Molecular Phylogenetics and Evolution 107 (2017) 10–15

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomic analysis of yellowjackets and hornets (Hymenoptera: Vespidae, Vespinae)

Federico Lopez-Osorio [a,b,*], Kurt M. Pickett [a,1], James M. Carpenter [b], Bryan A. Ballif [a], Ingi Agnarsson [a,c]

[a] Department of Biology, University of Vermont, Room 120A Marsh Life Science Building, 109 Carrigan Drive, Burlington, VT 05405, USA
[b] Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10023, USA
[c] Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA

ARTICLE INFO

Article history:
Received 16 June 2016
Revised 24 September 2016
Accepted 10 October 2016
Available online 11 October 2016

Keywords:
Phylogenomics
RNA-seq
Transcriptomics
Evolution
Sociality
Vespinae

ABSTRACT

The phylogenetic relationships among genera of the subfamily Vespinae (yellowjackets and hornets) remain unclear. Yellowjackets and hornets constitute one of the only two lineages of highly eusocial wasps, and the distribution of key behavioral traits correlates closely with the current classification of the group. The potential of the Vespinae to elucidate the evolution of social life, however, remains limited due to ambiguous genus-level relationships. Here, we address the relationships among genera within the Vespinae using transcriptomic (RNA-seq) data. We sequenced the transcriptomes of six vespid wasps, including three of the four genera recognized in the Vespinae, combined our data with publicly available transcriptomes, and assembled two matrices comprising 1,507 and 3,356 putative single-copy genes. The results of our phylogenomic analyses recover *Dolichovespula* as more closely related to *Vespa* than to *Vespula*, therefore challenging the prevailing hypothesis of yellowjacket (*Vespula* + *Dolichovespula*) monophyly. This suggests that traits such as large colony size and high paternity arose in the genus *Vespula* following its early divergence from the remaining vespine genera.

1. Introduction

Eusocial groups consist of overlapping generations of workers collectively caring for the offspring of the queen caste. Among wasps, eusociality is thought to have evolved once (Carpenter, 1982; Pickett and Carpenter, 2010) or twice (Hines et al., 2007) in the family Vespidae. Within the eusocial vespids, the paper wasp genus *Polistes* and the subfamily Vespinae, which includes the yellowjackets (*Vespula* and *Dolichovespula*) and hornets (*Vespa* and *Provespa*), are perhaps the most familiar. Vespine colonies generally comprise a single queen; show varying levels of morphological caste differentiation; live in enclosed, sometimes subterranean, nests built from paper-like material; construct cells used exclusively to raise future queens; and vary considerably in size (Evans and West-Eberhard, 1970). Ranges of colony size (i.e., number of workers) overlap in many vespine species, but members of the *Vespula vulgaris* and *V. squamosa* species groups typically have the largest societies (more than 2500 cells and 500 workers; Akre et al., 1981; Loope et al., 2014). Colony size can be viewed as a determinant of social interactions and life history characteristics (Anderson and McShea, 2001; Bourke, 1999). Indeed, in vespine wasps, colony size correlates with traits such as paternity (single or multiple mating by queens), reproductive potential of workers, the nature of conflict among colony members, and degree of caste differentiation (Akre and Davis, 1978; Foster and Ratnieks, 2001a, b; Loope et al., 2014), among others.

For example, species of *Dolichovespula* build small colonies with low paternity and workers that lay eggs in the presence of the queen, thereby instigating queen-worker conflict over the production of males, which develop only from unfertilized eggs (Foster and Ratnieks, 2001a,b; Foster et al., 2001; Freiburger et al., 2004; Wenseleers et al., 2005b). In contrast, colonies of large-colony species in the *Vespula vulgaris* and *squamosa* groups have the greatest degree of caste dimorphism (Greene, 1979), few workers with functional ovaries (Foster and Ratnieks, 2001a,b; Ross, 1985) and production of males exclusively by queens (Akre et al., 1976; Ross, 1986; Foster and Ratnieks, 2001a; Kovacs and Goodisman, 2007). In these large-colony vespines, queen-worker conflict over male production is typically resolved by means of policing; that is, the removal of worker-laid eggs that maintains the reproductive primacy of the queen (Ratnieks and Visscher, 1989; Wenseleers and Ratnieks, 2006). Conflict may occur between species, too. Such is the case of queens of socially parasitic species that exploit the worker force and colony resources of a host species – a behavior that among vespines has evolved primarily in *Vespula* and

* Corresponding author at: Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10023, USA.
E-mail address: flopezosorio@amnh.org (F. Lopez-Osorio).
[1] Deceased.

http://dx.doi.org/10.1016/j.ympev.2016.10.006
1055-7903/

Dolichovespula. These social parasites, lacking the worker caste, seize the nest of a host species and trick the resident workers into raising the parasitic offspring (Greene et al., 1978; MacDonald and Matthews, 1975; Reed and Akre, 1983). The monophyly of *Vespula* plus *Dolichovespula* would suggest, for example, that social parasitism is more likely to evolve in this particular clade rather in distant lineages within the Vespinae.

The Vespinae comprises 70 described species classified in four genera and distributed throughout tropical areas of the Oriental region and northern temperate latitudes (Akre and Davis, 1978; Carpenter and Kojima, 1997; Kimsey and Carpenter, 2012). *Vespula* and *Dolichovespula* are primarily temperate, *Vespa* occurs in both tropical and temperate regions, and *Provespa* is endemic to the Oriental tropics. Southeast Asia has been speculated as the 'center of origin' of the Vespinae on the basis of the sister relationship of *Vespa* to the remaining vespine genera, the species richness of the genus in that region, and because hornets are not native to the Western Hemisphere (van der Vecht, 1957; Matsuura and Yamane, 1990 p. 240). A common origin in the northern latitudes, however, has also been proposed for the subfamily (Bequaert, 1932).

Given the phylogenetic distribution of a suite of key behavioral traits and the relevance of genus-level relationships to the biogeography of yellowjackets and hornets, one of the primary goals in vespine phylogeny is elucidating deep-level relationships, which have been contradictory across studies (Carpenter, 1987; Lopez-Osorio et al., 2014; Perrard et al., 2016; Pickett and Carpenter, 2010). Previous analyses have recovered a yellowjacket clade (*Vespula* + *Dolichovespula*) sister to either *Provespa* (Carpenter, 1987; Saito and Kojima, 2011) or *Vespa* plus *Provespa* (Pickett and Carpenter, 2010), whereas non-monophyly of yellowjackets, placing *Dolichovespula* as sister group of the hornets (*Vespa* + *Provespa*), has been reported in studies relying exclusively on molecular data (Lopez-Osorio et al., 2014, 2015). The results of Lopez-Osorio et al. (2014), however, showed discordance between mitochondrial and nuclear gene fragments. Specifically, Lopez-Osorio et al. (2014) found that mitochondrial genes support the monophyly of yellowjackets (*Vespula* + *Dolichovespula*), but nuclear genes and the concatenated data indicated a sister group relationship between *Dolichovespula* and the hornet clade (*Vespa* + *Provespa*). Furthermore, in the first comprehensive phylogenetic analysis of vespine wasps based on morphological and DNA sequence data combined, Perrard et al. (2016) recovered *Vespa* as sister to *Vespula* + *Dolichovespula*, although relationships among genera were very poorly supported.

In this study, we address the genus-level relationships in the Vespinae and examine the monophyly of yellowjackets using a phylogenomic approach based on transcriptomic (RNA-seq) data. Our phylogenomic analysis includes three of the four genera recognized in the Vespinae and a total of nine transcriptomes, six of which are novel to this study: the solitary potter wasp *Ancistrocerus catskill*, the primitively eusocial *Polistes dominula*, and the highly eusocial *Vespa crabro*, *Dolichovespula maculata*, *D. arenaria*, and *Vespula vidua*. We conduct *de novo* transcriptome assemblies, identify putative single-copy genes, and use these candidate orthologs to test the sister-group relationship between *Vespula* and *Dolichovespula*. Our findings challenge previous phylogenetic hypotheses and provide a new framework for future comparative studies on yellowjackets and hornets.

2. Materials and methods

2.1. Sample collection, RNA isolation, library preparation and sequencing

We collected specimens of *A. catskill*, *D. arenaria*, *D. maculata*, *V. vidua*, and *P. dominula* at localities in the vicinity of Burlington, Vermont, USA, and specimens of *V. crabro* in Slovenia; the genus *Provespa* was not included because of lack of high-quality source material. Specimens were flash frozen in liquid nitrogen and stored at −80 °C. We isolated total RNA from single, whole specimens using the TRIzol® reagent (Invitrogen). Quality assessment of RNA samples, preparation of cDNA libraries, Roche 454 pyrosequencing of *A. catskill*, and paired-end 2×100 bp Illumina sequencing of the remaining species were outsourced to Beckman Coulter Genomics (Danvers, MA). We combined our data with publicly available transcriptomes from the cuckoo wasp *Argochrysis armilla*, the pollen wasp *Pseudomasaris vespoides*, and the paper wasp *Mischocyttarus flavitarsis* (NCBI SRA accessions SRX262928, SRX262920, and SRX259759; Johnson et al., 2013).

2.2. Processing of reads, de novo transcriptome assembly, and translation of transcripts

We cut adapters, trimmed low-quality bases and discarded reads below 36 bases in length from Illumina reads using Trimmomatic v. 0.32 (Bolger et al., 2014; Lohse et al., 2012) with default settings, except for a threshold of 20 for average base quality within the sliding window. Using the reads remaining after trimming, transcriptomes were assembled *de novo* using Trinity v. 2013-11-10 (Grabherr et al., 2011; Haas et al., 2013). We removed possible (human, bacterial, and viral) contamination sequences using the standalone release of DeconSeq v. 0.4.3 (Schmieder and Edwards, 2011), and used riboPicker v. 0.4.3 (Schmieder et al., 2012) to discard rRNA-like transcripts. In these two *in silico* sanitation steps an identity score of 90 and a coverage value of 15 were used.

We used TransDecoder r20131110 (Haas et al., 2013) to identify candidate coding regions within transcript sequences and CD-HIT (Fu et al., 2012) to cluster peptides using a stringent identity threshold (–c 1.0 –n 5). Translated vespid transcriptomes were submitted to BLASTP searches against the NCBI RefSeq database of protein reference sequences. BLASTP results were then used to remove any previously undetected contaminant transcripts.

2.3. Matrix construction and phylogenomic analyses

2.3.1. Homology inference

To identify groups of putative homologous sequences and orthologs, we followed a procedure based on sequence similarity and phylogenetic analysis (Yang and Smith, 2014; Supplementary Fig. 1). We assembled two data matrices, one including all nine species and another excluding the transcriptome of *A. catskill* due to its comparatively small size. All-by-all BLASTP searches were conducted with an E value cutoff of 10 and keeping a maximum of 500 aligned sequences (max_target_seqs 500). Sequence ends not covered by any BLASTP hits from other taxa were removed. BLASTP hits with query coverage greater than 0.4 were used for homology inference. We identified clusters of homologous sequences using the Markov Clustering Algorithm (MCL v. 14-137; Enright et al., 2002) tool with an E value cutoff of 10^{-5} and an inflation value of 2.0. The sequences of each cluster were aligned and alignments were cleaned using Phyutility (Smith and Dunn, 2008) with a minimum site occupancy threshold of 0.1. Clusters with less than one thousand sequences were aligned with MAFFT v. 7 (Katoh and Standley, 2013) using the options 'genafpair' and 'maxiterate 1000', whereas larger clusters were aligned with PASTA (Mirarab et al., 2014). We used RAxML 8 (Stamatakis, 2014) to infer an initial maximum likelihood phylogenetic tree for each aligned cluster of homologous sequences with the model PROTCATWAG. Terminal branches ten times longer than their sisters or longer than 0.8 were trimmed. Monophyletic and paraphyletic sequence isoforms from the same taxon were

removed, keeping only the sequence with less ambiguous characters as the representative. Moreover, internal branches longer than 1.0 were cut to break deep paralogs, thus generating two or more subtrees. This process of cluster refinement, consisting of sequence alignment, cleaning of alignments, and trimming of spurious branches was then repeated using a cutoff of 0.6 for tips and 0.7 for internal branches. We then conducted a third round of alignment and tree inference with 200 fast bootstrap pseudoreplicates to generate homolog trees used to identify orthologs.

2.3.2. Orthology inference, phylogenetic analysis and topology tests

We used the maximum inclusion method (Dunn et al., 2008; Hejnol et al., 2009) to prune homolog trees into subtrees with no more than one sequence per taxon. The sequences from each resulting set of orthologs were aligned with MAFFT and alignments were trimmed using Gblocks v0.91 (Castresana, 2000) with settings −b3 = 8 −b4 = 10 and −b5 = h. We concatenated trimmed alignments with full taxon sets and length greater than or equal to 300 sites.

We analyzed the two matrices using parsimony and maximum likelihood methods. We carried out parsimony searches in TNT (Goloboff et al., 2008) with 1,000 replications of RAS + TBR holding two trees per replicate. Moreover, group support was evaluated with 1,000 replications of symmetric resampling (Goloboff et al., 2003), summarizing the results as the difference in frequency between a group and the most frequent contradictory group (GC values). We chose models of amino acid substitution for each ortholog using the RAxML model selection Perl script 'ProteinModelSelection.pl' (available at http://sco.h-its.org/exelixis/web/software/raxml/). Partitioned ML searches and 1,000 rapid bootstrap inferences were carried out in RAxML on CIPRES (Miller et al., 2010). We evaluated uncertainty of edges and conflict between gene trees and species trees in two ways: first, we performed ML analyses of 200 jackknife pseudoreplicates obtained by resampling 30% of the total number of orthologs; second, based on ML results, we used PhyParts (Smith et al., 2015) to evaluate concordance and conflict by comparing ingroup (Vespinae) clades from each gene tree with the species tree topology, and to calculate internode certainty scores (ICA; Salichos and Rokas, 2013; Salichos et al., 2014) under a bootstrap filter of 50%. ICA values close to 1 indicate strong certainty in a bipartition of interest, whereas ICA values close to 0 indicate similar frequency of conflicting bipartitions, and negative values indicate higher frequency of one or more bipartitions conflicting with the internode of interest (Salichos et al., 2014; Smith et al., 2015). Lastly, we performed species tree analyses in MP-EST (Liu et al., 2010) with default settings on the STRAW web server (Shaw et al., 2013).

2.3.3. Hypothesis testing

We evaluated the significance of differences in log-likelihoods between ML trees and an alternative hypothesis of yellowjacket monophyly using the test developed by Shimodaira and Hasegawa (1999; hereafter SH test). SH tests were performed as implemented in RAxML 8 using the -f H option to re-estimate parameters for all trees.

3. Results

3.1. Transcriptome sequencing and de novo assembly

The five newly-sequenced, Illumina transcriptomes had an average of approximately 207 million passing filter (PF) reads (Supplementary Table 1), and the transcriptome of *A. catskill* had 1,379,816 Roche 454 reads. After quality trimming of Illumina transcriptomes, the percentage of surviving read pairs ranged from 80.76% to 91.58% (Supplementary Table 2). The six transcriptomes generated in this study had an average of 129,357 transcripts, an N50 of 3,186, and 51,786 potential coding regions (Table 1). After reducing redundancy, the average number of amino acid sequences (excluding *A. catskill*) was 14,896 (Table 1). Raw 454 and Illumina reads were deposited on the Sequence Read Archive (SRA) database under accession numbers SRR4301484–SRR4301489.

3.2. Homology and orthology inferences and phylogenetic analyses

The nine-taxon matrix comprised 1,507 putative orthologs, 933,533 aligned sites, and had 91% amino acid completeness (Table 1). The eight-taxon matrix comprised 3,356 putative orthologs, 2,285,441 aligned sites, and had 94% amino acid occupancy (Table 1). The two matrices had 100% ortholog coverage; that is, all orthologs had sequence data for all taxa. Transcriptome assemblies and matrices are available in Dryad repository doi: http://dx.doi.org/10.5061/dryad.7h0g0. Parsimony and ML analyses of the two matrices resulted in the same fully resolved topology, which had symmetric resampling, bootstrap, and 30% gene-jackknife support values of 100 for all nodes (Fig. 1, Supplementary Figs. 2 and 3). Moreover, we found the same topology in the species tree analyses conducted in MP-EST (Supplementary Fig. 4). In this topology, *P. vespoides* was sister to the remaining vespid species, and *A. catskill* was sister to the monophyletic subfamilies Polistinae and Vespinae. Within the Vespinae, *Vespula vidua* was recovered as sister to a clade including the hornet *Vespa crabro* and the monophyletic genus *Dolichovespula* (Fig. 1).

We compared bipartitions from each gene tree with the species tree to evaluate concordance and conflict, and found support for the sister relationship between *Vespa* and *Dolichovespula*. We extracted 1,197 ingroup (Vespinae) clades from trees built for each gene in the set of 1,507 orthologs and compared these clades with the species tree in Fig. 1. The *Vespa* + *Dolichovespula* node had 625 concordant gene trees and an ICA score of 0.47 (Table 2). In the case of the eight-taxon dataset with 3,356 orthologs, we extracted 2,757 Vespinae clades from gene trees, compared these clades with the species tree, and found that the *Vespa* + *Dolichovespula* clade was supported by 1,425 gene trees and had a 0.47 ICA value (Table 2). Additionally, we compared ingroup (Vespinae) clades from gene trees with an alternative species tree in which yellowjackets (*Vespula* + *Dolichovespula*) were monophyletic. We found that for the set of 1,507 orthologs, the *Vespula* plus *Dolichovespula* clade in the alternative species tree was supported by 127 gene trees and in conflict with 802 gene trees. Likewise, for the set of 3,356 orthologs, the comparison of gene trees with a species tree including the *Vespula* plus *Dolichovespula* clade showed that this clade was supported by 308 gene trees and in conflict with 1,839 gene trees (Table 2). Lastly, the topology recovered here (Fig. 1) was significantly different from the hypothesis in which *Vespula* and *Dolichovespula* were sister groups (Table 3).

4. Discussion

The deep-level phylogenetic relationships of vespine wasps have been elusive, as previous studies supported alternative genus-level groupings (Carpenter, 1987; Lopez-Osorio et al., 2015, 2014; Perrard et al., 2016; Pickett and Carpenter, 2010; Saito and Kojima, 2011). The lack of consensus regarding the backbone nodes of the Vespinae phylogeny hampers the use of a comparative framework in studies of, for example, evolution of behavioral traits and of the molecular traits associated with the evolution of sociality (Fischman et al., 2011; Rehan and Toth, 2015; Robinson et al., 2005). In this study, we provide, for the first time, a hypothesis of genus-level relationships of vespine wasps based on transcriptomic data.

Table 1
Characteristics of transcriptome assemblies; descriptive statistics are based on all transcript contigs.

	Total transcripts	GC%	Contig N50	Average contig length	ORFs	CD-HIT clusters
D. arenaria	130,448	37.50	2817	1,393.00	38,669	14,489
D. maculata	131,905	34.72	3775	2,049.22	61,099	13,944
V. crabro	201,718	35.21	4006	1,862.25	79,625	16,361
V. vidua	146,729	34.40	3688	2,048.18	65,754	14,919
P. dominula	155,861	32.47	3396	1,798.09	58,569	14,768
A. catskill	9481	36.03	1434	1,247.36	7002	–
Average	129,357	35.06	3186	1,733.02	51,786	14,896
SD	64,215.73	1.68	950.90	338.1485	25,614.52	899.32

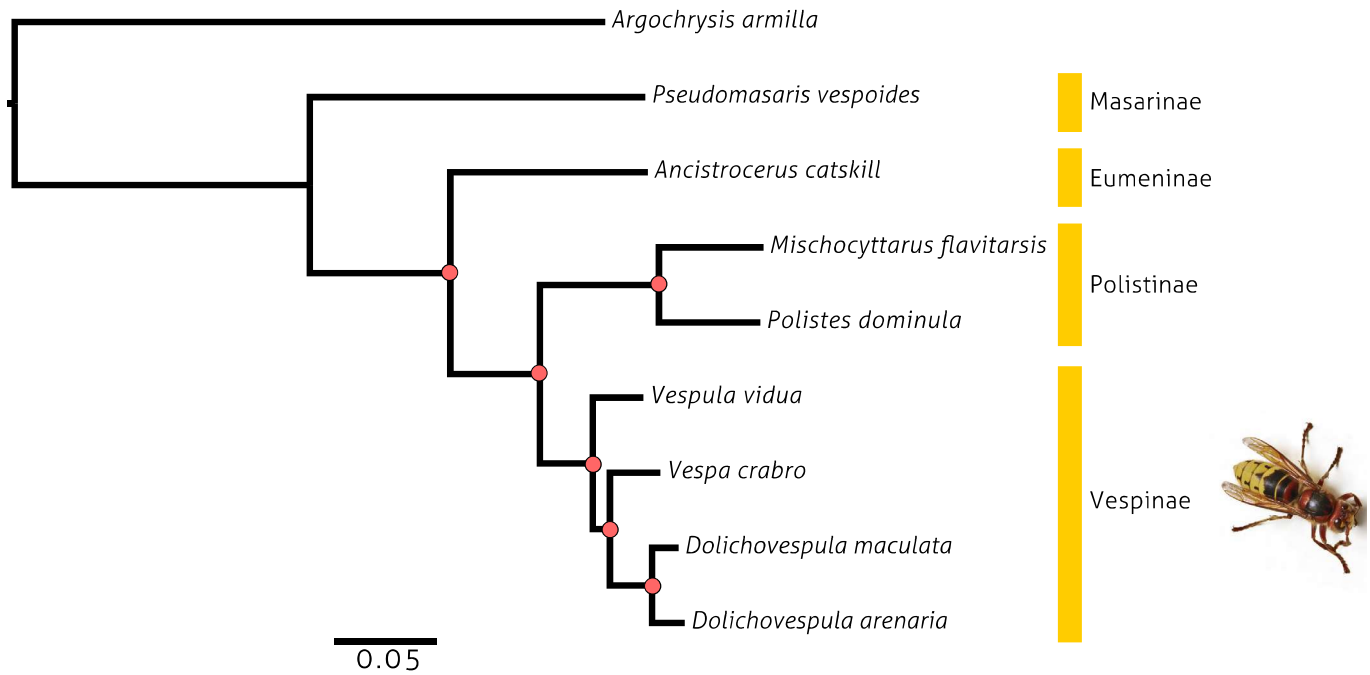

Fig. 1. Maximum-likelihood phylogeny of vespid wasps based on the analysis of 1,507 genes. Phylogeny inferred in RAxML from a 1,507-gene concatenated matrix with 933,533 amino acid sites. Circles on nodes indicate 100% bootstrap support derived from 1,000 replicates estimated using RAxML. Picture of *Vespa crabro* taken by Bernie Kohl and available at http://commons.wikimedia.org.

Table 2
Number of gene trees concordant or conflicting with the *Vespa* + *Dolichovespula* clade in the species tree (Fig. 1), or with a yellowjacket (*Vespula* + *Dolichovespula*) clade in an alternative species tree. Internode certainty (ICA) scores near 1 indicate strong certainty. *Vpl.* = *Vespula*.

Clade	Concordant	Conflicting	ICA score
9 taxa			
(*V. crabro*, (*D. maculata*, *D. arenaria*))	625	302	0.4742
(*Vpl. vidua*, (*D. maculata*, *D. arenaria*))	127	802	-0.447
8 taxa			
(*V. crabro*, (*D. maculata*, *D. arenaria*))	1,425	717	0.4729
(*Vpl. vidua*, (*D. maculata*, *D. arenaria*))	308	1,839	-0.446

Transcriptomic data challenge the relationships among genera found in previous phylogenetic analyses of vespine wasps (e.g., Carpenter, 1987; Perrard et al., 2016; Pickett and Carpenter, 2010). The prevailing hypothesis of vespine phylogeny indicates that *Vespa* is the sister group of the remaining Vespinae, and the monophyletic yellowjackets (*Vespula* and *Dolichovespula*) are sister to *Provespa* (Carpenter, 1987). Moreover, a recent study, based on comprehensive taxon sampling and the combined analysis of morphological characters and nine genes, found a sister-group relationship, albeit poorly supported, between *Vespa* and the yellowjackets (Perrard et al., 2016). That is, most previous studies have recovered yellowjackets as a monophyletic group (Carpenter, 1987; Perrard et al., 2016; Pickett and Carpenter, 2010; Saito and Kojima, 2011). Our transcriptomic data did not recover a yellowjacket clade. Instead, we found that the hornet genus *Vespa* is sister to the yellowjacket genus *Dolichovespula* (Fig. 1). A sister group relationship between *Vespa* + *Provespa* and *Dolichovespula* was previously reported in phylogenetic analyses of targeted gene fragments, although mitochondrial and nuclear genes had conflicting

Table 3
Results of SH test estimated for a hypothesis of yellowjacket monophyly tested against the best ML tree, showing the likelihood (LH) of the alternative tree, difference in likelihood D(LH), and standard deviation (SD) for each test. Asterisks indicate that the alternative tree is significantly worse (1% level).

	Best tree LH	LH	D(LH)	SD
9 taxa	−4978865.370233	−4981168.490189	−2303.119956**	156.172506
8 taxa	−11581322.050153	−11582672.438202	−1350.388049 **	231.471426

phylogenetic signals (Lopez-Osorio et al., 2015, 2014). The number of genes used in our analysis was orders of magnitude higher than in any previous phylogenetic study of vespine wasps, and these genome-scale data clearly support the sister relationship between *Vespa* and *Dolichovespula*.

The *Vespa* plus *Dolichovespula* clade was supported in our analyses of the full sets of putative orthologs, and in analyses of random samples of genes, suggesting strong support from independently evolving genes. However, we also found evidence of topological incongruence among gene histories. Considering that traditional measures of support, such as the standard bootstrap (Felsenstein, 1985), are less informative for concatenated genome-scale data sets (Rokas and Carroll, 2006; Siddall, 2010; Smith et al., 2015), we applied alternative procedures to evaluate the robustness and uncertainty of internal edges in both the eight- and nine-taxon datasets. Jackknife resampling of 30% of the total number of genes resulted in frequencies of 100 for all nodes. ICA values, however, were lower than 1.0 for focal nodes (Table 2), indicating conflicting groupings of vespine genera. In particular, the short internal branch subtending the grouping of *Vespa* plus *Dolichovespula* reflects a limited amount of phylogenetic signal, which may explain the conflict at the base of this clade (Fig. 1, Table 2) (Philippe et al., 2011; Regier et al., 2008; Salichos and Rokas, 2013). Conflict at the base of this clade suggests that biological processes such as gene duplication and extinction and incomplete lineage sorting might have influenced the origin of these wasps (Jeffroy et al., 2006; Maddison, 1997). Moreover, previous phylogenetic studies suggest that the Vespinae has experienced a period of early rapid radiation (Perrard et al., 2016), leaving little time for the accumulation of informative characters (Whitfield and Kjer, 2008; Whitfield and Lockhart, 2007). Further work is required on the sources of phylogenetic conflict in the Vespinae. As more genome-scale data become available, the relationship between *Vespa* and *Dolichovespula* will be more powerfully tested by the inclusion of more taxa.

The phylogeny inferred here can lead to different conclusions on the evolution of behavioral traits in the Vespinae. Large-colony species in the Vespinae usually have high paternity, which reduces relatedness between workers and, therefore, workers are predicted to police each other's reproduction (Ratnieks, 1988). This is the case for large-colony species of the *Vespula vulgaris* and *squamosa* groups (Bonckaert et al., 2008; Helanterä et al., 2006; Oi et al., 2015; Wenseleers et al., 2005a). In contrast, small-colony species of *Dolichovespula* usually have low paternity and worker reproduction (Foster and Ratnieks, 2001b; Foster et al., 2001; Wenseleers et al., 2005b; Bonckaert et al., 2011; van Zweden et al., 2013; Loope et al., 2014). Phylogenetically informed comparative analyses reveal that in vespine wasps, workers suppress each other's reproduction more frequently in species with high paternity, where workers are more related to the queen's sons than to sons of workers (Wenseleers and Ratnieks, 2006). Moreover, taking phylogeny into account, colony size predicts average intracolony relatedness and correlates positively with paternity frequency in vespine wasps (Loope et al., 2014). Colony size is a trait that may be considered both a cause and effect of reproductive conflict (Bourke, 1999). That is, effective policing in *Vespula* may have driven the evolution of large colony size or, alternatively, large colony size may have increased the benefits of worker policing (Foster and Ratnieks, 2001a,b). The comparative studies aforementioned, however, relied on a hypothesis of yellowjacket monophyly, but it remains to be determined whether the sister relationship between *Vespa* and *Dolichovespula* may change the understanding of the evolution of traits in the Vespinae. The phylogenetic framework proposed here implies, for example, that traits such as large colony size and high paternity may arise more frequently in a lineage (*Vespula*) distantly related to other vespine genera.

Acknowledgements

We thank Matjaz Kuntner and Adrien Perrard for providing specimens appropriately preserved for RNA extraction. We thank two anonymous reviewers for comments that helped improve our manuscript. This project was made possible by funds generously provided by National Science Foundation grant DEB-0843505 (to K.M.P., J.M.C., and B.A.B.). The computational resources provided by the Vermont Advanced Computing Core, which is supported by NASA (NNX-08AO96G), are gratefully acknowledged.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.10.006.

References

Akre, R.D., Davis, H.G., 1978. Biology and pest status of venomous wasps. Annu. Rev. Entomol. 23, 215–238.
Akre, R.D., Garnett, W.B., Donald, J.F.Mac., Greene, A., Landolt, P., 1976. Behavior and colony development of *Vespula pensylvanica* and *V. atropilosa* (Hymenoptera: Vespidae). J. Kansas Entomol. Soc. 49, 63–84.
Akre, R.D., Greene, A., MacDonald, J.F., Landolt, P.J., Davis, H.G., 1981. The Yellowjackets of America North of Mexico. United States Department of Agriculture, Washington, DC.
Anderson, C., McShea, D.W., 2001. Individual versus social complexity, with particular reference to ant colonies. Biol. Rev. 76, 211–237.
Bequaert, J., 1932. A tentative synopsis of the hornets and yellow-jackets (Vespinae; Hymenoptera) of America. Entomol. Am. 12, 71–138.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114–2120.
Bonckaert, W., Tofilski, A., Nascimento, F.S., Billen, J., Ratnieks, F.L.W., Wenseleers, T., 2011. Co-occurrence of three types of egg policing in the Norwegian wasp *Dolichovespula norwegica*. Behav. Ecol. Sociobiol. 65, 633–640.
Bonckaert, W., Vuerinckx, K., Billen, J., Hammond, R.L., Keller, L., Wenseleers, T., 2008. Worker policing in the German wasp *Vespula germanica*. Behav. Ecol. 19, 272–278.
Bourke, A.F.G., 1999. Colony size, social complexity and reproductive conflict in social insects. J. Evol. Biol. 12, 245–257.
Carpenter, J.M., 1987. Phylogenetic relationships and classification of the Vespinae (Hymenoptera: Vespidae). Syst. Entomol. 12, 413–431.
Carpenter, J.M., 1982. The phylogenetic relationships and natural classificition of the Vespoidea (Hymenoptera). Syst. Entomol. 7, 11–38.
Carpenter, J.M., Kojima, J., 1997. Checklist of the species in the subfamily Vespinae (Insecta: Hymenoptera: Vespidae). Nat. Hist. Bull. Ibaraki Univ. 1, 51–92.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., Sorensen, M.V., Haddock, S.H.D., Schmidt-Rhaesa, A., Okusu, A., Kristensen, R.M., Wheeler, W.C., Martindale, M. Q., Giribet, G., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745–749.
Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.
Evans, H.E., West-Eberhard, M.J., 1970. The Wasps. University of Michigan Press, Ann Arbor, Michigan.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Fischman, B.J., Woodard, S.H., Robinson, G.E., 2011. Molecular evolutionary analyses of insect societies. Proc. Natl. Acad. Sci. U.S.A. 108 (Suppl.), 10847–10854.
Foster, K.R., Ratnieks, F.L.W., 2001a. Paternity, reproduction and conflict in vespine wasps: a model system for testing kin selection predictions. Behav. Ecol. Sociobiol. 50, 1–8.
Foster, K.R., Ratnieks, F.L.W., 2001b. Convergent evolution of worker policing by egg eating in the honeybee and common wasp. Proc. Biol. Sci. 268, 169–174.
Foster, K.R., Ratnieks, F.L.W., Gyllenstrand, N., Thorén, P.A., 2001. Colony kin structure and male production in *Dolichovespula* wasps. Mol. Ecol. 10, 1003–1010.
Freiburger, B.J., Breed, M.D., Metcalf, J.L., 2004. Mating frequency, within-colony relatedness and male production in a yellow jacket wasp, *Dolichovespula arenaria*. Mol. Ecol. 13, 3703–3707.
Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.
Goloboff, P.A., Farris, J.S., Källersjö, M., Oxelman, B., Ramirez, M.J., Szumik, C.A., 2003. Improvements to resampling measures of group support. Cladistics 19, 324–332.
Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. Cladistics 24, 774–786.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotech. 29, 644–652.
Greene, A., 1979. Behavioral characters as indicators of yellowjacket phylogeny (Hymenoptera: Vespidae). Ann. Entomol. Soc. Am. 72, 614–619.
Greene, A., Akre, R.D., Landolt, P.J., 1978. Behavior of the yellowjacket social parasite, *Dolichovespula arctica* (Rohwer) (Hymenoptera: Vespidae). Melanderia 29, 1–28.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.
Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., Wiens, M., Müller, W.E.G., Seaver, E., Wheeler, W.C., Martindale, M.Q., Giribet, G., Dunn, C.W., 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. Biol. Sci. 276, 4261–4270.
Helanterä, H., Tofilski, A., Wenseleers, T., Ratnieks, F.L.W., 2006. Worker policing in the common wasp *Vespula vulgaris* is not aimed at improving colony hygiene. Insectes Soc. 53, 399–402.
Hines, H.M., Hunt, J.H., O'Connor, T.K., Gillespie, J.J., Cameron, S.A., 2007. Multigene phylogeny reveals eusociality evolved twice in vespid wasps. Proc. Natl. Acad. Sci. 104, 3295–3299.
Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.
Johnson, B.R., Borowiec, M.L., Chiu, J.C., Lee, E.K., Atallah, J., Ward, P.S., 2013. Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. Curr. Biol. 23, 2058–2062.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kimsey, L., Carpenter, J., 2012. The Vespinae of North America (Vespidae, Hymenoptera). J. Hymenopt. Res. 28, 37–65.
Kovacs, J.L., Goodisman, M.A.D., 2007. Irregular brood patterns and worker reproduction in social wasps. Naturwissenschaften 94, 1011–1014.
Liu, L., Yu, L., Edwards, S.V., 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10, 302.
Lohse, M., Bolger, A.M., Nagel, A., Fernie, A.R., Lunn, J.E., Stitt, M., Usadel, B., 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res. 40, 622–627.
Loope, K.J., Chien, C., Juhl, M., 2014. Colony size is linked to paternity frequency and paternity skew in yellowjacket wasps and hornets. BMC Evol. Biol. 14, 2625.
Lopez-Osorio, F., Perrard, A., Pickett, K.M., Carpenter, J.M., Agnarsson, I., 2015. Phylogenetic tests reject Emery's rule in the evolution of social parasitism in yellowjackets and hornets (Hymenoptera: Vespidae, Vespinae). R. Soc. Open Sci. 2, 150159.
Lopez-Osorio, F., Pickett, K.M., Carpenter, J.M., Ballif, B.A., Agnarsson, I., 2014. Phylogenetic relationships of yellowjackets inferred from nine loci (Hymenoptera: Vespidae, Vespinae, *Vespula* and *Dolichovespula*). Mol. Phylogenet. Evol. 73, 190–201.
MacDonald, J.F., Matthews, R.W., 1975. *Vespula squamosa*: a yellow jacket wasp evolving toward parasitism. Science 190, 1003–1004.
Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.
Matsuura, M., Yamane, S., 1990. Biology of the Vespine Wasps. Springer-Verlag.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gatew. Comput. Environ. Work. (GCE), 2010.
Mirarab, S., Nguyen, N., Warnow, T., 2014. PASTA: ultra-large multiple sequence alignment. In: Sharan, R. (Ed.), Research in Computational Molecular Biology SE – 15, Lecture Notes in Computer Science. Springer International Publishing, pp. 177–191.
Oi, C.A., Van Oystaeyen, A., Caliari Oliveira, R., Millar, J.G., Verstrepen, K.J., van Zweden, J.S., Wenseleers, T., 2015. Dual effect of wasp queen pheromone in regulating insect sociality. Curr. Biol. 25, 1638–1640.
Perrard, A., Lopez-Osorio, F., Carpenter, J.M., 2016. Phylogeny, landmark analysis and the use of wing venation to study the evolution of social wasps (Hymenoptera: Vespidae: Vespinae). Cladistics 32, 406–425.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.
Pickett, K.M., Carpenter, J.M., 2010. Simultaneous analysis and the origin of eusociality in the Vespidae (Insecta: Hymenoptera). Arthropod Syst. Phylogeny 68, 3–33.
Ratnieks, F.L.W., 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. Am. Nat. 132, 217–236.
Ratnieks, F.L.W., Visscher, P.K., 1989. Worker policing in the honeybee. Nature 342, 796–797.
Reed, H.C., Akre, R.D., 1983. Colony behavior of the obligate social parasite *Vespula austriaca* (Panzer) (Hymenoptera: Vespidae). Insectes Soc. 30, 259–273.
Regier, J.C., Shultz, J.W., Ganley, A.R.D., Hussey, A., Shi, D., Ball, B., Zwick, A., Stajich, J. E., Cummings, M.P., Martin, J.W., Cunningham, C.W., 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst. Biol. 57, 920–938.
Rehan, S.M., Toth, A.L., 2015. Climbing the social ladder: the molecular evolution of sociality. Trends Ecol. Evol. 30, 426–433.
Robinson, G.E., Grozinger, C.M., Whitfield, C.W., 2005. Sociogenomics: social life in molecular terms. Nat. Rev. Genet. 6, 257–270.
Rokas, A., Carroll, S.B., 2006. Bushes in the tree of life. PLoS Biol. 4, e352.
Ross, K.G., 1986. Kin selection and the problem of sperm utilization in social insects. Nature 323, 798–800.
Ross, K.G., 1985. Aspects of worker reproduction in four social wasp species (Insecta: Hymenoptera: Vespidae). J. Zool. 205, 411–424.
Saito, F., Kojima, J., 2011. Phylogenetic analysis and biogeography of the nocturnal hornets, *Provespa* (Insecta: Hymenoptera: Vespidae: Vespinae). Species Divers. 16, 65–74.
Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331.
Salichos, L., Stamatakis, A., Rokas, A., 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. 31, 1261–1271.
Schmieder, R., Edwards, R., 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. PLoS ONE 6, e17288.
Schmieder, R., Lim, Y.W., Edwards, R., 2012. Identification and removal of ribosomal RNA sequences from metatranscriptomes. Bioinformatics 28, 433–435.
Shaw, T.I., Ruan, Z., Glenn, T.C., Liu, L., 2013. STRAW: Species TRee Analysis Web server. Nucleic Acids Res. 41, 238–241.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Siddall, M.E., 2010. Unringing a bell: metazoan phylogenomics and the partition bootstrap. Cladistics 26, 444–452.
Smith, S.A., Dunn, C.W., 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24, 715–716.
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, 150.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
van der Vecht, J., 1957. The Vespinae of the Indo-Malayan and Papuan areas (Hymenoptera, Vespidae). Zool. Verh. 34, 1–83.
van Zweden, J.S., Bonckaert, W., Wenseleers, T., D'Ettorre, P., 2013. Queen signalling in social wasps. Evolution 68, 976–986.
Wenseleers, T., Badcock, N.S., Erven, K., Tofilski, A., Nascimento, F.S., Hart, A.G., Burke, T.A., Archer, M.E., Ratnieks, F.L.W., 2005a. A test of worker policing theory in an advanced eusocial wasp, *Vespula rufa*. Evolution 59, 1306–1314.
Wenseleers, T., Ratnieks, F.W., 2006. Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. Am. Nat. 168, E163–E179.
Wenseleers, T., Tofilski, A., Ratnieks, F.L.W., 2005b. Queen and worker policing in the tree wasp *Dolichovespula sylvestris*. Behav. Ecol. Sociobiol. 58, 80–86.
Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. Annu. Rev. Entomol. 53, 449–472.
Whitfield, J.B., Lockhart, P.J., 2007. Deciphering ancient rapid radiations. Trends Ecol. Evol. 22, 258–265.
Yang, Y., Smith, S.A., 2014. Orthology inference in non-model organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol. Biol. Evol. 31, 3081–3092.