crabronidae. We also demonstrate that taxon choice can fundamentally impact tree topology and clade support in phylogenomic inference.

RESULTS AND DISCUSSION

Phylogenomic Analysis

To resolve relationships among major stinging wasp lineages (superfamilies and families), we employed a phylogenomic approach that combines the targeted enrichment of ultraconserved elements (UCEs) with multiplexed next-generation sequencing (NGS) [12]. The UCE approach relies on DNA and allows for the efficient sequencing of hundreds of loci from both fresh and museum-preserved specimens. We followed published lab protocols [7, 12] (see also the Supplemental Experimental Procedures) and used a Hymenoptera-specific bait set designed to enrich 1,510 UCE loci [7]. We sequenced new molecular data from 139 taxa and combined these with data from 16 taxa previously sequenced [7] and UCE loci harvested from 32 genomes, resulting in a final dataset containing 187 taxa (see Data S1 for sample information).

We included 136 samples of stinging wasps, representing 30 out of 31 recognized families, missing only Scolebythidae. We sampled densely within the bees and apoid wasps (Apoidea), including 53 species from 23 out of 25 recognized bee subfamilies, and 16 species from outside bees, including the phylogenetically enigmatic families Ampulicidae and Heterogynaidae. We also sampled 14 species from four out of eight subfamilies within Crabronidae, including two subfamilies hypothesized to be closely related to bees (Pemphredoninae+Philanthinae) [5]. For outgroups, we sampled all superfamilies within the sawfly grade (Symphyta), and eight out of 12 non-aculeate superfamilies from the Apocrita (Parasitica), including Trigonaloidea, Evanioidea, Ichneumonoidea, and Ceraphronoidea, which previous analyses suggested are closely related to Aculeata [8, 10, 13–15].

Among the taxa from which we sequenced enriched UCE loci, we captured an average of 966 UCE contigs per sample, with a mean contig length of 801 bp and an average coverage per UCE contig of 80× (see Data S1 for assembly information). We evaluated the effects of filtering alignments for various levels of taxon occupancy (percentage of taxa required to be present in a given locus) and selected the 75% filtered locus set ("Hym-187T-F75") as the primary locus set for analysis. This dataset included 854 loci and 203,095 bp of sequence data, of which 143,608 sites were informative (see Data S1 for alignment matrix information).

We analyzed the Hym-187T-F75 dataset using maximum-likelihood (ML; RAxML v8 [16]), Bayesian (BI; ExaBayes v1.4 [17]), and species-tree (ST; ASTRAL-II [18]) approaches. For ML analyses, we compared several different data-partitioning schemes (see Data S1 for more information) and two approaches designed to mitigate phylogenetic error caused by base composition heterogeneity and/or substitution saturation. For the latter approaches, we created one dataset in which we converted the entire matrix to RY coding and one in which we removed loci exhibiting signs of base composition heterogeneity among taxa (47 loci removed). For ST analysis, we employed weighted statistical binning to reduce error from loci with low information content [19].

We recovered a robust phylogeny of the Aculeata, with topologies being nearly identical across all analyses (Figures 1, S1, and S2). We observed topological conflict at eight nodes, with the most important difference concerning relationships among families of the Chrysidoidea (cuckoo wasps and relatives). We recovered the Trigonaloidea as sister to stinging wasps (Aculeata) with maximum support in all analyses. Although we lacked several parasitoid superfamilies in our dataset, this result is congruent with most recent molecular analyses [8, 10, 15]. Importantly, we did not recover the Ichneumonoidea, a long-standing candidate as the sister group to Aculeata [13], to be closely related to the stinging wasps in any analysis. Within Aculeata, we found Chrysidoidea to form a paraphyletic grade, with the clade containing Sclerogibbidae+[Embolemidae+Dryinidae] recovered as the sister group to remaining non-chrysidoid lineages in most analyses. The rest of the aculeate superfamilies divided into two major clades that were each highly supported in all analyses. Overall, relationships among superfamilies largely agree with a recent transcriptome-based study [6], except for the placement of Vespoidea. Ants (Formicidae) were inferred to be the sister group to bees and apoid wasps with maximum support in all analyses, except for the RY-coded ML analysis (96%) and the ST analysis (90%).

Within the clade containing bees and apoid wasps (Apoidea), our results for relationships among families and higher-level clades are identical across analyses and largely agree with those of Debevec et al. [5]. Most significantly, we found the bees (Anthophila) to be nested inside of a paraphyletic Crabronidae and sister to Pemphredoninae+Philanthinae, confirming the finding first reported in [5], which was based on only four molecular markers and received only moderate support. Within Apoidea, we also found Ampulicidae to be the sister group to all other apoid families and Heterogynaidae, a phylogenetically enigmatic family, to be the sister group to Crabroninae+Sphecidae. Among bees (Anthophila), our results are largely congruent with previous studies of higher-level relationships [20]. Most notably, we found

Melittidae to be the sister group to the remaining bee families, and we recovered monophyly of the eusocial corbiculate tribes (Apini, Bombini, and Meliponini) in all concatenated analyses (not recovered in the ST analysis).

Testing the Phylogenetic Position of Ants and Bees

Identifying the phylogenetic positions of ants and bees within the stinging wasps is of critical importance. Ants are ecologically dominant social insects in virtually all terrestrial ecosystems, and bees are among the most important pollinators of the largest lineage of vascular plants on earth—the angiosperms. To evaluate the robustness of our phylogenetic results, we performed two types of analyses. First, we evaluated a range of previous phylogenetic hypotheses for both ants and bees using the Shimodaira-Hasegawa (SH) test [21]. For the SH tests, we analyzed nine alternate positions for ants and 14 alternate positions for bees (Table 1), taking into account previous phylogenetic hypotheses for both taxa (ants [4, 6, 8, 9]; bees [5, 22–25]). We performed the tests with the most taxon-rich 187-taxon dataset and with a taxonomically balanced 100-taxon dataset. In all cases, the alternative topologies were rejected (p < 0.01), providing unequivocal support for the preferred topology presented here. Based on these analyses and our current level of taxon sampling, ants are the sister group to Apoidea, and bees are clearly highly derived crabronid wasps.

Second, we analyzed the impact of taxon sampling on our phylogenetic results. Previous phylogenomic studies based on far fewer taxa than we included here have obtained conflicting results regarding the placement of ants vis-à-vis Apoidea. The transcriptome-based study of Johnson et al. [6] found ants to be sister to Apoidea (the result here), whereas the UCE-based study of Faircloth et al. [7] found ants to be sister to all other aculeates, with the exception of Chrysidoidea, which was not included in their analyses. We divided the taxon-sampling experiments into the following categories (Figure 2): (1) variations of Johnson et al. [6], (2) variations of Faircloth et al. [7], and (3) variations of the current taxon set. In the first category, we generated two datasets, one with exactly the same taxon sampling as [6] ("Johnson-19T") and one with the chrysidoid *Argochrysis armilla* removed ("Johnson-18T"). We included this particular manipulation because the major difference between [6] and [7] was the presence or absence of Chrysidoidea, which is the sister taxon to all other aculeate groups. For the Faircloth et al. [7] manipulations, we recreated the original 45-taxon matrix ("Faircloth-45T") and several alternative taxon sets. First, we added a single chrysidoid ("Faircloth-46T"), and then we continued to add additional aculeates to balance taxa across major lineages ("Faircloth-52T," "Faircloth-56T," and "Faircloth-61T"). We also balanced the dataset by removing excessive ant taxa from the original dataset ("Faircloth-26T") and then adding in a chrysidoid ("Faircloth-27T"). For the third category, we generated a dataset with most outgroups removed ("Hym-147T"), leaving *Nasonia* as the earliest diverging outgroup and *Megaspilus* (Ceraphronoidea), Evanioidea, and Trigonaloidea as more recently diverging outgroups. From this taxon set, we removed chrysidoids ("Hym-133T") and chrysidoids plus trigonaloids ("Hym-131T"). We also created what we considered to be the most balanced dataset by removing excessive ant, bee, and wasp taxa ("Hym-100T").
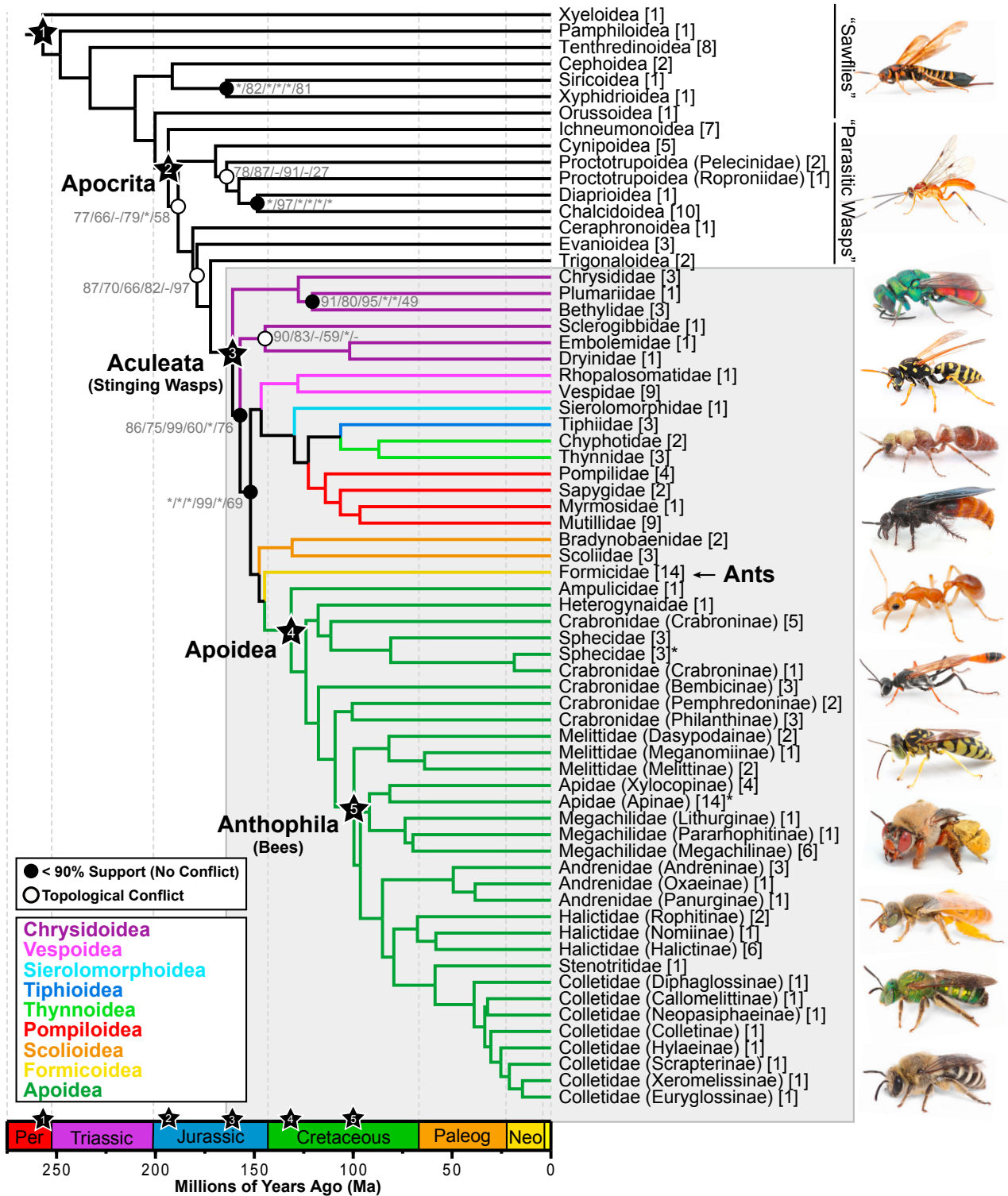
Figure 1. Dated Phylogeny of Aculeate Wasps and Outgroups

We inferred the topology by analyzing the Hym-187T-F75 matrix in RAxML (partitioned by k-means algorithm; 854 loci; 203,095 bp of sequence data) and estimated the dates in BEAST (50 random loci; fixed topology; 38 calibration points). Black dots indicate nodes that were recovered in all analyses but that received <90% support in at least one analysis. White dots indicate nodes with some topological conflict among analyses. Support values are provided for six analyses and are given in the following order: raxml-rcluster/raxml-kmeans/raxml-ry-coding/raxml-bcomp/exabayes-kmeans/astral. The asterisk and dash indicate 100% and 0% support, respectively. An asterisk by a terminal taxon name indicates paraphyly, and bracketed numbers indicate the number of samples. Sawfly, parasitoid wasp, and ant images are ©Alex Wild, used with permission. All other images are ©Joseph S. Wilson, used with permission. See also Figures S1 and S2 and Data S1.

Table 1. Results from SH Tests Comparing Our Favored Placement of Bees and Ants with 14 Alternative Positions for Bees and Nine Alternative Positions for Ants

| Tree	Position of Bees	100-Taxon Matrix				187-Taxon Matrix			
		Likelihood	D(LH)	SD	Significance	Likelihood	D(LH)	SD	Significance
Best tree	bees + [Pemphredoninae+Philanthinae]	−8,889,479	NA	NA	NA	−10,006,566	NA	NA	NA
Alt. tree 1	bees + Philanthinae	−8,889,755	−276	40	<0.01	−10,006,870	−304	38	<0.01
Alt. tree 2	bees + Pemphredoninae	−8,889,790	−311	38	<0.01	−10,006,904	−338	36	<0.01
Alt. tree 3	bees + Crabronidae (excl. Crabroninae)	−8,890,408	−928	74	<0.01	−10,007,072	−506	61	<0.01
Alt. tree 4	bees + Bembicinae	−8,890,609	−1,130	66	<0.01	−10,007,264	−698	53	<0.01
Alt. tree 5	bees + apoid wasps (excl. Ampulicidae)	−8,891,776	−2,297	117	<0.01	−10,007,885	−1,319	96	<0.01
Alt. tree 6	bees + [Heterogynaidae+[Crabroninae+ Sphecidae]]	−8,891,895	−2,416	113	<0.01	−10,008,008	−1,442	92	<0.01
Alt. tree 7	bees + Crabronidae (incl. Sphecidae)	−8,892,128	−2,649	121	<0.01	−10,008,220	−1,654	101	<0.01
Alt. tree 8	bees + Heterogynaidae	−8,892,298	−2,819	122	<0.01	−10,008,354	−1,788	101	<0.01
Alt. tree 9	bees + [Crabroninae+Sphecidae]	−8,892,303	−2,824	123	<0.01	−10,008,352	−1,786	102	<0.01
Alt. tree 10	bees + Crabronidae (excl. Sphecidae)	−8,892,355	−2,876	129	<0.01	−10,008,517	−1,951	110	<0.01
Alt. tree 11	bees + Crabroninae	−8,892,953	−3,474	140	<0.01	−10,008,952	−2,386	120	<0.01
Alt. tree 12	bees + Sphecidae	−8,892,966	−3,487	140	<0.01	−10,008,974	−2,408	119	<0.01
Alt. tree 13	bees + all apoid wasps	−8,893,198	−3,719	140	<0.01	−10,008,819	−2,253	115	<0.01
Alt. tree 14	bees + Ampulicidae	−8,893,210	−3,731	139	<0.01	−10,008,838	−2,272	114	<0.01
Tree	Position of Ants	Likelihood	D(LH)	SD	Significance	Likelihood	D(LH)	SD	Significance
Best tree	ants + Apoidea	−8,889,479	NA	NA	NA	−10,006,566	NA	NA	NA
Alt. tree 1	ants + [Scolioidea+Apoidea]	−8,889,710	−231	39	<0.01	−10,006,715	−149	32	<0.01
Alt. tree 2	ants + Scolioidea	−8,889,720	−241	37	<0.01	−10,006,677	−111	33	<0.01
Alt. tree 3	ants + Aculeata (excl. Chrysidoidea)	−8,890,755	−1,276	86	<0.01	−10,007,370	−804	73	<0.01
Alt. tree 4	ants + [Vespoidea+Tiphioidea]	−8,890,868	−1,389	83	<0.01	−10,007,479	−913	69	<0.01
Alt. tree 5	ants + Vespoidea	−8,891,195	−1,716	94	<0.01	−10,007,756	−1,191	81	<0.01
Alt. tree 6	ants + Tiphioidea	−8,891,245	−1,766	93	<0.01	−10,007,781	−1,215	81	<0.01
Alt. tree 7	ants + [Vespoidea+Scolioidea]	−8,891,453	−1,974	108	<0.01	−10,007,901	−1,335	90	<0.01
Alt. tree 8	ants + Aculeata	−8,892,268	−2,789	126	<0.01	−10,008,670	−2,104	113	<0.01
Alt. tree 9	ants + [Tiphioidea+Scolioidea]	−8,897,047	−7,568	205	<0.01	−10,011,698	−5,132	178	<0.01

We performed the analyses unpartitioned using the complete 187-taxon matrix and a taxonomically balanced 100-taxon matrix. Our favored topology was significantly better than the alternatives in all cases. D(LH), difference in likelihood scores; alt., alternative; excl., excluding; incl., including; NA, not applicable.

The results of the taxon-sampling experiments (Table 2) support the conclusion that, even with genome-scale data, both outgroup choice and taxonomic balance impact phylogenetic results. The Faircloth et al. [7] study suffered from both of these issues, and we suspect that the trees obtained in that study are incorrect with regard to the position of ants. Focusing on the placement of ants (Formicoidea: Formicidae), we recovered three alternative topologies (Figure 2; Table 2): ants sister to Apoidea (topology A); ants sister to all other groups, minus Chrysidoidea (topology B); and ants sister to Apoidea plus Scolioidea (topology C). In both of the Johnson et al. [6] matrices, we recovered topology A. Analysis of the original Faircloth et al. [7] taxon set (Faircloth-45T) produced topology B, as in the original study. For Faircloth-46T, Faircloth-52T, and Faircloth-56T, we also recovered topology B. However, in the Faircloth-61T analyses, the topology shifted to C, placing ants as sister to Scolioidea plus Apoidea. The difference between Faircloth-56T and Faircloth-61T was the addition of several chrysidoids (Embolemidae and Dryinidae), Rhopalosomatidae (Vespoidea), and Ampulicidae (Apoidea), with the latter two taxa breaking long branches. Reducing and balancing the taxa of Faircloth-45T also altered the resulting topology. Reducing the number of ant taxa from 22 in Faircloth-45T to three taxa in Faircloth-26T changed the topology to A. The Hym-147T matrix and variants (Hym-133T, Hym-131T, and Hym-100T) also produced topology A. For the Hym-100T matrix, in which we reduced the number of ant and bee taxa to balance the larger taxon set, all relationships, in addition to the placement of ants, were the same as those in the ML analysis of the Hym-187T matrix. In addition to the topological differences just described, removing outgroups from matrices (chrysidoids or trigonaloids) usually resulted in decreased bootstrap scores for the position of ants (Table 2).

Biological Implications

Our results resolve long-standing debates in aculeate phylogeny and provide a solid framework for understanding both the importance of pollenivory as a driver of bee diversification and the importance of eusociality as a driver of ant diversification and ecological dominance. Ants and bees are surprisingly closely related, which impacts how we view the evolution of important
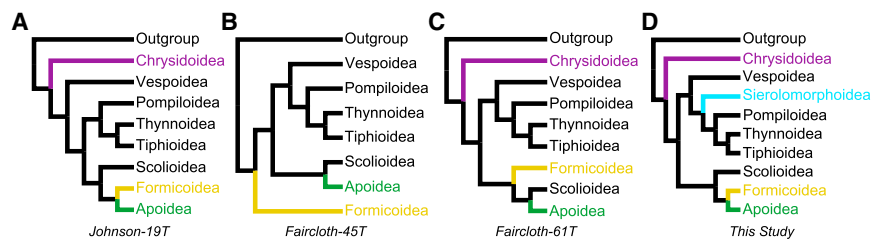
Figure 2. Alternative Hypotheses for Relationships among Aculeate Superfamilies
Topology from Johnson et al. [6] (A), topology from Faircloth et al. [7] (B), topology from analysis of the Faircloth-*61T* alignment supermatrix (C), and preferred topology inferred in this study (includes Sierolomorphoidea) (D). Topologies correspond to those reported in Table 2, except that topologies A and D are equivalent in terms of ants being sister to Apoidea.

behaviors, such as nest building, central place foraging, and eusociality in Aculeata [6]. It is important to highlight the fact that eusociality has evolved once at the origin of ants and at least six to eight times within bees [20], which means that the clade containing ants and bees may be particularly pre-disposed to becoming social. As discussed in [6], understanding the biology of all of the lineages within the Apoidea (bees and apoid wasps) will provide new insights into the biological factors that promote the evolution of social behavior.

Our results largely corroborate previous findings regarding relationships within Apoidea [5] and bees [20]. We confirm the placement of Ampulicidae as sister to the remaining Apoidea and the placement of bees as sister to the crabronid subfamilies Philanthinae+Pemphredoninae. The close affinities of bees to the crabronid subfamilies Philanthinae and Pemphredoninae have been suggested previously in studies based both on morphological and molecular data (reviewed in [5]). Philanthinae include ground-nesting wasps that hunt a variety of prey, including beetles, ants, and, ironically, bees. Pemphredoninae include small, mostly cavity-nesting wasps that hunt diverse prey, including Collembola (springtails), Thysanoptera (thrips), and an array of plant-feeding Hemiptera (aphids, scales, psyllids, cicadellids, cercopids, and membracids). Together, Pemphredoninae and Philanthinae comprise just over 2,200 described species [26]. That the bees, with over 20,000 described species, are sister to a group of just 2,200 hunting wasp species would suggest that the switch from predation to pollenivory was a significant driver of diversification in bees. Future studies should include an even broader sampling of Pemphredoninae and Philanthinae to test this hypothesis.

Within bees, our results provide further confirmation that Melittidae, previously thought to be sister to long-tongued bees (Apidae+Megachilidae) based on morphology [27], is monophyletic and sister to the remaining bee families. Family-level relationships in bees are fully congruent with previous studies [20]. It is notable that most of our analyses recovered the eusocial corbiculate bees (honeybees, bumblebees, and stingless bees) as monophyletic and sister to the weakly social Euglossini (orchid bees), thus favoring a single origin of eusociality within the group. Relationships among these taxa have been controversial, but our result agrees with a recent phylogenomic study that found that controlling for base-compositional heterogeneity favored monophyly of eusocial corbiculates [28].

Conclusions

The coupling of NGS with reduced representation phylogenomics has driven a revolution in molecular systematics, making it possible to generate large datasets at a fraction of the cost of traditional methods [29, 30]. Here, we further applied one promising approach, the targeted enrichment of UCEs [12], to the megadiverse insect order Hymenoptera, greatly extending a previous study that first employed the UCE method in arthropods [7]. We focused on family-level relationships of the stinging wasps (Aculeata) and produced a robust backbone phylogeny that provides many insights into the evolutionary history of this group. In addition, by carrying out a series of taxon-sampling experiments, we have demonstrated that even in the era of phylogenomics, careful taxon sampling can be of critical importance, with both outgroup choice and taxon evenness having a significant impact on topology and bootstrap support.

EXPERIMENTAL PROCEDURES

UCE Sequencing Workflow and Bioinformatics
The protocols for generating UCE data followed those reported in Faircloth et al. [7] and are described in detail in the Supplemental Experimental Procedures. For newly sampled taxa, we performed the following steps: DNA extraction, library preparation, sample pooling, UCE enrichment, enrichment verification, final pooling, and Illumina sequencing. For all bioinformatics steps, from read cleaning to alignment, we used the PHYLUCE v1.5 software package [31].

Phylogenomic Analyses of the Complete Taxon Set
Using the Hym-187T-F75 locus set, we carried out ML and BI analyses on the concatenated matrix with the programs RAxML v8 [16] and ExaBayes v1.4 [17], respectively (additional analytical details are in the Supplemental Experimental Procedures). For ML searches, we compared the following partitioning schemes: (1) unpartitioned, (2) partitioned by locus, (3) partitioned with the hcluster algorithm in PartitionFinder v1 [32] (data pre-partitioned by locus), (4) partitioned with the rcluster algorithm in PartitionFinder v2 [33], and (5) partitioned with the k-means algorithm [34] in PartitionFinder v2. For the BI analysis, we used the k-means partitioning scheme because this scheme resulted in the highest log likelihood in ML analyses (see Data S1 for partitioning results). For ST analysis, we used the summary method implemented in ASTRAL-II v4.8.0 [18] and employed weighted statistical binning [19], which was developed to reduce ST inference error caused by the inclusion of loci that have few informative sites.

To further test our results and to remove potential data biases, we carried out two additional analyses on the Hym-187T-75T locus set using RAxML. For the first analysis, we converted the concatenated matrix to RY coding, and for the second analysis, we used the program BaCoCa v1.1 [35] to identify and remove loci deviating significantly (p < 0.01) from base composition homogeneity among taxa. The latter analysis identified 47 offending loci, leaving 807 loci for concatenation and analysis ("Hym-187T-F75-BComp").

Divergence Dating
We employed node dating and used the program BEAST v1.8.2 [36] to estimate divergence dates on the complete 187-taxon tree. To calibrate the analysis, we used 37 fossils representing taxa from across Hymenoptera and one secondary calibration (252 Ma; taken from [37]) for the root node (see Data S1 for calibration information). For Aculeata, we selected a subset of the fossils used in two recent molecular studies [38, 39]. Importantly, we included the oldest known fossils for bees (*Melittosphex burmensis*), ants (*Haidomyrmex*

Table 2. Results of the Taxon Inclusion/Exclusion Experiments as Evidenced by Topological and Bootstrap Support Differences

Taxon Set	Topology	BS (Ants+Sister Group)	Outgroup	Notes
Johnson-18T	A	89	no chrysidoid	—
Johnson-19T	A	100	—	same taxon set as in [6]
Faircloth-26T	A	88	no chrysidoid	—
Faircloth-27T	A	97	—	—
Faircloth-45T	B	100	no chrysidoid	same taxon set as in [7]
Faircloth-46T	B	99	—	—
Faircloth-52T	B	98	—	—
Faircloth-56T	B	100	—	—
Faircloth-61T	C	100	—	—
Hym-100T	A	100	—	most balanced taxon set
Hym-131T	A	90	no chrysidoid or trigonaloid	—
Hym-133T	A	100	no chrysidoid	—
Hym-147T	A	100	—	—
Hym-187T-F75	A	100	—	this study

The results suggest that both outgroup choice (chrysidoid presence/absence) and taxon evenness can affect outcomes. The matrix name indicates whether the taxon set is a version of Johnson et al. [6], Faircloth et al. [7], or this study (Hym-). Three different topologies were recovered: ants sister to Apoidea (A); ants sister to all other aculeate superfamilies, except Chrysidoidea (B); and ants sister to Apoidea+Scoliodea (C). Bootstrap support indicates support for the clade that includes ants plus its sister group. Topologies correspond to those shown in Figures 1A–1C, in relation to the position of ants only. BS, bootstrap support.

cerberus and *Kyromyrma neffi*), Apoidea (Angarosphecidae), and Aculeata (*Sclerogibbodes embioleia*). We used the Fossilworks database [40] (http://www.fossilworks.org) to date fossils, and we followed best practices for node dating [41]. To decrease computation time in BEAST, we used a constraint topology, and we tested three different locus sets: (1) 25 best loci (where ''best'' indicates the highest mean gene-tree bootstrap score), (2) 50 best loci, and (3) 50 randomly selected loci. For additional details, see the Supplemental Experimental Procedures.

ACCESSION NUMBERS

For newly sequenced samples, raw sequence reads and contigs representing UCE loci are available from the NCBI Sequence Read Archive (SRA) and GenBank, respectively (NCBI BioProject: PRJNA379583). Additional data and results, including Trinity assemblies, alignments, alignment supermatrices, and all phylogenetic trees, are available from the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.r8d4q).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and one data file and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2017.03.027.

AUTHOR CONTRIBUTIONS

All authors conceived the ideas and designed methodology, M.G.B. collected and analyzed the data, and M.G.B. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

REFERENCES

1. Grimaldi, D., and Engel, M.S. (2005). Evolution of the Insects (Cambridge University Press).

2. Hölldobler, B., and Wilson, E.O. (1990). The Ants (Belknap Press).

3. Michener, C.D. (2007). The Bees of the World, Second Edition (The Johns Hopkins University Press).

4. Pilgrim, E.M., von Dohlen, C.D., and Pitts, J.P. (2008). Molecular phylogenetics of Vespoidea indicate paraphyly of the superfamily and novel relationships of its component families and subfamilies. Zool. Scr. 37, 539–560.

5. Debevec, A.H., Cardinal, S., and Danforth, B.N. (2012). Identifying the sister group to the bees: a molecular phylogeny of Aculeata with an emphasis on the superfamily Apoidea. Zool. Scr. 41, 527–535.

6. Johnson, B.R., Borowiec, M.L., Chiu, J.C., Lee, E.K., Atallah, J., and Ward, P.S. (2013). Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. Curr. Biol. 23, 2058–2062.

7. Faircloth, B.C., Branstetter, M.G., White, N.D., and Brady, S.G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Resour. 15, 489–501.

8. Heraty, J., Ronquist, F., Carpenter, J.M., Hawks, D., Schulmeister, S., Dowling, A.P., Murray, D., Munro, J., Wheeler, W.C., Schiff, N., and

Sharkey, M. (2011). Evolution of the hymenopteran megaradiation. Mol. Phylogenet. Evol. *60*, 73–88.

9. Brothers, D.J. (1975). Phylogeny and classification of the aculeate Hymenoptera, with special reference to Mutillidae. Univ. Kans. Sci. Bull. *50*, 483–648.

10. Klopfstein, S., Vilhelmsen, L., Heraty, J.M., Sharkey, M., and Ronquist, F. (2013). The hymenopteran tree of life: evidence from protein-coding genes and objectively aligned ribosomal data. PLoS ONE *8*, e69344.

11. Brothers, D.J., and Carpenter, J.M. (1993). Phylogeny of Aculeata: Chrysidoidea and Vespoidea (Hymenoptera). J. Hymenopt. Res. *2*, 227–304.

12. Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., and Glenn, T.C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. *61*, 717–726.

13. Sharkey, M.J. (2007). Phylogeny and classification of Hymenoptera. Zootaxa *548*, 521–548.

14. Sharkey, M.J., Carpenter, J.M., Vilhelmsen, L., Heraty, J., Liljeblad, J., Dowling, A.P.G., Schulmeister, S., Murray, D., Deans, A.R., Ronquist, F., et al. (2012). Phylogenetic relationships among superfamilies of Hymenoptera. Cladistics *28*, 80–112.

15. Castro, L.R., and Dowton, M. (2006). Molecular analyses of the Apocrita (Insecta: Hymenoptera) suggest that the Chalcidoidea are sister to the diaprioid complex. Invertebr. Syst. *20*, 603–614.

16. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

17. Aberer, A.J., Kobert, K., and Stamatakis, A. (2014). ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. *31*, 2553–2556.

18. Mirarab, S., and Warnow, T. (2015). ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics *31*, i44–i52.

19. Bayzid, M.S., Mirarab, S., Boussau, B., and Warnow, T. (2015). Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. PLoS ONE *10*, e0129183.

20. Danforth, B.N., Cardinal, S., Praz, C., Almeida, E.A.B., and Michez, D. (2013). The impact of molecular data on our understanding of bee phylogeny and evolution. Annu. Rev. Entomol. *58*, 57–78.

21. Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. *16*, 1114–1116.

22. Ohl, M., and Bleidorn, C. (2005). The phylogenetic position of the enigmatic wasp family Heterogynaidae based on molecular data, with description of a new, nocturnal species (Hymenoptera: Apoidea). Syst. Entomol. *31*, 321–337.

23. Alexander, B.A. (1992). An exploratory analysis of cladistic relationships within the superfamily Apoidea, with special reference to sphecid wasps (Hymenoptera). J. Hymenopt. Res. *1*, 25–61.

24. Prentice, M.A. (1998). The comparative morphology and phylogeny of apoid wasps (Hymenoptera: Apoidea). PhD thesis (University of California, Berkeley).

25. Melo, G.A.R. (1999). Phylogenetic relationships and classification of the major lineages of Apoidea (Hymenoptera), with emphasis on the crabronid wasps. Sci. Pap. Univ. Kansas Nat. Hist. Mus. *14*, 1–55.

26. Pulawski, W.J. (2017). Catalog of Sphecidae *sensu lato* (=Apoidea excluding Apidae). https://www.calacademy.org/scientists/projects/catalog-of-sphecidae.

27. Roig-Alsina, A., and Michener, C.D. (1993). Studies of the phylogeny and classification of long-tounged bees (Hymenoptera: Apoidea). Univ. Kans. Sci. Bull. *55*, 124–162.

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

29. Lemmon, E.M., and Lemmon, A.R. (2013). High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. *44*, 19.1–19.23.

30. McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., and Brumfield, R.T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. *66*, 526–538.

31. Faircloth, B.C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics *32*, 786–788.

32. Lanfear, R., Calcott, B., Ho, S.Y.W., and Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. *29*, 1695–1701.

33. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2016). PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. Published online December 23, 2016. http://dx.doi.org/10.1093/molbev/msw260.

34. Frandsen, P.B., Calcott, B., Mayer, C., and Lanfear, R. (2015). Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates. BMC Evol. Biol. *15*, 13.

35. Kück, P., and Struck, T.H. (2014). BaCoCa--a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. *70*, 94–98.

36. Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. *29*, 1969–1973.

37. Zhang, C., Stadler, T., Klopfstein, S., Heath, T.A., and Ronquist, F. (2016). Total-evidence dating under the fossilized birth-death process. Syst. Biol. *65*, 228–249.

38. Wilson, J.S., Dohlen, C.D., Forister, M.L., and Pitts, J.P. (2012). Family-level divergences in the stinging wasps (Hymenoptera: Aculeata), with correlations to angiosperm diversification. Evol. Biol. *40*, 101–107.

39. Cardinal, S., and Danforth, B.N. (2013). Bees diversified in the age of eudicots. Proc. Biol. Sci. *280*, 20122686.

40. Alroy, J. (2016). Fossilworks: gateway to the paleobiology database. http://www.fossilworks.org.

41. Magallon, S.A. (2004). Dating lineages: molecular and paleontological approaches to the temporal framework of clades. Int. J. Plant Sci. *165*, S7–S21.

42. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE) (IEEE), pp. 1–8.

Supplemental Information

Phylogenomic Insights into the Evolution

of Stinging Wasps and the Origins of Ants and Bees

Michael G. Branstetter, Bryan N. Danforth, James P. Pitts, Brant C. Faircloth, Philip S. Ward, Matthew L. Buffington, Michael W. Gates, Robert R. Kula, and Seán G. Brady
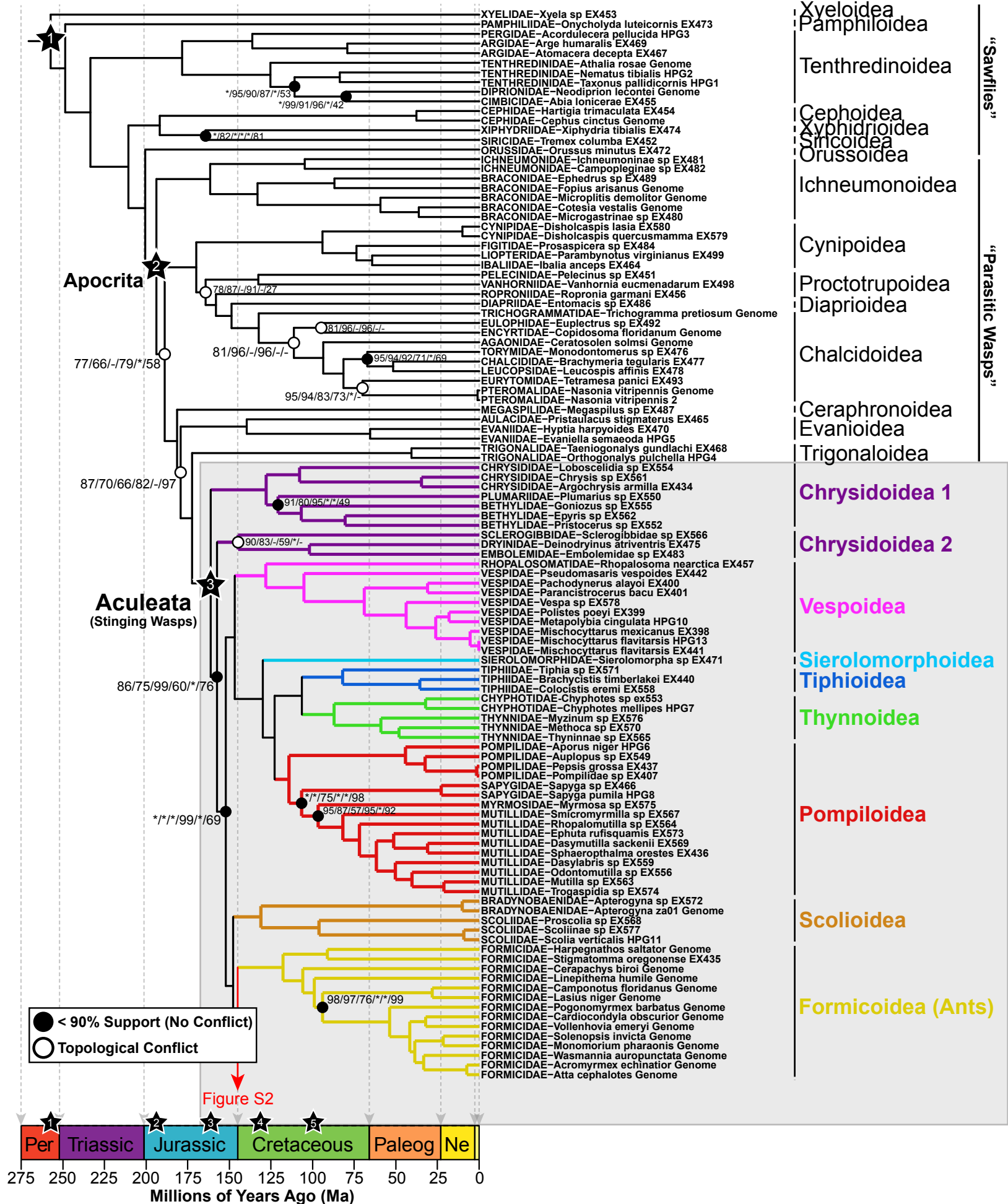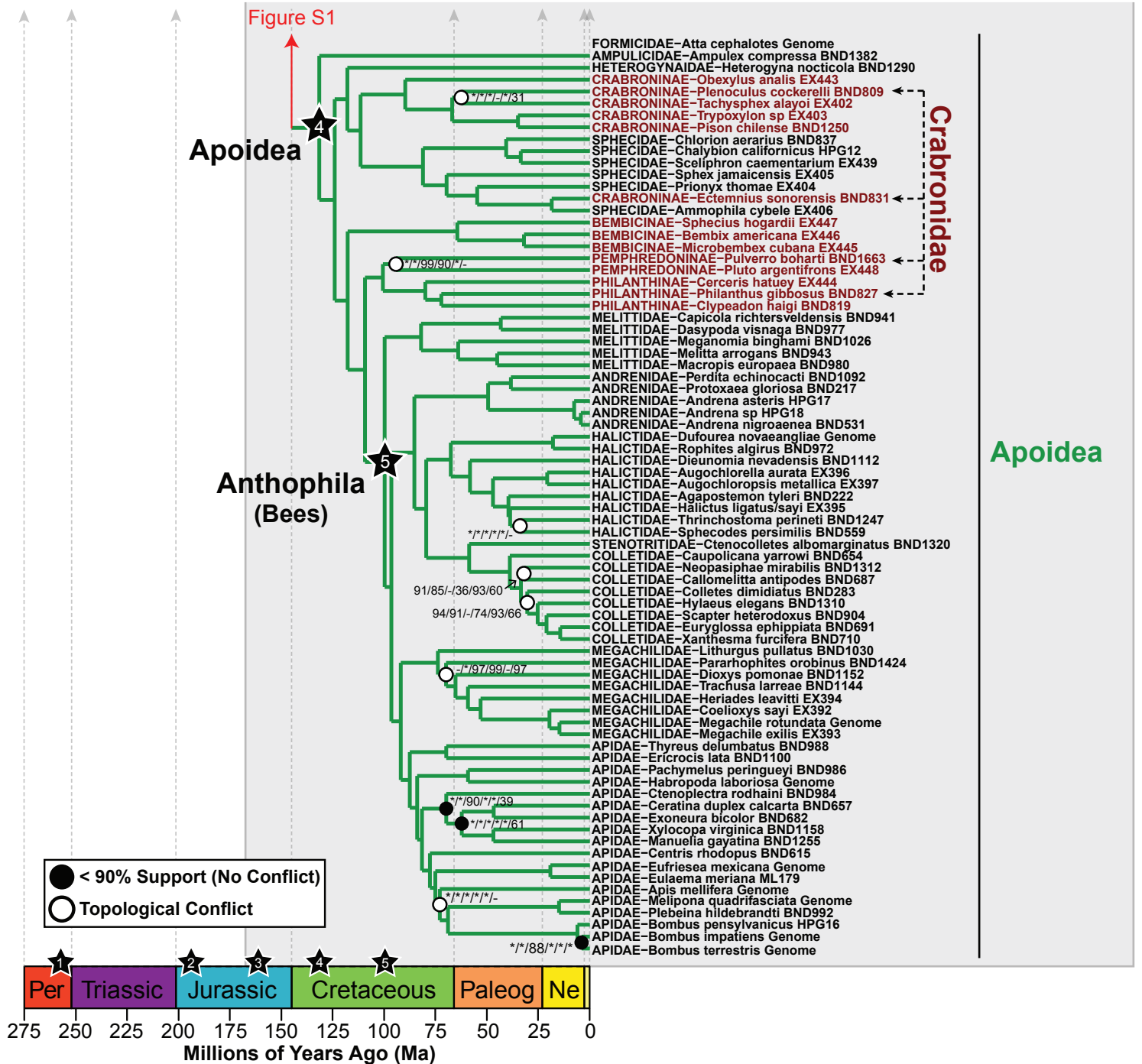
Figure S1. Detailed Phylogeny of Hymenoptera (Part One). Same tree as in Figure 1, except that all terminal taxa are shown. See also Figures 1 and S2 and Data S1.

Figure S2. Detailed Phylogeny of Hymenoptera (Part Two). Same tree as in Figure 1, except that all terminal taxa are shown. See also Figures 1 and S1 and Data S1.

SUPPLEMENTAL FIGURE AND DATA LEGENDS

Figure S1. Detailed Phylogeny of Hymenoptera (Part One). Same tree as in Figure 1, except that all terminal taxa are shown. See also Figures 1 and S2 and Data S1.

Figure S2. Detailed Phylogeny of Hymenoptera (Part Two). Same tree as in Figure 1, except that all terminal taxa are shown. See also Figures 1 and S1 and Data S1.

Data S1. Supplemental Data Tables. Includes additional information related to voucher specimens, genome samples, assembly statistics, alignment statistics, partitioning schemes, divergence dating calibrations, and divergence dating results. See also Figures 1, S1, and S2 and the Experimental Procedures.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

UCE Sequencing Pipeline

For all newly sampled taxa, we extracted DNA using Qiagen DNeasy Blood and Tissue kits (Qiagen Inc., Valencia, CA) and we fragmented up to 500 ng of input DNA to an average fragment distribution of 400-600 bp using a Qsonica Q800R sonicator (Qsonica LLC, Newton, CT). Following sonication, we constructed sequencing libraries using Kapa library preparation kits (Kapa Biosystems Inc., Wilmington, MA) and custom sample barcodes [S1]. We assessed success of library preparation following PCR amplification by measuring DNA concentration and visualizing libraries on an agarose gel. We purified reactions following PCR using 0.8 to 1.0X AMPure substitute [S2].

For UCE enrichment we pooled 6–10 libraries together at equimolar concentrations and adjusted pool concentrations to 147 ng/μl. For each enrichment we used a total of 500 ng of DNA (3.4 μl each pool), and we performed enrichments using a custom RNA bait library developed for Hymenoptera [S3] and synthesized by MYcroarray (MYcroarray, Ann Arbor, MI). The probe set includes 2,749 probes, targeting 1,510 UCE loci. We hybridized RNA bait libraries to sequencing libraries at 65ºC for a period of 24 hours, and we enriched each pool following a standardized protocol (version 1.5; protocol available from http://ultraconserved.org).

We verified enrichment success with qPCR (ViiA 7, Applied Biosystems, Waltham MA) by comparing amplification profiles of unenriched to enriched pools using PCR primers designed from several UCE loci. After verification, we used qPCR to measure the DNA concentration of each pool, and we combined all pools together at equimolar ratios to produce a final pool-of-pools. To remove overly large and small fragments, we size-selected the final pools to a range of 300–800 bp using a Blue Pippin size selection instrument (Sage Science, Beverly, MA). We mailed size-selected pools to either the UCLA Neuroscience Genomics Core or the Cornell University Biotechnology Resource Center (http://www.biotech.cornell.edu/brc/genomics-facility), where the samples were quality checked on a Bioanalyzer (Agilent Technologies, Santa Clara, CA), quantified with qPCR, and sequenced on an Illumina HiSeq 2500 (2x150 Rapid Run; Ilumina Inc, San Diego, CA).

Bioinformatics and Matrix Assembly

The sequencing facilities demultiplexed and converted raw data from BCl to FASTQ format using either BASESPACE or BCL2FASTQ (available at http://support.illumina.com/downloads/bcl2fastq_conversion_software_184.html). Using these files, we cleaned and trimmed raw reads using ILLUMIPROCESSOR [S4], which is a wrapper program around TRIMMOMATIC [S5,6]. We performed all initial bioinformatics steps, including read cleaning, assembly, and alignment, using the software package PHYLUCE v1.5. For sequenced samples, we assembled reads *de novo* using a wrapper script around TRINITY v2013-02-25 [S7]. After assembly, we used PHYLUCE to identify individual UCE loci from the bulk of assembled contigs while removing potential paralogs. We then used PHYLUCE to combine the UCE contigs from the sequenced taxa with the contigs from the 32 genome-enabled taxa into a single FASTA file. We aligned all loci individually using a wrapper around

MAFFT v7.130b [S8], and we trimmed the alignments using a wrapper around GBLOCKS v0.91b [S9,10], which we ran with reduced stringency settings (0.5, 0.5, 12, and 7 for b1–b4 settings, respectively).

To extract an equivalent set of UCE loci from 32 genome-enabled taxa (see Data S1 for genome sample information), we downloaded Hymenoptera genomes from NCBI and the Hymenoptera Genome Database [S11]. The genome of *Apterognya* za01 was provided by the authors of Johnson *et al.* [S12]. Using the software package PHYLUCE v1.5 [S13,14], we aligned our UCE probe sequences to each genome and then sliced out matching sequence along with 400 bp of flanking DNA on either side (*i.e.,* 180 bp target plus 800 bp total flanking sequence). We then used the resulting UCE "contigs" for input into the downstream bioinformatics and matrix assembly steps.

Phylogenomic Inference

We investigated the tradeoff between taxon occupancy and locus occupancy (=missing data) in order to select a set of loci to be used for all remaining analyses. Using PHYLUCE, we filtered the entire set of trimmed alignments for different amounts of taxon completeness (% of taxa that must be included in a given alignment for it to be retained). This resulted in six locus sets filtered at a taxon threshold of 0, 25, 50, 75, 90, and 95% taxon completeness. To evaluate these locus sets we generated concatenated matrices and inferred maximum likelihood trees in RAxML v8 [S15] (best tree search plus 100 rapid bootstrap replicates, GTR+Γ model of sequence evolution). We selected the best locus set by considering matrix completeness (more complete is better), topological consistency, and bootstrap support values (higher support is better). Using these criteria, we selected the 75% filtered set of alignments as the primary locus set for all subsequent analyses.

All maximum likelihood (ML) analyses were performed using the best-tree plus rapid bootstrapping search ("-f a" option) in RAxML with 200 bootstrap reps for the kmeans analysis and 100 for all others. We used the GTR+Γ model of sequence evolution for all analyses (best tree and bootstrap searches). For the partitioned-Bayesian inference (BI) search, we executed two independent runs, each with four coupled chains (one cold and three heated chains). We linked branch lengths across partitions, and we ran each partitioned search for one million generations. We assessed burn-in, convergence among runs, and run performance by examining parameter files with the program TRACER v1.6.0 [S16]. We computed consensus trees using the *consense* utility, which comes as part of EXABAYES.

To perform the ASTRAL species tree analyses we started by generating bootstrapped gene trees of all loci using RAxML (200 reps). To reduce error from loci with low information content, a problem that has been observed in other studies [S17–19], we employed weighted statistical binning, which bins loci together based on shared statistical properties and then weights bins by the number of included loci [S20]. To do this, we input all gene trees into the statistical binning pipeline using a support threshold of 75 (recommend for data sets with < 1000 loci). This grouped genes into 103 bins, comprising 73 bins of 8 loci and 30 bins of 9 loci. After binning we concatenated the genes into supergenes and used RAxML to infer supergene trees with bootstrap support (200 reps). We then input the resulting best trees, weighted by gene number, and the bootstrap trees, into ASTRAL and conducted a species tree analysis with 100 multi-locus bootstrap replicates [S21].

Topology Tests

We performed Shimodaira-Hasegawa tests (SH-test) using the "-f H" command in RAxML v8. In each case we compared a single best tree with a file containing all alternative trees. For both the 187- and 100-taxon matrices we performed the analyses without partitioning. We created the alternative topologies using Mesquite v3.03 [S22].

Taxon Sampling Experiments

For each taxon sampling experiment, we realigned the data after removing taxa, filtered alignments with GBLOCKS, filtered alignments for taxon completeness (using a 75% threshold), and generated a new

concatenated matrix. We then analyzed each matrix in RAxML using a best tree plus rapid bootstrap search (100 replicates) with GTR+ Γ as the model of sequence evolution.

Divergence Dating

To calibrate the BEAST v1.8.2 analyses, we selected 37 fossils representing taxa from across Hymenoptera and one secondary calibration taken from [S23] for the root node (see Data S1 for detailed calibration information). For fossil ages we used midpoint dates taken from date ranges provided on the Fossilworks website [S24] (www.fossilworks.org/). Due to computational challenges with BEAST, arising from having both a large number of taxa and a large amount of sequence data, we made the analysis feasible by inputting a starting tree (all nodes constrained), turning off tree-search operators, and using only a subset of the sequence data set rather than the entire concatenated matrix. We performed three separate analyses to compare the effects of different sets of loci on the final, dated results: (1) 25 loci that had the highest gene-tree bootstrap scores, (2) 50 loci that had the highest gene-tree bootstrap scores, and (3) 50 randomly selected loci. As the input topology for analyses, we used the best tree generated from the kmeans partitioned RAXML search of all loci. For each analysis, we concatenated the loci and analyzed the matrix without partitioning. We performed a total of four independent runs per analysis, with each run progressing for 200 million generations, sampling every 1,000 generations. We also performed one search with the data removed so that the MCMC sampled from the prior distribution only. For the clock and substitution models, we selected uncorrelated lognormal and GTR+Γ, respectively. For the tree prior, we used a birth-death model, and for the ucld.mean prior, we used an exponential distribution with the mean set to 1.0 and the initial value set to 0.003 (determined empirically from preliminary runs). We assessed burn-in, convergence among runs, and run performance by examining parameter files with the program TRACER v1.6.0 [S16].

SUPPLEMENTAL REFERENCES

S1.   Faircloth, B.C., and Glenn, T.C. (2012). Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. PLoS One *7*, e42543. doi:10.1371/journal.pone.0042543.

S2.   Rohland, N., and Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. *22*, 939–946. doi:10.1101/gr.128124.111.

S3.   Faircloth, B.C., Branstetter, M.G., White, N.D., and Brady, S.G. (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol. Ecol. Resour. *15*, 489–501. doi:10.1111/1755-0998.12328.

S4.   Faircloth, B.C. (2013). Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. doi:http://dx.doi.org/10.6079/J9ILL.

S5.   Lohse, M., Bolger, A.M., Nagel, A., Fernie, A.R., Lunn, J.E., Stitt, M., and Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-seq-based transcriptomics. Nucleic Acids Res. *40*, W622–W627. doi:10.1093/nar/gks540.

S6.   Del Fabbro, C., Scalabrin, S., Morgante, M., and Giorgi, F.M. (2013). An extensive evaluation of read trimming effects on illumina NGS data analysis. PLoS One *8*, e85024. doi:10.1371/journal.pone.0085024.

S7.   Grabherr, M.G., Hass, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–52. doi:10.1038/nbt.1883.

S8.   Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–66. doi:10.1093/nar/gkf436.

S9.   Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552. doi:10.1093/oxfordjournals.molbev.a026334.

S10.   Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564–577. doi:10.1080/10635150701472164.

S11.   Munoz-Torres, M.C., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., Childs, K.L., Anzola, J.M., Milshina, N., and Elsik, C.G. (2011). Hymenoptera Genome Database: integrated

community resources for insect species of the order Hymenoptera. Nucleic Acids Res. *39*, D658-62. doi:10.1093/nar/gkq1145.

S12. Johnson, B.R., Borowiec, M.L., Chiu, J.C., Lee, E.K., Atallah, J., and Ward, P.S. (2013). Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. Curr. Biol. *23*, 1–5. doi:10.1016/j.cub.2013.08.050.

S13. Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., and Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. *61*, 717–726. doi:10.1093/sysbio/sys004.

S14. Faircloth, B.C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics *32*, 786–788. doi:10.1093/bioinformatics/btv646.

S15. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313. doi:10.1093/bioinformatics/btu033.

S16. Rambaut, A., Suchard, M.A., Xie, D., and Drummond, A.J. (2014). Tracer v1.6. http://beast.bio.ed.ac.uk/Tracer.

S17. Meiklejohn, K.A., Faircloth, B.C., Glenn, Travis, C., Kimball, R.T., and Braun, E.L. (2016). Analysis of a rapid evolutionary radiation using ultraconserved elements (UCEs): evidence for a bias in some multispecies coalescent methods. Syst. Biol. *65*, 612–627. doi:10.1093/sysbio/syw014.

S18. Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., and Kimball, R.T. (2015). Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. *33*, 1110–1125. doi:10.1093/molbev/msv347.

S19. Manthey, J. D. (2016). Comparison of target-capture and restriction-site associated DNA sequencing for phylogenomics: a test in cardinalid tanagers (Aves, Genus: *Piranga*). Syst. Biol. *65*, 640–650. doi:10.1093/sysbio/syw005.

S20. Bayzid, M.S., Mirarab, S., Boussau, B., and Warnow, T. (2015). Weighted statistical binning: enabling statistically consistent genome-scale phylogenetic analyses. PLoS One *10*, e0129183. doi:10.1371/journal.pone.0129183.

S21. Seo, T.K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. *25*, 960–971. doi:10.1093/molbev/msn043.

S22. Maddison, W.P., and Maddison, D.R. (2016). Mesquite: a modular system for evolutionary analysis. Version 3.03. Available at http://mesquiteproject.org.

S23. Zhang, C., Stadler, T., Klopfstein, S., Heath, T.A., and Ronquist, F. (2015). Total-evidence dating under the fossilized birth-death process. Syst. Biol. *65*, 228–249. doi:10.1093/sysbio/syv080.

S24. Alroy, J. (2016). Fossilworks: gateway to the paleobiology database. Available at: www.fossilworks.org. [Accessed November 1, 2016].

S25. Nygaard, S., Zhang, G., Schiøtt, M., Li, C., Wurm, Y., Hu, H., Zhou, J., Ji, L., Qiu, F., Rasmussen, M., et al. (2011). The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. Genome Res. *21*, 1339–1348. doi:10.1101/gr.121392.111.10.

S26. Weinstock, G.M., Robinson, G.E., Gibbs, R.A., Worley, K.C., Evans, J.D., Maleszka, R., Robertson, H.M., Weaver, D.B., Beye, M., Bork, P., et al. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature *443*, 931–949. doi:10.1038/nature05260.

S27. Elsik, C.G., Worley, K.C., Bennett, A.K., Beye, M., Camara, F., Childers, C.P., Graaf, D.C. De, Debyser, G., Deng, J., Devreese, B., et al. (2014). Finding the missing honey bee genes: lessons learned from a genome upgrade. BMC Genomics *15*, 86. doi:10.1186/1471-2164-15-86.

S28. Suen, G., Teiling, C., Li, L., Holt, C., Abouheif, E., Bornberg-Bauer, E., Bouffard, P., Caldera, E.J., Cash, E., Cavanaugh, A., et al. (2011). The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. PLoS Genet. *7*, e1002007. doi:10.1371/journal.pgen.1002007.

S29. Sadd, B.M., Barribeau, S.M., Bloch, G., de Graaf, D.C., Dearden, P., Elsik, C.G., Gadau, J., Grimmelikhuijzen, C.J.P., Hasselmann, M., Lozier, J.D., et al. (2015). The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol. *16*, 76. doi:10.1186/s13059-015-0623-3.

S30. Bonasio, R., Zhang, G., Ye, C., Mutti, N.S., Fang, X., Qin, N., Donahue, G., Yang, P., Li, Q., Li, C., *et al.* (2010). Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. Science *329*, 1068–71. doi:10.1126/science.1192428.

S31.   Schrader, L., Kim, J.W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stökl, J., Schultner, E., et al. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. Nat. Commun. *5*, 5495. doi:10.1038/ncomms6495.

S32.   Oxley, P.R., Ji, L., Fetter-Pruneda, I., McKenzie, S.K., Li, C., Hu, H., Zhang, G., and Kronauer, D. J.C. (2014). The genome of the clonal raider ant *Cerapachys biroi*. Curr. Biol. *24*, 1–8. doi:10.1016/j.cub.2014.01.018.

S33.   Xiao, J.-H., Yue, Z., Jia, L.-Y., Yang, X.-H., Niu, L.-H., Wang, Z., Zhang, P., Sun, B.-F., He, S.-M., Li, Z., et al. (2013). Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. Genome Biol. *14*, R141. doi:10.1186/gb-2013-14-12-r141.

S34.   de Boer, J.G., Groenen, M.A.M., Pannebakker, B.A., Beukeboom, L.W., and Kraus, R.H.S. (2015). Population-level consequences of complementary sex determination in a solitary parasitoid. BMC Evol. Biol. *15*, 98. doi:10.1186/s12862-015-0340-2.

S35.   Kapheim, K.M., Pan, H., Li, C., Salzberg, S.L., Puiu, D., Magoc, T., Robertson, H.M., Hudson, M. E., and Venkat, A. (2015). Genomic signatures of evolutionary transitions from solitary to group living. Science. *348*, 1139–1144. doi:10.1126/science.aaa4788.

S36.   Smith, C.D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C.R., Elhaik, E., Elsik, C.G., et al. (2011). Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). Proc. Natl. Acad. Sci. USA *108*, 5673–5678. doi:10.1073/pnas.1008617108.

S37.   Burke, G.R., Walden, K.K.O., Whitfield, J.B., Robertson, H.M., and Strand, M.R. (2014). Widespread genome reorganization of an obligate virus mutualist. PLoS Genet. *10*, e1004660. doi:10.1371/journal.pgen.1004660.

S38.   Mikheyev, A.S., and Linksvayer, T.A. (2015). Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. Elife *4*, e04775. doi:10.7554/eLife.04775.

S39.   Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., and Colbourne, J.K. (2010). Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science. *327*, 343–348. doi:10.1126/science.1178028.

S40.   Smith, C.R., Smith, C.D., Robertson, H.M., Helmkampf, M., Zimin, A., Yandell, M., Holt, C., Hu, H., Abouheif, E., Benton, R., et al. (2011). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. Proc. Natl. Acad. Sci. USA *108*, 5667–72. doi:10.1073/pnas.1007901108.

S41.   Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B.G., Ingram, K.K., Falquet, L., Nipitwattanaphon, M., Gotzek, D., et al. (2011). The genome of the fire ant *Solenopsis invicta*. Proc. Natl. Acad. Sci. USA *108*, 5679–84. doi:10.1073/pnas.1009690108.

S42.   Rodriguez, J., Waichert, C., Von Dohlen, C.D., Poinar Jr., G., and Pitts, J.P. (2015). Eocene and not Cretaceous origin of spider wasps: fossil evidence from amber. Acta Palaeontol. Pol. *61*, 89–96. doi:10.4202/app.00073.2014.