2016

Phylogenetics And Molecular Evolution Of Highly Eusocial Wasps

Federico Lopez-Osorio
University of Vermont

PHYLOGENETICS AND MOLECULAR EVOLUTION OF HIGHLY EUSOCIAL WASPS.

A Dissertation Presented

by

Federico Lopez-Osorio

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Biology

May, 2016

Defense Date: March 22, 2016
Dissertation Examination Committee:

Ingi Agnarsson, Ph.D., Advisor
Stephen R. Keller, Ph.D., Chairperson
Bryan A. Ballif, Ph.D.
Sara Helms Cahan, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

ABSTRACT

Societies where workers sacrifice their own reproduction and cooperatively nurture the offspring of a reproductive queen caste have originated repeatedly across the Tree of Life. The attainment of such reproductive division of labor enabled the evolution of remarkable diversity in development, behavior, and social organization in the Hymenoptera (ants, bees, and wasps). Wasps of the family Vespidae exhibit a gamut of social levels, ranging from solitary to highly social behavior. The highly social yellowjackets and hornets (Vespinae) have well developed differences in form and function between queens and workers, large colony sizes, and intricate nest architecture. Moreover, certain socially parasitic species in the Vespinae have secondarily lost the worker caste and rely entirely on the workers of a host species to ensure the survival of parasitic offspring. Understanding the evolution of behavioral traits in the Vespinae over long periods of time would be greatly enhanced by a robust hypothesis of historical relationships.

In this study, I analyze targeted genes and transcriptomes to address three goals. First, infer phylogenetic relationships within yellowjackets (*Vespula* and *Dolichovespula*) and hornets (*Vespa* and *Provespa*). Second, test the hypothesis that social parasites are more closely related to their hosts than to any other species (Emery's rule). Third, test the protein evolution hypothesis, which states that accelerated evolution of protein coding genes and positive selection operated in the transition to highly eusocial behavior. The findings of this study challenge the predominant understanding of evolutionary relationships in the Vespinae. I show that yellowjacket genera are not sister lineages, instead recovering *Dolichovespula* as more closely related to the hornets, and placing *Vespula* as sister to all other vespine genera. This implies that traits such as large colony size and high paternity are mostly restricted to a particular evolutionary trajectory (*Vespula*) from an early split in the Vespinae. I demonstrate that obligate and facultative social parasites do not share immediate common ancestry with their hosts, indicating that socially parasitic behavior likely evolved independently of host species. Moreover, obligate social parasites share a unique evolutionary history, suggesting that their parasitic behavior might have a genetic component. Lastly, I analyze transcriptomic data to infer a phylogeny of vespid wasps and use this phylogeny to discover lineage-specific signatures of positive selection. I identify more than two hundred genes showing signatures of positive selection on the branch leading to the highly eusocial yellowjackets and hornets. These positively selected genes involve functions related mainly to carbohydrate metabolism and mitochondrial activity, in agreement with insights from studies of bees and ants. Parallels of functional categories for genes under positive selection suggests that at the molecular level the evolution of highly eusocial behavior across the Hymenoptera might have followed similar and narrow paths.

CITATIONS

Material from this dissertation has been published in the following form:

Lopez-Osorio, F., Pickett, K. M., Carpenter, J. M., Ballif, B. A., and Agnarsson, I.. (2014). Phylogenetic relationships of yellowjackets inferred from nine loci (Hymenoptera: Vespidae, Vespinae, *Vespula* and *Dolichovespula*). Molecular Phylogenetics and Evolution 73, 190–201.

Lopez-Osorio, F., Perrard, A., Pickett, K. M., Carpenter, J. M., and Agnarsson, I.. (2015). Phylogenetic tests reject Emery's rule in the evolution of social parasitism in yellowjackets and hornets (Hymenoptera: Vespidae, Vespinae). Royal Society Open Science 2, 150159.

ACKNOWLEDGMENTS

I am grateful to the late Kurt M. Pickett for his support and encouragement. Dr. Pickett's exceptional contributions to the study of phylogenetics and behavior of social wasps were an inspiration for the work I conducted throughout my doctoral studies. I am thankful to my advisor, Ingi Agnarsson, for his dedicated mentorship and guidance. I am grateful to my parents and brothers for their unconditional love. I am indebted to my committee members Bryan A. Ballif, Sara Helms Cahan, and Stephen Keller for their time and attention. James M. Carpenter kindly provided valuable advice for the completion of the work presented here. Kevin Loope, Adrien Perrard and Matjaz Kuntner generously offered me specimens. I am thankful to Marina Golivets for her love, support, and friendship. I thank my friends Fernando Gelin, Andrew Nguyen, Michael Herrmann, Nabil Nasseri, Yainna Hernaiz, and Lucia Orantes for all the great memories and conversations that enhanced my understanding of phylogenetics, evolutionary biology, and behavioral ecology.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1: INTRODUCTION

Cooperative behavior and self-sacrificing altruism evolved repeatedly in insects (Wilson 1971; Anderson 1984). In the social Hymenoptera (ants as well as certain bees and wasps), colony members belong to castes that perform specific tasks. The reproductive caste includes queens and males, whereas the mostly sterile workers raise offspring, forage, and guard the colony (Wilson 1971; Bourke and Franks 1995). Moreover, workers usually switch from nest activities to foraging through their lifespan; that is, workers exhibit temporal polyethisms (Wilson 1976; Seeley 1982; Jeanne 1991). This division of labor is considered a hallmark in the ecological success of hymenopteran societies (Wilson 1985). The Vespidae is a lineage of wasps that transitioned from solitary to social behavior once (Carpenter 1982; Pickett and Carpenter 2010). Within vespids, the paper wasp genus *Polistes*, yellowjackets (*Vespula* and *Dolichovespula*), and hornets (*Vespa* and *Provespa*) are among the most well-studied and all belong to eusocial subfamilies with reproductive division of labor, cooperative care of brood, and overlapping generations. Their social complexity, however, varies. In contrast to the primitively eusocial *Polistes*, colonies of the highly eusocial yellowjackets and hornets typically comprise hundreds to thousands of workers and have morphologically distinct castes (Evans and West-Eberhard 1970). These morphological differences between castes indicate a "point of no return" to a solitary or primitively eusocial condition (Wilson and Holldobler 2005; Wilson 2008).

The Vespinae have been the focus of numerous evolutionary studies due to the diversity of their natural history and behavioral traits (e.g., Foster et al. 1999, 2000, 2001; Foster and Ratnieks 2001a,b; Wenseleers et al. 2005a, b;

Helanterä et al. 2006; Goodisman et al. 2007a, b; Bonckaert et al. 2008; Kovacs et al. 2010; Loope et al. 2014; Oi et al. 2015). Vespine wasps occur throughout the oriental tropics and temperate regions in the Northern Hemisphere (Spradbery 1973). Across this geographic range, yellowjackets and hornets establish aerial or subterranean nests consisting of layers of paper that enclose combs (horizontal sections) suspended from one another (Evans and West-Eberhard 1970; Akre and Davis 1978). Vespine wasps have small- and large-colony species (Akre et al. 1981), and such variation in colony size correlates with patterns of reproduction and conflict in accordance with kin selection predictions (Foster and Ratnieks 2001b; Loope et al. 2014). Kin selection theory states that relatedness among individuals influences selection, and that an indirect fitness component received from effects on the reproduction of others favors the evolution of altruistic traits (Hamilton 1964a,b; Queller and Strassmann 2002; Foster et al. 2006; Strassmann et al. 2011). Kin selection has been particularly emphasized in the study of altruism in the social Hymenoptera because of their haplodiploid sex determination. In haplodiploid species, females developed from fertilized eggs, whereas males develop from haploid eggs. As a result, full sisters share more genes with each other (75%) than they would with their own offspring (50%). Workers, therefore, should invest in the survival of their sisters, rather than in the production of their own female offspring. Moreover, when queens mate only once, workers are more related to nephews (sons of workers) than to their brothers (queen's sons), and thus worker production of males is expected. On the contrary, multiple paternity results in little reproduction by workers (Ratnieks 1988). Among vespine wasps, large-colony species of *Vespula* exhibit high paternity, few workers with activated ovaries, and absence of worker

reproduction (Akre et al. 1976; Ross 1985, 1986; Foster and Ratnieks 2001b). Lack of worker reproduction in large-colony yellowjackets usually results from worker policing, where workers remove or eat worker-laid eggs to preserve the reproductive dominance of the queen (Foster and Ratnieks 2000, 2001a; Bonckaert et al. 2008). In contrast to large-colony species of *Vespula*, species of *Dolichovespula* have small colonies with low paternity and exhibit conflict between queens and workers over the production of males (Foster et al. 2001). Reproductive conflict in the Vespinae may occur between species, too. Vespine nests typically have a single founding queen devoted exclusively to reproduction for most of her lifespan. The growth of vespine colonies, however, may be interrupted by social parasites (MacDonald and Matthews 1975; Jeanne 1977). These socially parasitic species have secondarily lost the worker caste and rely on the workers of a host species to raise parasitic offspring. Altogether, the Vespinae show considerable variation in behavioral traits, and elucidating the origins of such traits over deep evolutionary time will benefit from a robust hypothesis of phylogenetic relationships.

The work I present here investigates the phylogenetic relationships of yellowjackets and hornets. The predominant hypothesis of vespine phylogeny was proposed by Carpenter (1987), who analyzed morphological characters and found that yellowjackets were monophyletic and sister to *Provespa*, and recovered *Vespa* as the sister group of the remaining Vespinae. Another phylogenetic study, however, reported that yellowjackets were more closely related to a hornet clade (*Vespa* and *Provespa*) (Pickett and Carpenter 2010). More recently, and subsequent to published findings from the first half of this dissertation, Perrard et al. (2015) analyzed morphological and molecular data

and found that yellowjackets were monophyletic and sister to *Vespa*, but also reported considerable lack of support for genus-level relationships. In Chapter One, I analyze nine targeted genes and evaluate the relationships of yellowjackets and hornets. In Chapter Two, I use a phylogenetic approach to test Emery's rule, which indicates that social parasites are more closely related to their hosts than to any other species.

Although the natural history of several vespid species is well known (Spradbery 1973; Ross and Matthews 1991; Hunt 2007), comprehensive genetic analyses of social wasps have burgeoned only in recent years, particularly focused on caste differences within *Polistes* (Toth et al. 2007; Berens et al. 2015a; Patalano et al. 2015). Caste determination has been thought to occur predominantly as a response to environmental factors, such as nutrition of larvae or rearing temperature, rather than due to genetic differences (Wilson 1976, 1985; O'Donnell 1998; Hölldobler and Wilson 2008). But division of labor also has a genetic component that varies in strength and is widespread across social insects (Anderson et al. 2008; Goodisman et al. 2008; Robinson et al. 2008; Smith et al. 2008; Schwander et al. 2010; Bloch and Grozinger 2011; Lattorff and Moritz 2013). Recently, queen–worker and worker–worker differences in form and function have been associated with patterns of gene expression, which originate from the interaction of genotype and environment during development (Evans and Wheeler 2001; Ben-Shahar et al. 2002; Whitfield et al. 2003, 2006; Ingram et al. 2005; Pereboom et al. 2005; Drapeau et al. 2006; Gräff et al. 2007; Grozinger et al. 2007). The study of a model species such as the honey bee has delivered tremendous genomic resources and enabled the identification of various genes underlying division of labor (Whitfield et al. 2003; Weinstock et al. 2006; Smith et

al. 2008), which is a prime endeavor of sociogenomics (Robinson et al. 2005; Rehan and Toth 2015; Kapheim 2016). Moreover, the advent of high-throughput RNA sequencing (RNA-Seq) or transcriptomics (Morozova et al. 2009; Wang et al. 2009; Cahais et al. 2012), availability of assembled genomes from bees, ants, and *Polistes* (Gadau et al. 2012; Simola et al. 2013; Oxley et al. 2014; Kapheim et al. 2015; Patalano et al. 2015; Sadd et al. 2015; Smith et al. 2015a), and novel findings of genes linked to caste-specific behavior (Hoffman and Goodisman 2007; Toth et al. 2007, 2010; Cardoen et al. 2011b; Ferreira et al. 2013; Feldmeyer et al. 2014; Woodard et al. 2014) now serve as impetus for elucidating broad molecular patterns in other social insects, such as the vespine wasps. A hypothesis of particular interest indicates that accelerated evolution of specific protein coding genes or gene families contributed to the origin of highly eusocial behavior (Fischman et al. 2011; Woodard et al. 2011; Simola et al. 2013). In Chapter Three, I analyze a transcriptomic data set of vespine wasps to reassess the phylogenetic relationships of vespine genera. In Chapter Four, I use single-copy genes to test the protein evolution hypothesis for the origin of highly eusocial behavior in the Vespinae.

CHAPTER 2: PHYLOGENETIC RELATIONSHIPS OF YELLOWJACKETS INFERRED FROM NINE GENES (HYMENOPTERA: VESPIDAE, VESPINAE, *VESPULA* AND *DOLICHOVESPULA*)

2.1. Introduction

Social wasps of the genera *Vespula* and *Dolichovespula*, or yellowjackets, have similar life history characteristics, but also exhibit considerable diversity in their social behavior (Greene, 1991). Throughout latitudes of the Northern Hemisphere, yellowjacket queens initiate colonies alone, and their elaborate nests, which might be aerial or subterranean, consist of levels or combs enclosed in layers of paper (Akre et al., 1981). Queens are larger than workers and both castes also differ in shape, physiology, and behavior (Akre and Davis, 1978; Jeanne, 1980). Larvae and adults frequently engage in mouth-to-mouth feeding (i.e., trophallaxis), a prominent behavior of most advanced wasp societies (Roskens et al., 2010; Spradbery, 1973). Their colonial life also displays considerable conflict and diversity in reproductive behavior (Foster and Ratnieks, 2001b). Conflict over male parentage among the queen and reproductive workers is commonly resolved by forceful prevention of worker reproduction, or policing (Ratnieks and Visscher, 1989; Wenseleers et al., 2004). The queen or workers can enforce policing by physical aggression toward egg-laying workers or by eating worker-laid eggs, maintaining the reproductive primacy of the queen (Bonckaert et al., 2011, 2008; Foster and Ratnieks, 2000; Freiburger et al., 2004; Goodisman et al., 2002; Helanterä et al., 2006; Wenseleers et al., 2005a,b). Furthermore, queens of obligate social parasite species, or inquilines, which lack the worker caste, instigate interspecific conflict by entering

the nest of a host species, killing the resident queen, and enslaving the host workers (Jeanne, 1977; Reed and Akre, 1983b). In contrast, facultative social parasites such as *Vespula squamosa* can produce their own workers, but frequently usurp colonies of other species (Hoffman et al., 2008; MacDonald and Matthews, 1975).

Given this wealth of sophisticated behavioral traits present in a relatively small number of species (about 48 species of yellowjackets are currently recognized), a robust phylogeny inferred from different sources of evidence is a high priority. However, only two formal studies have addressed the evolutionary history of yellowjackets. First, Carpenter (1987) conducted a cladistic analysis of the genera of the subfamily Vespinae, including *Vespa* (hornets) and *Provespa* (nocturnal hornets) in addition to *Vespula* and *Dolichovespula*. Using morphological and behavioral characters (including data of Yamane (1976) and Matsuura and Yamane (1984)), Carpenter's (1987) analysis supported yellowjackets as a clade sister to *Provespa*, placing *Vespa* sister to the remaining vespine genera (Fig. 2.1a). Second, Carpenter and Perera (2006), again using morphological and behavioral characters, also found that *Vespula* and *Dolichovespula* are monophyletic and presented relatively well-resolved relationships within each genus. Greene (1979) discussed yellowjacket relationships on the basis of behavioral characters, but his arguments were non-cladistic.

In contrast, the use of molecular characters to elucidate the evolutionary history of yellowjackets has been limited and, for the most part, peripheral. For example, Collins and Gardner (2001) analyzed a fragment of cytochrome b from six species of bees and wasps, including one hornet and two yellowjacket

species, which were recovered as sister taxa in their results. As part of a study of allergen characterization of paper wasp venom, Pantera et al. (2003) used amino acid sequences of antigen 5 and provided a neighbor-joining dendrogram, showing *Vespula* as closer to *Vespa* than to *Dolichovespula*. Hines et al. (2007) analyzed sequence data from four nuclear genes including three vespine wasps. Their results showed *Vespula squamosa* as more closely related to *Dolichovespula maculata* than to *Vespula maculifrons*. As part of a new species description, Landolt et al. (2010) performed parsimony and neighbor-joining analyses of 905 bp of mitochondrial DNA from seven yellowjacket species and found two sister clades corresponding to the *Vespula vulgaris* and *V. rufa* species groups. Pickett and Carpenter (2010) conducted a direct optimization (Wheeler, 1996) analysis of four loci, combined with morphology and behavior, to elucidate the phylogeny of the family Vespidae. Pickett and Carpenter's (2010) study included nine vespine species and found that *Vespa* is sister to *Provespa* and these two genera are sister to *Vespula* and *Dolichovespula* (Fig. 2.1b). More recently, Saito and Kojima (2011) investigated the relationships among species of *Provespa* using information from three loci and phenotypic data. The two yellowjacket species included in this study were sister taxa and more closely related to *Provespa* than to *Vespa*.

Here, we perform a phylogenetic analysis of yellowjackets on the basis of a new, comprehensive molecular data set. We generate 5.5 kb of DNA sequence from five mitochondrial and four nuclear loci, including both protein-coding and ribosomal fragments. The standard markers we chose have variable rates of evolution that are expected to provide resolution at different hierarchical levels.

Our taxon sampling encompasses part of the diversity of yellowjackets across their north temperate distribution. Using a strategy of marker concatenating and partitioning, we test the monophyly of *Vespula* and *Dolichovespula* and their sister relationship as yellowjackets. Furthermore, we infer the relationships within each yellowjacket genus and among species groups. Finally, we discuss the implications of the new phylogeny for the evolution of behavior and morphological features.

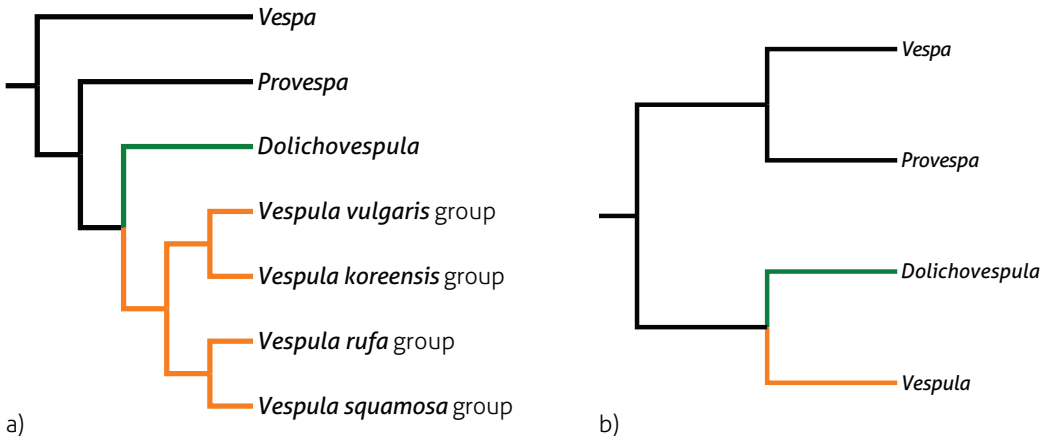

Figure 2.1: Previous hypotheses of generic and species-group relationships within Vespinae: (a) Carpenter's (1987) phylogeny in which *Vespula* and *Dolichovespula* are sister genera and both more closely related to *Provespa*; (b) Pickett and Carpenter's (2010) hypothesis also shows that *Vespula* and *Dolichovespula* are sister genera but more closely related to a clade that includes *Provespa* and *Vespa*.

2.2. Materials and methods

2.2.1. Taxon sampling

Five Holarctic species were recognized in the world checklist of Vespinae (Carpenter and Kojima, 1997). Recent taxonomic studies, however, have found diagnostic differences between the Old and New World forms of species previously considered Holarctic (Carpenter and Glare, 2010; Carpenter et al., 2011; Kimsey and Carpenter, 2012). These changes in the taxonomy of yellowjackets are adopted here, and thus Holarctic species are not recognized, with the exception of the European *Vespula germanica*, which is widely introduced. Sequences were obtained for a total of 28 species. Following the comprehensive analysis of Pickett and Carpenter (2010), four outgroup species were chosen from the Polistinae, the putative sister subfamily of Vespinae. Six more vespine outgroup species from *Provespa* and *Vespa* were included. Currently, there are 48 recognized species of yellowjackets (21 *Dolichovespula* and 27 *Vespula*), of which 18 were included in this analysis. Our analysis increases taxon sampling from five to 18 yellowjacket species in comparison to Pickett and Carpenter (2010). Moreover, the ingroup taxa in this study represent six species groups. These are the *Vespula rufa*, *V. vulgaris*, *V. squamosa*, *Dolichovespula maculata*, *D. norwegica*, and *D. sylvestris* groups (Archer 1999; Carpenter, 1987; Carpenter and Perera 2006).

2.2.2. DNA extraction, amplification, and sequencing

One leg and one antenna were removed from absolute ethanol-preserved specimens and the rest of each specimen was kept as a voucher. Legs and antennae were macerated with sterile plastic pestles and genomic DNA was

extracted using the DNeasy Blood & Tissue Kit (Qiagen) with an incubation period of 48 hours at 55ºC in lysis buffer and Proteinase K, and in other respects following the manufacturer's instructions. Loci from mitochondrial and nuclear genomes were selected for their variability and resolving power at different levels. The genes used in this study are 12S ribosomal DNA (12S), 16S ribosomal DNA (16S), cytochrome oxidase I (COI), cytochrome oxidase II (COII), cytochrome b (Cytb), 28S ribosomal DNA D2-D3 expansion regions (28S), elongation factor 1 α F2 copy (EF1α), RNA polymerase II (Pol II), and wingless (wg).

Fragments of these genes were amplified using the Polymerase Chain Reaction (PCR) on an Eppendorf Mastercycler Thermal Cycler and employing the primers listed on Table 2.1. Each PCR consisted of 22µL of nuclease-free dH_2O, 1µL of 10µM forward primer, 1µL of 10µM reverse primer, and 1µL of genomic DNA extract. The 25µL total volume was added to PuReTaq Ready-To-Go PCR beads (GE Healthcare). A typical PCR program started with 4 minutes of initial denaturation at 94 ºC, followed by 35-40 cycles of 30 seconds at 94 ºC, 45 seconds of annealing at 43–58 ºC, and 45 seconds of elongation at 72 ºC, and ended with a six minute period of final elongation at 72 ºC. PCR products were verified on 1% agarose/TBE electrophoresis gels. PCR product purification and standard Sanger sequencing were outsourced to Beckman Coulter Genomics and Macrogen USA. Sequencing was conducted with the same primers used for PCR amplification. Contigs were assembled from forward and reverse ABI chromatograms and trimmed of low-quality ends using Geneious 6 (Biomatters Ltd.). Upon inspection of agarose gels, chromatograms, and descriptive sequence statistics, no obvious symptoms of nuclear copies of mitochondrial genes (numts)

were found (e.g., PCR ghost bands, in-frame stop codons, unconstrained variability across codon positions (Bensasson et al., 2001; Calvignac et al., 2011)). All edited sequences were submitted to BLAST searches to screen for contamination.

Table 2.1: List of primer sequences (and their respective annealing temperatures) used for PCR amplification of yellowjackets and outgroups. The same primers were used for sequencing.

Primer	Sequence (5′ to 3′)	PCR temp ºC	Source
12S		43	
12S ai	AAA CTA GGA TTA GAT ACC CTA TTA T		Simon et al. (1994)
12S bi	AAG AGC GAC GGG CGA TGT GT		Simon et al. (1994)
16S		46	
16S ar	CGC CTG TTT ATC AAA AAC AT		Simon et al. (1994)
16S br	CTC CGG TTT GAA CTC AGA TCA		Simon et al. (1994)
COI		45	
LCO1490	GGT CAA CAA ATC ATA AAG ATA TTG G		Folmer et al. (1994)
HCO2198	TAA ACT TCA GGG TGA CCA AAA AAT CA		Folmer et al. (1994)
HCOoutout	GTA AAT ATA TGR TGD GCT C		Prendini et al. (2005)
Jerry	CAA CAT TTA TTT TGA TTT TTT GG		Simon et al. (1994)
COI-5	AAT TGC AAA TAC TGC ACC TAT TGA		Saito and Kojima (2011)
COII		45	
E2	GGC AGA ATA AGT GCA TTG		Garnery et al. (1992)
COII1-2	ATT TTA TAC CAC AAA TTT CTG AAC ATT G		Saito and Kojima (2011)
Cytb		46	
CB1	TAT GTA CTA CCA TGA GGA CAA ATA TC		Jermiin and Crozier (1994)
CB2	ATT ACA CCT CCT AAT TTA TTA GGA AT		Jermiin and Crozier (1994)
28S		48-52	
For28SVesp	AGA GAG AGT TCA AGA GTA CGT G		Hines et al. (2007)
Rev28SVesp	GGA ACC AGC TAC TAG ATG G		Hines et al. (2007)
EF1α		57	
F2-557F	GAA CGT GAA CGT GGT ATY ACS AT		Brady et al. (2006)
F2-1118R	TTA CCT GAA GGG GAA GAC GRA G		Brady et al. (2006)
HaF2For1	GGG YAA AGG WTC CTT CAA RTA TGC		Danforth et al. (1999)
F2-rev1	AAT CAG CAG CAC CTT TAG GTG G		Danforth et al. (1999)
Pol II		52	
polfor2a	AAY AAR CCV GTY ATG GGT ATT GTR CA		Danforth et al. (2006)
polrev2a	AGR TAN GAR TTC TCR ACG AAT CCT CT		Danforth et al. (2006)
wg			
beewgFor	TGC CAN GTS AAG ACC TGY TGG ATG AG	58	Danforth et al. (2004)
Lepwg2a	ACT CGC ARC ACC ART GGA ATG TRC A		Danforth et al. (2004)

2.2.3. Sequence alignment, data partitioning, and model selection

Each gene was aligned independently using MAFFT v.7 (Katoh, 2002; Katoh and Standley, 2013) with a gap opening penalty (--op) value of 1.53, an offset cost equal to 0.123, and automatic strategy selection. The data were partitioned corresponding to mitochondrial ("mtDNA": 12S, 16S, COI, COII, and Cytb) and nuclear ("nuDNA": 28S, EF1α, Pol II, wg) genes and combined into a single matrix ("AllData"). These matrices were assembled using SequenceMatrix (Vaidya et al., 2011). *Vespula flaviceps* was excluded from nuDNA because it was not possible to obtain nuclear, protein-coding sequences for this species. Considering that third codon positions are prone to substitution saturation (Swofford et al. 1996; see, however, Kälersjö et al. 1999), we separated all (both mitochondrial and nuclear) protein-coding genes into subsets including first and second codon positions on one hand ("Pos1&2") and third positions on the other ("Pos3"). Codon-position matrices were created using Mesquite 2.75 (Maddison and Maddison, 2011). To assess the influence of another potentially confounding factor in phylogenetic inference (Sanderson and Shaffer, 2002), nucleotide composition was evaluated for genes and partitions by conducting Chi-square tests of homogeneity of base frequencies across taxa using PAUP* 4.0b10 (Swofford, 2002). High AT bias was found in the mtDNA (AT 75.86% $p < 0.001$), Pos3 (AT 78.38% $p < 0.001$), and AllData (AT 65.47% $p < 0.001$) partitions (Table 2.2). In contrast, the null hypothesis of homogeneity in base composition was not rejected for the nuDNA (AT 52.05%, $p = 0.99$) and Pos1&2 (AT 60%, $p = 0.89$) partitions (Table 2.2). It is worth mentioning that nucleotide composition did not vary greatly from taxon to taxon, and thus any possible confounding effects of high AT content should be less drastic (Simon et al., 1994).

Highest AT content in Pos3 is in accordance with the prevailing use of codons ending in A and T in hymenopteran genomes (Behura and Severson, 2012). Likewise, other studies have found that AT composition bias in mitochondrial genes, either protein coding or ribosomal, is widespread in the Hymenoptera (Dowton and Austin, 1995, 1997). Moreover, in a recent phylogenetic analysis of hymenopteran superfamilies, Heraty et al. (2011) reported AT bias in the third codon position of EF1α and, to a much greater extent, COI (AT 90.3%). Base composition heterogeneity in part of our data motivated one more partition composed of first and second codon positions and nuclear rDNA ("Pos1&2+28S"); that is, excluding mitochondrial rDNA and third codon positions of nuclear and mitochondrial protein-coding genes. Pos1&2+28S showed homogeneous base frequencies (AT 56.27%, $p = 0.71$, Table 2.2). Lastly, the complete data set was also analyzed in a statistical framework (see below) partitioning by gene and codon position; that is, defining four partitions corresponding to three rDNA genes (12S, 16S, 28S) and 21bp of tRNALeu adjacent to COII, and 18 partitions in which each codon position of every protein-coding gene formed a partition. This last partitioning scheme is referred to as "AllData22" hereafter.

Models of nucleotide substitution were selected among 56 candidate models using jModelTest 2 (Darriba et al., 2012; Guindon and Gascuel, 2003) according to the Akaike Information Criterion corrected for sample size (AICc). The best-fit substitution models for mitochondrial genes were HKY + I + Γ and GTR + I + Γ (Table 2.2). The use of I + Γ, however, has been criticized due to strong correlation between the proportion of invariant sites and the gamma distribution, hence causing unreliable parameter estimation (Yang, 2006, p. 113-

114; see also Sullivan et al., 1999 and Stamatakis, 2006, RAxML 7.0.4 manual, section 6). In most analyses the simpler Γ model was preferred (that is, the proportion of invariant sites was excluded), since a gamma distribution with α smaller than one already accounts for sites with very low rates (Yang, 2006). Nonetheless, AllData and mtDNA were also analyzed using I + Γ to explore the influence of the I + Γ mixture on topology and clade posterior probabilities. None of the best-fit models chosen for nuclear genes included I + Γ (Table 2.2).

2.2.4. Phylogenetic inference

Parsimony analyses were performed using TNT (Goloboff et al., 2008) treating gaps as missing data (nstates nogaps). The heuristic search strategy consisted of 5000 random addition sequences with TBR branch swapping followed by ratchet (Nixon, 1999) saving two trees per replication (mult 5000 =tbr ratchet hold 2). Group support was calculated using 10000 replications of symmetric resampling (resample sym replic 10000) with default search settings and the results summarized as GC (Group present/Contradicted) values, which show the difference in frequency between a given group and the most frequent group that contradicts it (Goloboff et al., 2003). Parsimony analyses were conducted for single- and multi-gene matrices.

For model-based inference of phylogeny, the program MrBayes 3.2 (Ronquist et al., 2012) was accessed through the CIPRES Science Gateway (Miller et al., 2010) to run Bayesian analyses of each gene matrix, data subsets, and all genes combined. All analyses were run for 50M generations with sampling every 1000 generations, the number of runs was 4, and the default number of chains

was used (nruns=4 nchains=4). Likelihood models were set according to the AICc criterion, specifying the number of substitution types and model for among-site rate variation, and allowing MrBayes to estimate base frequencies, the substitution rates (GTR model) or transition/transversion ratio (HKY model), and the gamma distribution shape parameter. In the Bayesian analyses of mtDNA, nuDNA, and AllData, each gene formed a partition, models were specified for each gene, and parameters were unlinked across partitions (e.g., revmat, statefreq, and shape); in the analyses that included I + Γ the proportion of invariant sites (pinvar) was also unlinked. In the individual analyses of Pos1&2, Pos3, and Pos1&2+28S, each data set formed a single partition. For AllData22 each codon position of every protein-coding gene formed a partition and individual rDNA genes formed the remaining partitions. Stationarity of Markov chains was assessed by examining MrBayes' parameter output files in Tracer v1.5 (Rambaut and Drummond, 2007) as well as the Potential Scale Reduction Factor convergence diagnostic ~1.0 and average standard deviation of split frequencies <0.001. Moreover, the 'compare' command in AWTY (Nylander et al., 2008) was used to evaluate convergence of the posterior probabilities of all splits for paired MCMC runs. The default burn-in of 25% used in all analyses was adequate to discard samples before reaching convergence (usually within 2M to 4M generations). A maximum likelihood analysis on the complete concatenated data set performed in GARLI 2.0 (Zwickl, 2006) gave essentially the same results as the parsimony analysis of AllData and the Bayesian analysis of mtDNA, and therefore will not be discussed further.

2.3. Results

2.3.1. Sequence alignment and model selection

Out of the 252 possible DNA sequences, 242 (96%) were successfully amplified and sequenced. We deposited all sequences in GenBank under the following accession numbers: KJ147175 - KJ147201 (12S), KJ147202 - KJ147228 (16S), KJ147229 - KJ147256 (COI), KJ147257 - KJ147284 (COII), KJ147285 - KJ147312 (Cytb), KF981692 - KF981717 (28S), KF955639 - KF955665 (wg), KF981641 - KF981665 (Pol II), and KF981666 - KF981691 (EF1α). The protein coding COI, COII, Cytb, EF1α, Pol II, and wg aligned unambiguously without internal indels. For the 12S and 16S rDNA markers, differences in sequence length among species were generally small (8-32 bp) and alignments were thus unambiguous. The best-fit substitution models for individual genes were those with six substitution types and among-site rate variation (e.g., GTR + G), except for 12S and wg where the best fit models were HKY + G and K80 + G, respectively (Table 2.2). Simple models, with a single type of substitution and equal rates, characterized most first and second positions of nuclear, protein-coding genes (e.g., JC or F81), whereas the first and second codon positions of mitochondrial genes had more complex models (Table 2.2).

2.3.2. Inferred phylogeny

The single most-parsimonious tree (MPT) found for all nine genes combined (AllData) is presented in Fig. 2.2. According to this molecular "total evidence" hypothesis, yellowjackets (*Vespula* + *Dolichovespula*) form a natural group that is sister to the hornets (*Vespa*). Moreover, within *Vespula* two major clades correspond to species groups: the *rufa* group (with the *squamosa* group as

its sister) and the *vulgaris* group. Within *Dolichovespula*, the *maculata* group is a clade, while the *sylvestris* group (one species) is sister to the *norwegica* group. Simultaneous analyses generated fewer trees in comparison to single-gene analyses (Table 2.3). The same single MPT found for all concatenated genes (Fig. 2.2) was recovered independently for mtDNA and nuDNA. Symmetric-resampling support trees for mtDNA, nuDNA, and AllData are presented in Fig. 2.3; Fig. 2.3c summarizes the results of parsimony analyses for single genes and all data subsets. The three main analyses of concatenated data (Fig. 2.3a-c) provide strong support for the monophyly of each vespine genus, but relationships among genera are unresolved or poorly supported. In the support tree for all mitochondrial genes (Fig. 2.3a) *Vespula* and *Dolichovespula* are part of a trichotomy, and for nuclear genes (Fig. 2.3b) and all genes concatenated (Fig. 2.3c) the monophyly of *Vespula* + *Dolichovespula* is weakly supported. Similarly, individual genes also support the monophyly of each vespine genus, but do not provide clear resolution to the relationships among genera (Fig. 2.3c). For example, the COII gene tree was the only marker supporting the *Vespula* + *Dolichovespula* clade. In the gene tree for 28S (strict consensus of six equally parsimonious trees), *Vespa*, *Vespula* and *Dolichovespula* were found monophyletic, yet all were part of a polytomy and therefore the yellowjackets are unresolved for 28S (Fig. 2.3c).

Bayesian inference (BI) resulted in conflicting inferences between mtDNA and nuDNA (Fig. 2.4a,b), particularly in the resolution of supraspecific relationships. The Bayesian analysis of all mitochondrial genes supports *Vespula* and *Dolichovespula* as sister genera together sister to *Provespa* + *Vespa*. However, BI of the nuclear data set indicated that *Dolichovespula* is more closely related to

Provespa + *Vespa* than to *Vespula*. Likewise, BI of AllData (in which each gene formed a partition) depicted the ((*Provespa* + *Vespa*), *Dolichovespula*) clade (Fig. 2.4c), and with higher posterior probability (PP) in comparison to the Bayesian topology for nuDNA. Using only the Γ rate variation model did not cause any striking changes in comparison to using Γ + I. The Bayesian consensus trees found using Γ and Γ + I were the same for AllData and mtDNA, differing slightly in support values for relationships among genera.

The result of the most heavily partitioned Bayesian analysis, AllData22 (in which each codon position of every protein-coding gene formed a partition), is presented in Fig. 2.5. This majority consensus tree also shows *Dolichovespula* as more closely related to (*Provespa* + *Vespa*), although with much lower support (PP = 60) in comparison to partitioning only by gene (Fig 2.4c). There are some concerns about the analysis of AllData22 (Fig. 2.5), however, specifically related to estimation of the shape parameter of the gamma distribution for three partitions. After visually scrutinizing the estimates of all parameters from multiple runs of AllData22 using Tracer v1.5, considerably high mean values and large-scale fluctuations in the trace (suggesting poor mixing) were found for the gamma shape of partitions corresponding to the third codon position of EF1α, Pol II and wg; although the corresponding ESS values did not indicate problems. These findings suggest problems with 'over partitioning' the data (Brown and Lemmon, 2007; Leavitt et al., 2013; Rota and Wahlberg, 2012). Additional analyses were conducted altering various default settings in MrBayes for the problematic partitions (e.g., increasing the effort to update the gamma shape parameter (propset), augmenting the number of gamma categories, changing the starting values (startvals), placing a shorter uniform prior (shapepr)), but these

modifications did not improve the mixing behavior. Every other parameter in the analysis of AllData22 (and all other Bayesian analyses) reached convergence rapidly, showed adequate mixing, good ESS values, and plausible estimates (e.g., partition rate multipliers *m* higher for third positions than first or second positions). Furthermore, the AWTY 'compare' plots indicated convergence of posterior probabilities of all splits for AllData22. Because of this behavior of certain parameters in the AllData22 partitioning scheme, we prefer the results from BI partitioning only by gene.

Table 2.2: Summary statistics for genes and data subsets and their corresponding models of nucleotide substitution. Empirical base frequencies estimated in PAUP* (Swofford, 2002). Models of nucleotide substitution chosen according to the sample-size corrected Akaike Information Criterion as implemented in jModelTest 2 (Darriba et al., 2012).

Data set	Aligned sites	A (%)	C (%)	G (%)	T (%)	# Parsimony informative sites	AICc best-fit model
12S	378	43.18	3.95	12.39	40.48	152	HKY+I+G
16S	532	39.68	7.52	14.10	38.70	153	GTR+I+G
COI	1096	31.47	15.22	12.81	40.50	405	GTR+I+G
COI Pos 1	365	33.70	14.87	22.23	29.20	95	GTR+I+G
COI Pos 2	365	16.65	23.50	15.26	44.59	23	TVM+I
COI Pos 3	366	44.03	7.32	0.96	47.69	287	HKY+I+G
tRNA-Leu + COII	21 + 582	36.37	14.88	7.84	40.91	253	GTR+I+G
COII Pos 1	194	40.00	16.96	14.67	28.37	69	TrN+G
COII Pos 2	194	27.54	21.45	9.07	41.94	28	F81+G
COII Pos 3	194	42.19	7.38	0.52	49.91	146	HKY+I+G
Cytb	433	32.16	15.80	9.40	42.64	186	GTR+I+G
Cytb Pos 1	144	33.85	16.62	17.16	32.37	46	TPM1uf+I+G
Cytb Pos 2	144	23.93	20.91	10.39	44.77	17	TrN+I
Cytb Pos 3	145	38.64	9.93	0.69	50.74	123	HKY+G
28S	750	20.03	27.77	32.61	19.59	66	GTR+I
EF1α	517	29.14	21.92	23.43	25.51	109	TrN+G
EF1α Pos 1	172	31.33	15.18	36.56	16.93	7	F81
EF1α Pos 2	172	31.46	26.23	15.10	27.21	3	JC
EF1α Pos 3	173	24.67	24.34	18.65	32.34	99	HKY+G
Pol II	798	36.24	15.01	19.96	28.79	107	TrN+G
Pol II Pos 1	266	37.50	15.27	28.97	18.26	6	TrN
Pol II Pos 2	266	33.88	19.50	16.52	30.10	0	F81
Pol II Pos 3	266	37.35	10.26	14.38	38.01	101	TrN+G
wg	406	25.54	25.78	28.00	20.68	89	K80+G
wg Pos 1	135	27.13	24.21	32.55	16.11	13	JC
wg Pos 2	135	33.68	17.26	26.65	22.41	6	JC
wg Pos 3	136	15.86	35.81	24.83	23.50	70	K80+G
mtDNA	3042	35.26	12.68	11.46	40.6	1149	GTR+I+G
nuDNA	2471	28.08	22.12	25.83	23.97	371	GTR+I+G
Pos1&2	2552	30.36	19.15	19.89	30.60	313	GTR+I+G
Pos3	1280	36.36	13.39	8.23	42.02	826	GTR+G
Pos1&2+28S	3302	28.09	21.04	22.69	28.18	379	GTR+I+G
AllData	5513	32.10	16.80	17.73	33.37	1520	GTR+I+G

Table 2.3: Summary of parsimony and Bayesian analyses of single genes, major data subsets and all data. MPT = Most Parsimonious Tree(s), CI = Consistency Index, RI = Retention Index. CI and RI calculated for strict consensus trees when multiple equally parsimonious trees were found.

Data set	CI	RI	# MPT(s)	Length	-Ln L
12S	0.519	0.686	8	520	2752.69
16S	0.452	0.650	32	574	3267.196
COI	0.351	0.476	3	1805	8889.506
COII	0.380	0.527	7	1105	5420.516
Cytb	0.295	0.325	1	884	4159.107
28S	0.836	0.929	6	107	1691.682
EF1α	0.740	0.889	6	212	1811.12
Pol II	0.763	0.917	2	168	2046.669
wg	0.781	0.886	3	186	1559.394
mtDNA	0.345	0.445	1	4952	24636.411
nuDNA	0.554	0.734	1	681	7078.413
Pos1&2	0.476	0.672	2	1018	8799.368
Pos3	0.397	0.555	2	3337	14939.359
Pos1&2+28S	0.507	0.701	4	1136	10682.508
AllData	0.374	0.500	1	5646	31962.13

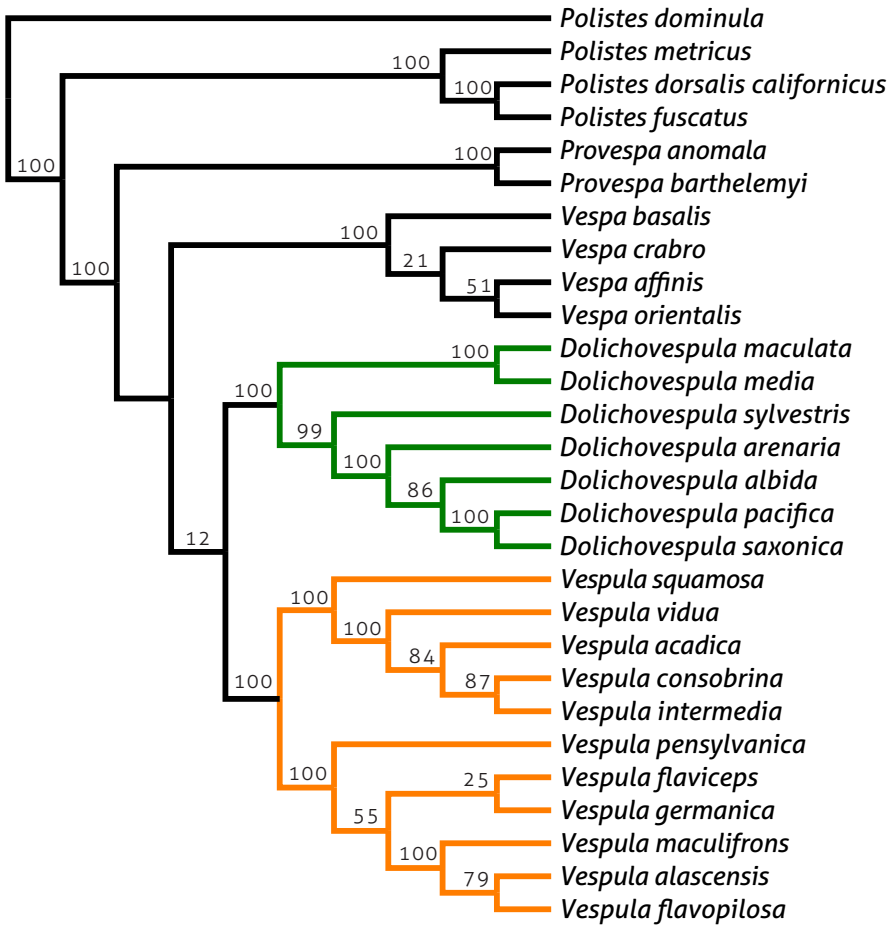

Figure 2.2: Single most-parsimonious tree found using a concatenated matrix of all data.

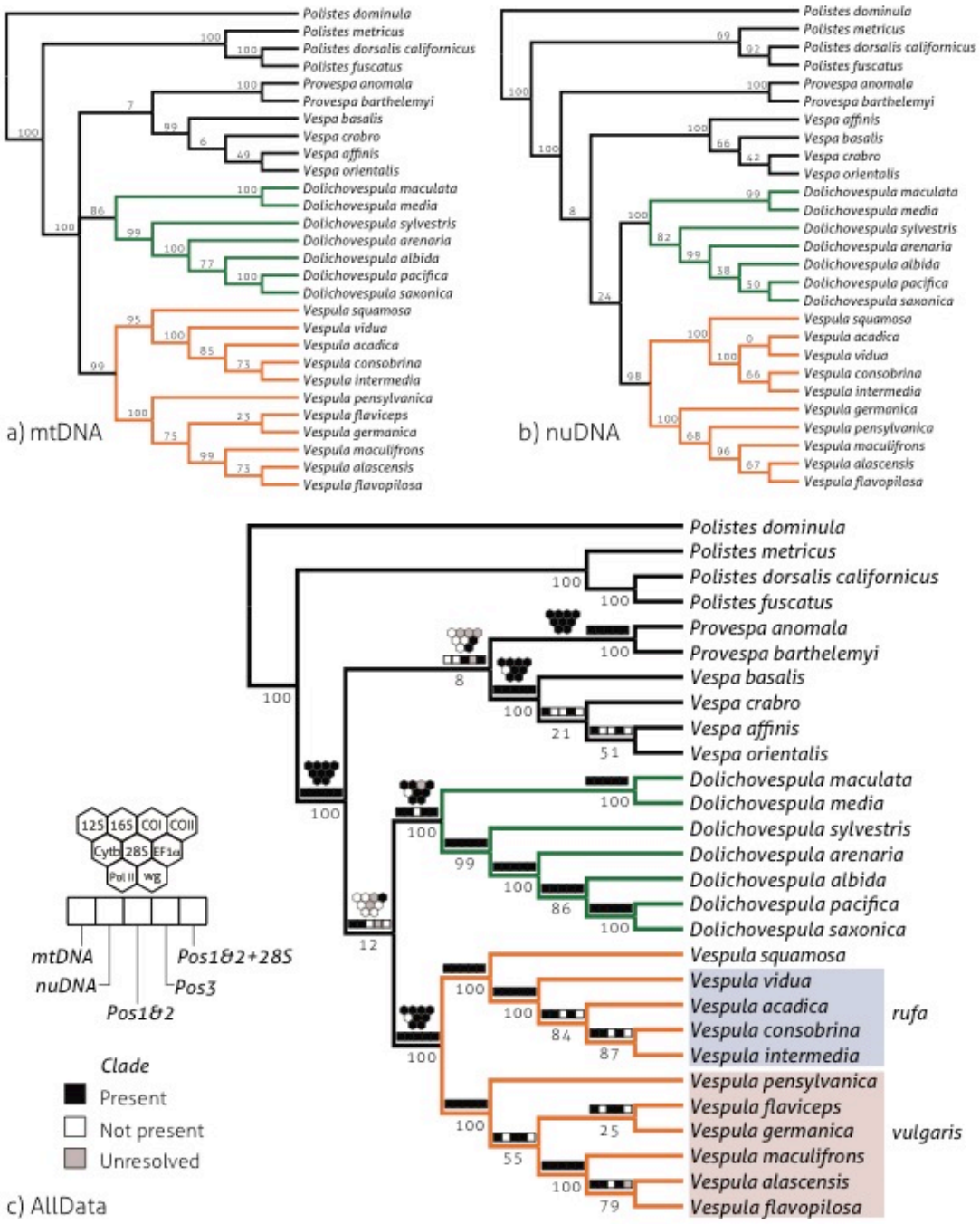

Figure 2.3: Parsimony support trees using symmetric resampling: (a) mtDNA, which includes all mitochondrial genes; (b) nuDNA, which includes all nuclear genes; and (c) AllData, in which all genes were concatenated into a single matrix. GC frequencies are presented above branches in (a) and (b) and below branches in (c). Squares and hexagons above branches in (c) are presented for each genus and above-genus level relationships, showing presence of a given clade for a particular data subset (e.g., mtDNA) or gene (e.g., 12S), respectively.

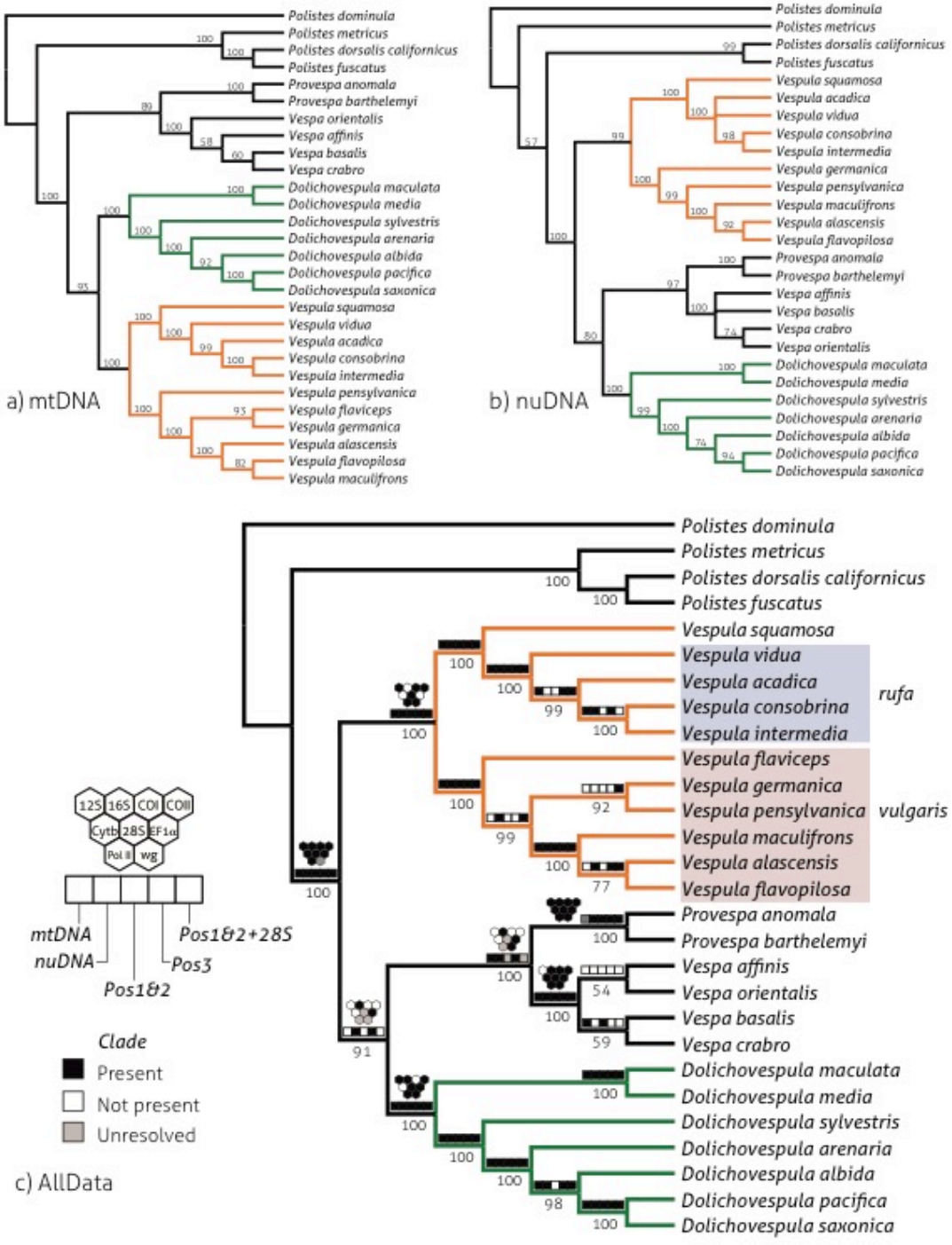

Figure 2.4: Bayesian majority consensus trees based on gene partitions: (a) mtDNA, which includes all mitochondrial characters; (b) nuDNA, which includes all nuclear genes; and (c) AllData, which includes the entire data. Values above branches are clade posterior probabilities (x 100). Squares and hexagons above branches in (c) are presented for each genus and above-genus level relationships, showing presence of a given clade for a particular data subset or gene, respectively.

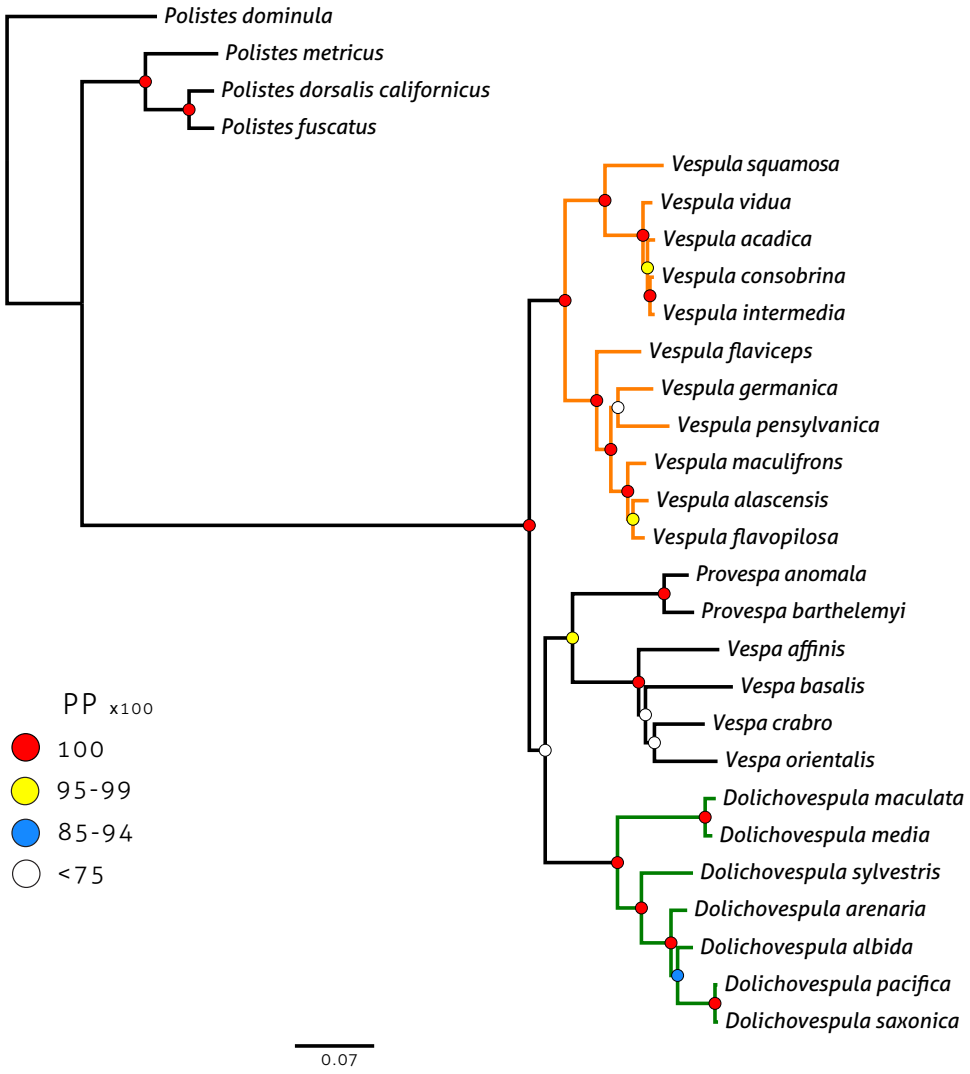

Figure 2.5: Bayesian phylogram based on the most heavily partitioned analysis (AllData22).

2.4. Discussion

2.4.1. Vespine phylogeny

Our molecular data and analyses strongly support the monophyly of the genera *Vespula* and *Dolichovespula* but either reject or weakly support a yellowjacket clade. The most parsimonious tree inferred using the entire data indicate that yellowjackets, *Vespula* + *Dolichovespula*, are monophyletic (Fig. 2.2). This grouping, however, is poorly supported and rarely recovered in the parsimony analyses of individual genes (Fig. 2.3c). The *Vespula* + *Dolichovespula* clade was also found in the Bayesian analysis of mtDNA, but this relationship is not recovered by BI based on nuclear genes and it erodes in partitioned Bayesian analyses of the entire data (Fig. 2.4c and Fig. 2.5). Incongruence between single- and multigene trees is expected, for single genes evolve under unique sets of characteristics and functional constraints (Miyamoto and Fitch, 1995). Although the simultaneous parsimony analysis of the entire data yielded a single, fully resolved MPT, poor support for the sister group relationship of *Vespula* and *Dolichovespula* suggest either conflicting character interactions (Ramirez, 2005), or lack of evidence. Whether poor group support from our molecular data set forecasts that the *Vespula* + *Dolichovespula* clade will be contradicted in subsequent studies of vespine phylogeny can only be answered by analyzing more sequences and other sources of evidence (e.g., morphology, behavior). Nevertheless, given the molecular data at hand here, we argue that the monophyly of *Vespula* + *Dolichovespula* requires further examination.

Partitioned Bayesian analyses of all genes indicate that *Dolichovespula* is sister to the hornets, *Provespa* + *Vespa*. Even though our data includes more mitochondrial than nuclear characters (Table 2.3), in Bayesian analyses the signal

of mitochondrial characters is overturned by a stronger signal from fewer nuclear characters. That is, nuclear genes contribute more to the resolution of deeper nodes, which seems to be a broad pattern in simultaneous phylogenetic analysis of mtDNA and nuDNA (Fisher-Reid and Wiens, 2011). This is not surprising given that mitochondrial genes evolve in concert and can be characterized as a single data partition, while the evolution of nuclear genes is more decoupled. Nonetheless, mtDNA alone provides valuable characters for inferring phylogenetic relationships. In our simultaneous analyses of mtDNA using different methods (Fig. 2.3a and Fig. 2.4a), the current classification of four vespine genera (Carpenter, 1987) is recovered with very strong support for the monophyly of each genus and well-resolved species-level relationships.

The sister group relationship between *Dolichovespula* and the hornets is novel for DNA sequence-based studies of vespine phylogeny (Fig. 2.3c). A sister relationship between *Dolichovespula* and *Vespa* was suggested by Greene (1979) based on behavioral traits and by Schmitz and Moritz (1990) using RFLP patterns, but their analyses had substantial drawbacks (Carpenter, 1987, 1992). Greene (1979) argued that *Dolichovespula* is more closely related to *Vespa* due to their shared, primitive social organization. According to Greene (1979), a closer relationship between *Dolichovespula* and *Vespa* seems plausible because both genera display low degree of queen-worker dimorphism, smaller worker:reproductive output ratios, frequent cell wall scraping by larvae, and a royal court of workers surrounding the queen. However, as stated by Greene (1979) himself, the characters he is using to support the arrangement of *Dolichovespula* + *Vespa* are symplesiomorphies in Vespinae, and therefore uninformative as evidence of kinship within the subfamily; shared primitive

traits do not inform on relationships (Hennig, 1965). Moreover, Carpenter (1987, pp. 416-421; see also Matsuura and Yamane 1984) did not find a closer relationship between *Dolichovespula* and *Vespa* in his phylogenetic study, which included colony size, royal court, and larval hunger signal (i.e., cell wall scraping) as binary characters, among other behavioral and morphological data. The sister group relationship between *Dolichovespula* and *Provespa + Vespa* recovered in our Bayesian analyses based on nuDNA and all genes (Fig. 2.4) disagrees with the vespine clade of Pickett and Carpenter (2010). In their total-evidence analysis, Pickett and Carpenter (2010) found support for yellowjacket monophyly (*Dolichovespula + Vespula*). The closer placement of *Dolichovespula* to hornets is moderately supported when the entire data is partitioned by gene (PP = 91) and poorly supported when partitioned by codon position (PP = 60) (Fig. 2.4c and Fig. 2.5).

The monophyly of genera *Vespula* and *Dolichovespula* and species-level relationships within each yellowjacket genus (Fig. 2.3c and Fig. 2.4c) are in agreement with the results of Carpenter (1987) and Carpenter and Perera (2006), and show improved resolution in comparison to the latter study. Furthermore, relationships within each genus are largely concordant between parsimony and BI. In the study of Carpenter and Perera (2006), the relationships within the *rufa* group (sister taxa of *V. squamosa*) were unresolved, but our results depict the following resolution: (*V. vidua* (*V. acadica* (*V. consobrina + V. intermedia*))). These relationships are well supported given our taxon sampling (Fig. 2.3c and Fig. 2.4c). Our results agree with Carpenter (1987) in the placement of the facultative social parasite *V. squamosa* as sister to the *rufa* group (Fig. 2.3c and Fig. 2.4c). Thus, our findings further contradict the hypothesis of MacDonald and

Matthews (1975, 1984) that *V. squamosa* is more closely related to the *vulgaris* group on the basis of nest architecture and behavioral characters. Carpenter (1987) showed that the features described by MacDonald and Matthews (1975, 1984) to support their hypothesis were either plesiomorphic in Vespinae or autapomorphic, and therefore phylogenetically uninformative. Relationships within the *vulgaris* group in our results are less clear, with differences in resolution between methods (Fig. 2.3c and Fig. 2.4c). Our results agree with Carpenter and Perera (2006) in that *V. flavopilosa*, *V. alascensis*, and *V. maculifrons* form a clade, but contradict their placement of *V. flaviceps* within the same group. Carpenter and Glare (2010) likewise found *V. flavopilosa*, *V. alascensis* and *V. maculifrons* forming a clade, together with *V. vulgaris*. The relationships within *Dolichovespula* are congruent between parsimony and BI and concordant with the results of Carpenter and Perera (2006), who found two subgeneric clades: the *maculata* group, which also includes *D. media*, and another group composed of the remaining species (Fig. 2.3c and Fig. 2.4c).

Our results from partitioned Bayesian analyses of all genes showed that increasing the number of partitions may lead to considerable changes in clade support, a phenomenon that has also been reported for other taxa (Castoe et al., 2004; Castoe and Parkinson, 2006; Dowton et al., 2009; Li et al., 2008; Nylander et al., 2004; Mueller et al., 2004; Powell et al., 2013). Along these lines, incrementing the number of partitions caused convergence and mixing problems for certain parameters in our most heavily partitioned Bayesian analysis (AllData22). But the influence of increasing the number of partitions on phylogeny was minimal, since most relationships found with nine and 22 partitions were the same, except for the resolution within *Vespa* (Fig. 2.4c, Fig. 2.5). Rota and Wahlberg (2012)

reported similar issues with convergence and mixing related to *a priori* partitioning in their phylogenetic study of metalmark moths.

2.4.2. Species groups

Within the *Dolichovespula* clade, two species groups can be recognized: the *maculata* and *norwegica* groups (Fig. 2.3 and Fig. 2.4). The *maculata* group, *maculata* + *media*, was recovered in gene trees as well as with multigene data sets, regardless of phylogeny inference method (Fig. 2.3 and Fig. 2.4). In the results of Carpenter and Perera (2006) (see also Carpenter, 1987) three synapomorphies are attributed to the *maculata* group: pronotal striae, emarginate apex of the seventh metasomal sternum in males, and aedeagal medial lobes. The other clade within *Dolichovespula* corresponds, for the most part, to the *norwegica* group *sensu* Archer (1999, 2006), although *D. pacifica*, which Archer (1999, 2006) places in a separate group, is also nested within the *norwegica* group (Fig. 2.3 and Fig. 2.4). Moreover, this *Dolichovespula* clade also includes the *sylvestris* group, as its sister. Archer (1999) indicated that females having a long oculo-malar space and lateroanterior clypeal angles with less prominent semicircular projections characterize the *norwegica* group. Within *Vespula*, the *rufa* and *vulgaris* groups are concordant between parsimony and BI based on mtDNA, nuDNA, and the entire data (Fig. 2.3 and Fig. 2.4). The *rufa* and *squamosa* groups are supported by four synapomorphies according to Carpenter and Perera (2006): dorsum of metasomal tergum I with slight depression behind anterior edge, shortened volsella, slender and fingerlike digitus, and dark hairs in metasomal tergum I. The *vulgaris* group shares at least nine derived characters, such as volsella with dorsal lobe,

aedeagus with subcircular apex, and "large" colony size (3,500 cells and 500 workers or more) (Carpenter and Perera, 2006).

2.4.3. Evolution of behavior

Variation in body or group size has important consequences on life history, physiology, and behavior across taxa (Bell and Mooers, 1997; Blueweiss et al., 1978; Bonner, 1988, 2004; Dornhaus et al., 2011; Karsai and Wenzel, 1998; McShea, 1996). For example, size tends to be positively related to organismal complexity, which refers to the number or functional specialization of parts (Bell and Mooers, 1997; Carroll, 2001; McShea, 1996). The relationship between size and complexity is analogous in both social insects and multicellular organisms, in which task specialization evolved from solitary or unicellular ancestors and covaries with size (Bell and Mooers, 1997; Bonner, 2004; Holbrook et al., 2011; Jeanson et al., 2007; Ratcliff et al., 2012; Simpson, 2012; Strassmann and Queller, 2007). For example, individual cells form colonies such as *Volvox*, comprising thousands of tightly linked cells, a small number of which specialize in reproduction; that is, *Volvox* colonies have some degree of division of labor. The cells forming the *Volvox* colony are interdependent to an extent that cells die in isolation and the organism cannot survive if the colony is disrupted. Similarly, among social insects, colony members subdivide labor, so that queens specialize in reproduction and workers carry out other duties, and the differentiation between both castes is more striking in larger colonies (Bourke 1999; but see Wenzel 1992).

Although their colonies can contain hundreds of thousands of individuals (Pickett et al., 2001 and references therein), yellowjackets have been

traditionally divided into two categories of colony size (Akre et al., 1981; Greene, 1991, p. 269). First, a small-colony category of mature nests with fewer than 2,500 cells and 75-400 workers encompasses most species of *Dolichovespula* and all species in the *rufa* group (Akre and Davis, 1978). Most species in the small-colony category are characterized by rearing of workers on a single comb, short colony life span, and larval nutrition strictly based on live arthropod prey (Akre et al., 1981; Greene, 1991; Reed and Akre, 1983a; see also Carpenter 1989). In contrast, the large-colony category includes the *V. vulgaris* group and *V. squamosa* (Greene, 1991, p. 270). Yellowjackets in the large-colony category build nests containing more than 2,500 cells and have population sizes of 500 to 5,000 workers or more (Akre and Davis, 1978; Spradbery, 1971). Moreover, large-colony yellowjackets build several worker-cell combs, have longer colony duration, and feed their brood with various food sources including live prey, fruit, and, perhaps more distinctively, carrion (Akre and Davis, 1978; Akre et al., 1981; Greene, 1991). Colony size is a key trait in the evolution of social hymenopterans because it explains a large amount of social complexity, including high degree of caste dimorphism and task specialization, lack of queen-worker conflict over reproduction, and reduced potential for worker reproduction (Bourke, 1999; Anderson and McShea, 2001).

However, colony size is not the only key determinant of social complexity, since kin structure is equally important. In vespine wasps, kin structure can be described by a single variable: effective paternity, which is defined by queen mating frequency and distribution of sperm (Foster and Ratnieks, 2001b; hereafter we use 'paternity' as shorthand for effective paternity). In colonies of *Dolichovespula* and *Vespula rufa* (and perhaps other species in the

rufa group), low paternity (< 2) is the norm, whereas colonies of species in the *V. vulgaris* and *squamosa* groups are characterized by high paternity (> 2) (Foster and Ratnieks, 2001c; Wenseleers et al., 2005a). High paternity in eusocial Hymenoptera is a prominent derived trait that lowers relatedness among workers and is associated with exclusive production of males by queens, lack of active ovaries in workers, and worker policing (Akre et al., 1976; Bonckaert et al., 2008; Boomsma and Ratnieks, 1996; Foster et al., 1999; Foster and Ratnieks, 2001a; Goodisman et al., 2002; Helanterä et al., 2006; Kovacs and Goodisman, 2007; Ratnieks, 1988; Ross, 1985; Strassmann, 2001). Moreover, high paternity promotes colony productivity (Cole, 1999; Goodisman et al., 2007; Mattila and Seeley, 2007) and induces workers to rear their brothers (queen's sons) rather than their nephews (workers' sons) (Foster and Ratnieks, 2001b; Ratnieks, 1988). Large colony size and high mating frequency (paternity > 2) in yellowjackets might be convergent traits that have evolved in the branch leading to the *vulgaris* group and in *V. squamosa* (Fig. 2.3 and Fig. 2.4). Alternatively, both traits might have evolved in the most common recent ancestor of all *Vespula* species and lost or suppressed in the *rufa* group. The positive association between colony size and paternity seems to hold across all social Hymenoptera when controlling for phylogeny (Jaffe et al., 2012). Bourke (1999) proposes the following feedback between colony size and low reproductive potential of workers, the latter being an outcome of high paternity. To begin with, in large colonies worker policing is common and workers have a low chance of reproducing. It follows that workers engage exclusively in tasks beneficial to the colony instead of attempting to reproduce, thus increasing colony performance (Ratnieks, 1988). Rising productivity then favors selection for worker policing. Consequently, queen

specialization in reproduction is enhanced and greater worker productivity allows the rearing of more individuals, eventually leading to larger colonies.

2.5. Future research

In general, our analyses provide moderate support for the monophyly of *Dolichovespula* and *Vespula* and suggest that the sister relationship between both genera warrants further examination. A thorough evaluation of the possible close relationship of *Dolichovespula* to the hornets should be addressed including more species of *Vespa*, which in the present study might have been an underrepresented outgroup. Recently, Perrard et al. (2013) have analyzed the phylogeny of the genus *Vespa* in considerable detail. Since some of the molecular markers employed by Perrard et al. (2013) are shared with our study, merging and analyzing characters from both studies may be a new starting point for further investigations of vespine phylogeny. Ideally, new studies should increasingly focus on nuclear markers, since these seem to provide more resolution of deeper nodes and thus should help the inference of generic relationships within the Vespinae. In pilot work, we found that 18S is easily amplified but uninformative within Vespinae and long-wavelength rhodopsin is difficult to amplify and sequence in some taxa. Among genes commonly used in hymenopteran phylogenetics, CAD (rudimentary) seems to be particularly reliable and informative, and therefore a good candidate for new phylogenetic studies of yellowjackets. Regarding the evolution of behavior, colony size is an interesting trait that has been phylogenetically analyzed as a discrete, binary character based on *ad hoc* character states. However, a more powerful approach would be to analyze colony size as well as paternity as continuous variables, thus

giving greater emphasis to the trait values of each species. This is possible using parsimony (Goloboff et al., 2006) or statistical comparative methods (Harvey and Pagel, 1991; O'Meara, 2012).

CHAPTER 3: PHYLOGENETIC TESTS REJECT EMERY'S RULE IN THE EVOLUTION OF SOCIAL PARASITISM IN YELLOWJACKETS AND HORNETS (HYMENOPTERA: VESPIDAE, VESPINAE)

3.1. Introduction

Division of labor and elaborate brood care are hallmarks of insect societies (Wilson 1971, 1985). Societies of ants, bees, and wasps typically comprise a reproductive queen, sterile (or less reproductive) workers and males. The worker caste specializes in provisioning the larvae and foraging, among other tasks (Oster and Wilson 1978). Cooperative brood care underlies the success of social hymenopterans, but is also vulnerable to exploitation. For example, lycaenid butterfly larvae employ chemical and sound mimicry to dupe worker ants into carrying them into the brood chambers of the ant nests, where the workers feed the caterpillars (Akino et al. 1999; Als et al. 2004; Barbero et al. 2009). This type of exploitation may be more easily enabled between close relatives because of their compatible communication systems and kin recognition cues. In an intriguing offshoot of sociality, socially parasitic hymenopterans have evolved a variety of strategies to deceive other species into caring for their young (Wheeler 1919; Buschinger 1986, 1990, 2009; Wcislo 1987; Davies et al. 1989; Hölldobler and Wilson 1990; Bourke and Franks 1991; Lenoir et al. 2001; Brandt et al. 2005; Cervo 2006; Huang and Dornhaus 2008; Kilner and Langmore 2011). Queens of facultative social parasites generally usurp established nests, kill the resident queen and produce workers to gradually replace the host worker force. In contrast, most obligate social parasites, or inquilines, lack the worker caste altogether. Inquiline queens, unable to found their own colonies, invade the nests

of other species and trick the conquered occupants into raising the parasitic brood, which develops into queens and males.

The evolution of social parasitism has been linked with close phylogenetic relationships. Motivated by the observed morphological affinities between parasitic species and their hosts, Emery (1909) conjectured that socially parasitic ants are more closely related to their hosts than to any other species. This generalization, which has since become known as Emery's rule, has been explained according to two evolutionary scenarios. On the one hand, the intraspecific or sympatric speciation hypothesis proposes that social parasites may originate directly from their hosts (West-Eberhard 1986; Buschinger 1990; Bourke and Franks 1991). Alternatively, the interspecific or social deception hypothesis claims that two species may evolve from geographically isolated populations (i.e., allopatrically) and parasitic habits develop when the populations come back together (Wilson 1971; Ward 1989; Hölldobler and Wilson 1990). In testing these two hypotheses, finding that social parasites and their hosts are sister taxa would be a necessary condition for invoking sympatric speciation. Moreover, lack of immediate common ancestry between social parasites and their hosts would be sufficient to rule out sympatric speciation. The validity of the sympatric speciation model of social parasitism remains contentious, with studies of certain ants favoring the model (Savolainen and Vepsäläinen 2003; Jansen et al. 2010; Rabeling et al. 2014), and absence of support for Emery's rule in other social Hymenoptera (Ward 1996, 1989; Carpenter et al. 1993; Agosti 1994; Choudhary et al. 1994; Carpenter 1997,Sumner et al. 2004a; Carpenter and Perera 2006; Hines and Cameron 2010; Gibbs et al. 2012; Smith et

al. 2013). Some of the latter studies, however, support a relaxed version of Emery's rule, that is, parasites and hosts are close relatives, but not sister taxa.

Phylogenetic analyses of inquiline wasps and their hosts seldom support the strict Emery's rule, instead finding that inquilines are monophyletic (Carpenter et al. 1993; Choudhary et al. 1994; Carpenter 1997; Carpenter and Perera 2006). In social wasps, parasitic behavior has been documented in paper wasps (Polistinae) and yellowjackets and hornets (Vespinae). The subfamily Vespinae, among its 70 recognized species, includes five species of inquilines and two facultative social parasites, most of which occur in the yellowjacket genera, *Dolichovespula* and *Vespula*. Two previous studies have assessed the veracity of Emery's rule in yellowjackets. First, Varvio-Aho et al. (1984; see also Pamilo et al. 1981) analyzed allozymes from eight species and reported that the inquilines *Vespula austriaca* and *Dolichovespula omissa* were sister to their hosts, therefore supporting Emery's rule. Upon reanalysis of Varvio-Aho et al.'s (1984) data, however, Carpenter (1987) found that the characters were largely uninformative and *D. omissa* was not sister to its host. Second, Carpenter and Perera (2006) performed a cladistic analysis of yellowjackets based on morphological and behavioral characters and recovered the inquilines *Dolichovespula adulterina* and *D. omissa* as sister taxa, thus rejecting Emery's rule. Similarly, the obligate and facultative social parasites of *Vespula* were not sister to their respective hosts (Carpenter and Perera 2006).

However, these previous phylogenetic studies of parasites and their hosts in vespine wasps were based on relatively few data and lacked resolution. For example, the analysis of Carpenter and Perera (2006) resulted in an inquiline clade as part of a polytomy with other *Dolichovespula* species. A well-resolved

phylogeny is essential for elucidating the evolution of predisposing traits that may explain why inquilinism occurs primarily in certain taxa. Such traits can be size of reproductives, nestmate recognition signals (van Wilgenburg et al. 2011), mating frequency (Sumner et al. 2004b), and sterility-inducing queen pheromones (Van Oystaeyen et al. 2014), to name a few. Here, we carry out the first molecular phylogenetic analysis of social parasites and their hosts in yellowjackets and hornets. Our study includes the inquilines *Dolichovespula adulterina*, *D. arctica* and *D. omissa*, and the facultative social parasites *Vespula squamosa* and *Vespa dybowskii*. These are five of the seven known social parasites in the Vespinae. We infer the relationships among these taxa and their hosts based on the analysis of 12 gene fragments to test two mutually exclusive hypotheses. First, social parasites evolved sympatrically from their hosts, and therefore Emery's rule in its strict sense is applicable in vespine wasps. Second, inquilinism has evolved only once in *Dolichovespula*, and thus the three inquiline species of *Dolichovespula* are monophyletic. Moreover, we discuss our results in terms of a 'relaxed Emery's rule' in which for any clade of social parasites the most closely related outgroup clade includes the host species (Buschinger 1990; Ward 1996).

3.2. Materials and methods

3.2.1. Taxonomic sampling

We assembled a set of 38 species from all genera in the Vespinae and spanning the distribution range of the subfamily. We included the following parasitic species and their hosts, which are enclosed in parentheses: the Palearctic *Dolichovespula adulterina* (*D. saxonica*, *D. norwegica*; Weyrauch 1937; Dvořák 2007),

D. omissa (*D. sylvestris*; Weyrauch 1937), and *Vespa dybowskii* (*V. simillima*, *V. crabro*; Sakagami and Fukushima 1957; Archer 1992), and the Nearctic *D. arctica* (*D. arenaria*, *D. alpicola*; Wheeler and Taylor 1921; Taylor 1939; Jeanne 1977; Greene et al. 1978; Wagner 1978) and *Vespula squamosa* (*V. maculifrons*, *V. vidua, V. flavopilosa*, *V. germanica*; MacDonald and Matthews 1975, 1984; Matthews and Matthews 1979; MacDonald et al. 1980; Hoffman et al. 2008).

3.2.2. DNA extraction, amplification, and sequencing

Extraction, amplification and sequencing protocols follow Lopez-Osorio et al. (2014). Briefly, we extracted genomic DNA using the DNeasy Blood & Tissue Kit (Qiagen) and conducted PCR amplification using PuReTaq Ready-To-Go PCR beads (GE Healthcare). We sequenced fragments of seven mitochondrial genes and five nuclear markers: 12S and 16S ribosomal DNA (*12S*, *16S*), cytochrome oxidase I and II (*COI*, *COII*), ATPase subunit 8 and 6 (*ATP8*, *ATP6*), cytochrome b (*Cytb*), 28S ribosomal DNA D2-D3 expansion regions (*28S*), elongation factor 1 alpha F2 copy (*EF1*), RNA polymerase II (*Pol II*), wingless (*Wg*), and rudimentary (*CAD*). Three of these genes (*CAD*, *ATP8*, *ATP6*) were not used in Lopez-Osorio et al. (2014). We amplified *CAD* with primers CD892F and CD1491R from Ward et al. (2010) and developed primers C2-J3661 (5′ – TTG GWC AAT GYT CWG AAA TTT GTG G) and A6-N4543 (5′ – CCA GCA WTT ATW TTA GCT GAT AAT CG) to amplify a region spanning the mitochondrial genes *ATP8* and *ATP6* – primers were labeled according to their positions in the *D. yakuba* mitogenome (Clary and Wolstenholme 1985). The PCR program for this primer pair was 35 cycles of 30s at 94ºC, 30s at 55ºC and 45s at 72ºC, preceded by 4min at 94ºC and followed by 6min at 72ºC.

Forward and reverse sequences were assembled into contigs and trimmed of low-quality ends in Geneious 6.1.7 (Biomatters Ltd). The sequences generated with the new primer pair were annotated using the MITOS WebServer (Bernt et al. 2013). Although the region amplified with primers C2-J3661 and A6-N4543 also spans the *trnK* and *trnD* genes, these sequences were not included in downstream analyses because of their short length and lack of variability. We aligned sequences with MAFFT v7.017 using the automatic strategy selection (Katoh and Standley 2013), removed introns of *CAD* and indel regions of *ATP8* and *Wg*, and concatenated gene matrices using SequenceMatrix (Vaidya et al. 2011). The concatenated alignment used in all analyses contains 418 sequences; 238 of these were previously published (Lopez-Osorio et al. 2014) and the remaining sequences were generated for this study (GenBank accessions KT225582–KT225591, KT250513–KT250524, KT257109–KT257164 and KT273417–KT273481).

3.2.3. Phylogenetic analyses

We performed parsimony analyses of single genes and the concatenated data using TNT (Goloboff et al. 2008). The search strategy in all cases consisted of 5000 replicates using random sectorial searches, drifting, ratchet and fusing combined (xmult=rss fuse 5 drift 5 ratchet 10). In all searches gaps were treated as missing data. Group support was calculated with 5000 replicates of symmetric resampling and the results were summarized with GC (Group present / Contradicted) frequencies.

We employed three partitioning strategies in maximum likelihood (ML) and Bayesian analyses of the concatenated data: 1) assigning each gene to a

separate subset; 2) defining each codon position in each protein-coding gene as a character set, in addition to three blocks of rDNA genes, resulting in 30 subsets; and 3) submitting these 30 predefined subsets to PartitionFinder v1.0.1 (Lanfear et al. 2012) to find the best-fit partitioning scheme and choose substitution models. In the greedy search with PartitionFinder, branch lengths were set to unlinked, 56 different models were compared for each subset, and models were selected according to the Akaike Information Criterion corrected for sample size (AICc). In the former two partitioning methods, substitution models were chosen with the AICc as implemented in jModeltest v2.1.7 (Darriba et al. 2012). In all cases, when the model chosen was not compatible with MrBayes, the closest available model was used.

ML analyses of the concatenated data were carried out using the OpenMP and MPI versions of GARLI v2.01 (Zwickl 2006). ML analyses consisted of 100 search replicates with default settings except for topoweight = 0.01 and brlenweight = 0.002. These two deviations from default settings were also employed in ML bootstrap analyses, which consisted of 500 pseudoreplicates.

Bayesian analyses of single genes and the concatenated data were conducted using MrBayes v3.2.3 (Ronquist et al. 2012) on CIPRES (Miller et al. 2010) with nucmodel = 4by4, nruns = 2, nchains = 8, and samplefreq = 1000. Unconstrained MCMC analyses were run for 40 M generations using the different partitioning schemes, whereas constrained analyses (see below) were carried out for 20 M generations employing the character subsets identified by PartitionFinder. Base frequencies, substitution rates, the gamma shape parameter, and proportion of invariable sites were unlinked across subsets. We set a shorter prior on the mean branch length – brlenspr=unconstrained:exp(100)

– to address the long-tree problem of partitioned analyses in MrBayes (Marshall 2010). We assessed convergence by examining effective sample size (ESS) values with Tracer v1.6 (Rambaut et al. 2013) and the potential scale reduction factor (PSRF) for all parameters in MrBayes. In all analyses of the concatenated data, stationarity was reached in less than four million generations.

3.2.4. Constraint analyses and topology tests

We conducted constraint analyses to quantify the difference in likelihoods between unconstrained and constrained topologies. Eight constraints enforcing host-parasite monophyly were evaluated: each social parasite sister to its primary host in separate topologies, resulting in five constraint trees; all five parasites sister to their respective hosts; all inquilines sister to their corresponding hosts; and an unresolved clade of inquilines and hosts. Mean marginal likelihoods of unconstrained and constrained models were calculated using stepping-stone sampling (Xie et al. 2011) in MrBayes and employing the partitioning scheme identified by PartitionFinder. Stepping-stone analyses consisted of 31 M total cycles, four independent runs of four parallel chains each, sampling every 1000 generations and using 30 steps to yield 1000 samples within each step (α=0.4). The first 25% samples of each step were discarded as burn-in. Log-likelihoods were compared using Bayes factors (Kass and Raftery 1995) calculated as $2(H_0 - H_A)$, where H_0 and H_A are the log-likelihoods of the unconstrained and constrained outcomes, respectively.

3.3. Results

3.3.1. Phylogenetic relationships

The entire DNA sequence alignment included 6568 sites and 30% of these were parsimony-informative (Table 3.1). The best-fit partitioning scheme identified by PartitionFinder consisted of eight subsets (Table 3.2). We found that phylogenetic relationships were stable across methods of phylogenetic inference and partitioning strategies, although with varying levels of group support (Fig. 3.1). Regardless of method of analysis or partitioning scheme, Emery's rule was rejected in yellowjackets and hornets (Fig. 3.1). Likewise, a loose form of Emery's rule in which for any clade of parasites the nearest nonparasitic outgroup is a clade of host species (Buschinger 1990; Ward 1996) was not supported. Instead, the hosts of inquilines were scattered within a clade sister to *Dolichovespula maculata* and *D. media* (Fig 3.1). Inquilines were consistently recovered as monophyletic with strong support – Bayesian posterior probability (PP), ML bootstrap frequency (BS), and GC = 100 (Fig. 3.1). Moreover, the facultative social parasites *Vespula squamosa* and *Vespa dybowskii* did not share immediate common ancestry with their respective host species.

In the single most parsimonious tree found with the concatenated data, *D. arenaria* is sister to the inquiline clade (*Dolichovespula omissa*, (*D. adulterina*, *D. arctica*)), but this group was poorly supported (GC = 53; Fig. 3.1a). Using the best-fit partitioning scheme, the ML analysis of all data recovered the inquiline clade as sister to a group of three *Dolichovespula* species (Fig. 3.1b), whereas in the Bayesian consensus tree the inquilines were part of a polytomy (Fig. 3.1c), which included *D. arenaria* and (*D. albida*, (*D. pacifica*, *D. saxonica*)). However, *D. arenaria* was also sister to the inquiline clade in the Bayesian consensus trees using gene

and codon partitions, although this grouping had low support (PP = 87 and 73). In the case of *V. squamosa*, this facultative social parasite was sister to a clade of five species including two of its hosts, *V. vidua* and *V. flavopilosa*, but its primary host, *V. maculifrons*, was grouped with another species group (Fig. 3.1). Similarly, the facultative parasite *Vespa dybowskii* was placed in a clade separate from its main host, *V. simillima*; although *V. dybowskii* was sister to another host species, *V. crabro*, in the ML result.

3.3.2. Hypothesis testing

Interpretation of Bayes factors follows Kass and Raftery (1995), and thus values greater than 150 indicate very strong evidence against the constrained topologies. Comparisons of likelihoods between the unconstrained topology and those forcing host-parasite monophyly indicated that the evidence was strongly against all constrained hypotheses (Table 3.3).

Table 3.1: Sequence characteristics of the complete data matrix and chosen substitution models. PI = Parsimony informative.

Gene	Number of sites	PI sites	Model
12S	384	157	HKY+I+G
16S	532	156	GTR+I+G
28S	750	67	GTR+I
CAD	517	125	TIM1+G
COII	582	255	TVM+I+G
COI	1096	419	GTR+I+G
Cytb	433	197	GTR+I+G
EF1aF2	517	109	TrN+G
Pol II	814	110	TrN+I+G
ATP6	441	206	TVM+I+G
ATP8	111	80	HKY+G
Wg	391	91	K80+G
Total	6568	1972	

Table 3.2: Best-fit partitioning scheme identified by PartitionFinder.

Subset	Best Model	Subset Partitions
1	GTR+I+G	12S, 16S
2	GTR+I+G	28S, CAD pos1, EF1aF2 pos1, Pol2 pos1, wg pos1, wg pos2
3	TrN+G	CAD pos3, EF1aF2 pos3, Pol2 pos3, wg pos3
4	TrN+I	CAD pos2, COI pos2, EF1aF2 pos2, Pol2 pos2
5	GTR+I+G	COII pos1, COI pos1, Cytb pos1
6	TVM+I+G	COII pos2, Cytb pos2, atp6 pos2
7	TrN+I+G	COII pos3, COI pos3, Cytb pos3, atp6 pos3, atp8 pos3
8	TIM+I+G	atp6 pos1, atp8 pos1, atp8 pos2

Table 3.3: Stepping-stone estimates of marginal likelihoods and Bayes factors estimated as 2(H_0 − H_1), where H_0 and H_1 are the log-likelihoods of the unconstrained topology (-44246.01) and an alternative hypothesis, respectively.

Constraints (H_1)	lnL	Bayes factors
(*D. adulterina*, *D. saxonica*)	-44688.47	884.92
(*D. omissa*, *D. sylvestris*)	-44332.11	172.2
(*D. arctica*, *D. arenaria*)	-44364.99	237.96
(*V. dybowskii*, *V. simillima*)	-44366.16	240.3
(*V. squamosa*, *V. maculifrons*)	-44540.12	588.22
(*D. adulterina*, *D. saxonica*), (*D. omissa*, *D. sylvestris*), (*D. arctica*, *D. arenaria*), (*V. dybowskii*, *V. simillima*), (*V. squamosa*, *V. maculifrons*)	-45202.14	1912.26
(*D. adulterina*, *D. saxonica*), (*D. omissa*, *D. sylvestris*), (*D. arctica*, *D. arenaria*)	-44789.44	1086.86
(*D. adulterina*, *D. saxonica*, *D. omissa*, *D. sylvestris*, *D. arctica*, *D. arenaria*)	-44538.58	585.14

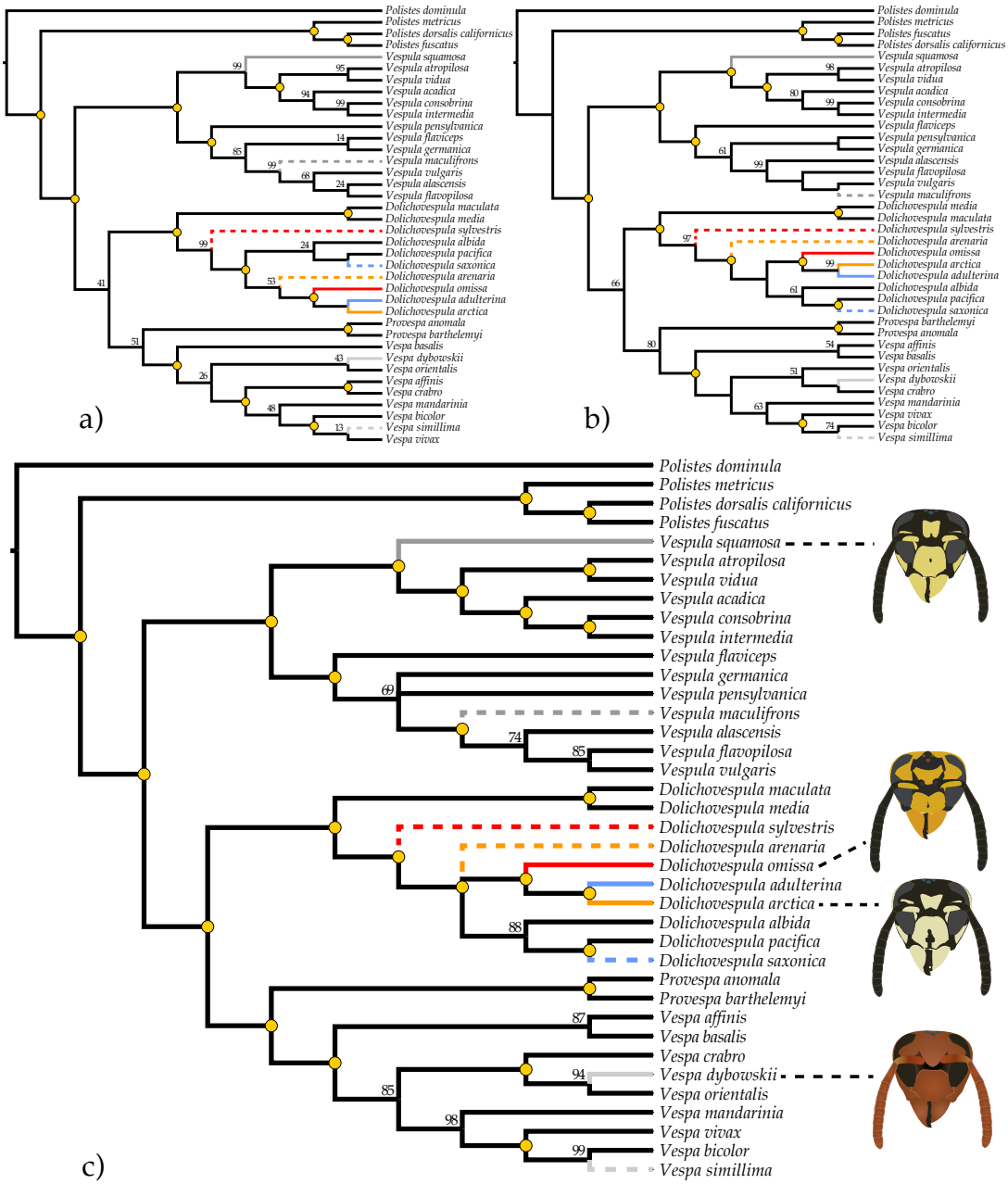

Figure 3.1: Phylogenetic relationships of social parasites, their hosts, and other vespines based on the concatenated data: a) single most parsimonious tree and GC values; b) maximum likelihood tree and bootstrap frequencies; c) Bayesian consensus tree and clade posterior probabilities. ML and Bayesian results obtained using the best-fit partitioning scheme. Yellow dots indicate node support equal to 100. Colored and grey solid branches indicate inquiline species and facultative social parasites, respectively. Dashed branches matching in color indicate the corresponding hosts.

3.4. Discussion

This study shows that social parasites among yellowjackets and hornets are not the closest relatives of their hosts, therefore rejecting Emery's rule in its strict form. Furthermore, monophyly of *Dolichovespula* inquilines, suggesting a single origin of the parasitic strategy in this genus, is strongly supported by all our analyses. In contrast to the results of Carpenter and Perera (2006), we find that the inquiline clade is not sister to *D. sylvestris*. Instead, *Dolichovespula* inquilines may be more closely related to either *D. arenaria* or a clade encompassing *D. albida*, *D. pacifica*, and *D. saxonica* (Fig. 3.1). Inquiline monophyly has also been found in *Polistes* paper wasps (Choudhary et al. 1994; Carpenter 1997). Vespine parasites usurp host societies by means of physical combat and kill the resident queen, whereas paper wasps employ chemical camouflage and coexist with the host queen (Cervo 2006; Lorenzi 2006; Cini et al. 2011), but these alternative usurpation strategies have resulted in the same pattern of inquiline monophyly. Our study adds to a growing body of examples where intraspecific or sympatric speciation has not occurred in the evolution of social parasitism (e.g., Agosti, 1994; Ward 1996; Choudhary et al. 1994; Carpenter and Perera 2006; Hines and Cameron 2010; Gibbs et al. 2012). In no case parasite and host formed a monophyletic group (Fig. 3.1). Thus, our analyses suggest that speciation occurred independently of the evolution of social parasitism. Berlocher (Berlocher 2003) argues that observing all possible intermediate forms of parasitism may be used to test hypotheses of allopatric speciation. These intermediate forms may be intra- and interspecific usurpation (Taylor 1939). In vespines, queens usurp nests of the same species as well as different species (Akre and Davis 1978; Greene 1991), but the latter type of usurpation is much

less frequent. Within *Dolichovespula*, *D. arenaria* usurps *V. vulgaris* (O'Rourke and Kurczewski 1983). Thus, it is possible that inquilinism in *Dolichovespula* evolved from facultative, temporary usurpation in *D. arenaria* (Fig. 3.1).

In addition to lack of phylogenetic support, the characteristics of yellowjacket societies seem incompatible with a key condition of the sympatric route to new inquiline species, namely the presence of multiple laying queens per colony (i.e., polygyny) (Bourke and Franks 1991; Buschinger 2009; Boomsma and Nash 2014; Rabeling et al. 2014). Certain authors (e.g., Alloway 1980; Buschinger 1986, 2009) argue that polygyny might be a precursor of social parasitism because it would provide the opportunity for some queens of the host species to specialize in producing reproductives, while other queens focus on producing workers. Furthermore, the adoption of conspecific young queens resembles the series of events in nest usurpation by socially parasitic queens. Yellowjacket colonies, however, typically include a single queen and have annual cycles (Spradbery 1973; Akre and Davis 1978), and polygyny is a rare deviation restricted to large-colony species of *Vespula* in warm climates; for example, *V. germanica*, *V. pensylvanica*, *V. vulgaris*, *V. maculifrons* (Greene 1991, and references therein). But the phylogenetic distribution of social parasitism shows that inquilinism is mostly limited to species of *Dolichovespula* (Fig. 3.1). If polygyny enables the sympatric speciation route in the evolution of social parasitism, more social parasites that follow Emery's rule would be expected in *Vespula*.

However, the tolerance of multiple egg-laying queens in large-colony species of *Vespula* may be associated with an increased vulnerability to parasitism by *V. squamosa*, which usurps several large-colony species. *Vespula squamosa* is considered a species crossing the threshold from free-living to

parasitism (MacDonald and Matthews 1975), capable of exploiting multiple host species in the *V. vulgaris* species group rather than in the more closely related *rufa* group (Fig. 3.1, see also Carpenter and Perera 2006; Lopez-Osorio et al. 2014). This suggests that strong phylogenetic affinities may not be imperative to pass easily through the defenses of host species by *V. squamosa*. It may be possible that social parasitism begins as a generalist strategy followed by host specialization. If facultative social parasitism is a necessary step in the path leading to inquiline behavior, then inquilines might have evolved from host species in sympatry but subsequently switched and specialized on a particular host, such that phylogenetic relationships of extant hosts and parasites would not be sufficient to reject a speciation model. This illustrates the difficulty in using phylogenetic analyses to test modes of speciation, as has been noted in studies of bees (Smith et al. 2007). A factor that has been thought to explain the rampant parasitism exerted by *V. squamosa* is its delayed release from diapause and subsequent spreading into the ranges of potential hosts (Taylor 1939).

Although Emery's rule in its strict form is here rejected for vespines, relatively close phylogenetic relationships seem to play a key role in the evolution of social parasitism, particularly for inquilines and their hosts nested within the same *Dolichovespula* clade (Fig. 3.1). Social parasitism in the Hymenoptera involves the exploitation not only of brood care but also the colony's intricate social structure. A mixed society thus must have compatible communication systems and pheromones for nestmate recognition (Buschinger 2009) as well as similar mechanisms of queen control. Cell-construction may be a trait of particular importance in the evolution of inquilinism in yellowjackets. In vespines, caste differentiation is physiologically determined, and eggs destined

to become queens typically develop in large cells. Cell size may function as a cue for workers to provide more food to certain larvae, which are thus launched on a queen developmental pathway (Jeanne and Suryanarayanan 2011). For example, in honeybees, larvae housed in royal cells are maintained on a diet of royal jelly, and its major active factor, royalactin, induces their development as queens (Kamakura 2011). If the colony's queen in part controls the construction of large cells, the parasitic queen must be able to mimic or circumvent this aspect of the host queen's behavior to avoid the production of workers (Greene 1991).

With the exception of *D. arctica* (Jeanne 1977), social parasites in Vespinae rely on physical attacks to subdue the host queen and her colony, but the mechanisms preventing the removal of parasitic eggs are largely unknown. Acceptance of parasitic eggs may be achieved by means of chemical mimicry, such as in the ant *Polyergus breviceps* (Johnson et al. 2004). Alternatively, parasitic eggs may be tolerated due to lack of cuticular chemicals used for nestmate recognition or usage of chemical deterrents (Ruano et al. 2005; Lambardi et al. 2007; Martin et al. 2007). To our knowledge, only a single study has investigated the chemical characteristics of parasitic eggs in vespines. Martin et al. (2008) identified compounds from the surface of eggs of *Vespa dybowskii* and suggested that this species employs a chemical transparency strategy. That is, parasitic eggs of *V. dybowskii* contain external chemicals that are either undetected or unused as recognition cues. Furthermore, these authors found that the chemical profile of *V. dybowskii*, including adults, shows more significant differences in comparison to its main host, *V. simillima*, than to *V. crabro* (Martin et al. 2008). Therefore, chemical mimicry does not seem to be involved in the parasitism of *V. simillima* by *V. dybowskii*. The similarities in chemical profiles in Martin et al. (2008) reflect

the relationships recovered in our Bayesian analysis (figure 1c), in which *V. crabro* is sister to *V. dybowskii* plus *V. orientalis*, but *V. simillima* is in a separate clade (see also Perrard et al. 2013).

To summarize, Emery's rule is a broad generalization about the evolution of a trait regardless of specific preconditions. Evidence from different groups indicates that the sympatric speciation model is a plausible explanation in *Myrmica* and *Mycocepurus* ants (Savolainen and Vepsäläinen 2003; Rabeling et al. 2014), but it is not applicable in bees (Hines and Cameron 2010; Gibbs et al. 2012; Smith et al. 2013) and social wasps (Choudhary et al. 1994; Carpenter 1997; Carpenter and Perera 2006). Even if Emery's rule is rejected in yellowjackets and hornets, it is clear that relatively close phylogenetic relationships, especially in inquilines, are important in the evolution of social parasitism (Fig. 3.1). Moreover, the monophyly of inquilines of *Dolichovespula* suggests an underlying genetic basis of parasitic habits.

CHAPTER 4: PHYLOGENOMIC ANALYSIS OF YELLOWJACKETS AND HORNETS (HYMENOPTERA: VESPIDAE, VESPINAE)

4.1. 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 in the family Vespidae (Carpenter 1982; Pickett and Carpenter 2010). 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 usually comprise a single, morphologically distinct queen; 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 2,500 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 (Bourke 1999; Anderson and McShea 2001). 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 2001; 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; Foster et al. 2001; Freiburger et al. 2004, Wenseleers et al. 2005c). In contrast, the 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 (Ross 1985; Foster and Ratnieks 2001a) 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 control of the queen (Ratnieks and Visscher 1989; Wenseleers and Ratnieks 2006). Conflict may occur between species as well. 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 yellowjackets. These social parasites, lacking the worker caste, seize the nest of a host species and trick the resident workers into raising the parasitic offspring (MacDonald and Matthews 1975; Greene et al. 1978; Reed and Akre 1983).

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 1987b; Pickett and Carpenter 2010; Lopez-Osorio et al. 2014). Previous analyses have recovered a yellowjacket clade sister to *Provespa* (Carpenter 1987b; 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, has been reported relying exclusively on molecular data (Lopez-Osorio et al. 2014, 2015). The results of Lopez-Osorio et al. (2014), however, were discordant 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 indicate 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. (2015) recovered poorly supported relationships among genera.

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 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 *Dolichovespula* and *Vespula*. Our findings challenge previous phylogenetic hypotheses and provide a robust framework for future comparative studies on yellowjackets and hornets.

4.2. Materials and methods

4.2.1. 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 -80ºC. 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 x 100 bp Illumina sequencing of the remaining species were outsourced to Beckman Coulter Genomics (Danvers, MA). We combined our data with publically 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). All transcriptomes were processed as described below.

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

We cut adapters, trimmed low-quality bases and discarded short reads from Illumina reads using Trimmomatic v. 0.32 (Lohse et al. 2012; Bolger et al. 2014) 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 contaminant and rRNA-like transcripts using the standalone releases of DeconSeq v. 0.4.3 (Schmieder and Edwards 2011) and riboPicker v. 0.4.3 (Schmieder et al. 2012), both of which use a modified version of the BWA-SW aligner (Li and Durbin 2009). 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 redundant 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.

4.2.3. Matrix construction, phylogenomic analyses, and hypothesis testing

To identify groups of putative homologous sequences and orthologs, we followed a procedure based on sequence similarity and phylogenetic analysis (Yang and Smith 2014). We analyzed two sets of taxa, one including all nine

species and another excluding the transcriptome of *A. catskill* due to its comparatively small size. An all-by-all BLASTP search was 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.

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. Models of amino acid substitution were chosen for each ortholog using the RAxML model selection script. We concatenated ortholog alignments with full taxon sets and number of sites greater than or equal to 300 in trimmed alignments. We then conducted partitioned maximum likelihood and rapid bootstrap analyses of "supermatrices" in RAxML on CIPRES (Miller et al. 2010). We evaluated uncertainty of edges and conflict between ortholog trees and species trees in two ways: first, by performing 200 jackknife pseudoreplicates, randomly resampling 30% of the total number of orthologs from each supermatrix; and second, after extracting ingroup clades from gene trees, we used PhyParts (Smith et al. 2015b) to examine concordance and conflict with respect to species trees, and to calculate internode certainty scores on the species tree (ICA; Salichos and Rokas 2013; Salichos et al. 2014) under a bootstrap filter of 50 %. ICA values close to 1 indicate strong certainty in the bipartition of interest, whereas ICA values close to 0 indicate similar frequency of conflicting bipartitions (Smith et al. 2015b). Lastly, for the eight- and nine-taxon data sets, we performed species tree analyses in MP-EST (Liu et al. 2010) with default settings on the STRAW web server (Shaw et al. 2013).

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.

4.3. Results

4.3.1. Transcriptome sequencing and *de novo* assembly

The five newly-sequenced, Illumina transcriptomes had an average of approx. 207 million passing filter (PF) reads, 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%. 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 4.1). After reducing redundancy, the average number of amino acid sequences (excluding *A. catskill*) was 14,896 (Table 4.1).

4.3.2. Homology and orthology inferences and phylogenetic analyses

We followed a procedure based on sequence similarity (*E* values from BLASTP) and phylogenetic inference to identify groups of putative homologs and orthologs (Yang and Smith 2014). An all-by-all BLASTP search using amino acid sequences from all transcriptomes was conducted, and similarity scores were used to identify clusters of homologous sequences. The sequences of each cluster were aligned, a phylogeny was inferred, and spurious branches were trimmed. This process of sequence alignment, phylogeny inference, and trimming of branches was repeated to obtain groups of refined homologs and then extract maximum inclusion orthologs. The nine-taxon dataset comprised 1,507 putative orthologs, 933,533 aligned sites, and had 91% amino acid

completeness. The eight-taxon dataset comprised 3,356 putative orthologs, 2,285,441 aligned sites, and had 94% amino acid occupancy. The two "supermatrices" had full ortholog coverage. Phylogenetic analyses of the two "supermatrices" in RAxML resulted in the same fully resolved topology, which had bootstrap and 30% gene-jackknife support values of 100 for all nodes (Fig. 4.1). Moreover, we found the same topology in the species tree analyses conducted in MP-EST. 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 two species of *Dolichovespula*, which were monophyletic, and the hornet *Vespa crabro* (Fig. 4.1).

Regarding the concordance and conflict between gene trees and species trees, we found considerable support for the sister relationship between *Dolichovespula* and *Vespa* (Fig. 4.1). The *Dolichovespula* + *Vespa* node had 1032 concordant gene trees and an ICA score of 0.82 in the analysis of the nine-taxon dataset (Table 4.2). In the case of the eight-taxon dataset, the *Dolichovespula* + *Vespa* node had 2383 concordant gene trees and a 0.83 ICA value (Table 4.2). Lastly, the topology recovered here (Fig. 4.1) was significantly different from the traditional hypothesis in which *Vespula* and *Dolichovespula* are sister groups (Table 4.3).

Table 4.1: 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	2,817	1,393.00	38,669	14,489
D. maculata	131,905	34.72	3,775	2,049.22	61,099	13,944
V. crabro	201,718	35.21	4,006	1,862.25	79,625	16,361
V. vidua	146,729	34.40	3,688	2,048.18	65,754	14,919
P. dominula	155,861	32.47	3,396	1,798.09	58,569	14,768
A. catskill	9,481	36.03	1,434	1,247.36	7,002	–
Average	129,357	35.06	3,186	1,733.02	51,786	14,896
SD	64,215.73	1.68	950.90	338.1484879	25,614.52	899.32

Table 4.2: Number of gene trees concordant and conflicting with vespine clades. Internode certainty (ICA) scores near 0 indicate maximum conflict and values near 1 indicate strong certainty.

Clade	Concordant	Conflicting	ICA score
9 taxa			
(*V. crabro*, (*D. maculata*, *D. arenaria*))	1032	64	0.8238
(*V. vidua*, (*V. crabro*, (*D. maculata*, *D. arenaria*)))	625	302	0.4742
8 taxa			
(*V. crabro*, (*D. maculata*, *D. arenaria*)	2383	137	0.8301
(*V. vidua*, (*V. crabro*, (*D. maculata*, *D. arenaria*))	1425	717	0.4729

Table 4.3: 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

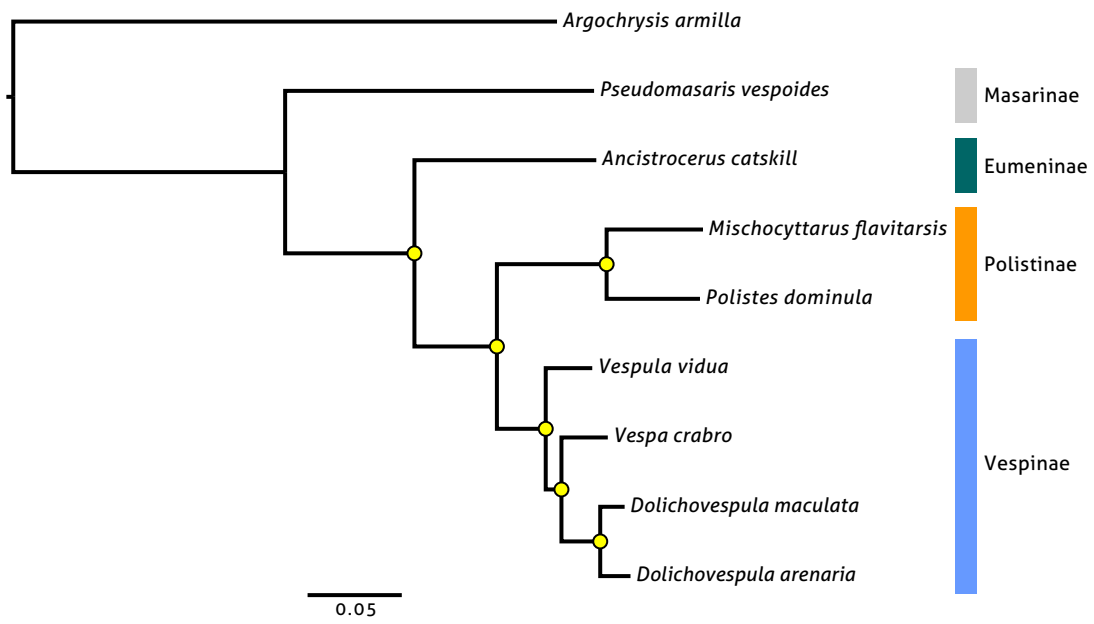

Figure 4.1: Phylogeny of vespid wasps based on the analysis of 1,507 single-copy genes.

4.4. Discussion

The deep-level phylogenetic relationships of vespine wasps have been elusive, as previous studies support alternative genus-level groupings (Carpenter 1987b; Pickett and Carpenter 2010; Saito and Kojima 2011; Lopez-Osorio et al. 2014, 2015; Perrard et al. 2015). 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 molecular evolution of sociality (Robinson et al. 2005; Fischman et al. 2011; Rehan and Toth 2015). In this study, we provide, for the first time, a robust hypothesis of genus-level relationships of vespine wasps based on transcriptomic data.

Transcriptomic data challenge the relationships among genera found in previous phylogenetic analyses of vespine wasps (e.g., Carpenter 1987; Pickett and Carpenter 2010; Perrard et al. 2015). The prevailing hypothesis of vespine phylogeny indicates that *Vespa* is the sister group of the remaining Vespinae, and the monophyletic yellowjackets (*Dolichovespula* and *Vespula*) are sister to the nocturnal hornets (*Provespa*) (Carpenter 1987b). 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. 2015). That is, most previous studies have recovered yellowjackets as a monophyletic group (Carpenter 1987b; Pickett and Carpenter 2010; Saito and Kojima 2011; Perrard et al. 2015). 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. 4.1). The sister group relationship between *Vespa* and

Dolichovespula was previously reported in phylogenetic analyses including data from up to eleven gene fragments, although mitochondrial and nuclear genes had conflicting phylogenetic signals (Lopez-Osorio et al. 2014, 2015). The number of genes used in this analysis was orders of magnitude higher than in any previous phylogenetic study of vespine wasps, and these genome-scale data further support the shared most recent common ancestry between *Vespa* and *Dolichovespula*.

Our analyses revealed that the sister grouping of *Vespa* and *Dolichovespula* remains stable whether the full set of putative orthologs or a random sample of genes is analyzed, but 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. 2015b), 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 4.2), indicating conflict at the groupings of vespine genera. Contrary to what may be expected from previous findings (Regier et al. 2008; Salichos and Rokas 2013), incongruence was higher for the internode subtending the vespine clade rather than for the shorter branch subtending the *Vespa* and *Dolichovespula* group (Fig. 4.1, Table 4.2). Conflict at the base of the Vespinae clade suggests that biological processes such as gene duplication and extinction and incomplete lineage sorting have influenced the origin of these wasps (Maddison 1997; Jeffroy et al. 2006; Philippe et al. 2011). Moreover, the origin of

the Vespinae might have been the result of a rapid radiation, where branching events are characterized by paucity of informative characters (Whitfield and Lockhart 2007; Whitfield and Kjer 2008). Further work is required on the sources of phylogenetic conflict in the Vespinae. As more genome-scale data becomes available, it also remains to be seen whether or not the sister-group relationship between *Vespa* and *Dolichovespula* is influenced 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 (Wenseleers et al. 2005b; Helanterä et al. 2006; Bonckaert et al. 2008; Oi et al. 2015). 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 2001b).

The comparative studies aforementioned, however, relied on a hypothesis of yellowjacket monophyly. Our results suggest that inferences of trait evolution in the Vespinae should not be based exclusively on an assumed position of *Vespula* as the sister group to the *Dolichovespula*. The phylogenetic framework proposed here implies, for example, that the evolution of large colony size and high paternity may be unique to species in a lineage (*Vespula*) distantly related from the remaining Vespinae.

CHAPTER 5: PATTERNS OF POSITIVE SELECTION IN SOCIAL WASP TRANSCRIPTOMES (HYMENOPTERA: VESPIDAE, VESPINAE)

5.1. Introduction

Convergent transitions in social evolution enable the discovery of shared genomic features associated with these transitions (Smith et al. 2008; Fischman et al. 2011; Simola et al. 2013; Stern 2013; Rehan and Toth 2015). Eusociality, where female workers cooperatively raise the offspring of the reproductive queen caste, originated repeatedly in the Hymenoptera – ants, bees, and wasps (Wilson 1971; Wilson and Holldobler 2005). Various mechanistic hypotheses have been proposed to explain transitions to eusociality and between alternative social phenotypes in molecular terms (Robinson et al. 2005; Rehan and Toth 2015). Two major hypotheses relate the evolution of castes to either changes in gene expression or changes in genomic sequence, although these mechanisms are not necessarily mutually exclusive (Rehan and Toth 2015). From the perspective of gene expression, the genetic toolkit hypothesis, for example, proposes that regulation of sets of genes with conserved roles underlie the evolution of castes across taxa (Toth and Robinson 2007; Toth et al. 2010). In the context of changes in genomic sequence, however, studies of ants and bees suggest that novel (i.e., taxonomically restricted) protein-coding genes (Khalturin et al. 2009; Tautz and Domazet-Loso 2011; Long et al. 2013) have influenced the attainment and elaboration of eusociality (Johnson and Tsutsui 2011; Simola et al. 2013; Feldmeyer et al. 2014; Berens et al. 2015a; Jasper et al. 2015). Furthermore, the protein evolution hypothesis proposes that the origin of social phenotypes is associated with positive selection acting on genes related to functional categories such as carbohydrate metabolism, immunity, neurogenesis, and olfaction, among

others (Fischman et al. 2011; Woodard et al. 2011; Harpur and Zayed 2013; Harpur et al. 2014; Roux et al. 2014; Kapheim et al. 2015). On a molecular level, therefore, the history of social insects may have been shaped by the birth of novel genes and rapid evolution of genes or gene families (Sumner 2014).

Transitions to eusociality likely involved intermediate stages, ranging from solitary to communal living (Evans and West-Eberhard 1970; West-Eberhard 1978; Carpenter 1989), with variation in the molecular mechanisms operating at different transitional stages. For example, primitively eusocial species of the paper wasp genus *Polistes* have rudimentary caste differences in morphology and their workers have the potential to become replacement queens (West 1967; Reeve 1991; Jandt et al. 2013). Similarly, in *Dinoponera* ants, where the distinct queen caste has been secondarily lost, young workers compete for reproductive primacy (Monnin and Peeters 1998; Lenhart et al. 2013). Comparative genomics of *Polistes canadensis* and *Dinoponera quadriceps* show that in these species both conserved toolkit genes and novel genes play a similar role in the reproductive plasticity that characterizes their simple societies (Patalano et al. 2015). Comparisons spanning other levels of social complexity, therefore, provide further insights into understanding social evolution in molecular terms. In particular, lineages that display a full range of lifestyles, and where eusociality has a relatively recent origin, may provide a more informative view into the evolution of eusociality and its genomics basis (Danforth 2002; Rehan and Toth 2015).

The Vespidae is a lineage of wasps exhibiting a full spectrum of social traits, including solitary as well as primitively and advanced eusocial species (Evans and West-Eberhard 1970; Jeanne 1980; O'Donnell 1998). Eusociality in the

Vespidae originated sometime in the mid-Cretaceous, with a minimum age of approximately 63 Mya (Wenzel 1990), later than in other eusocial groups such as ants and termites (Cardinal and Danforth 2011). Within vespids, the paper wasp genus *Polistes* and yellowjackets (*Vespula* and *Dolichovespula*) and hornets (*Vespa*) are perhaps the most well known, and all belong to eusocial subfamilies. In contrast to the primitively eusocial *Polistes*, the highly or advanced eusocial colonies of yellowjackets and hornets have morphologically distinct castes and their colonies comprise hundreds to thousands of workers (Evans and West-Eberhard 1970). Although the natural history of several vespid species has been well documented (Richards 1971; Spradbery 1973; Akre and Davis 1978; Ross and Matthews 1991; Hunt 2007; Gadagkar 2009), sociogenomic analyses have been conducted only recently (Jandt and Toth 2015).

Previous sociogenomic studies in the Vespidae have focused on comparisons of gene expression, particularly in paper wasps, whereas tests of alternative hypotheses and encompassing multiple social levels are wanting. In *Polistes metricus*, for example, individuals exhibiting maternal care (workers and foundresses) have more similar patterns of gene expression in comparison to individuals that do not (queens and gynes) (Toth et al. 2007). By contrast, transcriptomic analyses of *Polistes canadensis* suggest that caste differences derive from novel genes that are differentially expressed (Ferreira et al. 2013; see, however, Berens et al. 2015). Although certain insights have been gathered from studies of *Polistes* wasps, discovering broad genomic patterns in the Vespidae requires the inclusion of highly eusocial species such as yellowjackets and hornets. The ongoing synthesis of sociogenomics (Rehan and Toth 2015; Kapheim 2016) suggests that protein evolution and positive selection may be the

primary mechanisms influencing the transition to highly eusocial behavior. This protein evolution hypothesis has been supported in studies of bees (Woodard et al. 2011; Harpur et al. 2014; Jasper et al. 2015; Kapheim et al. 2015) and ants (Simola et al. 2013), but remains unexplored in the highly eusocial social wasps. Here, we conduct a comparative transcriptomic analysis of vespid wasps to test the protein evolution hypothesis. Specifically, we perform phylogeny-based comparisons to identify lineage-specific signatures of positive selection in transcriptomes of highly eusocial wasps, and compare these results to primitively eusocial and solitary wasp lineages.

5.2. Materials and methods

5.2.1. Transcriptome assembly

We sampled solitary as well as primitively and advanced eusocial species of the family Vespidae and extracted total RNA from single, whole specimens using the TRIzol® reagent (Invitrogen) (Table 5.1). We outsourced the preparation of cDNA libraries, Roche 454 pyrosequencing of *A. catskill*, and paired-end 2 x 100 bp Illumina sequencing of the remaining species to Beckman Coulter Genomics (Danvers, MA). Our data were analyzed in combination with published transcriptomes from the solitary cuckoo wasp *Argochrysis armilla* and the primitively eusocial vespids *Pseudomasaris vespoides* and *Mischocyttarus flavitarsis* (NCBI SRA accessions SRX262928, SRX262920, and SRX259759; Johnson et al. 2013). We trimmed adapters and low-quality bases from Illumina reads using Trimmomatic v. 0.32 (Lohse et al. 2012; Bolger et al. 2014) with a sliding window threshold for average base quality equal to 20. Transcriptomes were assembled *de novo* using Trinity v. 2013-11-10 (Grabherr et al. 2011; Haas et al.

2013) with default settings. We identified candidate, coding regions using TransDecoder r20131110 (Haas et al. 2013) and clustered redundant proteins using CD-HIT (Fu et al. 2012) with an identity threshold of 1.0 (–c 1.0 –n 5). Amino acid sequences were used for orthology inference and phylogenetic analysis.

5.2.2. Orthology inference and phylogenetic analysis

We followed a procedure based on sequence similarity and phylogenetic inference to identify putative orthologs (Yang and Smith 2014). Orthology inference was conducted for a taxon set including transcriptomes from nine species, and another set excluding *A. catskill*. We carried out all-by-all BLASTP searches with default settings and retained BLASTP hits with query coverage greater than 0.4. Putative homologs were inferred using Markov clustering (MCL v. 14-137; Enright et al. 2002) with an *E* value cutoff of 10^{-5} and an inflation value of 2.0. We aligned clusters using MAFFT v. 7 (Katoh and Standley 2013) and cleaned alignments using Phyutility (Smith and Dunn 2008) with minimum site occupancy of 0.1. MAFFT alignments were performed using the options 'genafpair' and 'maxiterate 1000'. Clusters with more than one thousand sequences were aligned using PASTA (Mirarab et al. 2014). We used RAxML 8 (Stamatakis 2014) to infer a maximum likelihood phylogenetic tree for each aligned cluster of homologous sequences with the model PROTCATWAG. We trimmed terminal branches ten times longer than their sisters or longer than 0.8, removed monophyletic and paraphyletic sequences belonging to the same taxon, and cut internal branches longer than 1.0. This process of sequence alignment, cleaning of alignments and trimming of branches was 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 to generate trees of homologs for orthology inference. Orthologs were determined using the maximum inclusion criterion (Dunn et al. 2008; Hejnol et al. 2009). We aligned orthologs using MAFFT and alignments trimmed with Gblocks v0.91 (Castresana 2000). Models of amino acid substitution were chosen for each ortholog using the model selection Perl script provided with RAxML. We concatenated orthologs with complete taxon sets and at least 300 sites. Partitioned maximum likelihood and rapid bootstrap analyses were carried out using RAxML on CIPRES (Miller et al. 2010).

5.2.3. Tests of positive selection

We realigned orthologs using PRANK (Löytynoja and Goldman 2005, 2008) to mitigate the influence of alignment errors that mislead branch-specific tests of positive selection (Fletcher and Yang 2010; Markova-Raina and Petrov 2011). PRANK protein alignments were conducted with default settings. We converted protein sequence alignments into their corresponding codon-based DNA alignments using PAL2NAL v14 (Suyama et al. 2006), enabling the option to remove columns with gaps. Codon alignments with more than 100 sites were considered for tests of positive selection. We performed tests of positive selection in CODEML employing branch tests (Yang 1998) to evaluate the *a priori* hypothesis that the highly eusocial yellowjackets and hornets have experienced accelerated evolution relative to primitively- or non-eusocial lineages. We performed tests on single branches or whole clades to detect signatures of positive selection on unrooted trees (Fig. 1). The alternative branch model assumes that branches of interest, or foreground branches, have a ratio of

synonymous (d_S) and nonsynonymous (d_N) substitutions (d_N/d_S or ω) that is different from the background ratio. This two-ratio model was contrasted with the null model of the same d_N/d_S for all branches. We compared alternative and null models using the likelihood-ratio test (LRT) and calculated *p*-values using the base R function pchisq (df = 1). *P*-values from all LRTs were then used to estimate *q*-values, which measure significance in terms of the false discovery rate (FDR) (Storey 2002; Storey and Tibshirani 2003), using the R package qvalue with an FDR level equal to 0.05. Considering that the branch test approach is quite conservative, since positive selection can be detected only when the average d_N/d_S over all sites is greater than one, we compared d_N/d_S values for foreground branches to find differences suggesting relaxed stabilizing selection.

5.2.4. Annotation of orthologs

Profiles of potential protein function were determined for each ortholog using InterProScan v5 (Jones et al. 2014), limited to identification of protein domains (Pfam; Punta et al. 2012) based on searches with HMMER v3.1 (hmmer.org) and prediction of signal peptides (SignalP v4.0; Petersen et al. 2011) and transmembrane regions (TMHMM v2.0; Krogh et al. 2001). We carried out BLASTP searches against the NCBI Reference Sequence Database (RefSeq, release 74) with an *E* value cutoff of 10^{-5} and restricted to Hymenoptera and *Drosophila* matches (NCBI taxonomy identifiers 7399 and 7215). InterProScan profiles and the top 20 BLASTP hits for each ortholog were mapped to gene ontology (GO) terms and annotated using BLAST2GO v3.2 (Conesa and Götz 2008) under default settings. Orthologs with d_N/d_S ratios significantly greater than one were additionally submitted to BLASTP searches against *Drosophila melanogaster*

proteins with UniProt (The UniProt Consortium 2015) as the target database. We performed functional enrichment analyses between all orthologs and subsets of orthologs with signatures of positive selection using Fisher's exact test correcting for multiple comparisons (FDR) in BLAST2GO.

5.3. Results

5.3.1. Orthology inference and phylogenetic analysis

Transcriptomes from six vespid species had an average of 129,357 transcripts and 51,786 potential coding regions (Table 5.1). After reducing sequence redundancy, the average number of amino acid sequences was 14,896, excluding *A. catskill* (Table 5.1). We identified 1,507 (9 taxa) and 3,356 (8 taxa) putative orthologs with full taxon coverage, 90-93% amino acid occupancy, and alignment length greater than 300 sites for phylogenetic inference. The maximum-likelihood phylogeny inferred with each matrix of concatenated orthologs had 100 bootstrap support values for all nodes. The topology recovered shows the highly eusocial yellowjackets and hornets (Vespinae) as monophyletic and sister to the primitively eusocial paper wasps (Polistinae), and, depending on the taxon set, either the solitary *A. catskill* or *P. vespoides* as sister to the eusocial clade (Fig. 5.1).

5.3.2. Patterns of positive selection

After converting protein sequence to codon-based DNA alignments and filtering by alignment length, 1,391 and 3291 orthologs remained for tests of positive selection. Estimates of d_N/d_s using the basic model (one ratio for all branches) did not show evidence of positive selection. Altogether, comparisons

between the single and two-ratio models resulted in up to 6.62% (218/3,291) orthologs displaying significant signals of accelerated evolution (foreground branch $d_N/d_S > 1$, LRT *q-value* < 0.05). For the two sets of orthologs, the proportion of positively selected genes varied across the branches tested. We found evidence of episodic positive selection predominantly on the branch subtending the highly eusocial yellowjackets and hornets (Table 5.2, branch 1). The branch test specific to the primitively eusocial lineage (branch 2) had 0.3-2.92% positively selected orthologs and the branch leading to the clade of eusocial vespids (branch 3) had up to 1.22% genes with signatures of positive selection (Table 5.2). For the set of 1,391 orthologs, we found an overlap of two positively selected genes between the highly eusocial lineage (branch 1) and along the branch subtending the eusocial clade (branch 3), whereas orthologs from the primitively eusocial lineage were exclusive. By contrast, our analyses of the 3,291 ortholog set revealed shared genes only between the highly and primitively eusocial lineages; 22 positively selected orthologs in common. The terminal branch leading to the solitary *P. vespoides* had no orthologs with signatures of positive selection. The clade-specific tests (branch labels 5 and 6) resulted in a single gene with a d_N/d_S ratio significantly greater than one, suggesting that the evolution of levels of eusociality in the Vespidae might have been episodic.

Comparisons of d_N/d_S estimates smaller than one and from significant LRTs suggested that relaxed selection might have operated in the evolution of highly and primitively eusocial lineages, but not along the branch leading to the eusocial clade (Figure 5.2). Mann-Whitney *U* tests using foreground and background branch labels as factors revealed significant differences between

d_N/d_S ratios for the highly (W = 15179, p-value < 0.00) and primitively (W = 85276, p-value < 0.00) eusocial lineages.

5.3.3. Functional targets of episodic positive selection

The sets of 1,391 and 3,291 orthologs had 67% (8,412) and 61% (16,138) sequences annotated, respectively. Orthologs with signatures of positive selection in the highly eusocial lineage corresponded to GO biological processes involved in the metabolism of sugars, oxidation of fatty acids, transport of calcium ions in mitochondria, glycolysis, and protein phosphorylation, among others (Table 5.3). Positively selected genes in the primitively eusocial lineage involved processes related to embryonic development, transport of phospholipids, binding of small nucleolar RNA, oxidation of fatty acids, phosphorylation of carbohydrates, and neural development (Table 5.4). Genes with signals of positive selection along the branch leading to the entire eusocial clade were associated with GO processes such as cellular responses to amino acid and starvation, biosynthesis of phospholipids, DNA and protein catabolism, and RNA processing.

We evaluated which GO terms were enriched in those orthologs showing significant signatures of positive selection. GO terms overrepresented in the highly eusocial lineage included transferase activity, ATPase complex, nucleotide-sugar metabolic process, catalytic activity, pyruvate metabolic process, sodium ion transport, and potassium ion transport (Table 5.5). GO terms enriched in the primitively eusocial lineage corresponded to functions such as binding of magnesium ions and DNA and translocation of phospholipids, cellular components related to ATPase dependent transmembrane transport, and

processes such as transport of proteins as well as sodium and calcium ions (Table 5.6).

Table 5.1: Characteristics of transcriptome assemblies.

Behavior	Species	Total transcripts	GC %	Contig N50	Average contig length	ORFs	CD-HIT clusters
Highly eusocial	*D. arenaria*	130,448	37.50	2,817	1,393.00	38,669	14,489
	D. maculata	131,905	34.72	3,775	2,049.22	61,099	13,944
	V. crabro	201,718	35.21	4,006	1,862.25	79,625	16,361
	V. vidua	146,729	34.40	3,688	2,048.18	65,754	14,919
Primitively eusocial	*P. dominula*	155,861	32.47	3,396	1,798.09	58,569	14,768
Solitary	*A. catskill*	9,481	36.03	1,434	1,247.36	7,002	–
Average		129,357	35.06	3,186	1,733.02	51,786	14,896
SD		64,215.73	1.68	950.90	338.15	25,614.5	899.32

Table 5.2: Amount of positive selection detected in lineage-specific branch tests. Branch labels correspond to designations in Fig. 5.1. Number of orthologs with d_N/d_S ratios significantly greater than 1 (*q-value* < 0.05) for the 1,391 and 3,291 ortholog sets are separated by slashes.

Branch label	Lineage of interest	Number of orthologs with $\omega > 1$
1	Highly eusocial	38/218
2	Primitively eusocial	4/96
3	Eusocial	17/7

Table 5.3: Overview of *D. melanogaster* matches for orthologs positively selected along the branch leading to the highly eusocial lineage; genes ranked according to their LRT significance values. ω values of 999 represent cases in which d_s equals 0, and therefore ω is undefined. Results based on the set of 3,291 genes.

Rank	*Drosophila melanogaster* gene	FlyBase ID	UniProt entry	GO biological process	*q*-value	ω
1	*BTB (POZ) domain-containing protein 9, BTBD9*	FBgn0030228	Q9W2S3	Adult locomotory behavior; positive regulation of circadian sleep/wake cycle, sleep; regulation of synaptic transmission, dopaminergic	1.02E-07	1.03
2	*Trehalase, Treh*	FBgn0003748	A5XCQ7	Trehalose metabolic process	1.44E-07	3.05
4	*Isocitrate dehydrogenase, Idh*	FBgn0001248	Q9VSI6	Fatty acid alpha-oxidation; isocitrate metabolic process; tricarboxylic acid cycle	2.39E-06	3.70
5	*Letm1*	FBgn0019886	P91927	Cellular response to hypoxia; mitochondrial calcium ion transport; mitochondrion morphogenesis; neurotransmitter secretion; potassium ion transmembrane transport; sodium ion transmembrane transport	4.70E-06	999
7	*CG6330*	FBgn0039464	Q9VBA0	Gravitaxis; nucleoside metabolic process; nucleotide catabolic process	0.0000	999
8	*Aldolase, Ald-PJ*	FBpp0297612	F3YDB5	Glycolytic process	2.16E-05	3.33
9	*Glycerol 3 phosphate dehydrogenase, Gpdh*	FBgn0001128	B5RIM9	Carbohydrate metabolic process; glycerol-3-phosphate catabolic process	2.39E-05	1.34
10	*auxilin, aux*	FBgn0037218	Q9VMY8	Compound eye morphogenesis; negative regulation of neuron death; Notch signaling pathway; protein phosphorylation; sperm individualization; synaptic vesicle uncoating	2.43E-05	999
11	*Inositol-requiring enzyme-1, Ire1*	FBgn0261984	A8JR46	Compound eye photoreceptor cell differentiation; endoplasmic reticulum unfolded protein response; Golgi organization; mRNA catabolic process; mRNA endonucleolytic cleavage involved in unfolded protein response; protein phosphorylation; regulation of RNA splicing	2.52E-05	1.94
13	*Myosin 61F,*	FBgn0010246	H8F4R0	ATP binding; motor	3.08E-05	3.69

	Myo61F-RA			activity		
14	*Vacuolar H+-ATPase SFD subunit, VhaSFD*	FBgn0027779	Q9V3J1	ATP hydrolysis coupled proton transport; determination of adult lifespan; dsRNA transport	3.54E-05	999
18	*axotactin, axo*	FBgn0262870	Q9VZ96	Transmission of nerve impulse	4.26E-05	999
19	*gartenzwerg, garz*	FBgn0264560	A1Z8W8	Cell morphogenesis; epithelial cell development; Golgi organization; lumen formation, open tracheal system; phagocytosis; positive regulation of GTPase activity	4.85E-05	999
20	*Vacuolar protein sorting 33B, Vps33B*	FBgn0039335	Q9VBR1	Endosomal transport; imaginal disc-derived wing morphogenesis; immune response; mitotic spindle assembly	5.16E-05	999
21	*Palmitoyl-protein thioesterase 1, Ppt1*	FBgn0030057	Q9W3C7	Determination of adult lifespan; endocytosis; macromolecule depalmitoylation; neuron fate specification	5.91E-05	999
23	*eIF3-S10*	FBgn0037249	Q9VN25	Formation of translation preinitiation complex; mitotic spindle elongation; mitotic spindle organization; regulation of translational initiation; translational initiation	7.66E-05	999
24	*CG5002*	FBgn0034275	Q7K4I4	Bicarbonate transport; chloride transmembrane transport; oxalate transport; regulation of intracellular pH	8.40E-05	999
30	*Sarcosine dehydrogenase, Sardh-PA*	FBpp0088528	T2FFP0	Oxidation-reduction process	0.0002	999
32	*Kua*	FBgn0032850	Q9V3B5	Oxidation-reduction process; protein ubiquitination	0.0002	999
33	*CG9674*	FBgn0036663	Q9VVA4	Ammonia assimilation cycle; glutamate biosynthetic process	0.0002	3.83
42	*Transferrin 2, Tsf2*	FBgn0036299	Q9VTZ5	Septate junction assembly	0.0004	999
44	*Suppressor of variegation 2-10, Su(var)2-10*	FBgn0003612	Q5BIG7	Chromosome condensation; chromosome organization; compound eye development; defense response to Gram-negative bacterium; hemopoiesis; imaginal disc growth; mitotic G2 DNA damage checkpoint; negative regulation of JAK-STAT cascade; neurogenesis;	0.0005	999

				positive regulation of innate immune response		
47	*β subunit of type II geranylgeranyl transferase, betaggt-II*	FBgn0028970	Q9XZ68	Protein geranylgeranylation	0.0005	1.98
57	*Malate dehydrogenase 2, Mdh2-PA*	FBpp0082985	Q9VEB1	Activation of cysteine-type endopeptidase activity involved in apoptotic process; carbohydrate metabolic process; larval midgut cell programmed cell death; malate metabolic process; positive regulation of programmed cell death; pupal development; regulation of programmed cell death; salivary gland cell autophagic cell death; salivary gland histolysis	0.0008	1.22
58	*Mitochondrial ribosomal protein L37, mRpL37*	FBgn0261380	Q9VGW9	Translation; structural constituent of ribosome	0.0009	5.54
60	*UDP-galactose 4'-epimerase, Gale-PB*	FBgn0035147	Q9W0P5	Galactose metabolic process; larval lymph gland hemopoiesis	0.0010	999
64	*UGP*	FBgn0035978	A5XCL5	UDP-glucose metabolic process	0.0012	999
75	*Insulin-like receptor, InR-2*	FBgn0283499	G2J5R2 (*A. mellifera*)	Transmembrane receptor protein tyrosine kinase signaling pathway; insulin-activated receptor activity	0.0018	999
100	*nervana, nrv3*	FBgn0032946	Q86NM2	Potassium ion transport; response to auditory stimulus; sensory perception of sound; sodium ion transport	0.0050	999
105	*Vitellogenin, vg*	-	Q868N5 (*A. mellifera*)	Lipid transport	0.0058	1.01
107	*Facilitated trehalose transporter Tret1-1, Tret1-1*	FBgn0050035	A1Z8N1	Glucose import; glucose transmembrane transport; hexose transmembrane transport; proton transport; trehalose transport	0.0062	2.07
110	*Pyruvate carboxylase, PCB*	FBgn0027580	Q0E9E2	Gluconeogenesis; pyruvate metabolic process	0.0067	999
151	*Aminolevulinate synthase, Alas*	FBgn0020764	O18680	Chitin-based cuticle development; protoporphyrinogen IX biosynthetic process	0.016	999
159	*GDP-4-keto-6-deoxy-D-mannose 3,5-epimerase/4-reductase,*	FBgn0267823	Q9W1X8	*'De novo'* GDP-L-fucose biosynthetic process; dsRNA transport; GDP-L-fucose biosynthetic process	0.0198	999

	Gmer					
162	*nervana, nrv2*	FBgn0015777	A4V0B5	Potassium ion transport; sodium ion transport; ATP hydrolysis coupled transmembrane transport	0.0213	999
168	*Rhodopsin 2, Rh2-PA*	FBpp0083111	P08099	Detection of visible light; G-protein coupled receptor signaling pathway; phototransduction; protein-chromophore linkage; visual perception	0.0234	999
186	*Enolase, Eno*	FBgn0000579	P15007	Glycolytic process	0.03	999
187	*CG7920*	FBgn0039737	Q9VAC1	Acetyl-CoA metabolic process; neurogenesis	0.0303	1.05
211	*Inositol 1,4,5-triphosphate kinase 1, IP3K1-PA*	FBpp0079465	M9ND56	Response to oxidative stress	0.042	999

Table 5.4: Overview of *D. melanogaster* matches for orthologs positively selected along the branch leading to the primitively eusocial lineage; orthologs ranked according to their LRT significance values. Extreme ω values of 999 represent cases where there is a lack of synonymous substitutions along the branch, and therefore ω is undefined. Results based on the set of 3,291 genes.

Rank	*Drosophila melanogaster* gene	FlyBase ID	UniProt entry	GO biological process	*q*-value	ω
5	*CG33298*	FBgn0032120	Q7KTG6	Golgi organization; intracellular protein transport; phospholipid translocation	3.81E-09	999
8	*CG8064*	FBgn0038597	Q9VE98	snoRNA binding	6.17E-08	1.16
12	*Isocitrate dehydrogenase, Idh*	FBgn0001248	Q9VSI6	Fatty acid alpha-oxidation; isocitrate metabolic process; tricarboxylic acid cycle	9.12E-07	999
13	*FI22366p1, Ppt1*	FBgn0030057	S5WMX5	Palmitoyl-(protein) hydrolase activity	2.66E-06	2.01
15	*EG:115C2.1*	FBgn0025640	O77425	Carbohydrate phosphorylation; D-ribose metabolic process	4.82E-06	1.67
19	*Acyl-CoA synthetase long-chain, Acsl*	FBgn0263120	A1Z7H2	Axon guidance; early endosome to recycling endosome transport; long-chain fatty acid metabolic process; negative regulation of BMP signaling pathway; negative regulation of synaptic growth at neuromuscular junction; nervous system development; neurogenesis	2.30E-05	999
20	*CG7379*	FBgn0038546	Q9VEF5	Chromatin modification	3.99E-05	999
23	*nervana 3, nrv3*	FBgn0032946	Q86NM2	Potassium ion transport; response to auditory stimulus; sensory perception of sound; sodium ion transport	8.48E-05	999
25	*Keap1*	FBgn0038475	Q9VEN5	Protein ubiquitination involved in ubiquitin-dependent protein catabolic process; response to oxidative stress	0.0001	999
26	*Light, lt*	FBgn0002566	O76248	Autophagosome maturation; cellular response to starvation; determination of adult lifespan; dsRNA transport; endocytosis; intracellular transport; lysosomal transport; negative regulation of Notch signaling pathway; Notch receptor processing	0.0002	999
31	*CG1882*	FBgn0033226	Q5U191	Lipid metabolic process; lipid storage	0.0003	1.06
34	*Vacuolar*	FBgn0038593	Q9VEA2	Autophagosome	0.0003	2.20

	protein sorting 39, Vps39			maturation; cellular response to starvation; intracellular protein transport; neuron projection morphogenesis; regulation of Notch signaling pathway; regulation of SNARE complex assembly		
39	*CG11665*	FBgn0033028	Q7JWI7	Monocarboxylic acid transport	0.0007	1.40
46	*LKRSDH*	FBpp0310050	Q6NP53	Oxidation-reduction process	0.0017	999
48	*Gliotactin, Gli*	FBgn0001987	Q9NK80	Border follicle cell migration; establishment of blood-nerve barrier; female meiosis chromosome segregation; maintenance of imaginal disc-derived wing hair orientation; modulation of synaptic transmission; neuron cell-cell adhesion; postsynaptic membrane assembly; presynaptic membrane assembly; regulation of tube size, open tracheal system; septate junction assembly	0.0021	999
49	*Probable cytochrome P450 9f2, Cyp9f2*	FBgn0038037	Q9VG82	Heme binding; iron ion binding; monooxygenase activity; oxidoreductase activity	0.0022	4.74
50	*CG5525*	FBgn0032444	Q9VK69	Cytoplasmic microtubule organization; mitotic spindle assembly; mitotic spindle organization; neurogenesis; protein folding	0.0023	8.08
55	*CG10512-RD*	FBgn0037057	Q8IPT9	Oxidation-reduction process	0.0033	999
63	*CG8005*	FBgn0035854	Q9VSF4	Deoxyhypusine biosynthetic process from spermidine; peptidyl-lysine modification to peptidyl-hypusine	0.0060	1.11
67	*Malate dehydrogenase 1, Mdh1*	FBgn0262782	Q9VKX2	Carbohydrate metabolic process; lateral inhibition; malate metabolic process; tricarboxylic acid cycle	0.0084	999

Table 5.5: Top GO categories enriched in genes revealing significant signatures of positive selection in the highly eusocial wasp lineage (branch test 1). GO categories are molecular function (F), cell component (C), and biological process (P). Results based on the set of 3,291 genes.

GO-ID	Term	Category	FDR	*P*-value	Test set
GO:0016740	Transferase activity	F	9.12E-19	2.75E-22	321
GO:1904949	ATPase complex	C	1.34E-16	2.02E-19	16
GO:0098533	ATPase dependent transmembrane transport complex	C	1.34E-16	2.02E-19	16
GO:0090533	Cation-transporting ATPase complex	C	1.34E-16	2.02E-19	16
GO:0005890	Sodium:potassium-exchanging ATPase complex	C	1.34E-16	2.02E-19	16
GO:1902495	Transmembrane transporter complex	C	7.61E-14	1.61E-16	16
GO:1990351	Transporter complex	C	7.61E-14	1.61E-16	16
GO:0051649	Establishment of localization in cell	P	2.90E-12	7.01E-15	88
GO:0009225	Nucleotide-sugar metabolic process	P	3.24E-11	8.79E-14	16
GO:0008318	Protein prenyltransferase activity	F	3.93E-11	1.18E-13	12
GO:0003824	Catalytic activity	F	5.07E-11	1.68E-13	638
GO:0051641	Cellular localization	P	5.71E-11	2.07E-13	88
GO:0006090	Pyruvate metabolic process	P	1.66E-10	6.52E-13	24
GO:0046907	Intracellular transport	P	4.12E-09	1.74E-11	72
GO:0006814	Sodium ion transport	P	5.02E-09	2.27E-11	16
GO:0006886	Intracellular protein transport	P	7.36E-09	4.00E-11	58
GO:0006813	Potassium ion transport	P	1.91E-08	1.37E-10	16
GO:0009451	RNA modification	P	1.91E-08	2.25E-10	40
GO:0034613	Cellular protein localization	P	1.91E-08	2.26E-10	58
GO:0034708	Methyltransferase complex	C	1.91E-08	4.62E-10	8
GO:0016040	Glutamate synthase (NADH) activity	F	1.91E-08	4.62E-10	8
GO:0016832	Aldehyde-lyase activity	F	1.91E-08	4.62E-10	8
GO:0042350	GDP-L-fucose biosynthetic process	P	1.91E-08	4.62E-10	8
GO:0042351	'De novo' GDP-L-fucose biosynthetic process	P	1.91E-08	4.62E-10	8
GO:0004661	Protein geranylgeranyltransferase activity	F	1.91E-08	4.62E-10	8
GO:0004663	Rab geranylgeranyltransferase activity	F	1.91E-08	4.62E-10	8

Table 5.6: Top GO categories enriched in genes with significant signatures of positive selection in the primitively eusocial lineage (branch test 2). GO categories are molecular function (F), cell component (C), and biological process (P). Results based on the set of 3,291 genes.

GO-ID	Term	Category	FDR	*P*-value	Test set
GO:0000287	Magnesium ion binding	F	7.91E-17	2.39E-20	32
GO:0003690	Double-stranded DNA binding	F	1.29E-11	7.76E-15	16
GO:0004012	Phospholipid-translocating ATPase activity	F	4.25E-11	2.57E-13	8
GO:0045332	Phospholipid translocation	P	4.25E-11	2.57E-13	8
GO:0001012	RNA polymerase II regulatory region DNA binding	F	4.25E-11	2.57E-13	8
GO:0031463	Cul3-RING ubiquitin ligase complex	C	4.25E-11	2.57E-13	8
GO:0098599	Palmitoyl hydrolase activity	F	4.25E-11	2.57E-13	8
GO:0004450	Isocitrate dehydrogenase (NADP+) activity	F	4.25E-11	2.57E-13	8
GO:0015914	Phospholipid transport	P	4.25E-11	2.57E-13	8
GO:0034204	Lipid translocation	P	4.25E-11	2.57E-13	8
GO:0000978	RNA polymerase II core promoter proximal region sequence-specific DNA binding	F	4.25E-11	2.57E-13	8
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	F	4.25E-11	2.57E-13	8
GO:0000976	Transcription regulatory region sequence-specific DNA binding	F	4.25E-11	2.57E-13	8
GO:0000987	Core promoter proximal region sequence-specific DNA binding	F	4.25E-11	2.57E-13	8
GO:0000982	Transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding	F	4.25E-11	2.57E-13	8
GO:0097035	Regulation of membrane lipid distribution	P	4.25E-11	2.57E-13	8
GO:0001159	Core promoter proximal region DNA binding	F	4.25E-11	2.57E-13	8
GO:0005548	Phospholipid transporter activity	F	4.25E-11	2.57E-13	8
GO:0015748	Organophosphate ester transport	P	4.25E-11	2.57E-13	8
GO:0006102	Isocitrate metabolic process	P	4.25E-11	2.57E-13	8
GO:0042393	Histone binding	F	3.56E-10	2.26E-12	8
GO:0052855	ADP-dependent NAD(P)H-hydrate dehydratase activity	F	1.46E-09	9.70E-12	7
GO:0009060	Aerobic respiration	P	8.13E-09	5.89E-11	16

GO:0006099	Tricarboxylic acid cycle	P	8.13E-09	5.89E-11	16
GO:0006820	Anion transport	P	2.09E-08	1.58E-10	24
GO:0072350	Tricarboxylic acid metabolic process	P	3.75E-08	3.17E-10	16
GO:0006101	Citrate metabolic process	P	3.75E-08	3.17E-10	16
GO:0004747	Ribokinase activity	F	3.91E-08	3.66E-10	6
GO:0019321	Pentose metabolic process	P	3.91E-08	3.66E-10	6
GO:0006014	D-ribose metabolic process	P	3.91E-08	3.66E-10	6
GO:0045333	Cellular respiration	P	3.99E-08	3.85E-10	16

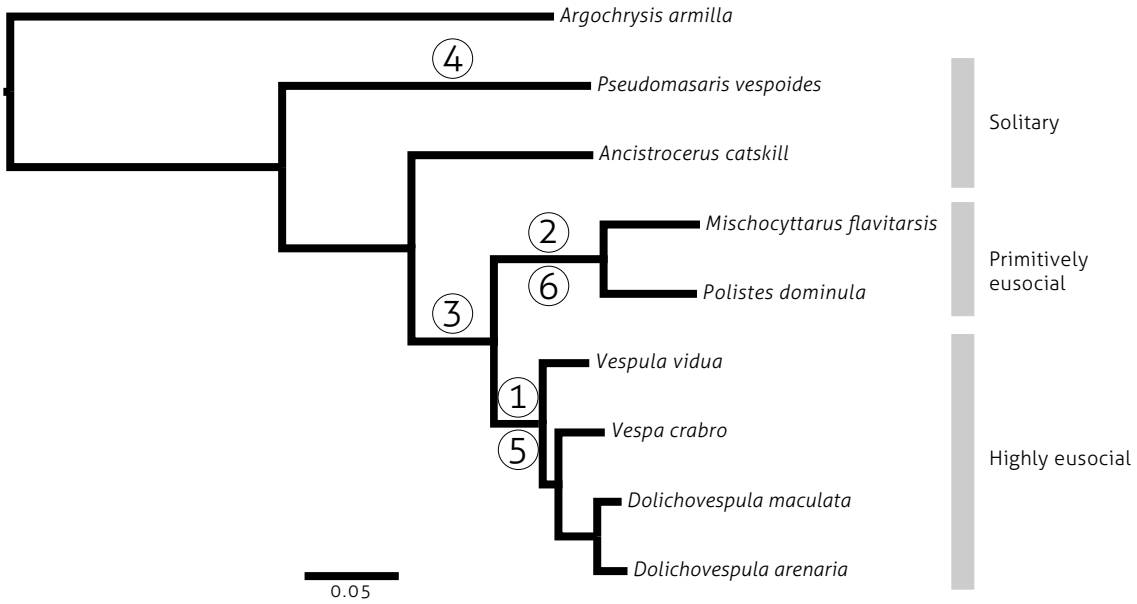

Figure 5.1: Phylogeny of vespid wasps showing branches labeled for positive selection analyses. Numbers above branches indicate lineage-specific branch tests, whereas numbers below branches indicate clade-specific tests.

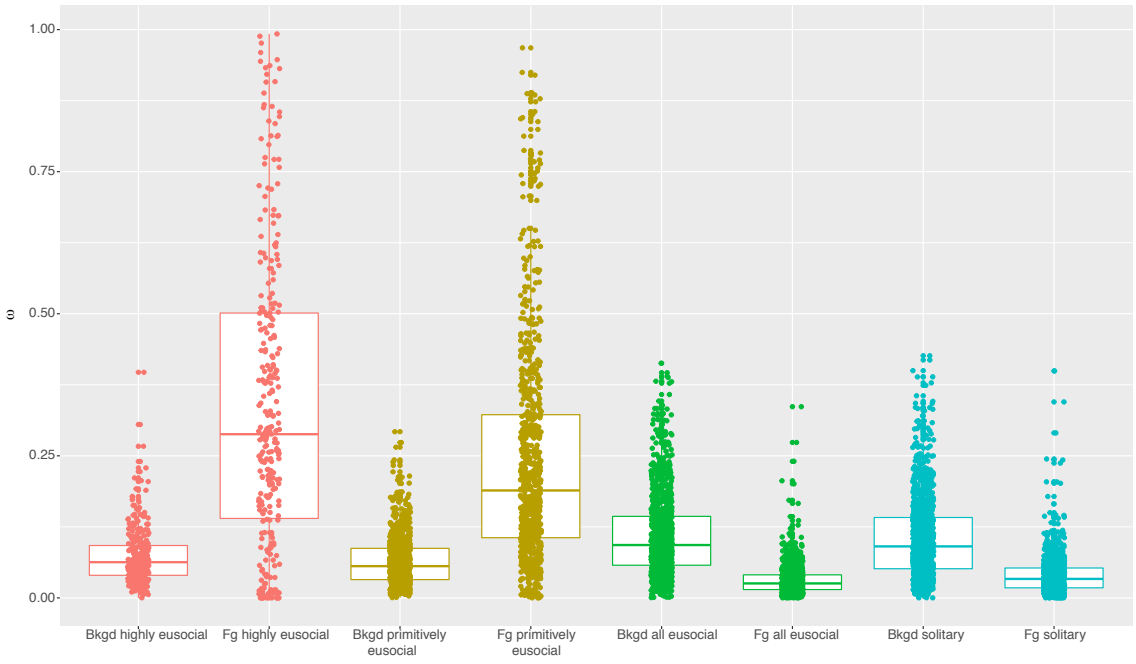

Figure 5.2: Comparison of ω values significantly less than one for background (Bkgd) and foreground (Fg) branches across lineage-specific tests. Boxplots of ω values overlaid with the actual data points, 'jittered' horizontally.

5.1. Discussion

On a molecular level, various mechanisms operate across transitional stages of social evolution (Rehan and Toth 2015). For example, the evolution of highly eusocial behavior in bees, and in ants, has been linked to accelerated evolution of protein coding genes (Hunt et al. 2010a; Fischman et al. 2011; Woodard et al. 2011; Harpur and Zayed 2013; Harpur et al. 2014; Roux et al. 2014; Kapheim et al. 2015). In this study, we find that episodic positive selection is associated with the origin of highly eusocial behavior in wasps. We identified up to 218 genes showing signatures of accelerated evolution in the Vespinae lineage (Table 5.3). Fewer genes were positively selected along branches leading to the eusocial clade and the primitively eusocial lineage (Table 5.2). Our findings suggest that genes involved in metabolism of carbohydrates and functioning in mitochondria (Table 5.3, 5.6) have been primary targets of selection in the origin of advanced eusociality. Below, we provide possible links between the rapid evolution of certain positively selected genes and the transition to advanced eusociality.

Traits defining advanced eusociality include morphologically distinct castes in colonies with hundreds to thousands of individuals. The growth of these large colonies is partially determined by specialized workers foraging for prey to provision larvae and wood pulp to build nests (Greene 1991; Richter 2000). Accelerated protein evolution in the yellowjackets and hornets might be related to worker foraging activity, which likely involved molecular changes in the metabolism of energy precursors such as sugars. The sugar trehalose exists in abundance in the hemolymph (circulatory fluid) of insects. We found that the gene *Trehalase* had significant signatures of positive selection only in the highly

eusocial lineage (Table 5.3), and contributed to the enriched category 'Catalytic activity' (Table 5.5). The trehalase protein catalyzes the conversion of trehalose into glucose, providing energy for the activity of flight muscles (Becker et al. 1996), as well as playing a crucial role in development, stress recovery, and synthesis of chitin in insects (Shukla et al. 2015). Changes in the metabolism of sugars might have been particularly important in the social Hymenoptera. For example, foraging is metabolically expensive in bees (Harrison and Fewell 2002; Schippers et al. 2010), which exclusively use sugars to fuel their flight muscles (Blatt and Roces 2001; Suarez et al. 2005) and show variation in flight performance depending on worker role and age of foragers (Roberts and Elekonich 2005; Vance et al. 2009). Flight muscle activity, therefore, allows selection for specialization to act within the worker caste. Moreover, *D. melanogaster* mutants of the enzyme that synthesizes trehalose *(Tps1)* show severe growth defects on a low-protein diet (Matsuda et al. 2015). Trehalose may also be involved in cognitive functions, since old honeybee foragers show brain overexpression of *Tps1* (Whitfield et al. 2003).

Positive selection acting on genes associated with metabolism and transport of trehalose might be related, in particular, to wasp queens entering a dormant state and living underground during the winter. In addition to *Treh*, we found evidence of positive selection for the gene *Facilitated trehalose transporter Tret1-1* (Table 5.3), which regulates levels of trehalose in the hemolymph and its incorporation into tissues (Kikawada et al. 2007; Kanamori et al. 2010). Accumulation of trehalose provides tolerance to cold and stability of protein structure in insects (Sinclair et al. 2003, 2013; Chen and Haddad 2004; Andersen et al. 2011; Koštál et al. 2012). Overwintering behavior might have selected for

extreme caste dimorphism in the Vespinae (West-Eberhard 1978; O'Donnell 1998), which are distributed throughout northern temperate regions and have queens with large body sizes. Elevated levels of trehalose after exposure to cold have been reported in the primitively eusocial *Polistes* (Strassmann et al. 1984). Furthermore, high survival of overwintering queens and queen body size correlate positively in the yellowjacket *Vespula maculifrons* (Kovacs and Goodisman 2012) and the paper wasp *Polistes gallicus* (Dani 1994).

Other signals of positive selection concerning metabolism of carbohydrates were detected for genes in the glycolysis pathway. We found accelerated rates of evolution for the glycolytic enzymes *Aldolase* and *Enolase* (Table 5.3), both contributing to the enriched category 'Catalytic Activity' (Table 5.5), among others. Changes in expression of *Enolase* relate to caste determination in *Apis mellifera*, where queen-destined larvae show increased levels of enolase protein relative to larvae that will become workers (Li et al. 2010). Comparative proteomic analyses reveal that foraging bees have higher levels of aldolase in comparison to workers performing within-hive activities (Schippers et al. 2006; Wolschin and Amdam 2007). Moreover, sterile honeybee workers have higher levels of aldolase protein compared to reproductive workers (Cardoen et al. 2011a). Adult caste differences in the abundance of glycolytic enzymes, therefore, likely result from the activity of flight muscles in foragers (Suarez et al. 2005). Our findings suggest that in addition to changes in the regulation of these glycolytic enzymes, changes in their coding sequences might have influenced the origin of advanced eusociality.

Accelerated evolution of glycolytic enzymes may relate to post-copulatory physiology. Insects rely largely on glycolysis to power sperm motility

(Werner and Simmons 2008). Social insects usually mate only once and maintain a lifelong supply of sperm in a specialized storage organ, the spermatheca. Aldolase and enolase have been detected in spermathecal fluid (Baer et al. 2009) as well as sperm fluid and seminal vesicles (Collins et al. 2006) in *A. mellifera*. Studies of *D. melanogaster* show that sperm and seminal fluid proteins transferred during mating launch females on a series of behavioral and physiological changes, such as decreased receptivity to mating and elevated egg laying (Wolfner 1997; Chapman 2001; Gillott 2003; McGraw et al. 2004). Furthermore, at the molecular level, mated females of *D. melanogaster* overexpress aldolase in their reproductive tract tissues (Mack et al. 2006). Likewise, behavioral and physiological post-mating changes in queens of *A. mellifera* correlate with gene expression patterns in brains and ovaries (Kocher et al. 2008). In contrast to ants and bees (Gobin et al. 2006; Gotoh et al. 2009), caste differences related to female reproductive anatomy seem implausible in the Vespinae, since queens and workers of the genus *Vespa*, for example, have strikingly similar spermathecae (Gotoh et al. 2008). Molecular mechanisms involved in post-mating sperm selection, storage, and expenditure seem more likely to have fitness consequences in social wasps (Boomsma 2013; Beani et al. 2014), as it is expected based on insights from ants and bees (Baer et al. 2006; den Boer et al. 2009, 2010; Jaffe et al. 2012). Seminal fluids that incapacitate rival sperm, however, provide another mechanism for differences in reproductive success in social insects (den Boer et al. 2010; Avila et al. 2011). Among yellowjackets, queens of *Vespula squamosa* mate with multiple males and show skewed sperm usage (Hoffman et al. 2008; see, however, Ross 1986). Additionally, colony size correlates negatively with paternity skew in other yellowjacket species (Loope et al. 2014). Highly

variable counts of stored sperm among queens of the yellowjacket *Dolichovespula maculata* suggest that sperm quantity may limit nest growth and affect queen fitness (Stein et al. 1996). Perhaps processes involving glycolytic enzymes influence the usage of sperm and successful founding of colonies in the Vespinae. Although glycolytic enzymes have been identified in sperm of *D. melanogaster*, these lacked signatures of accelerated evolution (Dorus et al. 2006), whereas proteins from the male accessory gland showed evidence of positive selection (Swanson et al. 2001; Mueller et al. 2005). This test of positive selection on glycolytic enzymes, however, was based on pairwise comparisons between two species of *Drosophila* (Dorus et al. 2006). Strong stabilizing selection acting on glycolytic enzymes in *Drosophila* may be expected considering that, for example, an *Aldolase* mutant with a single amino acid change in *D. melanogaster* shows decreased levels of ATP, reduced lifespan, and neurodegeneration (Miller et al. 2012). Such functional constraints for enzymes involved in glycolysis might have been relaxed in other taxa. For example, mammals possess multiple tissue-specific enolases, and an enolase copy unique to sperm (Edwards and Grootegoed 1983; Tracy and Hedges 2000).

In addition to glycolytic enzymes, we found evidence of accelerated evolution for other genes related to energy production. Specifically, we detected signals of positive selection for genes involved in metabolism of pyruvate and the tricarboxylic acid cycle, such as *Pyruvate carboxylase (PCB)*, *Isocitrate dehydrogenase (Idh)*, *Malate dehydrogenase 2 (Mdh2)*, and *CG7920* (Table 5.3). These findings further implicate the evolution of highly eusocial behavior to changes in metabolic function (Hunt et al. 2010b; Woodard et al. 2011; Roux et al. 2014). This association may result from different selection pressures experienced by castes;

queens are selected for their fecundity and longevity, and the molecular machinery underlying these traits, whereas workers are selected for their specialized, non-reproductive roles (Strassmann and Queller 2007; Smith et al. 2008; Bloch and Grozinger 2011; Sumner 2014). Caste differences in energy expenditure have been demonstrated in other taxa, such as eusocial mammals. Colonies of the Damaraland mole rat, for example, have a caste of infrequent workers that specialize in building up body reserves for potential dispersal when environmental conditions seem suitable (Scantlebury et al. 2006).

The origin and elaboration of eusociality likely involved demands in sensory and cognitive functions and changes in brain regions related to these functions (Chittka and Niven 2009; O'Donnell et al. 2011, 2013; Muscedere et al. 2014). We found evidence of positive selection for genes related to sound and visual perception (Table 5.3). The genes *Nervana 2* and *3 (nrv2, nrv3)* mediate the transport of sodium and potassium ions across membranes and therefore influence a variety of processes. In *D. melanogaster*, where *nrv2* and *nrv3* show expression specific to cell types in the auditory Johnston's organ, knocking down *nrv2* causes severe deafness (Roy et al. 2013). Furthermore, *Drosophila* flies with a reduced copy number of *nrv3* show increased sensitivity to noise trauma (Christie et al. 2013). Vibrations and sounds perceived through Johnston's organs transmit signals of alarm, recruitment, and larval provisioning cues in social insects (Kirchner 1997; Hunt and Richard 2013; Leonhardt et al. 2016). Larvae of the hornet *Vespa orientalis*, for example, produce scraping sounds that are interpreted by workers as hunger signals (Ishay and Landau 1972). Moreover, the phylogenetic distribution of vibrational signaling suggests that this trait characterized the hypothetical common ancestor of eusocial wasps (Jeanne and

Suryanarayanan 2011). These signals might have influenced the evolution of eusociality in the Vespidae, where larval nutrition underlies female caste determination (O'Donnell 1998). For example, larvae of the paper wasp *Polistes metricus* exhibit partial shifts in expression of caste-related genes depending on nourishment (Berens et al. 2015b).

Regarding changes in genes related to visual perception, we found evidence of positive selection for the light-sensitive protein encoded by *Rhodopsin 2* and found in photoreceptor cells (Table 5.3). Analyses of neuroanatomical data from social hymenopterans suggest that selection for visual acuity probably increases in lineages with large-bodied, aboveground foragers (Muscedere et al. 2014), such as the vespine wasps. Changes related to processing of visual information may result from the environmental challenges of foraging (Greiner et al. 2007). Another possible link may be recognition of signals of individual quality, such as facial markings in *Polistes* paper wasps (Tibbetts and Dale 2004; Tannure-Nascimento et al. 2008; Sheehan and Tibbetts 2011). An additional possibility may be the development among social insects of nocturnal foraging (Warrant 2008). Among the Vespinae, yellowjackets usually forage during daytime, hornets show facultative nocturnal activity, and the genus *Provespa* forage exclusively at night (Matsuura and Yamane 1990; Greene 1991). Although *Vespa*, for example, lack the eye morphology traits associated with nocturnal foraging (Kelber et al. 2011), differences in foraging lifestyles might be concomitant with molecular changes in photoreception within the worker caste. Additional changes related to the visual system were detected in genes determining compound eye morphogenesis, such as *auxilin* (Eun et al. 2008) and *Inositol-requiring enzyme-1* (Coelho et al. 2013), although these genes are involved

in a variety of processes. For example, *auxilin* plays a general role in the Notch signaling pathway (Hagedorn et al. 2006), which controls the fate of cells during animal development (Artavanis-Tsakonas et al. 1999; Hori et al. 2013).

In comparison to genes known to have effects on foraging behavior in the social Hymenoptera, we found positively selected orthologs corresponding to *Insulin-like receptor* and *vitellogenin* (Table 5.3). The reproductive ground plan hypothesis for the origin of eusociality states that pathways regulating reproduction in solitary insects have been rewired to influence age-related changes in the behavior of workers (Amdam et al. 2004, 2006). Insulin-like signaling and vitellogenin are two of such pathways, and both have been associated with division of labor and queen longevity in bees and ants (Corona et al. 2007; Ament et al. 2008; Lu and Pietrantonio 2011; Libbrecht et al. 2013). Perhaps the interplay of these pathways has also influenced the evolution of division of labor in wasps. Our transcriptomic data also included a cGMP-dependent kinase matching the putative foraging gene *Vvfor* (UniProt entry A1YTU8) of *Vespula vulgaris* (Tobback et al. 2008), with BLASTP coverage and identity greater than 97% using vespine sequences as queries, and the *for* genes of the bees *Bombus terrestris* and *Apis cerana* (UniProt entries C6GBY7 and A0A0H3WKE9, respectively), among others. However, the ortholog in our data set that matched the *foraging* gene (Pereira and Sokolowski 1993; Ben-Shahar 2003) did not show evidence of positive selection.

Our results are generally in agreement with patterns of positive selection in other eusocial insects. Studies of *Apis mellifera* have found that caste differences are associated with expression of proteins that metabolize carbohydrates (Sen Sarma et al. 2007; Cardoen et al. 2011b). Comparisons of

different levels of sociality in bees show that genes related to carbohydrate metabolism were evolving more rapidly in highly eusocial lineages (Woodard et al. 2011). In comparison to molecular patterns in bees (Woodard et al. 2011), however, gland development and signal transduction were not prominent functional categories in our analyses. Positive selection acting on genes functioning in mitochondria was particularly important during the evolution of ants, and this pattern has been attributed to the increased lifespan of queens (Roux et al. 2014). Although we found signals of accelerated evolution for genes functioning in mitochondria, queens in vespine wasps lack the extraordinarily long lifespans of ant queens. The 'genetic toolkit' hypothesis for the origin eusociality proposes that social behavior evolved from regulatory changes in conserved genes shared with solitary species (Toth and Robinson 2007). A revised of version of this hypothesis suggests that a 'loose toolkit' of crucial pathways rather than specific genes influence the evolution of eusociality (Berens et al. 2015a). Overlap of enriched functional categories between our analyses and studies of bees and ants seem to support the 'loose toolkit' hypothesis at the level of genomic sequence. However, we also found specific, positively selected genes shared with previous studies. For example, Woodard et al. (2011) and Roux et al. (2014) found, respectively, evidence of positive selection for the genes *Enolase* and *Mitochondrial ribosomal protein L37*, both of which had signatures of positive selection in our results (Table 5.3).

In conclusion, this study provides, for the first time, a test of the protein evolution hypothesis for vespid wasps, with emphasis on the molecular changes associated with the origin of highly eusocial behavior in yellowjackets and hornets. In particular, we found that genes contributing to carbohydrate

metabolic pathways and functioning in mitochondria might have been important to attain highly eusocial behavior. Our results may serve as a starting point for future work focused on finding specific sites under positive selection and experimentally investigating their effects on phenotypes.

References

Agosti, D.. (1994). A new inquiline ant (Hymenoptera: Formicidae) in *Cataglyphis* and its phylogenetic relationship. *Journal of Natural History*, *28*(4), 913–919.

Akino, T., Knapp, J. J., Thomas, J. A., & Elmes, G. W.. (1999). Chemical mimicry and host specificity in the butterfly *Maculinea rebeli*, a social parasite of *Myrmica* ant colonies. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1427), 1419–1426.

Akre, R. D., & Davis, H. G.. (1978). Biology and pest status of venomous wasps. *Annual Review of Entomology*, *23*(1), 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). *Journal of the Kansas Entomological Society*, *49*(1), 63–84.

Akre, R. D., Greene, A., MacDonald, J. F., Landolt, P. J., & Davis, H. G.. (1981). *The yellowjackets of America north of Mexico* (2nd ed.). Washington, DC: United States Department of Agriculture.

Alloway, T. M.. (1980). The origins of slavery in leptothoracine ants (Hymenoptera: Formicidae). *The American Naturalist*, *115*(2), 247–261.

Als, T. D., Vila, R., Kandul, N. P., Nash, D. R., Yen, S.-H., Hsu, Y.-F., & Pierce, N. E.. (2004). The evolution of alternative parasitic life histories in large blue butterflies. *Nature*, *432*(7015), 386–390.

Amdam, G. V, Csondes, A., Fondrk, M. K., & Page, R. E.. (2006). Complex social behaviour derived from maternal reproductive traits. *Nature*, *439*(7072), 76–78.

Amdam, G. V., Norberg, K., Fondrk, M. K., & Page, R. E.. (2004). Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(31), 11350–11355.

Ament, S. A., Corona, M., Pollock, H. S., & Robinson, G. E.. (2008). Insulin signaling is involved in the regulation of worker division of labor in honey bee colonies. *Proceedings of the National Academy of Sciences of the United States*

of America, *105*(11), 4226–4231.

Andersen, H. D., Wang, C., Arleth, L., Peters, G. H., & Westh, P.. (2011). Reconciliation of opposing views on membrane-sugar interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(5), 1874–1878.

Anderson, C., & McShea, D. W.. (2001). Individual versus social complexity, with particular reference to ant colonies. *Biological Reviews*, *76*(2), 211–237.

Anderson, K. E., Linksvayer, T. A., & Smith, C. R.. (2008). The causes and consequences of genetic caste determination in ants (Hymenoptera: Formicidae). *Myrmecological News*, *11*, 119–132.

Anderson, M.. (1984). The evolution of eusociality. *Annual Review of Ecology and Systematics*, *15*(1), 165–189.

Archer, M. E.. (1992). The taxonomy of *Vespa crabro* L. and *V. dybowskii* Andre (Hym., Vespinae). *Entomologist's Monthly Magazine*, *128*, 157–163.

Archer, M. E.. (1999). Taxonomy and world distribution of the Euro-Asian species of *Dolichovespula* (Hym., Vespinae). *Entomologist's Monthly Magazine*, *135*, 153–160.

Archer, M. E.. (2006). Taxonomy, distribution and nesting biology of species of the genus *Dolichovespula* (Hymenoptera, Vespidae). *Entomological Science*, *9*(3), 281–293.

Artavanis-Tsakonas, S., Rand, M. D., & Lake, R. J.. (1999). Notch signaling: cell fate control and signal integration in development. *Science*, *284*(5415), 770–776.

Avila, F. W., Sirot, L. K., LaFlamme, B. A., Rubinstein, C. D., & Wolfner, M. F.. (2011). Insect seminal fluid proteins: identification and function. *Annual Review of Entomology*, *56*, 21–40.

Baer, B., Armitage, S. A. O., & Boomsma, J. J.. (2006). Sperm storage induces an immunity cost in ants. *Nature*, *441*(7095), 872–875.

Baer, B., Eubel, H., Taylor, N. L., O'Toole, N., & Millar, A. H.. (2009). Insights into

female sperm storage from the spermathecal fluid proteome of the honeybee *Apis mellifera*. *Genome Biology*, *10*(6), R67.

Barbero, F., Bonelli, S., Thomas, J. A., Balletto, E., & Schönrogge, K.. (2009). Acoustical mimicry in a predatory social parasite of ants. *The Journal of Experimental Biology*, *212*(24), 4084–4090.

Beani, L., Dessì-Fulgheri, F., Cappa, F., & Toth, A.. (2014). The trap of sex in social insects: from the female to the male perspective. *Neuroscience and Biobehavioral Reviews*, *46*(Pt. 4), 519–533.

Becker, A., Schlöder, P., Steele, J. E., & Wegener, G.. (1996). The regulation of trehalose metabolism in insects. *Experientia*, *52*(5), 433–439.

Behura, S. K., & Severson, D. W.. (2012). Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS ONE*, 7(8), e43111.

Bell, G., & Mooers, A. O.. (1997). Size and complexity among multicellular organisms. *Biological Journal of the Linnean Society*, *60*(3), 345–363.

Bensasson, D., Zhang, D.-X., Hartl, D. L., & Hewitt, G. M.. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, *16*(6), 314–321.

Ben-Shahar, Y.. (2003). cGMP-dependent changes in phototaxis: a possible role for the foraging gene in honey bee division of labor. *Journal of Experimental Biology*, *206*(14), 2507–2515.

Ben-Shahar, Y., Robichon, A., Sokolowski, M. B., & Robinson, G. E.. (2002). Influence of gene action across different time scales on behavior. *Science*, *296*(5568), 741–744.

Bequaert, J.. (1932). A tentative synopsis of the hornets and yellow-jackets (Vespinae; Hymenoptera) of America. *Entomologica Americana*, *12*(2), 71–138.

Berens, A. J., Hunt, J. H., & Toth, A. L.. (2015). Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Molecular Biology and Evolution*, *32*(3), 690–703.

Berens, A. J., Hunt, J. H., & Toth, A. L.. (2015). Nourishment level affects caste-related gene expression in *Polistes* wasps. *BMC Genomics*, *16*(1), 235.

Berlocher, S. H.. (2003). When houseguests become parasites: Sympatric speciation in ants. *Proceedings of the National Academy of Sciences*, *100*(12), 6896–6898.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F.. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319.

Blatt, J., & Roces, F.. (2001). Haemolymph sugar levels in foraging honeybees (*Apis mellifera carnica*): dependence on metabolic rate and in vivo measurement of maximal rates of trehalose synthesis. *Journal of Experimental Biology*, *204*(15), 2709–2716.

Bloch, G., & Grozinger, C. M.. (2011). Social molecular pathways and the evolution of bee societies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1574), 2155–2170.

Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R., & Sams, S.. (1978). Relationships between body size and some life history parameters. *Oecologia*, *37*(2), 257–272.

Bolger, A. M., Lohse, M., & Usadel, B.. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

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*. *Behavioral Ecology and Sociobiology*, *65*(4), 633–640.

Bonckaert, W., Vuerinckx, K., Billen, J., Hammond, R. L., Keller, L., & Wenseleers, T.. (2008). Worker policing in the German wasp *Vespula germanica*. *Behavioral Ecology*, *19*(2), 272–278.

Bonner, J. T.. (1988). *The evolution of complexity by means of natural selection*. Princeton, New Jersey: Princeton University Press.

Bonner, J. T.. (2004). The size-complexity rule. *Evolution*, *58*(9), 1883–1890.

Boomsma, J. J.. (2013). Beyond promiscuity: mate-choice commitments in social breeding. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *368*(1613), 20120050.

Boomsma, J. J., & Nash, D. R.. (2014). Evolution: sympatric speciation the eusocial way. *Current Biology*, *24*(17), R798–800.

Boomsma, J. J., & Ratnieks, F. L. W.. (1996). Paternity in eusocial Hymenoptera. *Philosophical Transactions: Biological Sciences*, *351*(1342), 947–975.

Bourke, A. F. G.. (1999). Colony size, social complexity and reproductive conflict in social insects. *Journal of Evolutionary Biology*, *12*(2), 245–257.

Bourke, A. F. G., & Franks, N. R.. (1991). Alternative adaptations, sympatric speciation and the evolution of parasitic, inquiline ants. *Biological Journal of the Linnean Society*, *43*(3), 157–178.

Bourke, A. F. G., & Franks, N. R.. (1995). *Social evolution in ants*. Princeton, New Jersey: Princeton University Press.

Brady, S. G., Schultz, T. R., Fisher, B. L., & Ward, P. S.. (2006). Evaluating alternative hypotheses for the early evolution and diversification of ants. *Proceedings of the National Academy of Sciences*, *103*(48), 18172–18177.

Brandt, M., Foitzik, S., Fischer-Blass, B., & Heinze, J.. (2005). The coevolutionary dynamics of obligate ant social parasite systems — between prudence and antagonism. *Biological Reviews*, *80*(2), 251–267.

Brown, J. M., & Lemmon, A. R.. (2007). The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systematic Biology*, *56*(4), 643–655.

Buschinger, A.. (1986). Evolution of social parasitism in ants. *Trends in Ecology & Evolution*, *1*(6), 155–160.

Buschinger, A.. (1990). Sympatric speciation and radiative evolution of socially parasitic ants - Heretic hypotheses and their factual background. *Journal of Zoological Systematics and Evolutionary Research*, *28*(4), 241–260.

Buschinger, A.. (2009). Social parasitism among ants: a review (Hymenoptera:

Formicidae). *Myrmecological News*, *12*, 219–235.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., … Galtier, N.. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, *12*(5), 834–845.

Calvignac, S., Konecny, L., Malard, F., & Douady, C. J.. (2011). Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*, *11*(2), 246–254.

Cardinal, S., & Danforth, B. N.. (2011). The antiquity and evolutionary history of social behavior in bees. *PLoS ONE*, *6*(6), e21086.

Cardoen, D., Ernst, U. R., Van Vaerenbergh, M., Boerjan, B., de Graaf, D. C., Wenseleers, T., … Verleyen, P.. (2011). Differential proteomics in dequeened honeybee colonies reveals lower viral load in hemolymph of fertile worker bees. *PloS One*, *6*(6), e20043.

Cardoen, D., Wenseleers, T., Ernst, U. R., Danneels, E. L., Laget, D., DE Graaf, D. C., … Verleyen, P.. (2011). Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Molecular Ecology*, *20*(19), 4070–4084.

Carpenter, J. M.. (1982). The phylogenetic relationships and natural classificition of the Vespoidea (Hymenoptera). *Systematic Entomology*, *7*(1), 11–38.

Carpenter, J. M.. (1987a). On "The evolutionary genetics of social wasps" and the phylogeny of the Vespinae (Hymenoptera, Vespidae). *Insectes Sociaux*, *34*(1), 58–64.

Carpenter, J. M.. (1987b). Phylogenetic relationships and classification of the Vespinae (Hymenoptera: Vespidae). *Systematic Entomology*, *12*(4), 413–431.

Carpenter, J. M.. (1989). Testing scenarios: wasp social behavior. *Cladistics*, *5*(2), 131–144.

Carpenter, J. M.. (1997). Phylogenetic relationships among European *Polistes* and the evolution of social parasitism (Hymenoptera: Vespidae, Polistinae). *Memoires Du Museum National d'Histoire Naturelle*, *173*, 135–161.

Carpenter, J. M., Dvořák, L., & Pickett, K. M.. (2011). *Dolichovespula albida* (Sladen), a valid species, not a synonym of *D. norwegica* (Fabricius) (Hymenoptera: Vespidae, Vespinae). *Entomologica Americana*, *117*(3), 113–116.

Carpenter, J. M., & Glare, T. R.. (2010). Misidentification of *Vespula alascensis* as *V. vulgaris* in North America (Hymenoptera: Vespidae; Vespinae). *American Museum Novitates*, 1–7.

Carpenter, J. M., & Kojima, J.. (1997). Checklist of the species in the subfamily Vespinae (Insecta: Hymenoptera: Vespidae). *Natural History Bulletin of Ibaraki University*, *1*, 51–92.

Carpenter, J. M., Strassmann, J. E., Turillazzi, S., Hughes, C. R., Solis, C. R., & Cervo, R.. (1993). Phylogenetic relationships among paper wasp social parasites and their hosts (Hymenoptera: Vespidae; Polistinae). *Cladistics*, *9*(2), 129–146.

Carpenter, J., & Perera, E.. (2006). Phylogenetic relationships among yellowjackets and the evolution of social parasitism (Hymenoptera : Vespidae, Vespinae). *American Museum Novitates*, *31*(4), 375–386.

Castoe, T. A., Doan, T. M., & Parkinson, C. L.. (2004). Data partitions and complex models in Bayesian analysis: The phylogeny of gymnophthalmid lizards. *Systematic Biology*, *53*(3), 448–469.

Castoe, T. A., & Parkinson, C. L.. (2006). Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). *Molecular Phylogenetics and Evolution*, *39*(1), 91–110.

Castresana, J.. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552.

Cervo, R.. (2006). *Polistes* wasps and their social parasites: an overview. *Annales Zoologici Fennici*, *43*(5-6), 531–549.

Chapman, T.. (2001). Seminal fluid-mediated fitness traits in *Drosophila*. *Heredity*, *87*(5), 511–521.

Chen, Q., & Haddad, G. G.. (2004). Role of trehalose phosphate synthase and trehalose during hypoxia: from flies to mammals. *The Journal of Experimental Biology*, *207*(18), 3125–3129.

Chittka, L., & Niven, J.. (2009). Are bigger brains better? *Current Biology*, *19*(21), R995–R1008.

Choudhary, M., Strassmann, J. E., Queller, D. C., Turillazzi, S., & Cervo, R.. (1994). Social parasites in polistine wasps are monophyletic: implications for sympatric speciation. *Proceedings: Biological Sciences*, *257*(1348), 31–35.

Christie, K. W., Sivan-Loukianova, E., Smith, W. C., Aldrich, B. T., Schon, M. A., Roy, M., … Eberl, D. F.. (2013). Physiological, anatomical, and behavioral changes after acoustic trauma in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15449–15454.

Cini, A., Bruschini, C., Poggi, L., & Cervo, R.. (2011). Fight or fool? Physical strength, instead of sensory deception, matters in host nest invasion by a wasp social parasite. *Animal Behaviour*, *81*(6), 1139–1145.

Clary, D. O., & Wolstenholme, D. R.. (1985). The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, *22*(3), 252–271.

Coelho, D. S., Cairrão, F., Zeng, X., Pires, E., Coelho, A. V, Ron, D., … Domingos, P. M.. (2013). Xbp1-independent Ire1 signaling is required for photoreceptor differentiation and rhabdomere morphogenesis in *Drosophila*. *Cell Reports*, *5*(3), 791–801.

Cole, B. J.. (1999). The selective advantage of low relatedness. *Science*, *285*(5429), 891–893.

Collins, A. M., & Gardner, L. M.. (2001). Partial Cytochrome b sequences for six Hymenoptera of the Eastern United States. *Journal of Heredity*, *92*(6), 519–521.

Collins, A. M., Caperna, T. J., Williams, V., Garrett, W. M., & Evans, J. D.. (2006). Proteomic analyses of male contributions to honey bee sperm storage and mating. *Insect Molecular Biology*, *15*(5), 541–549.

Conesa, A., & Götz, S.. (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, *2008*, 619832.

Corona, M., Velarde, R. A., Remolina, S., Moran-Lauter, A., Wang, Y., Hughes, K. A., & Robinson, G. E.. (2007). Vitellogenin, juvenile hormone, insulin signaling, and queen honey bee longevity. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(17), 7128–7133.

Danforth, B. N.. (2002). Evolution of sociality in a primitively eusocial lineage of bees. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 286–290.

Danforth, B. N., Brady, S. G., Sipes, S. D., & Pearson, A.. (2004). Single-copy nuclear genes recover Cretaceous-age divergences in bees. *Systematic Biology*, *53*(2), 309–326.

Danforth, B. N., Fang, J., & Sipes, S.. (2006). Analysis of family-level relationships in bees (Hymenoptera: Apiformes) using 28S and two previously unexplored nuclear genes: CAD and RNA polymerase II. *Molecular Phylogenetics and Evolution*, *39*(2), 358–372.

Danforth, B. N., Sauquet, H., & Packer, L.. (1999). Phylogeny of the bee genus *Halictus* (Hymenoptera: Halictidae) based on parsimony and likelihood analyses of nuclear EF-1alpha sequence data. *Molecular Phylogenetics and Evolution*, *13*(3), 605–618.

Dani, F. R.. (1994). Caste size differences in *Polistes gallicus* (L.) (Hymenoptera Vespidae). *Ethology Ecology & Evolution*, *6*(Supplement 1), 67–73.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D.. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Davies, N. B., Bourke, A. F., & de L Brooke, M.. (1989). Cuckoos and parasitic ants: Interspecific brood parasitism as an evolutionary arms race. *Trends in Ecology & Evolution*, *4*(9), 274–278.

den Boer, S. P. A., Baer, B., & Boomsma, J. J.. (2010). Seminal fluid mediates ejaculate competition in social insects. *Science*, *327*(5972), 1506–1509.

den Boer, S. P. A., Baer, B., Dreier, S., Aron, S., Nash, D. R., & Boomsma, J. J.. (2009). Prudent sperm use by leaf-cutter ant queens. *Proceedings. Biological Sciences / The Royal Society*, *276*(1675), 3945–3953.

Dornhaus, A., Powell, S., & Bengston, S.. (2011). Group size and its effects on collective organization. *Annual Review of Entomology*, *57*(1), 123–141.

Dorus, S., Busby, S. A., Gerike, U., Shabanowitz, J., Hunt, D. F., & Karr, T. L.. (2006). Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nature Genetics*, *38*(12), 1440–1445.

Dowton, M., & Austin, A. D.. (1995). Increased genetic diversity in mitochondrial genes is correlated with the evolution of parasitism in the Hymenoptera. *Journal of Molecular Evolution*, *41*(6), 958–965.

Dowton, M., & Austin, A. D.. (1997). Evidence for AT-transversion bias in wasp (Hymenoptera: Symphyta) mitochondrial genes and its implications for the origin of parasitism. *Journal of Molecular Evolution*, *44*(4), 398–405.

Dowton, M., Cameron, S. L., Austin, A. D., & Whiting, M. F.. (2009). Phylogenetic approaches for the analysis of mitochondrial genome sequence data in the Hymenoptera, a lineage with both rapidly and slowly evolving mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *52*(2), 512–519.

Drapeau, M. D., Albert, S., Kucharski, R., Prusko, C., & Maleszka, R.. (2006). Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Research*, *16*(11), 1385–1394.

Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E., Smith, S. A., … Giribet, G.. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, *452*(7188), 745–749.

Dvořák, L.. (2007). Parasitism of *Dolichovespula norwegica* by *D. adulterina* (Hymenoptera: Vespidae). *Silva Gabreta*, *13*(1), 65–67.

Edwards, Y. H., & Grootegoed, J. A.. (1983). A sperm-specific enolase. *Journal of Reproduction and Fertility*, *68*(2), 305–310.

Emery, C.. (1909). Über den Ursprung der dulotischen, parasitischen und myrmekophilen Ameisen. *Biologische Centralblatt*, *29*, 352–362.

Enright, A. J., Van Dongen, S., & Ouzounis, C. A.. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*(7), 1575–1584.

Eun, S. H., Banks, S. M. L., & Fischer, J. A.. (2008). Auxilin is essential for Delta signaling. *Development*, *135*(6), 1089–1095.

Evans, H. E., & West-Eberhard, M. J.. (1970). *The Wasps*. Ann Arbor, Michigan: University of Michigan Press.

Evans, J. D., & Wheeler, D. E.. (2001). Expression profiles during honeybee caste determination. *Genome Biology*, *2*(1), 1–6.

Feldmeyer, B., Elsner, D., & Foitzik, S.. (2014). Gene expression patterns associated with caste and reproductive status in ants: worker-specific genes are more derived than queen-specific ones. *Molecular Ecology*, *23*(1), 151–161.

Felsenstein, J.. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*(4), 783–791.

Ferreira, P. G., Patalano, S., Chauhan, R., Ffrench-Constant, R., Gabaldón, T., Guigó, R., & Sumner, S.. (2013). Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biology*, *14*(2), R20.

Fisher-Reid, M. C., & Wiens, J.. (2011). What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from *Plethodon* salamanders and 13 other vertebrate clades. *BMC Evolutionary Biology*, *11*(1), 300.

Fischman, B. J., Woodard, S. H., & Robinson, G. E.. (2011). Molecular evolutionary analyses of insect societies. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(Supplement 2), 10847–10854.

Fletcher, W., & Yang, Z.. (2010). The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular Biology and Evolution*, *27*(10), 2257–2267.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R.. (1994). DNA primers for amplification of mitochondrial Cytochrome c oxidase subunit I from

diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Foster, K. R., & Ratnieks, F. L. W.. (2000). Social insects: Facultative worker policing in a wasp. *Nature*, *407*(6805), 692–693.

Foster, K. R., & Ratnieks, F. L. W.. (2001a). Convergent evolution of worker policing by egg eating in the honeybee and common wasp. *Proceedings: Biological Sciences*, *268*(1463), 169–174.

Foster, K. R., & Ratnieks, F.. (2001b). Paternity, reproduction and conflict in vespine wasps: a model system for testing kin selection predictions. *Behavioral Ecology and Sociobiology*, *50*(1), 1–8.

Foster, K. R., & Ratnieks, F. L. W.. (2001c). The effect of sex allocation biasing on the evolution of worker policing in hymenopteran societies. *The American Naturalist*, *158*(6), 615–623.

Foster, K. R., Ratnieks, F. L. W., Gyllenstrand, N., & Thorén, P. A.. (2001). Colony kin structure and male production in *Dolichovespula* wasps. *Molecular Ecology*, *10*(4), 1003–1010.

Foster, K. R., Ratnieks, F. L. W., & Raybould, A. F.. (2000). Do hornets have zombie workers? *Molecular Ecology*, *9*(6), 735–742.

Foster, K. R., Seppa, P., Ratnieks, F. L. W., & Thoren, P. A.. (1999). Low paternity in the hornet *Vespa crabro* indicates that multiple mating by queens is derived in vespine wasps. *Behavioral Ecology and Sociobiology*, *46*(4), 252–257.

Foster, K. R., Wenseleers, T., & Ratnieks, F. L. W.. (2006). Kin selection is the key to altruism. *Trends in Ecology & Evolution*, *21*(2), 57–60.

Foster, K. R., Seppa, P., Ratnieks, F. L. W., & Thoren, P. A.. (1999). Low paternity in the hornet *Vespa crabro* indicates that multiple mating by queens is derived in vespine wasps. *Behavioral Ecology and Sociobiology*, *46*(4), 252–257.

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*. *Molecular Ecology*, *13*(12), 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*(23), 3150–3152.

Gadagkar, R.. (2009). Interrogating an insect society. *Proceedings of the National Academy of Sciences*, *106*(26), 10407–10414.

Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D. F., Smith, C. R., … Smith, C. D.. (2012). The genomic impact of 100 million years of social evolution in seven ant species. *Trends in Genetics*, *28*(1), 14–21.

Garnery, L., Cornuet, J. M., & Solignac, M.. (1992). Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, *1*(3), 145–154.

Gibbs, J., Albert, J., & Packer, L.. (2012). Dual origins of social parasitism in North American *Dialictus* (Hymenoptera: Halictidae) confirmed using a phylogenetic approach. *Cladistics*, *28*(2), 195–207.

Gillott, C.. (2003). Male accessory gland secretions: modulators of female reproductive physiology and behavior. *Annual Review of Entomology*, *48*, 163–184.

Gobin, B., Ito, F., Peeters, C., & Billen, J.. (2006). Queen-worker differences in spermatheca reservoir of phylogenetically basal ants. *Cell and Tissue Research*, *326*(1), 169–178.

Goloboff, P. A., Farris, J. S., Källersjö, M., Oxelman, B., Ramıacute, M., … Szumik, C. A.. (2003). Improvements to resampling measures of group support. *Cladistics*, *19*(4), 324–332.

Goloboff, P. A., Farris, J. S., & Nixon, K. C.. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, *24*(5), 774–786.

Goloboff, P. A., Mattoni, C. I., & Quinteros, A. S.. (2006). Continuous characters analyzed as such. *Cladistics*, *22*(6), 589–601.

Goodisman, M. A. D., Kovacs, J. L., & Hoffman, E. A.. (2007a). Lack of conflict during queen production in the social wasp *Vespula maculifrons*. *Molecular Ecology*, *16*(12), 2589–2595.

Goodisman, M. A. D., Kovacs, J. L., & Hoffman, E. A.. (2007b). The significance of multiple mating in the social wasp *Vespula maculifrons*. *Evolution*, *61*(9), 2260–2267.

Goodisman, M. A. D., Kovacs, J. L., & Hunt, B. G.. (2008). Functional genetics and genomics in ants (Hymenoptera: Formicidae): The interplay of genes and social life. *Myrmecological News*, *11*, 107–117.

Goodisman, M. A. D., Matthews, R. W., & Crozier, R. H.. (2002). Mating and reproduction in the wasp *Vespula germanica*. *Behavioral Ecology and Sociobiology*, *51*(6), 497–502.

Gotoh, A., Billen, J., Hashim, R., & Ito, F.. (2008). Comparison of spermatheca morphology between reproductive and non-reproductive females in social wasps. *Arthropod Structure & Development*, *37*(3), 199–209.

Gotoh, A., Billen, J., Hashim, R., & Ito, F.. (2009). Evolution of specialized spermatheca morphology in ant queens: insight from comparative developmental biology between ants and polistine wasps. *Arthropod Structure & Development*, *38*(6), 521–525.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A.. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech*, *29*, 644–652.

Gräff, J., Jemielity, S., Parker, J. D., Parker, K. M., & Keller, L.. (2007). Differential gene expression between adult queens and workers in the ant *Lasius niger*. *Molecular Ecology*, *16*(3), 675–83.

Greene, A.. (1979). Behavioral characters as indicators of yellowjacket phylogeny (Hymenoptera: Vespidae). *Annals of the Entomological Society of America*, *72*(5), 614-619.

Greene, A.. (1991). *Dolichovespula* and *Vespula*. In K. G. Ross & R. W. Matthews (Eds.), *The social biology of wasps:* (pp. 263–305). Ithaca, NY: Cornell University Press.

Greene, A., Akre, R. D., & Landolt, P. J.. (1978). Behavior of the yellowjacket social parasite, *Dolichovespula arctica* (Rohwer) (Hymenoptera: Vespidae). *Melanderia*, *29*, 1–28.

Greiner, B., Narendra, A., Reid, S. F., Dacke, M., Ribi, W. A., & Zeil, J.. (2007). Eye structure correlates with distinct foraging-bout timing in primitive ants. *Current Biology*, *17*(20), R879–R880.

Grozinger, C. M., Fan, Y., Hoover, S. E. R., & Winston, M. L.. (2007). Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*). *Molecular Ecology*, *16*(22), 4837–4848.

Guindon, S., & Gascuel, O.. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*(5), 696–704.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., … Regev, A.. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*(8), 1494–1512.

Hagedorn, E. J., Bayraktar, J. L., Kandachar, V. R., Bai, T., Englert, D. M., & Chang, H. C.. (2006). *Drosophila melanogaster* auxilin regulates the internalization of Delta to control activity of the Notch signaling pathway. *The Journal of Cell Biology*, *173*(3), 443–452.

Hamilton, W. D.. (1964a). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, *7*(1), 1–16.

Hamilton, W. D.. (1964b). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, *7*(1), 17–52.

Harpur, B. A., Kent, C. F., Molodtsova, D., Lebon, J. M. D., Alqarni, A. S., Owayss, A. A., & Zayed, A.. (2014). Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proceedings of the National Academy of Sciences*, *111*(7), 2614–2619.

Harpur, B. A., & Zayed, A.. (2013). Accelerated evolution of innate immunity proteins in social insects: Adaptive evolution or relaxed constraint? *Molecular Biology and Evolution*, *30*(7), 1665–1674.

Harrison, J. F., & Fewell, J. H.. (2002). Environmental and genetic influences on flight metabolic rate in the honey bee, *Apis mellifera*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *133*(2), 323–333.

Harvey, P. H., & Pagel, M. D.. (1991). *The comparative method in evolutionary biology*. Oxford: Oxford University Press.

Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G. W., Edgecombe, G. D., … Dunn, C. W.. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proceedings. Biological Sciences / The Royal Society*, *276*(1677), 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 Sociaux*, *53*(4), 399–402.

Hennig, W.. (1965). Phylogenetic Systematics. *Annual Review of Entomology*, *10*(1), 97–116.

Heraty, J., Ronquist, F., Carpenter, J. M., Hawks, D., Schulmeister, S., Dowling, A. P., … Sharkey, M.. (2011). Evolution of the hymenopteran megaradiation. *Molecular Phylogenetics and Evolution*, *60*(1), 73–88.

Hines, H. M., & Cameron, S. A.. (2010). The phylogenetic position of the bumble bee inquiline *Bombus inexspectatus* and implications for the evolution of social parasitism. *Insectes Sociaux*, *57*(4), 379–383.

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. *Proceedings of the National Academy of Sciences*, *104*(9), 3295–3299.

Hoffman, E. A., & Goodisman, M. A. D.. (2007). Gene expression and the evolution of phenotypic diversity in social wasps. *BMC Biology*, *5*(1), 23.

Hoffman, E., Kovacs, J., & Goodisman, M.. (2008). Genetic structure and breeding system in a social wasp and its social parasite. *BMC Evolutionary Biology*, *8*(1), 1–13.

Holbrook, C. T., Barden, P. M., & Fewell, J. H.. (2011). Division of labor increases with colony size in the harvester ant *Pogonomyrmex californicus*. *Behavioral Ecology*, *22*(5), 960–966.

Hölldobler, B., & Wilson, E. O.. (1990). *The Ants*. Harvard University Press.

Hölldobler, B., & Wilson, E. O.. (2008). *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies*. W. W. Norton & Company.

Hori, K., Sen, A., & Artavanis-Tsakonas, S.. (2013). Notch signaling at a glance. *Journal of Cell Science*, *126*(Pt 10), 2135–2140.

Huang, M. H., & Dornhaus, A.. (2008). A meta-analysis of ant social parasitism: host characteristics of different parasitism types and a test of Emery's rule. *Ecological Entomology*, *33*(5), 589–596.

Hunt, B. G., Wyder, S., Elango, N., Werren, J. H., Zdobnov, E. M., Yi, S. V, & Goodisman, M. A. D.. (2010a). Sociality is linked to rates of protein evolution in a highly social insect. *Molecular Biology and Evolution*, *27*(3), 497–500.

Hunt, B. G., Wyder, S., Elango, N., Werren, J. H., Zdobnov, E. M., Yi, S. V, & Goodisman, M. A. D.. (2010b). Sociality is linked to rates of protein evolution in a highly social insect. *Molecular Biology and Evolution*, *27*(3), 497–500.

Hunt, J. H.. (2007). *The Evolution of Social Wasps*. Oxford University Press.

Hunt, J. H., & Richard, F.-J.. (2013). Intracolony vibroacoustic communication in social insects. *Insectes Sociaux*, *60*(4), 403–417.

Ingram, K. K., Oefner, P., & Gordon, D. M.. (2005). Task-specific expression of the foraging gene in harvester ants. *Molecular Ecology*, *14*(3), 813–818.

Ishay, J., & Landau, E. M.. (1972). Vespa larvae send out rhythmic hunger signals. *Nature*, *237*(5353), 286–287.

Jaffe, R., Garcia-Gonzalez, F., den Boer, S. P. A., Simmons, L. W., & Baer, B.. (2012). Patterns of paternity skew among polyandrous social insects: what can they tell us about the potential for sexual selection? *Evolution*, *66*(12), 3778–3788.

Jandt, J. M., Tibbetts, E. A., & Toth, A. L.. (2013). *Polistes* paper wasps: a model genus for the study of social dominance hierarchies. *Insectes Sociaux*, *61*(1), 11–27.

Jandt, J. M., & Toth, A. L.. (2015). Physiological and genomic mechanisms of social organization in wasps (Family: Vespidae). In A. Zayed & K. F.

Clement (Eds.), *Advances in Insect Physiology, Vol. 48* (pp. 95–130). Oxford: Academic Press.

Jansen, G., Savolainen, R., & Vepsalainen, K.. (2010). Phylogeny, divergence-time estimation, biogeography and social parasite-host relationships of the Holarctic ant genus *Myrmica* (Hymenoptera: Formicidae). *Molecular Phylogenetics and Evolution*, *56*(1), 294–304.

Jasper, W. C., Linksvayer, T. A., Atallah, J., Friedman, D., Chiu, J. C., & Johnson, B. R.. (2015). Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. *Molecular Biology and Evolution*, *32*(2), 334–346.

Jeanne, R. L.. (1977). Behavior of the obligate social parasite *Vespula arctica* (Hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, *50*(4), 541–557.

Jeanne, R. L.. (1980). Evolution of social behavior in the Vespidae. *Annual Review of Entomology*, *25*(1), 371–396.

Jeanne, R. L.. (1991). Polyethism. In K. G. Ross & R. W. Matthews (Eds.), *The Social Biology of Wasps:* (pp. 389–425). Cornell University Press.

Jeanne, R. L., & Suryanarayanan, S.. (2011). A new model for caste development in social wasps. *Communicative & Integrative Biology*, *4*(4), 373–377.

Jeanson, R., Fewell, J., Gorelick, R., & Bertram, S.. (2007). Emergence of increased division of labor as a function of group size. *Behavioral Ecology and Sociobiology*, *62*(2), 289–298.

Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H.. (2006). Phylogenomics: the beginning of incongruence? *Trends in Genetics*, *22*(4), 225–231.

Jermiin, L. S., & Crozier, R. H.. (1994). The Cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: Sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution*, *38*, 282–294.

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. *Current Biology*, *23*(20), 2058-2062.

Johnson, B., & Tsutsui, N.. (2011). Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics*, *12*(1), 164.

Johnson, C. A., Topoff, H., Vander Meer, R. K., & Lavine, B.. (2004). Do these eggs smell funny to you?: an experimental study of egg discrimination by hosts of the social parasite *Polyergus breviceps* (Hymenoptera: Formicidae). *Behavioral Ecology and Sociobiology*, *57*(3), 245–255.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S.. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.

Kälersjö, M., Albert, V. A., & Farris, J. S.. (1999). Homoplasy increases phylogenetic structure. *Cladistics*, *15*(1), 91–93.

Kamakura, M.. (2011). Royalactin induces queen differentiation in honeybees. *Nature*, *473*(7348), 478–483.

Kanamori, Y., Saito, A., Hagiwara-Komoda, Y., Tanaka, D., Mitsumasu, K., Kikuta, S., … Okuda, T.. (2010). The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involved in trehalose import into peripheral tissues. *Insect Biochemistry and Molecular Biology*, *40*(1), 30–37.

Kapheim, K. M.. (2016). Genomic sources of phenotypic novelty in the evolution of eusociality in insects. *Current Opinion in Insect Science*, *13*, 24–32.

Kapheim, K. M., Pan, H., Li, C., Salzberg, S. L., Puiu, D., Magoc, T., … Zhang, G.. (2015). Genomic signatures of evolutionary transitions from solitary to group living. *Science*, *348*(6239), 1139–1143.

Karsai, I., & Wenzel, J. W.. (1998). Productivity, individual-level and colony-level flexibility, and organization of work as consequences of colony size. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(15), 8665–8669.

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

Katoh, K.. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066.

Katoh, K., & Standley, D. M.. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kelber, A., Jonsson, F., Wallén, R., Warrant, E., Kornfeldt, T., & Baird, E.. (2011). Hornets can fly at night without obvious adaptations of eyes and ocelli. *PloS One*, *6*(7), e21892.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., & Bosch, T. C. G.. (2009). More than just orphans: are taxonomically-restricted genes important in evolution? *Trends in Genetics*, *25*(9), 404–413.

Kikawada, T., Saito, A., Kanamori, Y., Nakahara, Y., Iwata, K., Tanaka, D., … Okuda, T.. (2007). Trehalose transporter 1, a facilitated and high-capacity trehalose transporter, allows exogenous trehalose uptake into cells. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(28), 11585–11590.

Kilner, R. M., & Langmore, N. E.. (2011). Cuckoos versus hosts in insects and birds: adaptations, counter-adaptations and outcomes. *Biological Reviews*, *86*(4), 836–852.

Kimsey, L., & Carpenter, J.. (2012). The Vespinae of North America (Vespidae, Hymenoptera). *Journal of Hymenoptera Research*, *28*(0), 37–65.

Kirchner, W. H.. (1997). Acoustical communication in social insects BT - Orientation and Communication in Arthropods. In M. Lehrer (Ed.), (pp. 273–300). Basel: Birkhäuser Basel.

Kocher, S. D., Richard, F.-J., Tarpy, D. R., & Grozinger, C. M.. (2008). Genomic analysis of post-mating changes in the honey bee queen (*Apis mellifera*). *BMC Genomics*, *9*(1), 232.

Koštál, V., Šimek, P., Zahradníčková, H., Cimlová, J., & Štětina, T.. (2012). Conversion of the chill susceptible fruit fly larva (*Drosophila melanogaster*) to a freeze tolerant organism. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3270–3274.

Kovacs, J. L., & Goodisman, M. A. D.. (2007). Irregular brood patterns and worker reproduction in social wasps. *Naturwissenschaften*, *94*(12), 1011–1014.

Kovacs, J. L., & Goodisman, M. A. D.. (2012). Effects of size, shape, genotype, and mating status on queen overwintering survival in the social wasp *Vespula maculifrons*. *Environmental Entomology*, *41*(6), 1612–1620.

Kovacs, J. L., Hoffman, E. A., Marriner, S. M., & Goodisman, M. A. D.. (2010). Detecting selection on morphological traits in social insect castes: the case of the social wasp *Vespula maculifrons*. *Biological Journal of the Linnean Society*, *101*(1), 93–102.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L.. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580.

Lambardi, D., Dani, F., Turillazzi, S., & Boomsma, J.. (2007). Chemical mimicry in an incipient leaf-cutting ant social parasite. *Behavioral Ecology and Sociobiology*, *61*(6), 843–851.

Landolt, P. J., Monzón Sierra, J., Unruh, T. R., & Zack, R. R.. (2010). A new species of *Vespula*, and first record of *Vespa crabro* L.. (Hymenoptera: Vespidae) from Guatemala, Central America. *Zootaxa*, *2629*, 61–68.

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

Lattorff, H. M. G., & Moritz, R. F. A.. (2013). Genetic underpinnings of division of labor in the honeybee (*Apis mellifera*). *Trends in Genetics*, *29*(11), 641–648.

Leavitt, J. R., Hiatt, K. D., Whiting, M. F., & Song, H.. (2013). Searching for the optimal data partitioning strategy in mitochondrial phylogenomics: A phylogeny of Acridoidea (Insecta: Orthoptera: Caelifera) as a case study. *Molecular Phylogenetics and Evolution*, *67*(2), 508–494.

Lenhart, P., Dash, S. T., & Mackay, W. P.. (2013). A revision of the giant Amazonian ants of the genus *Dinoponera* (Hymenoptera, Formicidae). *Journal of Hymenoptera Research*, *31*, 119–164.

Lenoir, A., D'Ettorre, P., Errard, C., & Hefetz, A.. (2001). Chemical ecology and social parasitism in ants. *Annual Review of Entomology*, *46*(1), 573–599.

Leonhardt, S. D., Menzel, F., Nehring, V., & Schmitt, T.. (2016). Ecology and evolution of communication in social insects. *Cell*, *164*(6), 1277–1287.

Li, H., & Durbin, R.. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, C., Lu, G., & Ortí, G.. (2008). Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Systematic Biology*, *57*(4), 519–539.

Li, J., Wu, J., Begna Rundassa, D., Song, F., Zheng, A., & Fang, Y.. (2010). Differential protein expression in honeybee (*Apis mellifera* L.) larvae: underlying caste differentiation. *PloS One*, *5*(10), e13455.

Libbrecht, R., Corona, M., Wende, F., Azevedo, D. O., Serrão, J. E., & Keller, L.. (2013). Interplay between insulin signaling, juvenile hormone, and vitellogenin regulates maternal effects on polyphenism in ants. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(27), 11050–11055.

Liu, L., Yu, L., & Edwards, S. V.. (2010). A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology*, *10*(1), 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 Research*, *40*, W622–627.

Long, M., VanKuren, N. W., Chen, S., & Vibranovski, M. D.. (2013). New gene evolution: little did we know. *Annual Review of Genetics*, *47*, 307–333.

Loope, K. J., Chien, C., & Juhl, M.. (2014). Colony size is linked to paternity frequency and paternity skew in yellowjacket wasps and hornets. *BMC Evolutionary Biology*, *14*(1), 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). *Royal Society Open Science*, *2*(9), 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*). *Molecular Phylogenetics and Evolution*, *73*, 190–201.

Lorenzi, M. C.. (2006). The result of an arms race: the chemical strategies of Polistes social parasites. *Annales Zoologici Fennici*, *43*(5-6), 550–563.

Löytynoja, A., & Goldman, N.. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(30), 10557–10562.

Löytynoja, A., & Goldman, N.. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, *320*(5883), 1632–1635.

Lu, H.-L., & Pietrantonio, P. V.. (2011). Insect insulin receptors: insights from sequence and caste expression analyses of two cloned hymenopteran insulin receptor cDNAs from the fire ant. *Insect Molecular Biology*, *20*(5), 637–649.

MacDonald, J. F., & Matthews, R. W.. (1975). *Vespula squamosa*: a yellow jacket wasp evolving toward parasitism. *Science*, *190*(4218), 1003–1004.

MacDonald, J. F., & Matthews, R. W.. (1984). Nesting biology of the southern yellowjacket, *Vespula squamosa* (Hymenoptera: Vespidae): social parasitism and independent founding. *Journal of the Kansas Entomological Society*, *57*(1), 134–151.

MacDonald, J. F., Matthews, R. W., & Jacobson, R. S.. (1980). Nesting biology of the yellowjacket, *Vespula flavopilosa* (Hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, *53*(2), 448–458.

Mack, P. D., Kapelnikov, A., Heifetz, Y., & Bender, M.. (2006). Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(27), 10358–10363.

Maddison, W. P.. (1997). Gene trees in species trees. *Systematic Biology*, *46*(3), 523–536.

Maddison, W. P., & Maddison, D. R.. (2011). Mesquite: a modular system for evolutionary analysis. Version 2.75. http://mesquiteproject.org.

Markova-Raina, P., & Petrov, D.. (2011). High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Research*, *21*(6), 863–874.

Marshall, D. C.. (2010). Cryptic failure of partitioned Bayesian phylogenetic analyses: Lost in the land of long trees. *Systematic Biology*, *59*(1), 108–117.

Martin, S. J., Jenner, E. A., & Drijfhout, F. P.. (2007). Chemical deterrent enables a socially parasitic ant to invade multiple hosts. *Proceedings. Biological Sciences / The Royal Society*, *274*(1626), 2717–2721.

Martin, S. J., Takahashi, J., Ono, M., & Drijfhout, F. P.. (2008). Is the social parasite *Vespa dybowskii* using chemical transparency to get her eggs accepted? *Journal of Insect Physiology*, *54*(4), 700–707.

Matsuda, H., Yamada, T., Yoshida, M., & Nishimura, T.. (2015). Flies without trehalose. *The Journal of Biological Chemistry*, *290*(2), 1244–1255.

Matsuura, M., & Yamane, S.. (1984). *Comparative ethology of the vespine wasps (in Japanese)*. Sapporo, Japan: Hokkaido University Press.

Matsuura, M., & Yamane, S.. (1990). *Biology of the Vespine Wasps*. Springer-Verlag.

Matthews, R. W., & Matthews, J. R.. (1979). War of the yellow jacket queens. *Natural History*, *88*(8), 57–64.

Mattila, H. R., & Seeley, T. D.. (2007). Genetic diversity in honey bee colonies enhances productivity and fitness. *Science*, *317*(5836), 362–364.

McGraw, L. A., Gibson, G., Clark, A. G., & Wolfner, M. F.. (2004). Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*. *Current Biology*, *14*(16), 1509–1514.

McShea, D. W.. (1996). Metazoan complexity and evolution: Is there a trend? *Evolution*, *50*(2), 477–492.

Miller, D., Hannon, C., & Ganetzky, B.. (2012). A mutation in *Drosophila* Aldolase causes temperature-sensitive paralysis, shortened lifespan, and neurodegeneration. *Journal of Neurogenetics*, *26*(3-4), 317–327.

Miller, M. A., Pfeiffer, W., & Schwartz, T.. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gateway Computing Environments Workshop (GCE), 2010*.

Mirarab, S., Nguyen, N., & Warnow, T.. (2014). PASTA: Ultra-large multiple sequence alignment. In R. Sharan (Ed.), *Research in Computational Molecular Biology SE - 15* (Vol. 8394, pp. 177–191).

Miyamoto, M. M., & Fitch, W. M.. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology*, *44*(1), 64–76.

Monnin, T., & Peeters, C.. (1998). Monogyny and regulation of worker mating in the queenless ant *Dinoponera quadriceps*. *Animal Behaviour*, *55*(2), 299–306.

Morozova, O., Hirst, M., & Marra, M. A.. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics*, *10*(1), 135–151.

Mueller, R. L., Macey, J. R., Jaekel, M., Wake, D. B., & Boore, J. L.. (2004). Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(38), 13820–13825.

Mueller, J. L., Ravi Ram, K., McGraw, L. A., Bloch Qazi, M. C., Siggia, E. D., Clark, A. G., … Wolfner, M. F.. (2005). Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics*, *171*(1), 131–143.

Muscedere, M. L., Gronenberg, W., Moreau, C. S., & Traniello, J. F. A.. (2014). Investment in higher order central processing regions is not constrained by brain size in social insects. *Proceedings. Biological Sciences / The Royal Society*, *281*(1784), 20140217.

Nixon, K. C.. (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, *15*(4), 407–414.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., & Nieves-Aldrey, J. L.. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, *53*(1), 47–67.

Nylander, J. A. A., Wilgenbusch, J. C., Warren, D. L., & Swofford, D. L.. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, *24*(4), 581–583.

O'Donnell, S.. (1998). Reproductive caste determination in eusocial wasps (Hymenoptera : Vespidae). *Annual Review of Entomology*, *43*, 323–346.

O'Donnell, S., Clifford, M., & Molina, Y.. (2011). Comparative analysis of constraints and caste differences in brain investment among social paper wasps. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(17), 7107–7112.

O'Donnell, S., Clifford, M. R., DeLeon, S., Papa, C., Zahedi, N., & Bulova, S. J.. (2013). Brain size and visual environment predict species differences in paper wasp sensory processing brain regions (Hymenoptera: Vespidae, Polistinae). *Brain, Behavior and Evolution*, *82*(3), 177–184.

O'Meara, B. C.. (2012). Evolutionary inferences from phylogenies: a review of methods. *Annual Review of Ecology, Evolution, and Systematics*, *43*(1), 267–285.

O'Rourke, M. E., & Kurczewski, F. E.. (1983). Nest usurpation of *Vespula vulgaris* by *Dolichovespula arenaria* with successive parasitism of *Dolichovespula arenaria* by *Dolichovespula arctica* (Hymenoptera, Vespidae). *Entomological News*, *95*, 212–214.

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. *Current Biology*, *25*(12), 1638–1640.

Oster, G. F., & Wilson, E. O.. (1978). Caste and ecology in the social insects. *Monographs in Population Biology*, *12*, 1–352.

Oxley, P. R., Ji, L., Fetter-Pruneda, I., McKenzie, S. K., Li, C., Hu, H., … Kronauer,

D. J. C.. (2014). The genome of the clonal raider ant *Cerapachys biroi*. *Current Biology*, *24*(4), 451–458.

Pamilo, P., Pekkarinen, A., & Varvio-Aho, S.-L.. (1981). Phylogenetic relationships and origin of social parasitism in Vespidae and in *Bombus* and *Psithyrus* as revealed by enzyme genes. In *Biosystematics of social insects* (pp. 37–48). New York: Academic Press Inc.

Pantera, B., Hoffman, D. R., Carresi, L., Cappugi, G., Turillazzi, S., Manao, G., … Pazzagli, L.. (2003). Characterization of the major allergens purified from the venom of the paper wasp *Polistes gallicus*. *Biochimica et Biophysica Acta*, *1623*(2), 72–81.

Patalano, S., Vlasova, A., Wyatt, C., Ewels, P., Camara, F., Ferreira, P. G., … Sumner, S.. (2015). Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 13970–13975.

Pereboom, J. J. M., Jordan, W. C., Sumner, S., Hammond, R. L., & Bourke, A. F. G.. (2005). Differential gene expression in queen-worker caste determination in bumble-bees. *Proceedings. Biological Sciences / The Royal Society*, *272*(1568), 1145–1152.

Pereira, H. S., & Sokolowski, M. B.. (1993). Mutations in the larval foraging gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(11), 5044–5046.

Perrard, A., Lopez-Osorio, F., & Carpenter, J. M.. (2015). Phylogeny, landmark analysis and the use of wing venation to study the evolution of social wasps (Hymenoptera: Vespidae: Vespinae). *Cladistics*, doi:10.1111/cla.12138.

Perrard, A., Pickett, K., Villemant, C., Kojima, J., & Carpenter, J.. (2013). Phylogeny of hornets: a total evidence approach (Hymenoptera, Vespidae, Vespinae, Vespa). *Journal of Hymenoptera Research*, *32*, 1–15.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H.. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786.

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 Biology*, *9*(3), e1000602.

Pickett, K. M., & Carpenter, J. M.. (2010). Simultaneous analysis and the origin of eusociality in the Vespidae (Insecta: Hymenoptera). *Arthropod Systematics & Phylogeny*, *68*(1), 3–33.

Pickett, K. M., Osborne, D. M., Wahl, D., & Wenzel, J. W.. (2001). An enormous nest of *Vespula squamosa* from Florida, the largest social wasp nest reported from North America, with notes on colony cycle and reproduction. *Journal of the New York Entomological Society*, *109*(3), 408–415.

Powell, A. F. L. A., Barker, F. K., & Lanyon, S. M.. (2013). Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: An avian case study. *Molecular Phylogenetics and Evolution*, *66*(1), 69–79.

Prendini, L., Weygoldt, P., & Wheeler, W. C.. (2005). Systematics of the *Damon variegatus* group of African whip spiders (Chelicerata: Amblypygi): Evidence from behavior, morphology and DNA. *Organisms Diversity & Evolution*, *5*(3), 203–236.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., … Finn, R. D.. (2012). The Pfam protein families database. *Nucleic Acids Research*, *40*(Database issue), D290–301.

Queller, D. C., & Strassmann, J. E.. (2002). Kin selection. *Current Biology*, *12*(24), R832.

Rabeling, C., Schultz, T. R., Pierce, N. E., & Bacci, M.. (2014). A social parasite evolved reproductive isolation from its fungus-growing ant host in sympatry. *Current Biology*, *24*(17), 2047–2052.

Rambaut, A., Suchard, M. A., & Drummond, A. J.. (2013). Tracer v1.6, available from http://beast.bio.ed.ac.uk/tracer.

Ramírez, M. J.. (2005). Resampling measures of group support: a reply to Grant and Kluge. *Cladistics*, *21*(1), 83–89.

Ratcliff, W. C., Denison, R. F., Borrello, M., & Travisano, M.. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences of*

the United States of America, *109*(5), 1595–1600.

Ratnieks, F. L. W.. (1988). Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. *The American Naturalist*, *132*(2), 217–236.

Ratnieks, F. L. W., & Visscher, P. K.. (1989). Worker policing in the honeybee. *Nature*, *342*(6251), 796–797.

Reed, H. C., & Akre, R. D.. (1983a). Comparative colony behavior of the forest yellowjacket, *Vespula acadica* (Sladen) (Hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, *56*(4), 581–606.

Reed, H. C., & Akre, R. D.. (1983b). Usurpation behavior of the yellowjacket social parasite, *Vespula austriaca* (Panzer) (Hymenoptera: Vespidae). *American Midland Naturalist*, *110*(2), 419–432.

Reed, H. C., & Akre, R. D.. (1983c). Colony behavior of the obligate social parasite *Vespula austriaca* (Panzer) (Hymenoptera: Vespidae). *Insectes Sociaux*, *30*(3), 259–273.

Reeve, H. K.. (1991). *Polistes*. In *The Social Biology of Wasps* (pp. 99–148). Ithaca, New York: Comstock Publishing Associates.

Regier, J. C., Shultz, J. W., Ganley, A. R. D., Hussey, A., Shi, D., Ball, B., … Cunningham, C. W.. (2008). Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. *Systematic Biology*, *57*(6), 920–938.

Rehan, S. M., & Toth, A. L.. (2015). Climbing the social ladder: the molecular evolution of sociality. *Trends in Ecology & Evolution*, *30*(7), 426–433.

Richards, O. W.. (1971). The biology of the social wasps (Hymenoptera, Vespidae). *Biological Reviews*, *46*(4), 483–528.

Richter, M. R.. (2000). Social wasp (Hymenoptera: Vespidae) foraging behavior. *Annual Review of Entomology*, *45*(1), 121–150.

Roberts, S. P., & Elekonich, M. M.. (2005). Muscle biochemistry and the ontogeny of flight capacity during behavioral development in the honey bee, *Apis mellifera*. *The Journal of Experimental Biology*, *208*(Pt 22), 4193–4198.

Robinson, G. E., Fernald, R. D., & Clayton, D. F.. (2008). Genes and social behavior. *Science*, *322*(5903), 896–900.

Robinson, G. E., Grozinger, C. M., & Whitfield, C. W.. (2005). Sociogenomics: social life in molecular terms. *Nature Reviews. Genetics*, *6*(4), 257–270.

Rokas, A., & Carroll, S. B.. (2006). Bushes in the Tree of Life. *PLoS Biology*, *4*(11), e352.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P.. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542.

Roskens, V. A., Carpenter, J. M., Pickett, K. M., & Ballif, B. A.. (2010). Preservation of field samples for enzymatic and proteomic characterization: analysis of proteins from the trophallactic fluid of hornets and yellowjackets. *Journal of Proteome Research*, *9*(10), 5484–5491.

Ross, K. G.. (1985). Aspects of worker reproduction in four social wasp species (Insecta: Hymenoptera: Vespidae). *Journal of Zoology*, *205*(3), 411–424.

Ross, K. G.. (1986). Kin selection and the problem of sperm utilization in social insects. *Nature*, *323*(6091), 798–800.

Ross, K. G., & Matthews, R. W.. (1991). *The Social Biology of Wasps*. Ithaca, New York: Cornell University Press.

Rota, J., & Wahlberg, N.. (2012). Exploration of data partitioning in an eight-gene data set: phylogeny of metalmark moths (Lepidoptera, Choreutidae). *Zoologica Scripta*, *41*(5), 536–546.

Roux, J., Privman, E., Moretti, S., Daub, J. T., Robinson-Rechavi, M., & Keller, L.. (2014). Patterns of positive selection in seven ant genomes. *Molecular Biology and Evolution*, *31*(7), 1661–1685.

Roy, M., Sivan-Loukianova, E., & Eberl, D. F.. (2013). Cell-type-specific roles of Na+/K+ ATPase subunits in *Drosophila* auditory mechanosensation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(1), 181–186.

Ruano, F., Hefetz, A., Lenoir, A., Francke, W., & Tinaut, A.. (2005). Dufour's gland secretion as a repellent used during usurpation by the slave-maker ant *Rossomyrmex minuchae*. *Journal of Insect Physiology*, *51*(10), 1158–1164.

Sadd, B. M., Barribeau, S. M., Bloch, G., de Graaf, D. C., Dearden, P., Elsik, C. G., … Worley, K. C.. (2015). The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biology*, *16*(1), 76.

Saito, F., & Kojima, J.. (2011). Phylogenetic analysis and biogeography of the nocturnal hornets, *Provespa* (Insecta: Hymenoptera: Vespidae: Vespinae). *Species Diversity*, *16*, 65–74.

Sakagami, S. F., & Fukushima, K.. (1957). *Vespa dybowskii* Andre as a facultative temporary social parasite. *Insectes Sociaux*, *4*, 1–12.

Salichos, L., & Rokas, A.. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, *497*(7449), 327–331.

Salichos, L., Stamatakis, A., & Rokas, A.. (2014). Novel information theory-based measures for quantifying incongruence among phylogenetic trees. *Molecular Biology and Evolution*, *31*(5), 1261–1271.

Sanderson, M. J., & Shaffer, H. B.. (2002). Troubleshooting molecular phylogenetic analysis. *Annual Review of Ecology and Systematics*, *33*(1), 49–72.

Savolainen, R., & Vepsäläinen, K.. (2003). Sympatric speciation through intraspecific social parasitism. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12), 7169–7174.

Scantlebury, M., Speakman, J. R., Oosthuizen, M. K., Roper, T. J., & Bennett, N. C.. (2006). Energetics reveals physiologically distinct castes in a eusocial mammal. *Nature*, *440*(7085), 795–797.

Schippers, M.-P., Dukas, R., & McClelland, G. B.. (2010). Lifetime- and caste-specific changes in flight metabolic rate and muscle biochemistry of honeybees, *Apis mellifera*. *Journal of Comparative Physiology. B, Biochemical, Systemic, and Environmental Physiology*, *180*(1), 45–55.

Schippers, M.-P., Dukas, R., Smith, R. W., Wang, J., Smolen, K., & McClelland, G. B.. (2006). Lifetime performance in foraging honeybees: behaviour and

physiology. *The Journal of Experimental Biology*, *209*(Pt 19), 3828–3836.

Schmieder, R., & Edwards, R.. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PloS One*, *6*(3), e17288.

Schmieder, R., Lim, Y. W., & Edwards, R.. (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, *28*(3), 433–435.

Schmitz, J., & Moritz, R. F. A.. (1990). Mitochondrial DNA variation in social wasps (Hymenoptera, Vespidae). *Experientia*, *46*(10), 1068–1072.

Schwander, T., Lo, N., Beekman, M., Oldroyd, B. P., & Keller, L.. (2010). Nature versus nurture in social insect caste differentiation. *Trends in Ecology & Evolution*, *25*(5), 275–282.

Seeley, T. D.. (1982). Adaptive significance of the age polyethism schedule in honeybee colonies. *Behavioral Ecology and Sociobiology*, *11*(4), 287–293.

Sen Sarma, M., Whitfield, C. W., & Robinson, G. E.. (2007). Species differences in brain gene expression profiles associated with adult behavioral maturation in honey bees. *BMC Genomics*, *8*(1), 202.

Shaw, T. I., Ruan, Z., Glenn, T. C., & Liu, L.. (2013). STRAW: Species tree analysis web server. *Nucleic Acids Research*, *41*(Web Server issue), W238–241.

Sheehan, M. J., & Tibbetts, E. A.. (2011). Specialized face learning is associated with individual recognition in paper wasps. *Science*, *334*(6060), 1272–1275.

Shimodaira, H., & Hasegawa, M.. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution*, *16*(8), 1114.

Shukla, E., Thorat, L. J., Nath, B. B., & Gaikwad, S. M.. (2015). Insect trehalase: physiological significance and potential applications. *Glycobiology*, *25*(4), 357–367.

Siddall, M. E.. (2010). Unringing a bell: metazoan phylogenomics and the partition bootstrap. *Cladistics*, *26*(4), 444–452.

Simola, D. F., Wissler, L., Donahue, G., Waterhouse, R. M., Helmkampf, M., Roux, J., … Gadau, J.. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. *Genome Research*, *23*(8), 1235–1247.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Floors, P.. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, *87*(6), 651–701.

Simpson, C.. (2012). The evolutionary history of division of labour. *Proceedings. Biological Sciences / The Royal Society*, *279*(1726), 116–121.

Sinclair, B. J., Addo-Bediako, A., & Chown, S. L.. (2003). Climatic variability and the evolution of insect freeze tolerance. *Biological Reviews*, *78*(2), 181–195.

Sinclair, B. J., Ferguson, L. V, Salehipour-shirazi, G., & MacMillan, H. A.. (2013). Cross-tolerance and cross-talk in the cold: relating low temperatures to desiccation and immune stress in insects. *Integrative and Comparative Biology*, *53*(4), 545–556.

Smith, C. R., Helms Cahan, S., Kemena, C., Brady, S. G., Yang, W., Bornberg-Bauer, E., … Mikheyev, A.. (2015). How do genomes create novel phenotypes? Insights from the loss of the worker caste in ant social parasites. *Molecular Biology and Evolution*, *32*(11), msv165.

Smith, C. R., Toth, A. L., Suarez, A. V, & Robinson, G. E.. (2008). Genetic and genomic analyses of the division of labour in insect societies. *Nature Reviews. Genetics*, *9*(10), 735–748.

Smith, J. A., Chenoweth, L. B., Tierney, S. M., & Schwarz, M. P.. (2013). Repeated origins of social parasitism in allodapine bees indicate that the weak form of Emery's rule is widespread, yet sympatric speciation remains highly problematic. *Biological Journal of the Linnean Society*, *109*(2), 320–331.

Smith, J. A., Tierney, S. M., Park, Y. C., Fuller, S., & Schwarz, M. P.. (2007). Origins of social parasitism: The importance of divergence ages in phylogenetic studies. *Molecular Phylogenetics and Evolution*, *43*(3), 1131–1137.

Smith, S. A., & Dunn, C. W.. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics*, *24*(5), 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 Evolutionary Biology*, *15*(1), 150.

Spradbery, J. P.. (1971). Seasonal changes in the population structure of wasp colonies (Hymenoptera: Vespidae). *Journal of Animal Ecology*, *40*(2), 501–523.

Spradbery, J. P.. (1973). *Wasps. An account of the biology and natural history of solitary and social wasps*. London, Great Britain: Sidgwick and Jackson.

Stamatakis, A.. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690.

Stamatakis, A.. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, btu033.

Stein, K. J., Fell, R. D., & Holtzman, G. I.. (1996). Sperm use dynamics of the baldfaced hornet (Hymenoptera: Vespidae). *Environmental Entomology*, *25*(6), 1365–1370.

Stern, D. L.. (2013). The genetic causes of convergent evolution. *Nature Reviews. Genetics*, *14*(11), 751–764.

Storey, J. D.. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498.

Storey, J. D., & Tibshirani, R.. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(16), 9440–9445.

Strassmann, J.. (2001). The rarity of multiple mating by females in the social Hymenoptera. *Insectes Sociaux*, *48*(1), 1–13.

Strassmann, J. E., & Queller, D. C.. (2007). Insect societies as divided organisms: The complexities of purpose and cross-purpose. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(Suppl. 1), 8619–8626.

Strassmann, J. E., Lee, R. E., Rojas, R. R., & Baust, J. G.. (1984). Caste and sex

differences in cold-hardiness in the social wasps, *Polistes annularis* and *P. exclamans* (Hymenoptera: Vespidae). *Insectes Sociaux*, *31*(3), 291–301.

Strassmann, J. E., Page, R. E., Robinson, G. E., & Seeley, T. D.. (2011). Kin selection and eusociality. *Nature*, *471*(7339), E5–6.

Strassmann, J. E., & Queller, D. C.. (2007). Insect societies as divided organisms: The complexities of purpose and cross-purpose. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(Suppl. 1), 8619–8626.

Suarez, R. K., Darveau, C.-A., Welch, K. C., O'Brien, D. M., Roubik, D. W., & Hochachka, P. W.. (2005). Energy metabolism in orchid bee flight muscles: carbohydrate fuels all. *The Journal of Experimental Biology*, *208*(Pt 18), 3573–3579.

Sullivan, J., Swofford, D. L., & Naylor, G. J. P.. (1999). The effect of taxon sampling on estimating rate heterogeneity parameters of maximum-likelihood models. *Molecular Biology and Evolution*, *16*(10), 1347.

Sumner, S.. (2014). The importance of genomic novelty in social evolution. *Molecular Ecology*, *23*(1), 26–28.

Sumner, S., Aanen, D. K., Delabie, J., & Boomsma, J. J.. (2004). The evolution of social parasitism in *Acromyrmex* leaf-cutting ants: a test of Emery's rule. *Insectes Sociaux*, *51*(1), 37–42.

Sumner, S., Hughes, W. O. H., Pedersen, J. S., & Boomsma, J. J.. (2004). Ant parasite queens revert to mating singly. *Nature*, *428*(6978), 35–36.

Suyama, M., Torrents, D., & Bork, P.. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*, W609–612.

Swanson, W. J., Clark, A. G., Waldrip-Dail, H. M., Wolfner, M. F., & Aquadro, C. F.. (2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(13), 7375–7379.

Swofford, D. L.. (2002). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Sunderland, Massachusetts: Sinauer Associates.

Swofford, D. L., Olsen, G. L., Waddell, P. J., & Hillis, D. M.. (1996). Phylogenetic inference. In *Molecular Systematics* (Second ed., pp. 407–514). Sunderland, Massachusetts: Sinauer Associates.

Tannure-Nascimento, I. C., Nascimento, F. S., & Zucchi, R.. (2008). The look of royalty: visual and odour signals of reproductive status in a paper wasp. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1651), 2555–2561.

Tautz, D., & Domazet-Loso, T.. (2011). The evolutionary origin of orphan genes. *Nature Reviews. Genetics*, *12*(10), 692–702.

Taylor, L. H.. (1939). Observations on social parasitism in the genus *Vespula* Thomson. *Annals of the Entomological Society of America*, *32*(2), 304–315.

The UniProt Consortium.. (2015). UniProt: a hub for protein information. *Nucleic Acids Research*, *43*(D1), D204–212.

Tibbetts, E. A., & Dale, J.. (2004). A socially enforced signal of quality in a paper wasp. *Nature*, *432*(7014), 218–222.

Tobback, J., Heylen, K., Gobin, B., Wenseleers, T., Billen, J., Arckens, L., & Huybrechts, R.. (2008). Cloning and expression of PKG, a candidate foraging regulating gene in *Vespula vulgaris*. *Animal Biology*, *58*(4), 341–351.

Toth, A. L., & Robinson, G. E.. (2007). Evo-devo and the evolution of social behavior. *Trends in Genetics*, *23*(7), 334–341.

Toth, A. L., Varala, K., Henshaw, M. T., Rodriguez-Zas, S. L., Hudson, M. E., & Robinson, G. E.. (2010). Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1691), 2139–2148.

Toth, A. L., Varala, K., Newman, T. C., Miguez, F. E., Hutchison, S. K., Willoughby, D. A., … Robinson, G. E.. (2007). Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. *Science*, *318*(5849), 441–444.

Tracy, M. R., & Hedges, S. B.. (2000). Evolutionary history of the enolase gene family. *Gene*, *259*(1-2), 129–138.

Vaidya, G., Lohman, D. J., & Meier, R.. (2011). SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171–180.

van der Vecht, J.. (1957). The Vespinae of the Indo-Malayan and Papuan areas (Hymenoptera, Vespidae). *Zoologische Verhandelingen*, *34*, 1–83.

Van Oystaeyen, A., Oliveira, R. C., Holman, L., van Zweden, J. S., Romero, C., Oi, C. A., … Wenseleers, T.. (2014). Conserved class of queen pheromones stops social insect workers from reproducing. *Science*, *343*(6168), 287–290.

van Wilgenburg, E., Symonds, M. R. E., & Elgar, M. A.. (2011). Evolution of cuticular hydrocarbon diversity in ants. *Journal of Evolutionary Biology*, *24*(6), 1188–1198.

van Zweden, J. S., Bonckaert, W., Wenseleers, T., & D'Ettorre, P.. (2013). Queen signalling in social wasps. *Evolution, 68*(4), 976–986.

Vance, J. T., Williams, J. B., Elekonich, M. M., & Roberts, S. P.. (2009). The effects of age and behavioral development on honey bee (*Apis mellifera*) flight performance. *The Journal of Experimental Biology*, *212*(Pt 16), 2604–2611.

Varvio-Aho, S. L., Pamilo, P., & Pekkarinen, A.. (1984). Evolutionary genetics of social wasps (Hymenoptera, Vespidae, *Vespula*). *Insectes Sociaux*, *31*(4), 375–386.

Wagner, R. E.. (1978). The genus *Dolichovespula* and an addition to its known species of North America (Hymenoptera: Vespidae). *Pan-Pacific Entomology*, *54*, 131–142.

Wang, Z., Gerstein, M., & Snyder, M.. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, *10*(1), 57–63.

Ward, P.. (1989). Genetic and social changes associated with ant speciation. In M. D. Breed & R. G. Page (Eds.), *The genetics of social evolution* (pp. 123–148). Boulder, Colorado: Westview Press.

Ward, P. S.. (1996). A new workerless social parasite in the ant genus Pseudomyrmex (Hymenoptera: Formicidae), with a discussion of the origin of social parasitism in ants. *Systematic Entomology*, *21*(3), 253–263.

Ward, P. S., Brady, S. G., Fisher, B. L., & Schultz, T. R.. (2010). Phylogeny and biogeography of dolichoderine ants: effects of data partitioning and relict taxa on historical inference. *Systematic Biology*, *59*(3), 342–362.

Warrant, E. J.. (2008). Seeing in the dark: vision and visual behaviour in nocturnal bees and wasps. *The Journal of Experimental Biology*, *211*(Pt 11), 1737–1746.

Wcislo, W. T.. (1987). The roles of seasonality, host synchrony, and behaviour in the evolutions and distributions of nest parasites in Hymenoptera (Insecta), with special reference to bees (Apoidea). *Biological Reviews*, *62*(4), 515–542.

Weinstock, G. M., Robinson, G. E., Gibbs, R. A., Worley, K. C., Evans, J. D., Maleszka, R., … Wright, R.. (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, *443*(7114), 931–949.

Wenseleers, T., Badcock, N. S., Erven, K., Tofilski, A., Nascimento, F. S., Hart, A. G., … Ratnieks, F. L. W.. (2005a). A test of worker policing theory in an advanced eusocial wasp, *Vespula rufa*. *Evolution*, *59*(6), 1306–1314.

Wenseleers, T., Hart, A. G., & Ratnieks, F. L. W.. (2004). When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. *American Naturalist*, *164*(6), E154–E167.

Wenseleers, T., & Ratnieks, F. W.. (2006). Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. *The American Naturalist*, *168*(6), E163–E179.

Wenseleers, T., Tofilski, A., & Ratnieks, F. L. W.. (2005b). Queen and worker policing in the tree wasp *Dolichovespula sylvestris*. *Behavioral Ecology and Sociobiology*, *58*(1), 80–86.

Wenzel, J. W.. (1990). A social wasp's nest from the Cretaceous period, Utah, USA, and its biographical significance. *Psyche*, *97*(1-2), 21–30.

Wenzel, J. W.. (1992). Extreme queen-worker dimorphism in *Ropalidia ignobilis*, a small-colony wasp (Hymenoptera: Vespidae). *Insectes Sociaux*, *39*(1), 31–43.

Werner, M., & Simmons, L. W.. (2008). Insect sperm motility. *Biological Reviews*, *83*(2), 191–208.

West-Eberhard, M. J.. (1978). Polygyny and the evolution of social behavior in wasps. *Journal of the Kansas Entomological Society*, *51*(4), 832–856.

West-Eberhard, M. J.. (1986). Alternative adaptations, speciation, and phylogeny. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(5), 1388–1392.

West, M. J.. (1967). Foundress associations in polistine wasps: dominance hierarchies and the evolution of social behavior. *Science*, *157*(3796), 1584–1585.

Weyrauch, W.. (1937). Zur Systematik und Biologie der Kuckuckswespen *Pseudovespa*, *Pseudovespula* und *Pseudopolistes*. *Zoologische Jahrbücher (Systematik)*, *70*, 243– 290.

Wheeler, W. C.. (1996). Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics*, *12*(1), 1–9.

Wheeler, W. M.. (1919). The parasitic Aculeata, a study in evolution. *Proceedings of the American Philosophical Society*, *58*(1), 1–40.

Wheeler, W. M., & Taylor, L. H.. (1921). *Vespa arctica* Rohwer, a parasite of *Vespa diabolica* de Saussure. *Psyche: A Journal of Entomology*, *28*(5-6), 135–144.

Whitfield, C. W., Ben-Shahar, Y., Brillet, C., Leoncini, I., Crauser, D., Leconte, Y., … Robinson, G. E.. (2006). Genomic dissection of behavioral maturation in the honey bee. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16068–16075.

Whitfield, C. W., Cziko, A.-M., & Robinson, G. E.. (2003). Gene expression profiles in the brain predict behavior in individual honey bees. *Science*, *302*(5643), 296–299.

Whitfield, J. B., & Kjer, K. M.. (2008). Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annual Review of Entomology*, *53*(1), 449–472.

Whitfield, J. B., & Lockhart, P. J.. (2007). Deciphering ancient rapid radiations. *Trends in Ecology & Evolution*, *22*(5), 258–265.

Wilson, E. O.. (1971). *The insect societies*. Harvard University Press.

Wilson, E. O.. (1976). Behavioral discretization and the number of castes in an ant species. *Behavioral Ecology and Sociobiology*, *1*, 141–154.

Wilson, E. O.. (1985). The sociogenesis of insect colonies. *Science*, *228*(4707), 1489–1495.

Wilson, E. O.. (2008). One giant leap: how insects achieved altruism and colonial life. *BioScience*, *58*(1), 17–25.

Wilson, E. O., & Holldobler, B.. (2005). Eusociality: origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(38), 13367–13371.

Wolfner, M. F.. (1997). Tokens of love: Functions and regulation of *Drosophila* male accessory gland products. *Insect Biochemistry and Molecular Biology*, *27*(3), 179–192.

Wolschin, F., & Amdam, G. V.. (2007). Comparative proteomics reveal characteristics of life-history transitions in a social insect. *Proteome Science*, *5*(1), 10.

Woodard, S. H., Bloch, G. M., Band, M. R., & Robinson, G. E.. (2014). Molecular heterochrony and the evolution of sociality in bumblebees (*Bombus terrestris*). *Proceedings of the Royal Society B: Biological Sciences*, *281*(1780), 20132419–20132419.

Woodard, S. H., Fischman, B. J., Venkat, A., Hudson, M. E., Varala, K., Cameron, S. A., … Robinson, G. E.. (2011). Genes involved in convergent evolution of eusociality in bees. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7472–7477.

Xie, W., Lewis, P. O., Fan, Y., Kuo, L., & Chen, M.-H.. (2011). Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic Biology*, *60*(2), 150–160.

Yamane, S.. (1976). Morphological and taxonomic studies on vespine larvae, with reference to the phylogeny of the subfamily Vespinae (Hymenoptera : Vespidae). *Insecta Matsumurana. Series Entomology. New Series*, *8*, 1–45.

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. *Molecular Biology and Evolution*, *31*(11), 3081–3092.

Yang, Z.. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, *15*(5), 568–573.

Yang, Z.. (2006). *Computational Molecular Evolution*. New York: Oxford University Press.

Zwickl, D. J.. (2006). *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*. The University of Texas at Austin.