Expansion and accelerated evolution of 9-exon odorant receptors in *Polistes* paper wasps (Hymenoptera: Vespidae)

Andrew W. Legan[1*], Christopher M. Jernigan[1], Sara E. Miller[1], Matthieu F. Fuchs[1], Michael J. Sheehan[1*]

1. Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, USA
*correspondence to: awl75@cornell.edu, msheehan@cornell.edu

Abstract

Independent origins of sociality in bees and ants are associated with independent expansions of the odorant receptor (OR) gene family. In ants, one clade within the OR gene family, the 9-exon subfamily, has dramatically expanded. These receptors detect cuticular hydrocarbons (CHCs), key social signaling molecules in insects. It is unclear to what extent 9-exon OR subfamily expansion is associated with the independent evolution of sociality across Hymenoptera, warranting studies of taxa with independently derived social behavior. Here we describe OR gene family evolution in the northern paper wasp, *Polistes fuscatus*, and compare it to four additional paper wasp species spanning ~40 million years of divergence. We find 200 functional OR genes in *P. fuscatus* matching predictions from neuroanatomy, and more than half of these are in the 9-exon subfamily. Lineage-specific expansions of 9-exon subfamily ORs are tandemly arrayed in *Polistes* genomes and exhibit a breakdown in microsynteny relative to tandem arrays in other OR subfamilies. There is evidence of episodic positive diversifying selection shaping ORs in expanded subfamilies, including 9-exon, E, H, and L, but 9-exon ORs do not stand out as selectively diversified among *Polistes* species. Accelerated evolution has resulted in lower amino acid similarity and higher d_N/d_S among 9-exon ORs compared to other OR subfamilies. Patterns of OR evolution within *Polistes* are consistent with 9-exon OR function in CHC perception by combinatorial coding, with both selection and drift contributing to interspecies differences in copy number and sequence.

Keywords: odorant receptor, paper wasp, birth-and-death evolution, comparative genomics, tandem array, antennal lobe glomeruli

Introduction

Chemosensation is an organism's ability to sense chemicals in the environment, whether by olfaction (smell) or gustation (taste). Insects rely on chemosensation when identifying a compatible mate, avoiding toxic microbes, escaping predators, and finding food - whether a patch of nectar-rich flowers or a nutritious caterpillar to predate or parasitize (Wicher 2015; Fleischer et al. 2018; Yan et al. 2020). Social insects further use chemicals to communicate with each other about nest membership, caste, food locations, and alarms (Leonhardt et al. 2016). The evolution of the remarkable chemical communication abilities in social bees and ants is associated with expansion of the odorant receptor (OR) gene family, and the 9-exon OR subfamily in particular has experienced increased gene turnover and sequence evolution relative to other OR subfamilies (Engsontia et al. 2015; Kapheim et al. 2015; Zhou et al. 2015; Karpe et al. 2016, 2017; McKenzie et al. 2016). Comparative studies across species with different evolutionary histories and variable social behaviors can shed light on the molecular evolutionary processes underlying chemical signal perception (Tsutsui et al. 2013; Yan et al. 2020). The independent origins of sociality in wasps provide an opportunity to compare patterns of OR gene family evolution to those that have been observed within social bees and ants (Hines et al. 2007). At the same time, studies of closely related species with similar ecologies and life histories can reveal the dynamics of receptor evolution at finer timescales (Guo & Kim 2007; Brand et al. 2015; Karpe et al. 2016; Brand & Ramirez 2017; Miller CH et al. 2020). A better understanding of the short-term mechanisms of OR evolution provides additional insights into the molecular evolutionary dynamics shaping receptor diversity across more distantly related taxa.

The molecular evolution of the OR gene family is best described as a birth-and-death process, in which genes are duplicated and deleted over evolutionary time (Nei 2007; Nozawa & Nei 2007; Eirín-López et al. 2012). The combined forces of random drift and natural selection determine the extent of gene copy number variation and the rate of sequence evolution (Nei 2007; Nozawa & Nei 2007). The behavioral and molecular functions of ORs shape their molecular evolution. Organisms use odorant receptor proteins to detect stimuli and provide input to neural circuits that determine decision-making, here termed behavioral function (Yapici et al. 2014). At the molecular level, the specificity with which an OR binds chemical compounds or families of compounds (ligands) determines the tuning specificity of the olfactory receptor neuron (ORN) in which it is expressed (Hallem et al. 2004). ORN tuning specificity describes the relationship

between the change in the frequency of action potentials fired by the ORN across a range of different chemical stimuli. An OR's molecular function can be categorized as specialist if ligand binding is highly specific, i.e. one chemical, or generalist if ligand binding allows for detection of multiple chemicals or chemical types. Narrowly tuned ORNs express specialist ORs that bind a specific ligand, creating a dedicated channel of olfaction, while broadly tuned ORNs express generalist ORs that bind a larger spectrum of ligands (Touhara & Vosshall 2009; Andersson et al. 2015). Generalist ORs are important for combinatorial coding: the process of combining input from multiple ORs that bind an overlapping set of ligands in order to discriminate a large number of ligands (Malnic et al. 1999). In nature, ORs exist on a spectrum from specialist to generalist, and ORN responses are dependent upon a complex interaction of OR, ligand concentration, and odorant binding proteins (Vogt et al. 1991; Hallem et al. 2004; Hallem & Carlson 2006; Stensmyr et al. 2012; Mathew et al. 2013; Ebrahim et al. 2015; Dweck et al. 2015; Münch & Galizia 2016).

There are distinct sets of predictions for the molecular evolution of ORs depending on their behavioral and molecular functions. Negative selection (purifying selection) is expected to conserve specialist ORs, which are not resilient to mutations due to the tight link between receptor and ligand in a dedicated olfactory channel (Andersson et al. 2015). For example, a volatile emitted by toxic microbes triggers avoidance behavior in *Drosophila melanogaster* by binding a specialist OR that is conserved across the genus (Stensmyr et al. 2012). In contrast to dedicated olfactory channels, combinatorial coding is thought to involve ORs with overlapping responses to ligands (Andersson et al. 2015). Mutations that slightly alter the response profiles of functionally redundant ORs may not be eliminated by negative selection, since other ORs can help compensate (Fishilevich et al. 2005; Keller & Vosshall 2007). Copy number variation and relaxed selection allow ORs to gain mutations that might endow them with an adaptive behavioral function. Divergent chemosensory landscapes between species lead to increased copy number variation as ORs with new behavioral functions are gained and ORs that detect irrelevant ligands are lost (Ramdya & Benton 2010; Goldman-Huertas et al. 2015).

The molecular evolution of the OR gene family is dynamic among the Hymenoptera, with prevalent lineage-specific gene expansions and losses, especially in the 9-exon OR subfamily (Engsontia et al. 2015; Zhou et al. 2015; McKenzie & Kronauer 2018). The 9-exon ORs constitute about one third of all ant ORs, and have evolved rapidly in ants, frequently under positive selection, leading researchers to propose that 9-exon ORs facilitate recognition of cuticular hydrocarbons

(CHCs) (Smith CR, Smith CD et al. 2011; Smith CD, Zimin et al. 2011; Zhou et al. 2012; Engsontia et al. 2015; Zhou et al. 2015; McKenzie et al. 2016). CHCs are used by insects to waterproof the cuticle and to communicate with conspecifics (Blomquist & Bagnères 2010). While less pronounced than in ants, dynamic evolution is also characteristic of 9-exon OR evolution in social bees, which rely on CHCs in communication (Sadd et al. 2015; Karpe et al. 2016, 2017). Functional studies in which ORs were transfected into an empty *D. melanogaster* ORN have verified that at least some 9-exon ORs of the ant *Harpegnathos saltator* overlap in their responses to ligands, with multiple 9-exon ORs responding to the same CHC molecule and unique 9-exon ORs responding to multiple different CHC molecules (Pask et al. 2017; Slone et al. 2017). Functional ORs are necessary for normal nesting behavior and for nestmate recognition in ants, a process which involves detecting variation in the CHCs on the cuticles of conspecifics (Lavine et al. 1990; van Zweden & d'Ettorre 2010; Sturgis & Gordon 2012; Trible et al. 2017; Yan et al. 2017; Ferguson et al. 2020). Together these studies suggest that 9-exon ORs function in combinatorial coding of CHC perception.

Like other Hymenopterans, vespid wasps, including the genus *Polistes*, use CHCs in complex social behaviors (Gamboa et al. 1986, 1996; Dani & Turillazzi 2018). *Polistes* use chemicals as signals and cues in a variety of behaviors, including during mate attraction, mate compatibility recognition, queen recognition, dominance/fertility signaling, and nestmate recognition (Reed & Landolt 1990; Post & Jeanne 1984; Dapporto et al. 2007; Jandt et al. 2014; Sledge et al. 2001a, 2001b, 2004; Oi et al. 2019; Espelie et al. 1994). Recent efforts to sequence *Polistes* genomes provide an opportunity to resolve patterns of OR evolution among closely related species as an independent test of 9-exon OR gene subfamily expansion during social evolution and (Patalano et al. 2015; Standage et al. 2016; Miller SE et al. 2020). We annotated the OR repertoires of five *Polistes* species representing ~40 million years of evolution: *P. fuscatus*, *P. metricus*, *P. dorsalis*, *P. canadensis*, and *P. dominula* (Figure 1). Combining neuroanatomy, manual gene annotation, and molecular evolution, we examined the evolution of odorant receptors in light of the patterns predicted for different behavioral and molecular functions. We discover that social wasps, like ants, have an expanded subfamily of 9-exon ORs. Between *Polistes* species, 9-exon ORs exhibit dynamic evolution relative to ORs in other subfamilies, which are highly conserved. Expansion and birth-and-death evolution of the 9-exon OR subfamily in social wasps is consistent with a unique function in combinatorial coding perception of CHCs.

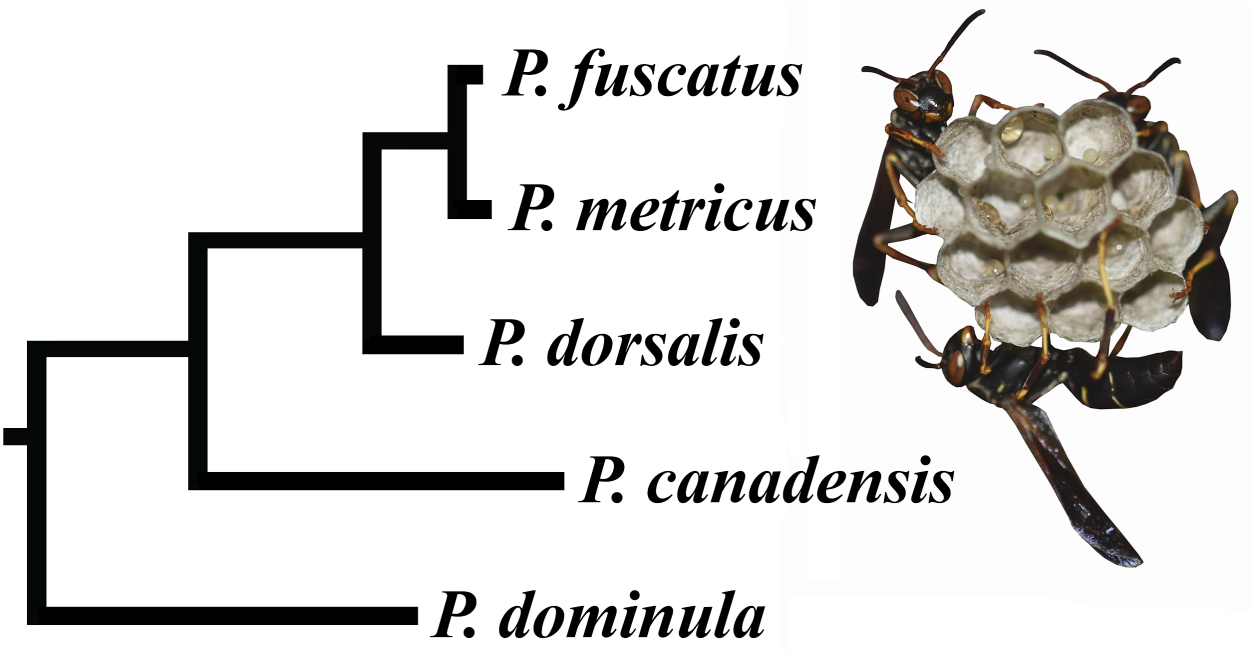


Fig. 1: Phylogeny of five *Polistes* species considered in this study: *P. fuscatus*, *P. metricus*, *P. dorsalis*, *P. canadensis,* and *P. dominula*. The photo to the right of the phylogeny shows multiple *P. fuscatus* foundresses on a nest. Phylogenetic tree built from 16S ribosomal RNA and cytochrome oxidase subunit I tree in Supplemental Fig 2 in Sheehan et al. 2015.

Results and Discussion

Antennal Lobe Neuroanatomy and Manual Gene Annotation Predict 200 ORs in *P. fuscatus*

In order to predict the OR repertoires of *P. fuscatus* and four other *Polistes* species, we combined fluorescent confocal microscopy of the *P. fuscatus* antennal lobe with manual genome annotation informed by antennal RNAseq. We found 229 glomeruli in the antennal lobe of an adult gyne (female reproductive) (Figure S1). Across a sample of insects, the number of intact OR genes in the genome correlates with the number of glomeruli in the antennal lobe, predicting 229 ORs in the *P. fuscatus* genome (Figure 2). Here we focus on the *P. fuscatus* genome because it has nearly chromosome level scaffolds and is the best assembled *Polistes* genome (Table S1; Patalano et al. 2015; Standage et al. 2016; Miller SE et al. 2020). Automated annotation using the MAKER pipeline (Holt & Yandell 2011) without guidance from antennal mRNA predicted 115 OR gene models in the *P. fuscatus* genome. A combined *P. fuscatus* male and gyne (reproductive female) antennal transcriptome generated using Trinity (Haas et al. 2013) yielded 89 OR genes greater than 900 nucleotides in length. Some long Trinity genes contain multiple 7-transmembrane domains and likely represent concatenated OR genes. The small fraction of the *P. fuscatus* OR repertoire predicted by transcriptome assembly is consistent with previous observations that annotation of OR repertoires using only transcriptome data typically fails to recover all ORs (Karpe et al. 2016, 2017, 2020).

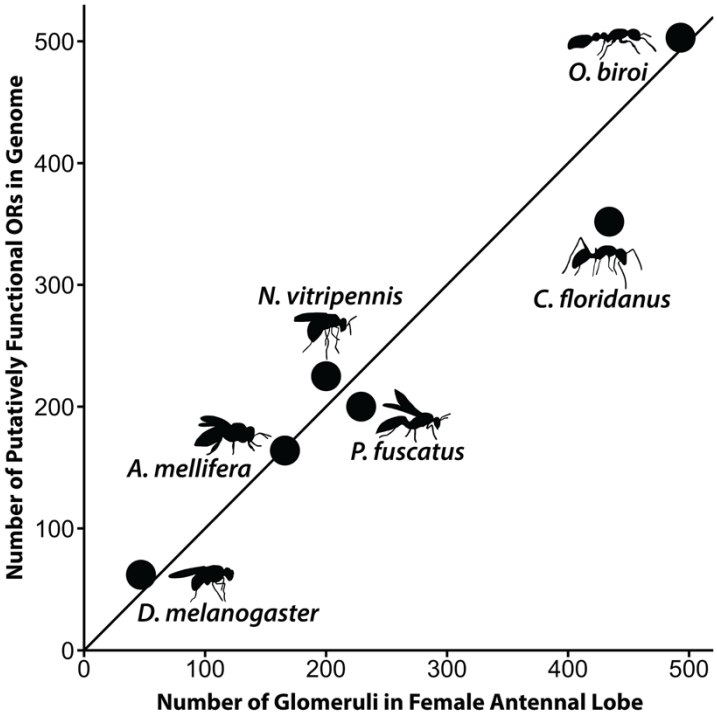


Fig. 2: The number of functional ORs is correlated with the number of antennal lobe glomeruli across insect species (50 glomeruli and 62 ORs in the genome of the common fruit fly *D. melanogaster*, Fishilevich & Vosshall 2005; 166 glomeruli in the worker, 103 glomeruli in the drone, and 163 functional ORs in the genome of the honey bee *A. mellifera*, Arnold et al. 1985, Robertson & Wanner 2006; ~200 glomeruli in females and 225 intact ORs in the genome of the parasitic wasp *N. vitripennis*, Groothuis et al. 2019, Robertson & Wanner 2006; ~434 glomeruli in the worker, 258 in the male, and 352 functional ORs in the genome of the ant *C. floridanus*, Zube & Rössler 2008, Zhou et al. 2012; 493 glomeruli in the worker, 119 glomeruli in the male, and 503 intact ORs in the genome of the ant *O. biroi*, McKenzie et al. 2016; McKenzie & Kronauer 2018; 229 glomeruli in a gyne and 200 putatively functional ORs in the genome of *P. fuscatus*). The diagonal line represents a line of equality with slope of 1. Insect silhouette sizes are not to scale.

Manual gene annotation of *P. fuscatus* ORs recovered 231 gene models across 28 scaffolds (Figure S2), of which 28 are pseudogenes and 10 are incomplete gene models (7 missing N termini, 2 missing C termini, and one missing both N and C termini). Since functional insect ORs are typically composed of 400 amino acids, we defined gene models as putatively functional if they coded for proteins greater than or equal to 300 amino acids in length. In *P. fuscatus*, the 200 putatively functional gene models encode protein sequences with an average length of 395 ± 15 (SD) amino acids, and 198 of these gene models encode protein sequences greater than 350 amino acids in length (Table 1). Odorant receptor proteins possess seven transmembrane domains. The putatively functional *P. fuscatus* OR proteins possess on average 5.95 ± 0.91 (SD) transmembrane domains as predicted by TMHMM version 2.0c (Sonnhammer et al. 1998) and 6.43 ± 1.13 (SD) as predicted by Phobius version 1.01 (Käll et al. 2004). For comparison, transmembrane domain prediction in 61 *D. melanogaster* ORs coding for proteins greater than 375 amino acids in length found on average 5.77 ± 1.12 (SD) transmembrane domains as predicted by TMHMM version 2.0c and 6.18 ± 1.09 (SD) as predicted by Phobius version 1.01 (sequences from Hopf et al. 2015 Supplemental Data 1). The close match between the number of ORs predicted by neuroanatomy and the number recovered from manual annotation suggests that we have identified nearly all of the OR genes in *P. fuscatus*. The number of transmembrane domains predicted are comparable to annotations of *D. melanogaster* and approach the 7 transmembrane domains expected for insect ORs. Manual OR gene annotation in *P. fuscatus* and four other *Polistes* genomes is summarized in Table 1:

Species	Putatively Functional ORs (>300 aa)	Mean Length in Amino Acids (± SD)	Mean Trans-membrane domains by TMHMM	Mean Trans-membrane domains by Phobius	Total Gene Models	PSE	Incomplete OR gene models
P. fuscatus	200	395 ± 15	5.95 ± 0.91	6.43 ± 1.13	231	28	10
P. metricus	204	396 ± 13	5.96 ± 0.85	6.45 ± 1.17	217	12	9
P. dorsalis	177	393 ± 20	5.90 ± 0.90	6.40 ± 1.21	203	16	24
P. canadensis	188	394 ± 17	5.95 ± 0.91	6.48 ± 1.20	235	13	59
P. dominula	180	392 ± 19	5.99 ± 0.88	6.59 ± 1.33	202	7	33

9-exon OR Subfamily Expanded During the Evolution of Social Wasps

We conducted a Hymenoptera-wide analysis of OR evolution to test the prediction that 9-exon subfamily ORs were independently expanded during the evolution of eusociality in vespid wasps. By comparing the *P. fuscatus* OR repertoire to other Hymenopterans, our findings reinforce previous results showing that across Hymenopteran families, ORs evolve with lineage-specific expansions of multiple OR subfamilies (Figure 3). Gene gain and loss events were predicted using NOTUNG (Chen et al. 2000) and mapped onto a species cladogram of 14 Hymenopterans (Figure 4). NOTUNG estimated an ancestral Apocritan repertoire of 56 ORs, which has expanded independently during the evolution of braconid wasps, ants, bees, and paper wasps (Figure 4). The 9-exon subfamily is commonly expanded across Hymenoptera (~90 genes on average), and comprises ~36% of social insect OR repertoires. The largest lineage-specific expansions of Hymenopteran 9-exon ORs have occurred independently during the evolution of ants and social wasps. In *P. fuscatus*, this clade has expanded to 105 genes, comprising 53% of the OR gene set (Figure 4). Given the well-documented use of CHCs as signal molecules in *Polistes* (Singer 1998; Dani et al. 2001; Dani 2009), it is not surprising to find expansions in the CHC-detecting 9-exon subfamily in this genus. Subfamilies L, T, H, E, and V have also expanded in *Polistes*, but not to the extent of the 9-exon OR subfamily.

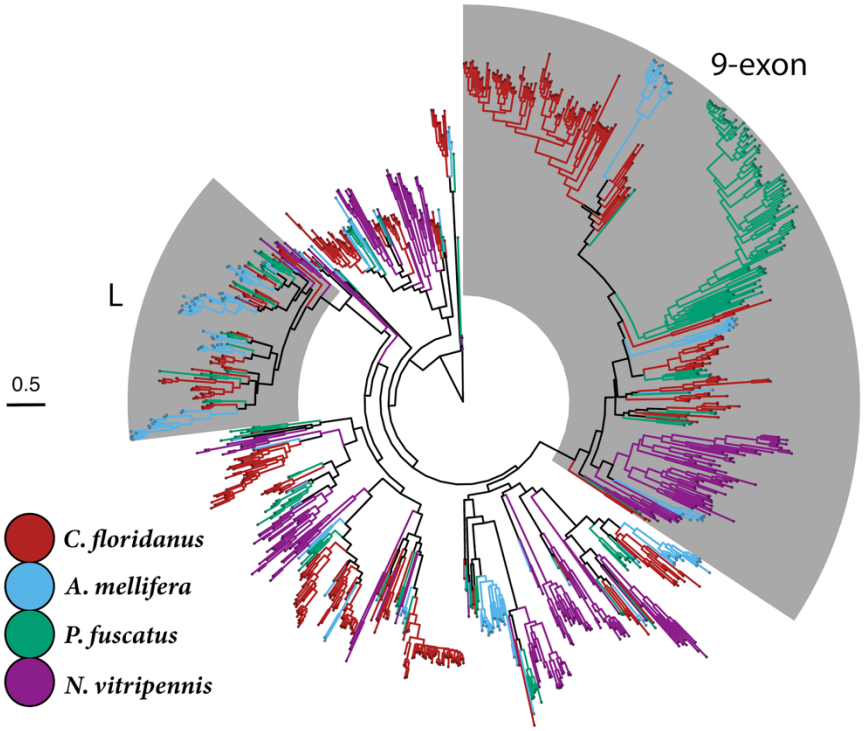


Fig. 3: Maximum likelihood OR protein tree constructed using data from four Hymenopterans (*Apis mellifera*, Robertson & Wanner 2006; *Camponotus floridanus*, Zhou et al. 2012; *Nasonia vitripennis*, Robertson et al. 2010). Branches are colored by species (Red: *Acromyrmex echinatior*; Light blue: *A. mellifera*; Green: *P. fuscatus*; Purple: *N. vitripennis*). The L and 9-exon subfamilies are highlighted. Scale bar represents mean substitutions per site.

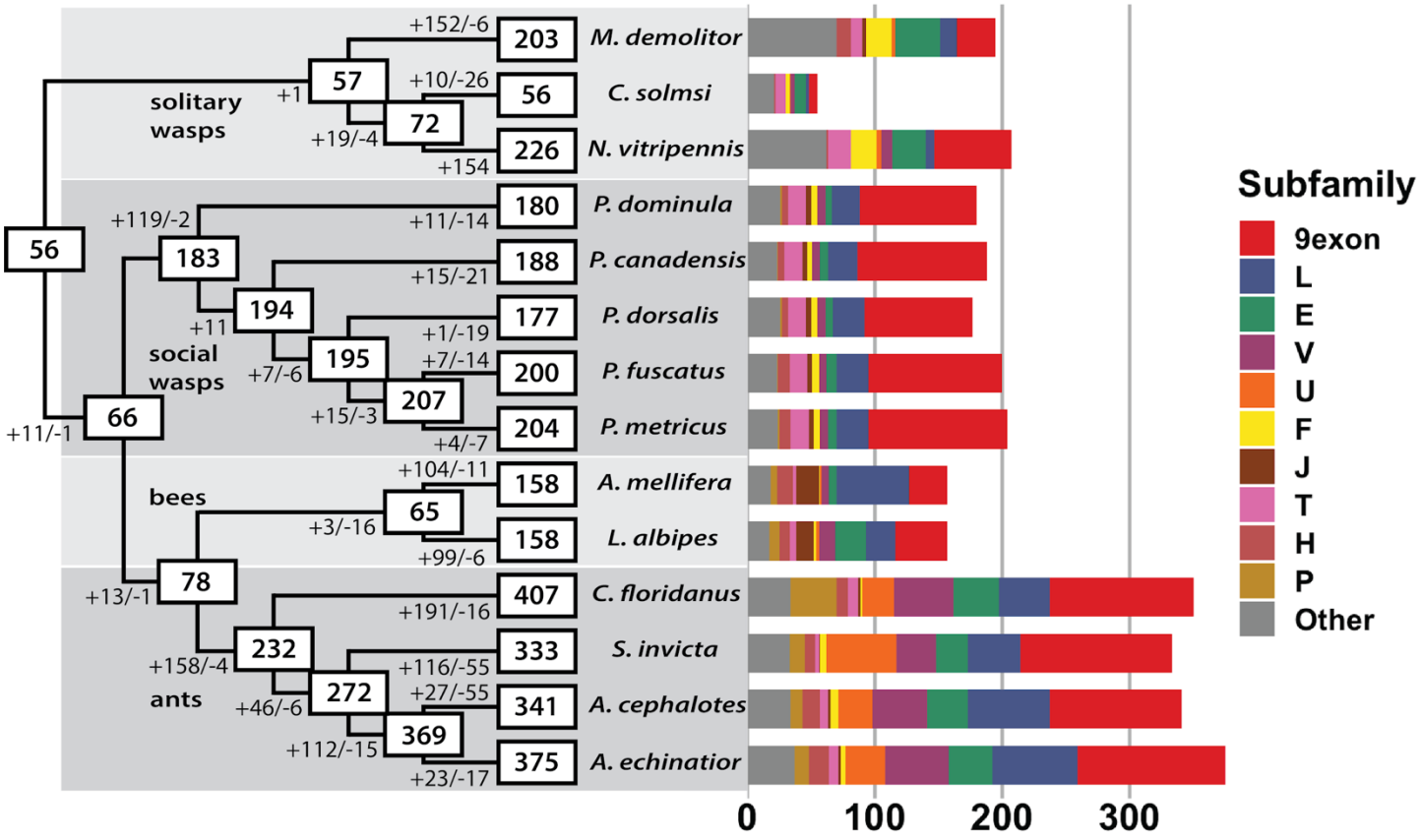


Fig. 4: Cladogram of Hymenoptera species showing estimated number of OR gene gain and loss events along branches and estimated size of ancestral and extant species OR repertoires in boxes. To the right is a bar chart showing numbers of ORs broken down by subfamily. Non-Polistine OR data are from Robertson et al. 2010 and Zhou et al. 2012, 2015. The set of intact ORs that are greater than 300 amino acids was used for gene gain and loss estimates. For *C. floridanus*, only ORs considered putatively functional by Zhou et al. 2012 were used to generate the bar chart showing numbers of ORs broken down by subfamily.

The 9-exon Subfamily Shows a Distinct Pattern of Orthology within *Polistes*

We next examined the evolutionary history of OR genes among the five *Polistes* species to reveal patterns of orthology and paralogy within subfamilies. Across the *Polistes* genus, most OR subfamilies are highly conserved (Figure 5). About 70% of non-9-exon family *P. fuscatus* ORs are in 1:1 orthology with all other *Polistes* species sampled as predicted by OrthoFinder (Emms & Kelly 2015; Table S3). The remaining orthologous groups contain an expansion in one or more species (Figure 6). Considering non-9-exon ORs, most ORs are shared by all five *Polistes* species examined, and most expansions are shared across all five species. Given that the species examined here span ~40 million years of divergence (Peters et al. 2017), the conservation of most of the OR repertoire is notable and may be related to the similarity of ecological and social niches found

among *Polistes* wasps. While a common evolutionary history has led to large 9-exon OR complements in all *Polistes* species examined, lineage-specific gains and losses of 9-exon ORs account for most of the variation in OR repertoires size across *Polistes* species (Figure 4). In contrast with the other OR subfamilies, the 9-exon OR subfamily shows more lineage specificity with only 32% of *P. fuscatus* 9-exon ORs showing simple 1:1 orthology across all five *Polistes* examined (Table S3). Most 9-exon subfamily orthologous groups contain gene copies from four or fewer species, and lineage-specific expansions are more common in 9-exon OR orthologous groups (Figure 6). Patterns of orthology (or rather the relative lack thereof) among 9-exon OR genes compared to the rest of the OR gene subfamilies suggest unique evolutionary processes shaping 9-exon ORs.

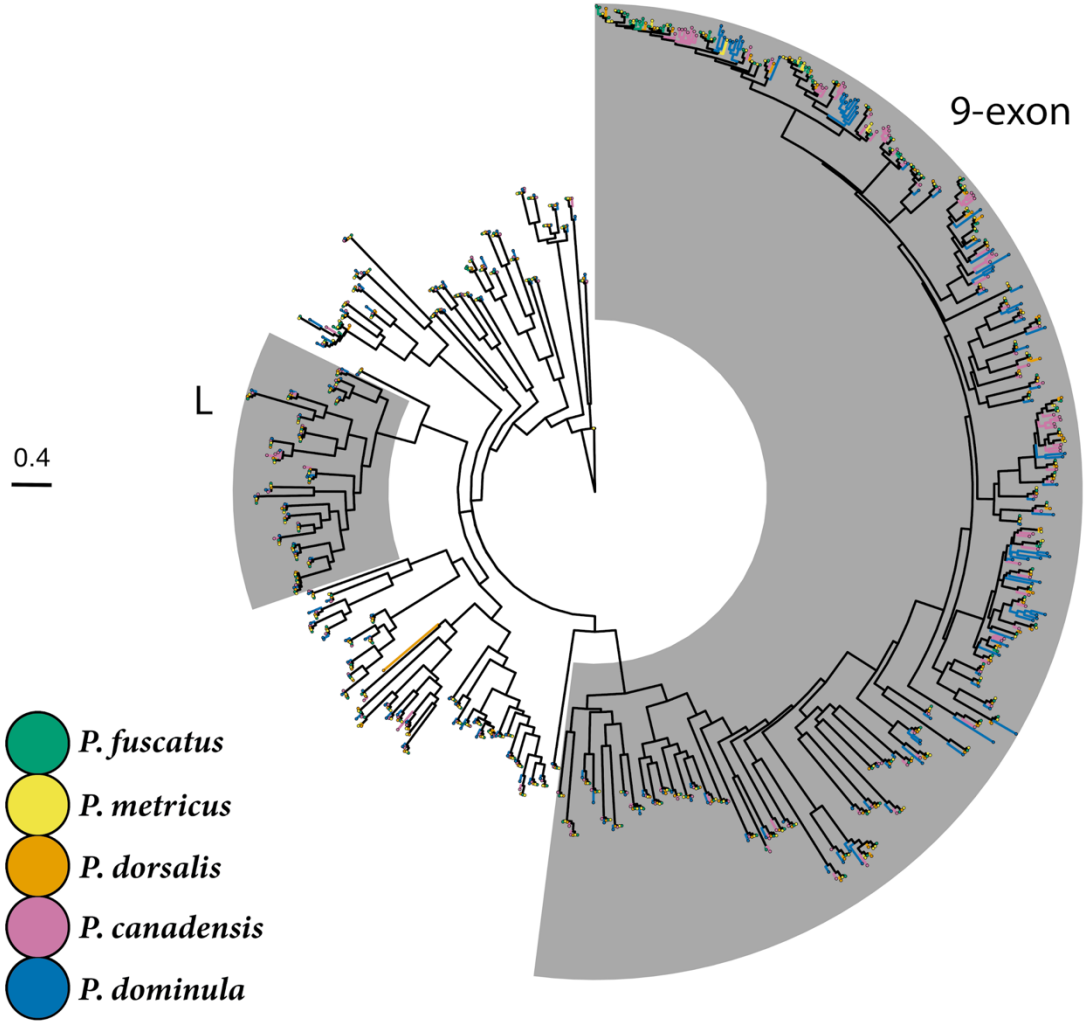


Fig. 5: Maximum likelihood OR protein tree with branches colored by species (Green: *P. fuscatus*; Yellow: *P. metricus*; Orange: *P. dorsalis*; Magenta: *P. canadensis*; Blue: *P. dominula*). The L and 9-exon subfamilies are highlighted. Scale bar represents mean substitutions per site.

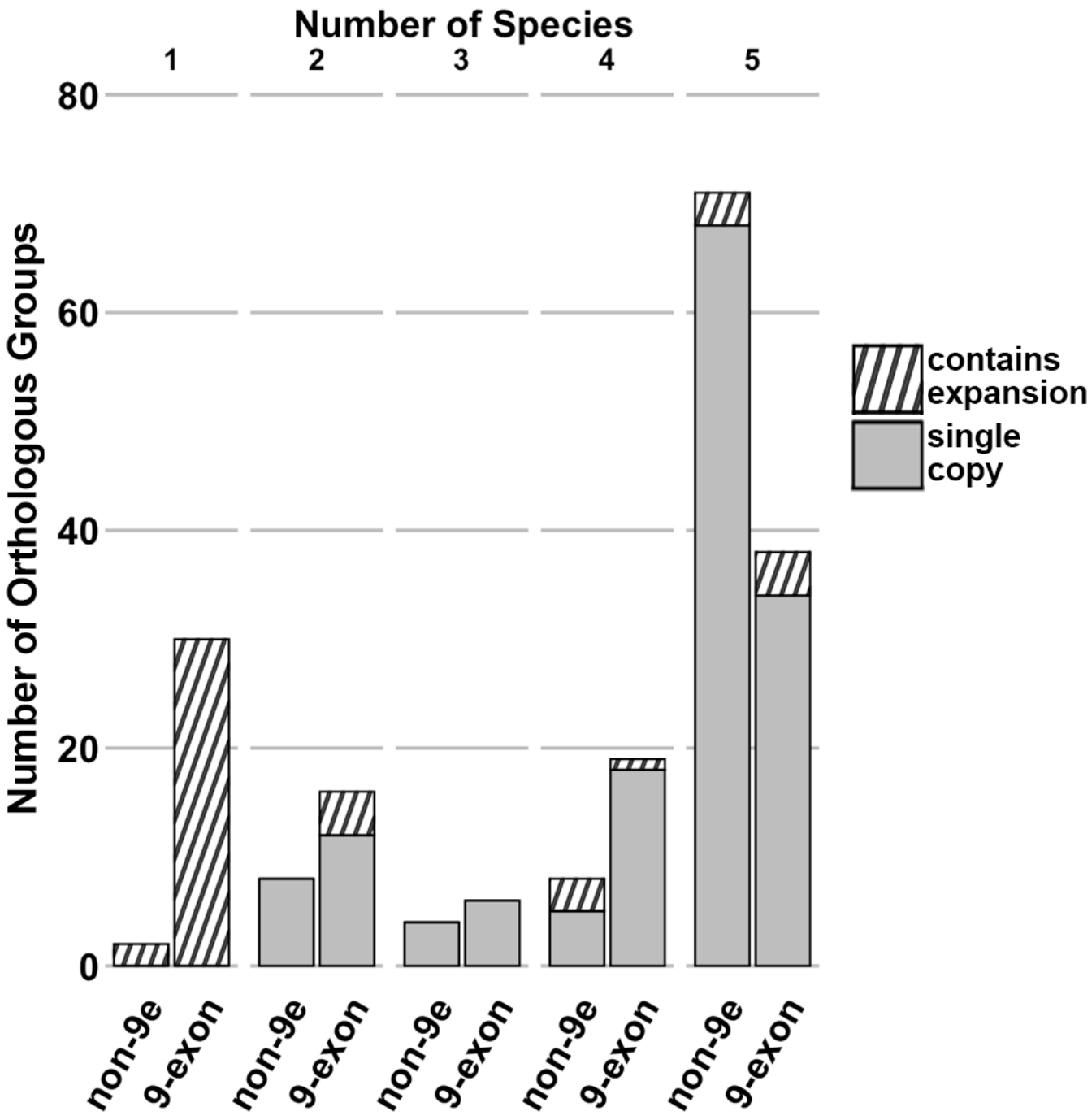
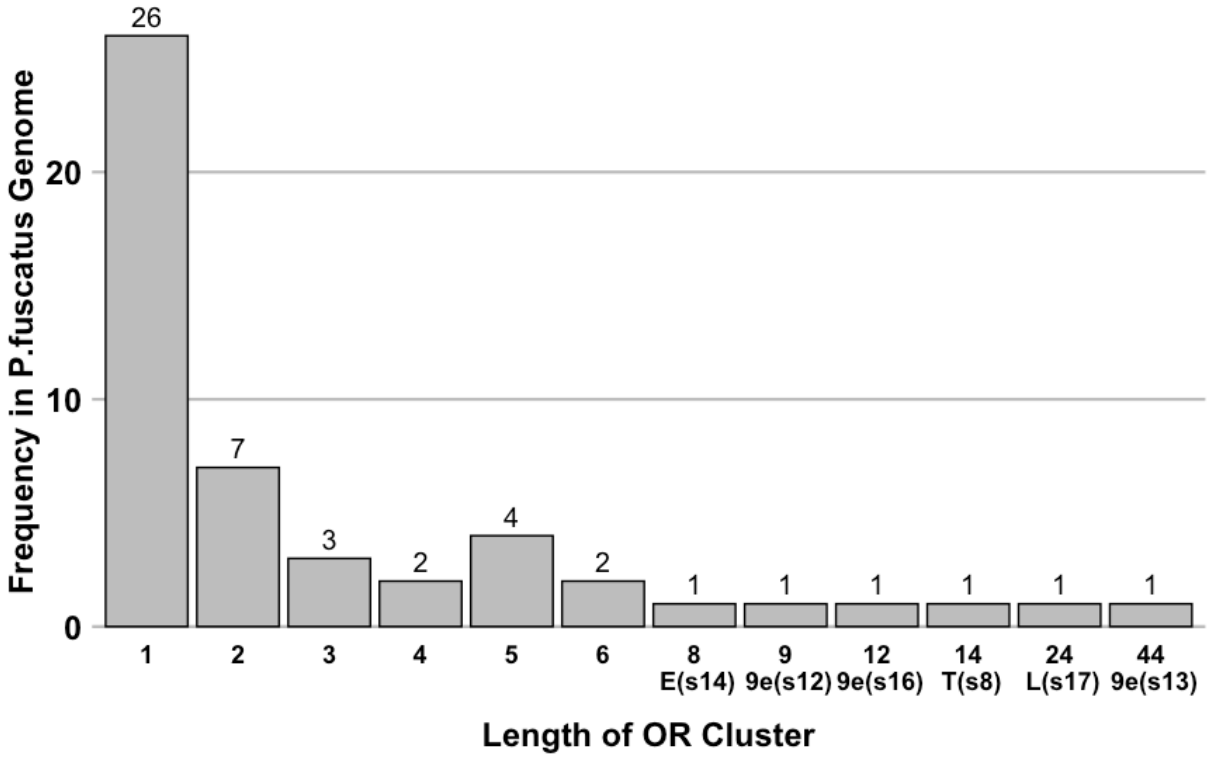


Fig. 6: Stacked bar chart showing the number of species (x-axis) represented in each orthologous group (y-axis), and whether or not each orthologous group is single copy (shaded bottom portion of bar) or contains an expansion in at least one species (top striped portion of bar). Orthologous groups are split into two categories: non-9-exon orthologous groups (left bar) and 9-exon orthologous groups (right bar).

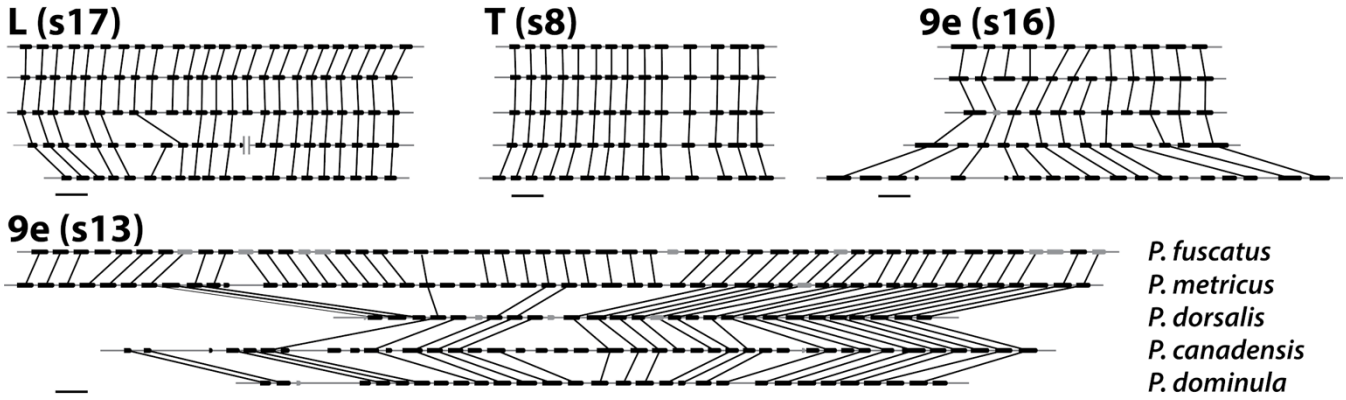
Microsynteny Reveals Recent Birth-and-Death Events in *Polistes* 9-exon Subfamily

Expanded gene families often occur as tandem arrays, a genomic architecture that can contribute to increased rates of gene birth and death, increasing copy number variation among species (Ohno 1970). Therefore, we examined how genomic organization varies between OR subfamilies in *Polistes* species to generate insights into the molecular evolutionary mechanisms shaping OR subfamily function. Genomic organization of ORs across *Polistes* is consistent with a model of birth-and-death evolution shaping OR repertoires. As in bees, gene gain and loss at a small number of loci containing tandem arrays is responsible for most copy number variation in the OR family across closely-related species (Brand & Ramírez 2017). In *P. fuscatus*, 62% of ORs occur in tandem arrays of 6 or more genes (Figure 7A). The frequency of tandem arrays and the tail-to-head orientations of neighboring genes point to tandem duplication as the primary mechanism of OR expansion, likely caused by non-allelic homologous recombination (Lynch 2007; Ramdya & Benton 2010). We examined microsynteny among genes and pseudogenes in the four longest tandem arrays of ORs in *Polistes* genomes (Figure 7B). There is marked decrease in OR synteny among genes in orthologous 9-exon OR arrays compared to tandem arrays of L and T subfamily ORs in the *Polistes* genus. The longest OR gene tandem array in *P. fuscatus* is comprised of 44 genes in the 9-exon subfamily on scaffold 13 (s13), which corresponds to homologous arrays of 50 genes *P. metricus*, 25 genes in *P. dorsalis*, 33 genes in *P. canadensis*, and 29 genes in *P. dominula*. Only 34% of *P. fuscatus* ORs in this array have orthologs across all *Polistes* species sampled (Figure 7B). The second longest OR gene tandem array in *P. fuscatus* contains 24 ORs in the L subfamily on scaffold 17 (s17), and these ORs show 1:1 orthology across *P. fuscatus*, *P. metricus*, and *P. dorsalis*, while *P. canadensis* possesses an array of ~23 genes split across two scaffolds, and *P. dominula* possesses an array of 21 ORs at this locus (Figure 7B). This tandem array, widely expanded across Hymenoptera, has been expanded and conserved across *Polistes*. The T subfamily, located on scaffold 8 (s8) of the *P. fuscatus* genome, is composed of 14 tandemly arrayed genes that show 1:1 orthology across five *Polistes* (Figure 7B). Differences in the extent of microsynteny among tandem arrays belonging to different OR subfamilies highlight the unique evolutionary processes shaping 9-exon OR evolution in paper wasps. At the same time, the

312 extreme conservation of L and T subfamily tandem arrays across species highlights the strong
313 conservation of the OR repertoire outside of the 9-exon subfamily among *Polistes* species.



314 **Fig. 7**: (A) Frequency of OR gene singletons and tandem arrays in the *P. fuscatus* genome. 62% of ORs
315 in *P. fuscatus* occur in tandem arrays of 6 or more genes. The longest tandem array is a 44 gene cluster on
316 scaffold 13 (s13) containing 9-exon subfamily ORs. The first row of x-axis labels is the number of OR
317 genes in a tandem array cluster, and the second row labels the OR subfamily and scaffold number
318 (abbreviated s# in parentheses) of the six longest tandem arrays.



319 (B) Genome alignments of four loci containing tandem arrays of OR genes in all *Polistes* species
320 examined. Each alignment is labeled with the corresponding subfamily and *P. fuscatus* scaffold number
321 (abbreviated s# in parentheses). Black boxes represent functional genes and gray boxes represent
322 pseudogenes. Directionality of genes is denoted by curved corners at their 3' (tail) end. Black lines
323 connect orthologous genes between species. Genomic scaffolds are represented by horizontal, gray lines,
324 and scaffold ends are represented by vertical gray lines. The black scale bars represent 5kb.

Microsynteny analysis suggests a process of ongoing gene turnover in 9-exon arrays but stasis in most other expanded subfamilies. More recent turnover should be associated with higher pairwise amino acid identity between neighboring genes in an array if they are the result of recent duplication events (Ohno 1970; Bohbot et al. 2007). To explore the relationship between amino acid divergence and tandem array locus, we compared the mean percent amino acid identity among neighboring genes within an array between the eight loci containing the longest tandem arrays of ORs in the *P. fuscatus* genome using one-way ANOVA (Figure 8). Mean percent amino acid identity of neighboring genes was significantly separated by OR array identity (DF = 7; F = 5.39; P = 2.67e-05). Differences between particular OR tandem arrays were identified using Tukey HSD post hoc tests. The mean percent amino acid identity among neighboring genes within one tandem array of nine 9-exon ORs on scaffold 12 (s12) of the *P. fuscatus* genome is higher than in the s13 9-exon array (P Adj = 0.04586), the s17 L array (P Adj = 0.00013), the s8 T array (P Adj = 0.00458), the s16 9-exon array (P Adj = 0.00124), and the s19 V array (P Adj = 0.02247). The s12 9-exon OR array is composed of a larger proportion of pseudogenes (5 PSE, 10 intact gene models) than the other two 9-exon arrays (s13: 9 PSE, 44 intact gene models; s16: 0 PSE, 12 intact gene models). ORs in the s12 9-exon array lack clear orthologous relationships with ORs in species other than *P. metricus*. Taken together, the high within array sequence similarity, high frequency of pseudogenes, and low orthology exhibited by this array indicate that it is the result of one or more recent gene duplication events since the divergence of *P. fuscatus* and *P. metricus*, the most closely related species analyzed here. The s19 H subfamily array also shows higher amino acid sequence identity among neighboring genes than the s17 L subfamily array (P Adj = 0.01625) and the s16 9-exon array (P Adj = 0.04038). Increased amino acid similarity may also occur within older tandem arrays as a result of gene conversion (Nagawa et al. 2002). However, we searched for gene conversion using GENECONV (Sawyer 1989) and did not detect gene conversion events within the s12 9-exon array or in the s19 H array after Bonferroni correction. Patterns of genomic organization of OR genes in *Polistes* genomes lead to the conclusion that gene gain and loss in the 9-exon OR subfamily is an ongoing process within this genus, in contrast to the stable and conserved tandem arrays in most other OR subfamilies.

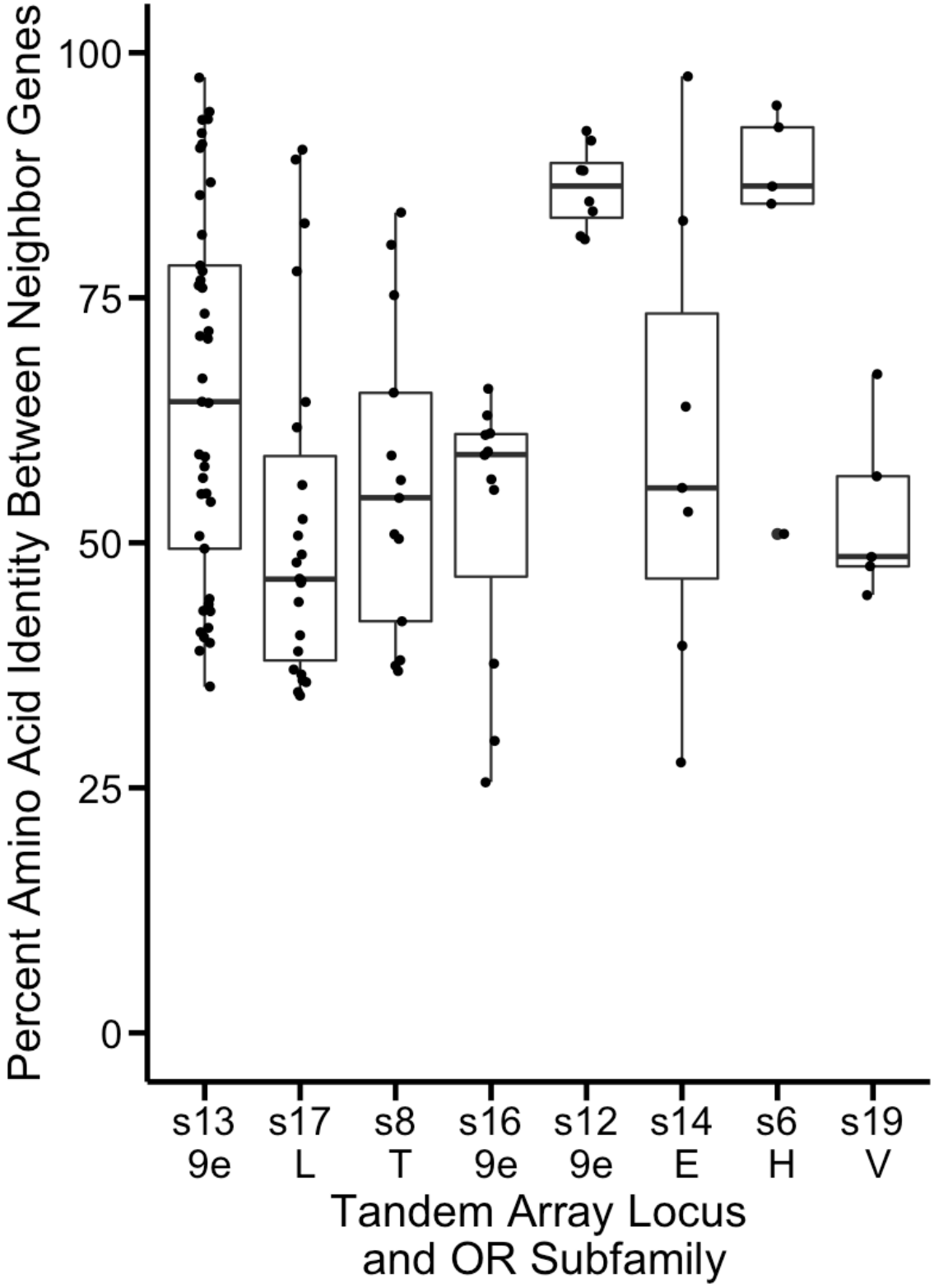


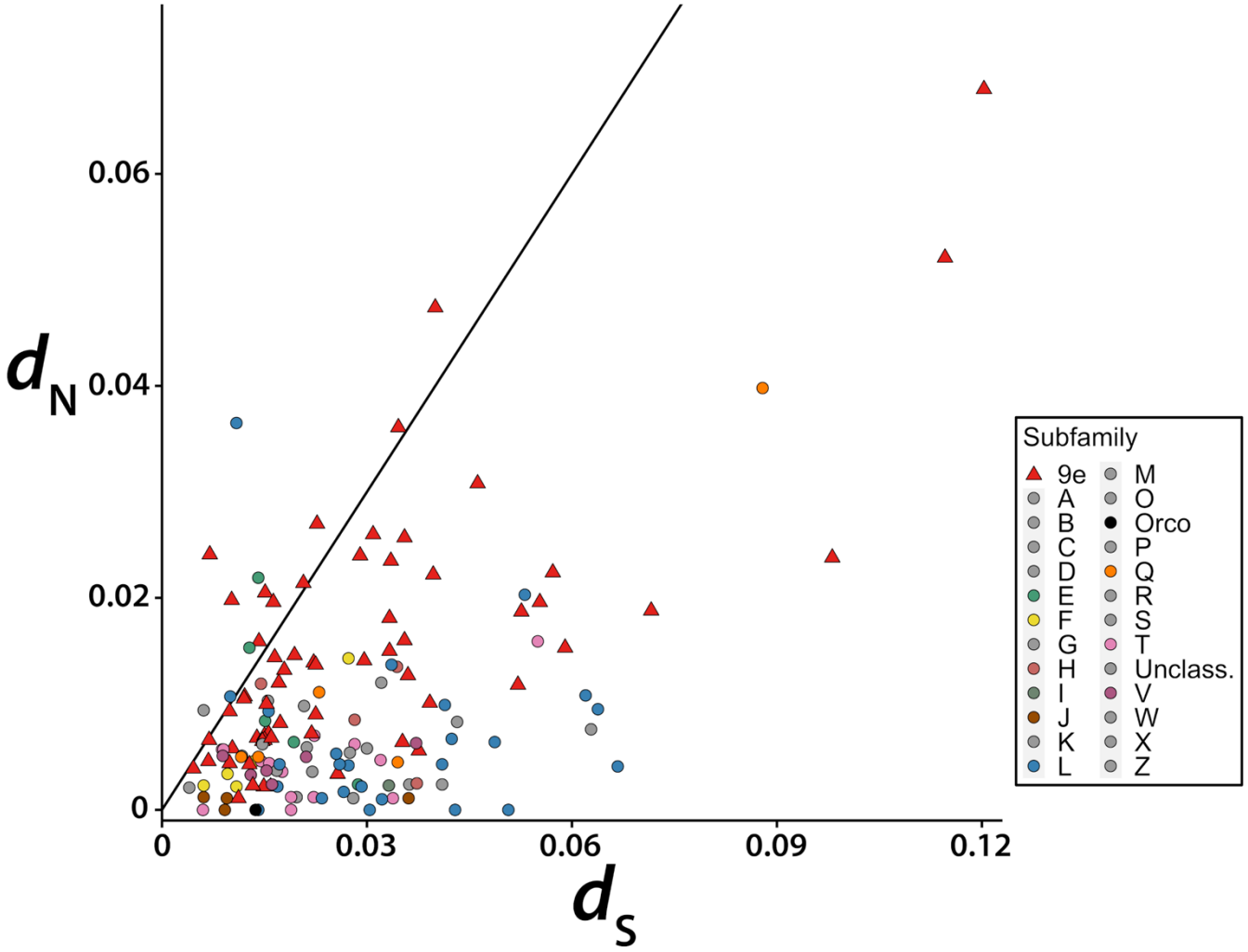
Fig. 8: Percent amino acid identity between neighboring genes at eight loci containing the longest OR gene tandem arrays in the *P. fuscatus* genome. Arrays are ordered by length in gene number, from longest (44 9-exon ORs) to shortest (6 H ORs and 6 V ORs).

403 Accelerated Evolution of 9-exon ORs and Positive Selection in Expanded OR Subfamilies

404 We examined patterns of OR evolution among five *Polistes* species to test the prediction that OR
405 subfamilies exhibit signatures of positive selection. We were especially interested in whether the
406 recent dynamism in 9-exon OR gene copy number has been accompanied by episodes of positive
407 selection in *Polistes*. HyPhy aBSREL (Smith MD et al. 2015) analyses of *Polistes* OR subfamilies
408 detected eight branches under episodic positive diversifying selection, all in OR subfamilies with
409 expansions: three branches in the 9-exon subfamily (0.33% of 918 9-exon subfamily branches;
410 Figure S4); three branches in the L subfamily (1.28% of 234 L subfamily branches; Figure S5);
411 one branch in the E subfamily (1.67% of 60 E subfamily branches; Figure S6); and one branch in
412 the H subfamily (1.54% of 65 H subfamily branches; Figure S7). This supports the hypothesis that
413 gene duplication releases duplicate genes from selective constraints, allowing duplicate sequences
414 to evolve towards other evolutionary optima (Ohno 1970). While the 9-exon OR subfamily is not
415 unique among expanded OR subfamilies in its instances of episodic positive selection as measured
416 by HyPhy aBSREL, the rate of amino acid divergence is higher among the 9-exon OR subfamily
417 as a whole. The 91% mean amino acid identity among 1:1 9-exon subfamily orthologs in *Polistes*
418 is significantly lower than the 95% mean amino acid identity among 1:1 orthologs in all other OR
419 subfamilies (Figure S8; Welch Two Sample t-test: P = 7.051e-05).

420 To further evaluate the patterns of nucleotide substitution driving accelerated amino acid
421 evolution of 9-exon ORs, we computed the values of d_N and d_S for pairwise alignments of 150
422 single copy orthologs between *P. fuscatus* and *P. dorsalis* (Figure 9) using model yn00 of PAML
423 (Yang 2007). Values of d_N are significantly higher in 9-exon (mean $d_N = 0.015$) compared to other
424 OR ortholog pairs (mean $d_N = 0.006$) (Welch Two Sample t-test, P = 5.317e-07). Values of d_S are
425 not significantly elevated among 9-exon ortholog pairs compared to other OR subfamilies (mean
426 $d_S = 0.029$ in 9-exon ORs and 0.025 in non-9exon ORs, P = 0.343). Omega values (d_N/d_S) greater
427 than 1 are often considered evidence of positive selection, while $d_N/d_S = 1$ corresponds to neutral
428 drift, and $d_N/d_S < 1$ is evidence of negative selection. The omega value (d_N/d_S) for all genes is less
429 than one, suggesting negative selection. However, omega is significantly higher in 9-exon ORs
430 than in non-9-exon ORs, indicating that negative selection is weaker on 9-exon ORs (Welch Two
431 Sample t-test: mean omega = 0.32 in non-9-exon ORs and 0.644 in 9-exon ORs, P = 8.027e-05).
432 In general, negative selection conserves ORs shared by *P. fuscatus* and *P. dorsalis* (mean omega
433 = 0.454), but an elevated rate of non-synonymous substitutions in 9-exon ortholog pairs imply

434 relaxed negative selection and more drift responsible for sequence evolution in the 9-exon relative
435 to other OR subfamilies.



436 **Fig. 9**: The values of d_S (x-axis) and d_N (y-axis) from pairwise alignments of *P. fuscatus* and *P. dorsalis*
437 orthologs are elevated in the 9-exon OR subfamily relative to other OR subfamilies. The diagonal line
438 represents a line of equality with slope of 1. Omega values of $d_N/d_S > 1$ may be evidence of positive
439 selection, while $d_N/d_S = 1$ may result from neutral drift, and $d_N/d_S < 1$ may result from negative selection.

Conclusions

By manually annotating the OR repertoires of five social wasp species spanning ~40 million years of divergence in the *Polistes* genus, this study adds a higher resolution lens to our view of the evolution of social insect odorant receptors. During the diversification of *Polistes,* evolutionary patterns show genus-wide conservation of the ~200 OR repertoire except for the 9-exon genes, which show elevated turnover and lower sequence conservation. The 9-exon OR subfamily has dramatically expanded in paper wasps, and now makes up over half of the *Polistes* OR gene set. Social and ecological niches are relatively conserved within *Polistes*, though there is considerable variation in social behavior and ecological niches among vespid wasps (Ross & Matthews 1991; O'Neill 2001). For example, an analysis of three new, high-quality hornet genomes suggested that the highly eusocial hornets have larger OR repertoires compared to the primitively eusocial *Polistes* (Harrop et al. 2020). That analysis recovered less than half of the ORs reported here for *Polistes*, likely due to a lack of antennal transcriptome data, suggesting that hornets may have even larger OR repertoires than has been reported. Evidence from the hornet *Vespa velutina*, including the discovery of 264 antennal lobe glomeruli, indicates that the hornet OR repertoire has expanded (Couto et al. 2016, 2017). Future analysis of additional high-quality genomes and antennal transcriptomes of diverse social and solitary wasps within Vespidae will allow for further examination of the relationship between social behavior and 9-exon as well as other OR subfamily expansions.

Social insect species differ in their level of sociality and extent of olfactory recognition abilities (d'Ettorre & Moore 2008; Rehan & Toth 2015). Some social aspects of the *Polistes* colony cycle vary across species. For example, the average number of cooperative foundresses varies from 1 to ~6, and average sizes of mature nests may vary ~60 to ~490 cells (Reeve 1991; Sheehan et al. 2015; Miller SE et al. 2018). Increased 9-exon OR copy number may facilitate complex olfactory recognition in species with larger colony sizes, higher cooperative nest-founding rates, and greater sympatry with related species. However, expansions of 9-exon ORs are not exclusive to social wasps, suggesting that the specific chemical ecology of an insect is a more influential factor shaping OR evolution than level of sociality (Karpe et al. 2017). Furthermore, a meta-analysis found that the complexity of CHC phenotypes does not differ between social and solitary Hymenopteran species (Kather & Martin 2015). The CHC profile of *Nasonia vitripennis* includes at least 52 CHC compounds, and detection of CHCs on prey items may help *Microplitis* identify

prey (Lewis et al. 1988; Niehuis et al. 2011). The need for solitary wasps to perceive CHCs could explain why *N. vitripennis* and *M. demolitor* exhibit expansions in the 9-exon OR subfamily.

Electrophysiological deorphanization studies of 9-exon ORs in the ant *Harpegnathos saltator* offer key insights into how 9-exon OR coding might relate to gene expansion. Through combinatorial coding, 9-exon ORs can detect a large variety of structurally diverse CHCs. Pask et al. (2017) examined 22 *H. saltator* 9-exon ORs, a subset of the 118 annotated 9-exon ORs in this species, and found that 9-exon ORs were responsive to CHCs, and overlapped in their responses to multiple CHC compounds. The combined responses of these 22 ORs to CHC extracts from different castes were sufficient to map the CHC profiles of males, workers, and reproductive females (gamergates) to separate regions of a 22-dimensional receptor space (Pask et al. 2017). This highlights the ability of 9-exon ORs to facilitate social recognition by combinatorial coding. In social insect colonies, CHC variation holds information at multiple levels of conspecific recognition, from inter-colony nestmate recognition to within colony individual recognition (d'Ettorre & Moore 2008). Expansion of the 9-exon OR subfamily might result from selection for more combinations of ORs that together can discriminate between subtle qualitative and quantitative variations in CHC blends of conspecifics. Nest-specific quantitative variation in CHCs has been documented across *Polistes* species (Espelie et al. 1990; Singer et al. 1992; Espelie et al. 1994; Layton et al. 1994), but the molecular mechanisms underlying nestmate recognition in *Polistes* are still obscure. Increased copy number of 9-exon ORs may not only expand the qualitative range of compounds perceived by paper wasps, but also the quantitative olfactory space, since wasps may be able to discern unique concentration differences between CHC blends as a result of the combined action of 9-exon ORs with various response thresholds. Gene duplication can also promote regulatory diversification (Kucharski et al. 2016; Dyson & Goodisman 2020). Between castes and across stages of the colony cycle, CHCs vary in *P. metricus* (Toth et al. 2014). Regulatory subfunctionalization of duplicate ORs could be responsible for caste- and colony phase-specific expression of ORs involved in detecting caste-specific and seasonally variable CHCs.

Most expanded OR subfamilies are highly conserved in copy number across five *Polistes* species, with the exception of the 9-exon OR subfamily. In particular, one portion of the 9-exon subfamily arranged in a single tandem array (*P. fuscatus* 9e s13) has experienced dynamic evolution. What might drive rapid gain and loss of 9-exon ORs? Divergent social chemical

landscapes between species may cause gene turnover as 9-exon OR evolution tracks evolutionarily labile chemical signals. *P. fuscatus* and *P. metricus* are closely related, and both species possess CHC profiles consisting of linear and methyl-branched alkanes (Espelie et al. 1990; Espelie et al. 1994). However, the *P. fuscatus* CHC profile includes a higher proportion of alkenes than *P. metricus* or *P. dominulus*, and the position of the methylated carbon of methyl-branched alkanes is sometimes shifted between species (Espelie et al. 1990; Singer et al. 1992; Espelie et al. 1994; Layton et al. 1994). Ant 9-exon ORs respond differently to subtle variations in CHC structure (Pask et al. 2017). Between closely related *Polistes* species, structural isomers of methyl-branched alkanes probably activate different ensembles of ORs.

If a chemical develops new behavioral relevance in a lineage, gene duplication would allow the olfactory system to explore chemical space in the direction of this compound. HyPhy aBSREL analyses identified eight branches in expanded OR subfamilies, including the 9-exon subfamily, that have undergone positive selection during the last ~40 million years, consistent with neofunctionalization or subfunctionalization of duplicated genes. Signatures of positive selection on OR genes may reflect directional selection to perceive species-specific chemical signals. Perception of species-specific CHCs might be important in mate compatibility recognition. In *Polistes*, mating occurs at territories defended by males and often frequented by multiple species (Post & Jeanne 1983; Reed & Landolt 1990). However, the frequency of interspecific mating is low, suggesting *Polistes* use vision and/or olfaction to inform their mating decisions (Miller SE et al. 2019). Duplication and deletion of ORs would facilitate evolution of species-specific chemical signaling systems that could contribute to reproductive isolation of sympatric species. If a chemical signal is lost in a species, the corresponding ORs may become obsolete, and would be expected to pseudogenize and be purged from the genome. Duplication and deletion of ORs could also lead to species-specific chemical signaling in the absence of evolutionary change in chemical signals (Cande et al. 2013). However, OR evolution is not strictly necessary for such a difference to evolve between species, and circuit-level changes can prescribe new valence to chemical signals that are shared between species and perceived by common peripheral receptors (Seeholzer et al. 2018). Neutral processes also contribute to birth-and-death events of ORs. There may be an advantage for a large copy number up to a point, followed by random gene duplication and deletion around this optimal copy number. This random genomic drift has been proposed to shape vertebrate

532 olfactory receptor evolution and copy number variation in all large multigene families (Nei 2007;
533 but see Hayden et al. 2010).

534 Aside from the 9-exon OR subfamily, gene expansions have occurred in subfamilies L, T,
535 H, E, and V (Figure 4). A larger variety of ORs relaying information through ORNs to a larger
536 number of antennal lobe glomeruli will increase sensory acuity in any olfactory discrimination
537 task, social or otherwise. An ancient locus of tandemly duplicated L subfamily ORs observed
538 across social insects has expanded in *Polistes*, although to a lesser extent than in other social insects
539 (~50 L subfamily ORs in honeybee and ants, 25 L subfamily ORs in a tandem array on *P. fuscatus*
540 scaffold 17). Odorant receptors in the L subfamily are thought to detect queen pheromone
541 components and fatty acids in bees as well as CHCs in ants (Wanner et al. 2007; Karpe et al. 2016;
542 Pask et al. 2017). The T subfamily has expanded to a greater degree in *P. fuscatus* (14 genes) than
543 in ants (~7 genes) and the honeybee (2 genes), but no ORs in this clade have been functionally
544 characterized. *P. fuscatus* has 9 H subfamily ORs, which are putative floral odorant detectors in
545 bees, and which also respond to CHCs and other general odorants in ants (Claudianos et al. 2014;
546 Slone et al. 2017). Fatty acids and volatile organic compounds are produced by flowers that wasps
547 rely on as a source of carbohydrates (Raguso 2008). Expansions in several OR subfamilies may
548 increase olfactory discrimination of chemicals with diverse behavioral relevance. *Polistes* species
549 are distributed globally in temperate and tropical regions, occupying similar social and ecological
550 niches as generalist predators and floral foragers that form primitively eusocial societies (Reeve
551 1991; Richter 2000). A conserved set of ORs may perform common functions in conserved
552 behaviors across paper wasp species. High levels of OR conservation are also consistent with a
553 specialist molecular function of an OR in a dedicated channel of olfaction. Patterns of molecular
554 evolution suggest conserved behavioral and molecular functions of most non-9-exon *Polistes* OR
555 subfamilies.

556 The differences between the conserved OR repertoires in *Drosophila* and the more
557 dynamic evolution of vertebrate OR gene families have given rise to speculation about the
558 relationship between OR function and evolution (Nozawa & Nei 2007; Andersson et al. 2015).
559 There is prevalent negative selection conserving odorant receptors across *Drosophila* species, and
560 the majority of *D. melanogaster* ORs form simple orthologous relationships across the genus
561 (Clark et al. 2007; Guo & Kim 2007; McBride & Arguello 2007; Nozawa & Nei 2007; Sánchez-
562 Gracia et al. 2009; Mansourian & Stensmyr 2015). In paper wasps we report both highly conserved

OR expansions similar to those seen in *Drosophila* as well as elevated gene turnover and drift among the 9-exon ORs, reminiscent of a more vertebrate-like evolutionary pattern. If the highly dynamic clades of 9-exon ORs of social wasps are involved in more combinatorial coding compared to other more conserved 9-exon or non-9-exon ORs, that would indicate a link between molecular evolution of receptors and neural coding. Further investigations into the relative tuning of 9-exon as well as more conserved ORs in social wasps and other social insects provide a promising research direction to investigate the links between molecular evolutionary patterns, receptor tuning, and neural coding.

Materials and Methods

Antennal lobe imaging

Immunocytochemistry: In the fall of 2018, 20 adult *P. fuscatus* wasps (10 male and 10 gyne) were collected in Ithaca, NY. Wasps were immobilized by cooling at 4°C, heads were removed, and brains were dissected from the head capsule and placed in ~1 ml of 4% paraformaldehyde in 0.01M phosphate buffered saline (PFA) overnight at 4°C. Mouse monoclonal anti-synapsin antibodies (SYNORF1, 3C11; Data Bank Hybridoma) were used as primary antibodies to label synaptically rich antennal lobe glomeruli. Phalloidin conjugated with TRITC (Tetramethylrhodamine Isothiocyanate; Invitrogen) was used to distinguish glomeruli borders and label antennal nerve tracts. F(ab')2 fragments of donkey anti-mouse antibodies conjugated to Alexa 488 were used as secondary antibodies to visualize anti-synapsin (Jackson ImmunoResearch Laboratories). After overnight fixation, brains were washed in 0.01M phosphate buffered saline with Triton-X 100 (PBS-TX) 6 times for 20 min. Brains were incubated for three nights at room temperature with anti-synapsin at 1:800 in PBS-TX solution. Brains were then washed again in PBS-TX (6 times for 20 min) and incubated for three nights at RT with donkey anti-mouse antibodies conjugated with Alexa 488 at 1:270, and then with Phalloidin-TRITC at 1:160 in PBS-TX. After staining with secondary antibodies, brains were post-fixed with 4% PFA for 10 min and dehydrated using increasing concentrations of methanol. Following full dehydration and two washes in 100% methanol, brains were cleared in methyl salicylate overnight. They were then mounted on slides in methyl salicylate.

Imaging: Images were collected using an inverted 880 confocal laser scanning microscope (Zeiss LSM880 Indimo, AxioObserver) with a C-Apochromat10x/0.45 water-immersion objective and a 1.8x digital zoom using the appropriate filter and laser setting for each fluorescent molecule and 8x line averaging. Image stacks were collected using 10-15 µm optical sections for both males and gynes. To individually label the glomeruli from confocal image stacks we used AVIZO and Amira software (Thermo Scientific, FEI) to make 3D reconstructions of the antennal lobe from whole mount preparations. Confocal stack files were imported into AVIZO software, and the voxel dimension outputs of each image stack were imported to provide correct dimensions (Thermo Scientific, FEI). Using the image segmentation function, we then identified glomeruli by hand throughout the entire image stack.

Antennal RNAseq

P. fuscatus mRNA library preparation and sequencing: In the fall of 2017, 12 adult *P. fuscatus* wasps (6 males and 6 gynes) from Ithaca, NY and Watkins Glen, NY were freeze-killed on dry ice and stored at -80°C. Each pair of antennae was cut at the base of the scape and weighed (paired gyne antennae = 1.27 mg ± 0.26 SD; paired male antennae = 1.57 mg ± 0.44 SD). RNA was isolated from 2 pooled antennae of each individual using the PureLink RNA Micro Kit (Ambion). Libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB #E7770L) with the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490) and NEBNext Multiplex Oligos for Illumina (NEB #E7600S). Libraries were analyzed on a fragment analyzer for quality control and sequenced on two lanes of an Illumina HiSeq (Illumina, San Diego, CA, USA).

Mapping mRNA reads of *P. fuscatus* and *P. dominula*: Raw reads from *P. fuscatus* antennae were trimmed using Trimmomatic v.0.39 to (1) remove adapter sequences, (2) cut leading and trailing bases below quality of 3, (3) cut 4-base wide windows when the average quality per base dropped below 20, and (4) remove reads below 35 base pairs (bp) long (Bolger et al. 2014). After trimming and filtering, more than 154 million paired end reads remained (75,846,826 male; 78,453,620 gyne) with an average length of 132 bp. Processed paired end reads were mapped, first to the *P. fuscatus* genome and subsequently to the genomes of 2 other *Polistes* species in the fuscopolistes clade (*P. metricus* and *P. dorsalis*), using STAR v2.6 to generate BAM files sorted by genome coordinate (Dobin et al. 2013). Each read was allowed to map to no more than one locus. Mapped reads from each sex were combined and assembled using Trinity genome-guided assembly (Haas et al. 2013). Summary statistics for the assembly are shown in Table S2. Raw reads from the whole-body transcriptome of a single individual *P. dominula* were downloaded from the Sequence Read Archive (Lopez-Osorio et al. 2017). To aid in manual annotation of ORs, reads were aligned to the *P. dominula* genome with STAR v2.7, using the same parameters as described above for *P. fuscatus*.

Gene annotation

OR annotation in *P. fuscatus*: The *Polistes fuscatus* genome was assembled as described in Miller SE et al. 2020. Genome scaffolds containing putative ORs were identified based on sequence similarity with published insect ORs from a range of species: *Atta cephalotes*, *Acromyrmex echinatior*, *Apis mellifera*, *Camponotus floridanus*, *Cardiocondyla obscurior*,

Ceratosolen solmsi, *Drosophila melanogaster*, *Eulaema bombiformis*, *Euglossa dilemma*, *Euglossa flammea*, *Euglossa imperialis*, *Eulaema meriana*, *Eufriesea mexicana*, *Lasioglossum albipes*, *Microplitis demolitor*, *Monomorium pharaonis*, *Melipona quadrifasciata*, *Nasonia vitripennis*, *Solenopsis invicta* (Robertson et al. 2003, Robertson et al. 2010, Zhou et al. 2012, Zhou et al. 2015, Brand & Ramirez 2017, McKenzie & Kronauer 2018). The *P. fuscatus* genome was queried with amino acid sequences from each of the listed species using the TBLASTN algorithm with an e-value cutoff of 1e-5 (Altschul et al. 1997). Putative coding regions were identified by examining TBLASTN match regions and chaining the match regions according to criteria similar to those used by Zhou et al. (2012): neighboring match regions were included in the same coding region if the match regions were in close proximity and if the upstream amino acid query sequence was N-terminal to that of the downstream match region. Coding sequence (CDS) was determined by manually curating exon-intron boundaries in Geneious v11.1.5 using evidence from MAKER automated predictions, TBLASTN homology, Trinity predicted transcripts (aligned to genome using BLAT) and mapped RNAseq reads. We consulted version 24 of the Havana annotation guidelines to inform manual annotation of exon-intron splice sites (March 2016 guidelines by the Human and Vertebrate Analysis and Annotation group at WTSI). "AG" was the only permitted acceptor site. "GT" and "GC" were both acceptable donor sites, although "GC" was only present in a small number of genes. Amino acid sequences of ORs generated by this pipeline were used as queries in an iterative round of TBLASTN with scoring matrix PAM30 (Pearson 2013), to search for species-specific expansions, especially in the highly divergent 9-exon receptor clade. The majority of apparently functional ORs that were not detected by the automated annotation and required extensive manual curation were 9-exon subfamily receptors (e.g. Figure S9).

OR annotation in 4 other *Polistes*: The *Polistes dorsalis* and *Polistes metricus* genomes were assembled and annotated as described in Miller SE et al. 2020. The *Polistes canadensis* and *Polistes dominula* genomes and annotations were accessed through NCBI (Patalano et al. 2015; Standage et al. 2016). Putative ORs were identified by using TBLASTN with the same sample of Hymenopteran protein query sequences used during manual annotation of *P. fuscatus* ORs, with the addition of *P. fuscatus* OR protein sequences. Uncertain gene models were aligned using Muscle version 3.8.425 with maximum 4 iterations and gene models were manually adjusted based on homology with functional ORs in *P. fuscatus* (Edgar 2004). Gene models were called

663 pseudogenes if they exhibited frame-shift mutation, premature stop codons, or unacceptable 5'
664 donor or 3' receptor splice sites.

665 Quality assessment of putatively functional ORs: All *Polistes* OR protein sequences greater
666 than 300 amino acids in length were analyzed using InterProScan version 5.17-56.0 to check for
667 the InterPro olfactory receptor annotation IPR004117, the 7tm odorant receptor Pfam accession
668 PF02949, and the odorant receptor PANTHER accession PTHR21137 (Jones P et al. 2014). For
669 each of 192 orthologous groups predicted by OrthoFinder, the orthologs were aligned using
670 Muscle version 3.8.425 with maximum 4 iterations, and alignments were inspected for anomalies
671 that could be the result of annotation error. Adjustments were made to the gene models in 29
672 orthologous groups. The majority of these adjustments were changes in the length of already
673 annotated exons with few or no mapped mRNA reads. Transmembrane helices of all putatively
674 functional (>300 aa) ORs were predicted using TMHMM version 2.0c (Sonnhammer et al. 1998)
675 and Phobius version 1.01 (Käll et al. 2004).

676 Naming ORs: All ORs were named with a species-specific prefix using the first letter of
677 the genus followed by the first three letters of the species epithet (e.g. PfusOR#). Orco was not
678 numbered (e.g. PfusOrco), and no OR was labeled OR1, since some previous studies refer to Orco
679 as OR1. All tuning ORs were numbered in the order in which they appeared in the genome,
680 beginning with OR2. Therefore, OR names with the same number do not necessarily indicate
681 orthology among *Polistes*. Incomplete gene models were named with a suffix to indicate which
682 features of the gene model are missing, following Oxley et al. (2014): "NTE" missing N-terminus,
683 "INT" missing interior sequence, "CTE" missing C terminus, "NI" missing N terminus and interior
684 sequence, "NC" missing N and C termini. All pseudogenes were given the suffix "PSE."

685 Search for ORs in unassembled genomic reads: Genome misassembly is a potential cause
686 of overestimation of pseudogenes, and is more common in repetitive regions of the genome. For
687 example, many ORs that were previously annotated as pseudogenes in the clonal raider ant were
688 not identified in a new, chromosome arm level assembly of the genome, which also yielded many
689 more functional OR gene models (McKenzie & Kronauer 2018). Given the completeness of the
690 *Polistes* genomes (Table S1) and the close correspondence between functional *P. fuscatus* ORs
691 and number of *P. fuscatus* antennal lobe glomeruli, this type of error likely accounts for minimal
692 annotation errors in this species. We searched for sequence similarity to ORs in the unassembled
693 genomic reads of *P. dorsalis*, the species in which we initially found the fewest functional ORs.

694 The 10x Genomics Supernova Assembler version 2.1.1 failed to assemble unmapped *P. dorsalis*
695 genomic reads greater than or equal to 150 nucleotides (~21.4 million paired-end reads). Therefore,
696 the reads themselves were formatted as a blast database and queried using TBLASTN at an e-value
697 cutoff of 1e-5. The same sample of Hymenopteran protein sequences used during manual
698 annotation, with the addition of the putatively functional OR protein sequences of all five *Polistes*
699 species, were used to search the unaligned *P. dorsalis* genomic reads. This resulted in significant
700 hits to 120 paired-end reads. Given that on average 159 ± 555 paired-end reads map to a given OR
701 gene in *P. fuscatus*, it is unlikely that many ORs were lost in unassembled genomic reads.
702 However, it cannot be ruled out that genome misassembly may account for overestimation of
703 pseudogenes in some cases.

704 Phylogenetic reconstruction of gene families

705 Hymenoptera-wide analysis: We used published OR sequences from *Apis mellifera*,
706 *Nasonia vitripennis* (Robertson & Wanner 2006; Robertson et al. 2010), and *Camponotus*
707 *floridanus* (Zhou et al. 2012), along with OR sequences of *P. fuscatus* (this study) to reconstruct
708 the phylogeny of the OR gene family. All amino acid (aa) sequences of chemoreceptors at least
709 300 aa in length were aligned using MAFFT version 7.453 with the high accuracy "E-INS-i"
710 method ("genafpair" option combined with 1000 cycles of iterative refinement ("maxiterate"))
711 (Katoh and Standley 2013). The alignment was trimmed using trimAl v3 with the "automated1"
712 option to automatically choose the parameters that best fit the qualities of the alignment (Capella-
713 Gutierrez et al. 2009), generating a phylip file with 1066 sequences of 231 aa length. Gene trees
714 were generated with RAxML version 8.2.12 using the Maximum Likelihood optimality criterion
715 with 100 rapid bootstrap inferences, CAT model of rate heterogeneity across sites (Stamatakis
716 2006), JTT amino acid replacement matrix, F option to use empirical base frequencies from the
717 input alignment ("-m PROTCATJTTF"), "-p 12345" to set initial random number for parsimony
718 search, *N. vitripennis* Orco as outgroup, "-f a" to combine bootstrap and maximum likelihood
719 searches in one run, "-x 12345" to set initial random number for bootstrap search, "-N 100" to
720 perform a bootstrap analysis informed by 100 unique maximum likelihood trees, and "-k" to add
721 branch lengths to trees (Stamatakis 2014, Jones DT et al. 1992). The tree was rerooted in FigTree
722 v1.4.4 (https://github.com/rambaut/figtree/releases/tag/v1.4.4) to visually clarify the position of
723 the Orco gene clade. The tree was visualized in R version 3.6.1 using ggtree (Figure 4A; Yu et al.
724 2017; R Core Team 2019).

Estimation of gene gain and loss events: Gene duplication and loss events were reconstructed by reconciling a gene tree with a species tree in NOTUNG version 2.9.1.3 (Durand et al. 2006; Vernot et al. 2008). A gene tree of putatively functional OR proteins at least 300 amino acids in length was generated using published protein sequences for 4 ants (*A. echinatior*, *A. cephalotes*, *S. invicta*, *C. floridanus*; Zhou et al. 2012, 2015), 2 bees (*A. mellifera*, *L. albipes*; Robertson et al. 2010, Zhou et al. 2015), 3 nonsocial wasps (*N. vitripennis*, *M. demolitor*, *C. solmsi*; Robertson et al. 2010, Zhou et al. 2015), and 5 social wasps (*P. fuscatus*, *P. metricus*, *P. dorsalis*, *P. canadensis*, *P. dominula*; annotated in this study). The gene tree was generated as described above for the Hymenoptera OR tree, without rerooting. Following Zhou et al. 2012, all internal branches with bootstrap support values below 70 were collapsed into polytomies in R using ggtree. Notung was run in rearrangement mode using an edge weight (bootstrap value) threshold of 90, with cost of "1.5" for a duplication and cost of "1" for a loss.

Analysis of *Polistes* ORs: The OR sequences of *P. fuscatus*, *P. metricus*, *P. dorsalis*, *P. canadensis*, and *P. dominula* were filtered to include only ORs of at least 300 amino acids in length. These were aligned using MAFFT ("E-INS-i" method) and trimmed in trimal using the parameters described above for the Hymenoptera OR tree. The tree was generated using RAxML, rerooted in FigTree, and visualized in R using ggtree as described above. Trees in the supplementary materials were visualized using Interactive Tree Of Life (iTOL) version 5.6.2 (Letunic & Bork 2019). Orthology of *Polistes* OR amino acid sequences was inferred using sequence similarity and gene collinearity in the genome. Orthologs were called based on similarity of amino acid sequences with OrthoFinder version 2.2.7 using DIAMOND for sequence similarity searches with an e-value cutoff of 0.001 (Emms & Kelly 2015). OrthoFinder predicted 135 orthologous groups that included gene copies from both *P. fuscatus* and *P. dorsalis*. For the purpose of pairwise d_N/d_S analysis, this set of orthologs was expanded by inspecting the gene tree (Figure 5; Figure S3) and Muscle alignments (see "Genomic organization" below) of syntenic regions. Gene copies within a clade supported by a bootstrap value of at least 70 and showing synteny were considered orthologs in the expanded ortholog set of 150 single copy orthologs between *P. fuscatus* and *P. dorsalis*.

Genomic organization

OR genes and pseudogenes were considered to be in a tandem array if they were uninterrupted by non-OR genes and were within 5kb of each other. The lengths of OR clusters

reflect the number of putatively functional ORs - gene models coding for proteins at least 300 amino acids in length - and exclude the pseudogenes contained within the array. The pairwise percent amino acid identity between neighbors in an array was calculated using only putatively functional ORs that neighbored another putatively functional OR within 5kb. Two functional ORs separated by a pseudogenized OR were considered neighbors if their coding sequences were within 5kb of each other. Neighboring OR protein sequences were aligned using Muscle v3.8.425 with maximum 4 iterations to determine percent amino acid identity. One-way ANOVA comparing percent amino acid identity of neighbors across eight tandem arrays was performed using the "aov" function, and Tukey HSD test with the "TukeyHSD" function, both of which are in the base R package stats version 3.6.1 (R Core Team 2019).

Sequence analyses

Tests for positive selection: All putatively functional *Polistes* ORs greater than 350 amino acids in length were used in analyses of episodic diversifying selection within OR subfamilies using aBSREL in HyPhy version 2.5.15 (Smith MD et al. 2015). Subfamily-specific alignments were generated using MAFFT version 7.453 ("E-INS-i") and trimmed in trimal with the "gappyout" option. Trees were generated using RAxML as above, except the "PROTGAMMAAUTO" model was used to automatically choose the best protein model given likelihood. CDS alignments were generated using trimal with the "backtranslate" option added to the above parameters. The significance threshold for each aBSREL analysis was Holm-Bonferroni corrected p=0.05. Values of pairwise d_N/d_S for orthologs shared by *P. fuscatus* and *P. dorsalis* were calculated using PAML version 4.9 (Yang 2007) program yn00. The values of d_N and d_S reported were calculated using the Yang & Nielsen (2000) method. Values of d_N, d_S, and omega were compared between 9-exon and non-9-exon orthologous pairs using the function "t.test" in the base R package stats version 3.6.1 (R Core Team 2019).

Test for gene conversion: To search for gene conversion events, all *Polistes* OR subfamily codon alignments were analyzed using GENECONV version 1.81a with options "/lp" to list all pairwise significant fragments, "/w123" to set initial random number, and "-nolog" to avoid a segmentation fault in UNIX (Sawyer 1989). Bonferroni-corrected KA p-values below 0.05 were considered significant.

Acknowledgements

This work was supported by the National Science Foundation [Graduate Research Fellowship Program grant number DGE-1650441 to AWL, CAREER grant number DEB-1750394 to MJS], National Institutes of Health [grant numbers DP2-GM128202 to MJS, S10OD018516 to the Cornell University Biotechnology Resource Center (BRC)], and New York State Stem Cell Science [grant number CO29155 to Cornell BRC]. We thank Qi Sun and the Computational Biology Service Unit of the Cornell Life Sciences Core Laboratories Center for making software available on BioHPC and for providing helpful advice regarding gene annotation. We also thank Brook Luers and the Cornell Statistical Consulting Unit for statistics consultation.

Author Contributions

AWL and MJS conceptualized the study. SEM and MJS sequenced and assembled the genomes. CMJ dissected and imaged antennal lobes. AWL sequenced mRNA and manually annotated ORs. AWL and MFF labeled antennal lobe glomeruli. AWL conducted all downstream analyses and wrote the paper. CMJ wrote the methods section on antennal lobe imaging. AWL, SEM, CMJ, and MJS edited the final manuscript.

Supplementary Materials

GFF files of OR models for 5 *Polistes* species
Fasta files of OR amino acid sequences and CDS for 5 *Polistes* species
Supplementary excel file with meta-data associated with ORs of 5 *Polistes* species
Table S1
Table S2
Table S3
Figure S1
Figure S2
Figure S3
Figure S4
Figure S5
Figure S6
Figure S7
Figure S8
Figure S9

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Andersson MN, Löfstedt C, Newcomb RD. 2015. Insect olfaction and the evolution of receptor tuning. *Front Ecol Evol.* 3:53.

Arnold G, Masson C, Budharugsa S. 1985. Comparative study of the antennal lobes and their afferent pathway in the worker bee and the drone (*Apis mellifera*). *Cell Tissue Res.* 242:593-605.

Blomquist GJ, Bagnères A-G. 2010. Insect Hydrocarbons. Cambridge University Press, UK.

Bohbot J, Pitts RJ, Kwon H-W, Rützler M, Robertson HM, Zwiebel LJ. 2007. Molecular characterization of the *Aedes aegypti* receptor gene family. *Insect Mol Biol.* 16:525-537.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30:2114–2120.

Brand P, Ramírez SR, Leese F, Quezada-Euan JG, Tollrian R, Eltz T. 2015. Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). *BMC Evol Biol.* 15:176.

Brand P, Ramirez SR. 2017. The evolutionary dynamics of the odorant receptor gene family in corbiculate bees. *Genome Biol Evol.* 9:2023-2036.

Butterwick JA, del Mármol J, Kim KH, Kahlson MA, Rogow JA, Walz T, Ruta V. 2018. Cryo-EM structure of the insect olfactory receptor Orco. *Nature.* 560:447-467.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25:1972–1973.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol.* 7:429-447.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 450:203-218.

Claudianos C, Lim J, Young M, Yan S, Cristino AS, Newcomb RD, Gunasekaran N, Reinhard J. 2014. Odor memories regulate olfactory receptor expression in the sensory periphery. *European J Neurosci.* 39:1642-1654.

Couto A, Lapeyre B, Thiéry D, Sandoz J-C. 2016. Olfactory pathway of the hornet Vespa velutina: new insights into the evolution of the Hymenopteran antennal lobe. *Journal Comp Neurol.* 524:2335-2359.

Couto A, Mitra A, Thiéry D, Marion-Poll F, Sandoz J-C. 2017. Hornets have it: a conserved olfactory subsystem for social recognition in Hymenoptera? *Front Neuroanat.* 11(48):1-12.

850 Dani FR, Jones GR, Destri S, Spence SH, Turillazzi S. 2001. Deciphering the recognition
851 signature within the cuticular chemical profile of paper wasps. *Anim Behav.* 62:165-171.

852 Dani FR. 2009. Cuticular lipids as semiochemicals in paper wasps and other social insects. *Ann*
853 *Zool Fennici.* 43:500-514.

854 Dani FR, Turillazzi S. 2018. Chemical communication and reproduction partitioning in social
855 wasps. *J Chem Ecol.* 44:796-804.

856 Dapporto L, Santini A, Dani FR, Turillazzi S. 2007. Workers of a *Polistes* paper wasp detect the
857 presence of their queen by chemical cues. *Chem Senses.* 32:795-802.

858 d'Ettorre P, Moore AJ. 2008. Chapter 5: Chemical communication and the coordination of social
859 interactions in insects. d'Ettorre P, Hughes DP, editors. Sociobiology of communication: an
860 interdisciplinary perspective. Oxford University Press, NY, USA. p. 81-96.

861 Dobin A, Davis CA, Schlesinger F, Drnkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
862 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2:15-21.

863 Durand D, Halldorsson BV, Vernot B. 2006. A Hybrid Micro-Macroevolutionary Approach to
864 Gene Tree Reconstruction. *J Comput Biol.* 13:320-335.

865 Dweck HKM, Ebrahim SAM, Farhan A, Hansson BS, Stensmyr MC. 2015. Olfactory proxy
866 detection of dietary antioxidants in *Drosophila*. *Curr Biol.* 25:455-466.

867 Dyson CJ, Goodisman MAD. 2020. Gene duplication in the honeybee: patterns of DNA
868 methylation, gene expression, and genomic environment. *Mol Biol Evol.* 37:2322-2331.

869 Ebrahim SAM, Dweck HKM, Stökl J, Hofferberth JE, Trona F, Weniger K, Rybak J, Seki Y,
870 Stensmyr MC, Sachse S, Hansson BS, Knaden M. 2015. *Drosophila* avoids parasitoids by
871 sensing their semiochemicals via a dedicated olfactory circuit. *PLoS Biol.* 13:e1002318.

872 Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
873 throughput. *Nucleic Acids Res.* 32:1792–1797.

874 Eirín-López JM, Rebordinos L, Rooney AP, Rozas J. 2012. The birth-and-death evolution of
875 multigene families revisited. *Genome Dyn.* 7:170-196.

876 Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome
877 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.

878 Engsontia P, Sangket U, Robertson HM, Stasook C. 2015. Diversification of the ant odorant
879 receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. *BMC*
880 *Res Notes.* 8:380.

881 Espelie KE, Wenzel JW, Chang G. 1990. Surface lipids of social wasp *Polistes metricus* Say and
882 its nest pedicel and their relation to nestmate recognition. *J Chem Ecol.* 16:2229-2241.

883 Espelie KE, Gamboa GJ, Grudzien BA, Bura EA. 1994. Cuticular hydrocarbons of the paper
884 wasp, *Polistes fuscatus*: a search for recognition pheromones. *J Chem Ecol.* 20:1677-1687.

885 Ferguson ST, Park KY, Ruff AA, Bakis I, Zwiebel LJ. 2020. Odor coding of nestmate
886 recognition in the eusocial ant *Camponotus floridanus*. *J Exp Biol.* 223, jeb215400.

887 Fishilevich E, Domingos AI, Asahina K, Naef F, Vosshall LB, Louis M. 2005. Chemotaxis
888 behavior mediated by single larval olfactory neurons in *Drosophila*. *Curr Biol.* 15:2086-2096.

889 Fishilevich E, Vosshall LB. 2005. Genetic and functional subdivision of the *Drosophila* antennal
890 lobe. *Curr Biol.* 15:1548-1553.

891 Fleischer J, Pregitzer P, Breer H, Krieger J. 2018. Access to the odor world: olfactory receptors
892 and their role for signal transduction in insects. *Cell Mol Life Sci.* 75:485-508.

893 Gamboa GJ, Reeve HK, Pfennig DW. 1986. The evolution and ontogeny of nestmate recognition
894 in social wasps. *Ann Rev Entomol.* 31:431-454.

895 Gamboa GJ, Grudzien TA, Espelie KE, Bura EA. 1996. Kin recognition in social wasps:
896 combining chemical and behavioural evidence. *Anim Behav.* 51:625-629.

897 Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK.
898 2015. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses,
899 odorant receptors, and ancestral diet. *Proc Natl Acad Sci USA.* 112:3026-3031.

900 Groothuis J, Pfeiffer K, el Jundi B, Smid HM. 2019. The jewel wasp standard brain: average
901 shape atlas and morphology of the female *Nasonia vitripennis* brain. *Arthropod Struct Dev.*
902 51:41-51.

903 Guo S, Kim J. 2007. Molecular evolution of *Drosophila* odorant receptor genes. *Mol Biol Evol.*
904 24:1198-1207.

905 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D,
906 Li B, Lieber M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the
907 Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494-1512.

908 Hallem EA, Ho MG, Carlson JR. 2004. The molecular basis of odor coding in the *Drosophila*
909 Antenna. *Cell.* 117:965-979.

910 Hallem EA, Carlson JR. 2006. Coding of odors by a receptor repertoire. *Cell.* 125:143-160.

911 Harrop TWR, Guhlin J, McLaughlin GM, Perminia E, Stockwell P, Gilligan J, Le Lec MF,
912 Gruber MAM, Quinn O, Lovegrove M, et al. 2020. High-quality assemblies for three invasive
913 social wasps from the Vespula genus. *G3-Genes Genom Genet. Early online August 28, 2020*;
914 https://doi.org/10.1534/g3.120.401579

915 Hayden S, Bekaert M, Crider TA, Mariani S, Murphy WJ, Teeling EC. 2010. Ecological
916 adaptation determines functional mammalian olfactory subgenomes. *Genome Res.* 20:1-9.

917 Hines HM, Hunt JH, O'Connor TK, Gillespie JJ, Cameron SA. 2007. Multigene phylogeny
918 reveals eusociality evolved twice in vespid wasps. *Proc Natl Acad Sci USA.* 104:3295-3299.

919 Hopf TA, Morinaga S, Ihara S, Touhara K, Marks DS, Benton R. 2015. Amino acid coevolution
920 reveals three-dimensional structure and functional domains of insect odorant receptors. *Nat*
921 *Commun.* 6:6077.

922 Jandt JM, Tibbetts EA, Toth AL. 2014. *Polistes* paper wasps: a model genus for the study of
923 social dominance hierarchies. *Insect Soc.* 61:11-27.

924 Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from
925 protein sequences. *Comput Appl Biosci.* 8:275–282.

926 Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell
927 A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification.
928 *Bioinformatics.* 30:1236-1240.

929 Käll L, Krogh A, Sonnhammer ELL. 2004. A combined transmembrane topology and signal
930 peptide prediction method. *J Mol Biol.* 338:1027-1036.

931 Kapheim KM, Pan H, Li C, Salzberg SL, Puiu D, Magoc T, Robertson HM, Hudson ME, Venkat
932 A, Fischman BJ, et al. 2015. Genomic signatures of evolutionary transitions from solitary to
933 group living. *Science.* 348:1139-1143.

934 Karpe SD, Jain R, Brockmann A, Sowdhamini R. 2016. Identification of complete repertoire of
935 *Apis florea* odorant receptors reveals complex orthologous relationships with *Apis mellifera*.
936 *Genome Biol Evol.* 8:2879-2895.

937 Karpe SD, Dhingra S, Brockmann A, Sowdhamini R. 2017. Computational genome-wide survey
938 of odorant receptors from two solitary bees *Dufourea novaeangliae* (Hymenoptera: Halictidae)
939 and *Habropoda laboriosa* (Hymenoptera: Apidae). *Sci Rep.* 7:10823.

940 Karpe SD, Tiwari V, Sowdhamini R. 2020. InsectOR - webserver for sensitive identification of
941 insect olfactory receptor genes from non-model genomes. bioRxiv doi:
942 https://doi.org/10.1101/2020.04.29.067470

943 Kather R, Martin SJ. 2015. Evolution of cuticular hydrocarbons in the Hymenoptera: a meta-
944 analysis. *J Chem Ecol.* 41:871-883.

945 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
946 improvements in performance and usability. *Mol Biol Evol.* 30: 772-780.

947 Keller A, Vosshall LB. 2007. Influence of odorant receptor repertoire on odor perception in
948 humans and fruit flies. *Proc Natl Acad Sci USA.* 104:5614-5619.

949 Kucharski R, Maleszka J, Maleszka R. 2016. A possible role of DNA methylation in functional
950 divergence of a fast evolving duplicate gene encoding odorant binding protein 11 in the
951 honeybee. *Proc R Soc B.* 283:20160558.

952 Layton JM, Camann MA, Espelie KE. 1994. Cuticular lipid profiles of queens, workers, and
953 males of social wasp *Polistes metricus* Say are colony-specific. *J Chem Ecol.* 20:2307-2321.

954 Lavine BK, Morel L, Vander Meer RK, Gunderson RW, Han JH, Bonanno A, Stine A. 1990.
955 Pattern recognition studies in chemical communication: nestmate recognition in *Camponotus*
956 *floridanus*. *Chemometr Intell Lab.* 9:107-114.

957 Leonhardt SD, Menzel F, Nehring V, Schmitt T. 2016. Ecology and evolution of communication
958 in social insects. *Cell.* 164:1277-1287.

959 Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
960 developments. *Nucleic Acids Res.* 47:W256-W259.

961 Lewis WJ, Sonnet PE, Nordlund DA. 1988. Responses of braconid parasitoids *Microplitis*
962 *croceipes* (Cresson) and *M. demolitor* Wilkonson to stereoisomers of kairomone 13-
963 methylhentriacontane. *J Chem Ecol.* 14:883-888.

964 Lopez-Osorio F, Pickett KM, Carpenter JM, Ballif BA, Agnarsson I. 2017. Phylogenomic
965 analysis of yellowjackets and hornets (Hymenoptera: Vespidae, Vespinae). *Mol Phylogenet Evol.*
966 107:10-15.

967 Lynch M. 2007. The origins of genome architecture. Sunderland, MA, USA. Sinauer Associates,
968 Sunderland, MA, USA.

969 Malnic B, Hirono J, Sato T, Buck LB. 1999. Combinatorial receptor codes for odors. *Cell.*
970 96:713-723.

971 Mansourian S, Stensmyr MC. 2015. The chemical ecology of the fly. *Current Opinion in*
972 *Neurobiology* 34:95-102.

973 Mathew D, Martelli C, Kelley-Swift E, Brusalis C, Gershow M, Samuel ADT, Emonet T,
974 Carlson JR. 2013. Functional diversity among sensory receptors in a *Drosophila* olfactory
975 circuit. *Proc Natl Acad Sci USA.* 110:E2134-E2143.

976 McBride CS, Arguello JR. 2007. Five *Drosophila* genomes reveal nonneutral evolution and the
977 signature of host specialization in the chemoreceptor superfamily. *Genetics.* 177:1395-1416.

978 McKenzie SK, Fetter-Pruneda I, Ruta V, Kronauer DJC. 2016. Transcriptomics and
979 neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in
980 chemical communication. *Proc Natl Acad Sci USA.* 49:14091-14096.

981 McKenzie SK, Kronauer DJC. 2018. The genomic architecture and molecular evolution of ant
982 odorant receptors. *Genome Res.* 28:1757-1765.

983 Miller CH, Campbell P, Sheehan MJ. 2020. Distinct evolutionary trajectories of V1R clades
984 across mouse species. *BMC Evol Biol.* 20:99.

985 Miller SE, Bluher SE, Bell E, Cini A, Da Silva RC, De Souza AR, Gandia KM, Jandt J, Loope
986 K, Prato A et al. 2018. WASPnest: a worldwide assessment of social Polistine nesting behavior.
987 *Ecology*. 99:2405.

988 Miller SE, Legan AW, Flores ZA, Ng HY, Sheehan MJ. 2019. Strong, but incomplete, mate
989 choice discrimination between two closely related species of paper wasp. *Biol J Linn Soc.*
990 126:614-622.

991 Miller SE, Legan AW, Henshaw M, Ostevik KL, Samuk K, Uy FMK, Sheehan MJ. 2020.
992 Evolutionary dynamics of recent selection for enhanced social cognition. *Proc Natl Acad Sci*
993 *USA*. 117:3045-3052.

994 Münch D, Galizia CG. 2016. DoOR 2.0 - comprehensive mapping of *Drosophila melanogaster*
995 odorant responses. *Sci Rep.* 6:21841.

996 Nagawa F, Yoshihara S-i, Tsuboi A, Serizawa S, Itoh K, Sakano H. 2002. Genomic analysis of
997 the murine odorant receptor *MOR28* cluster: a possible role of gene conversion in maintaining
998 the olfactory map. *Gene.* 292:73-80.

999 Nei M. 2007. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci USA*.
1000 104:12235-12242.

1001 Niehuis O, Büllesbach J, Judson AK, Schmitt T, Gadau J. 2011. Genetics of cuticular
1002 hydrocarbon differences between males of the parasitoid wasps *Nasonia giraulti* and *Nasonia*
1003 *vitripennis*. *Heredity.* 107:61-70.

1004 Nozawa M, Nei M. 2007. Evolutionary dynamics of olfactory receptor genes in *Drosophila*
1005 species. *Proc Natl Acad Sci USA*. 104:7122-7127.

1006 Ohno S. 1970. Evolution by gene duplication. Berlin, Springer-Verlag.

1007 Oi CA, Oliveira RC, van Zweden JS, Mateus S, Millar JG, Nascimento FS, Wenseleers T. 2019.
1008 Do primitively eusocial wasps use queen pheromones to regulate reproduction? A case study of
1009 the paper wasp *Polistes satan*. *Front Ecol Evol.* 7:199.

1010 O'Neill KM. 2001. Solitary wasps: behavior and natural history. Cornell University Press, Ithaca,
1011 NY.

1012 Oxley PR, Ji L, Fetter-Pruneda I, Mckenzie SK, Li C, Hu H, Zhang G, Kronauer DJC. 2014. The
1013 genome of the clonal raider ant *Cerapachys biroi*. *Curr Biol.* 24:451-458.

1014 Pask GM, Slone JD, Millar JG, Das P, Moreira JA, Zhou X, Bello J, Berger SL, Bonasio R,
1015 Desplan C, et al. 2017. Specialized odorant receptors in social insects that detect cuticular
1016 hydrocarbon cues and candidate pheromones. *Nat Commun.* 8:297.

1017 Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, Asher CL, Jurkowski TP,
1018 Segonds-Pichon A, Bachman M, et al. 2015. Molecular signatures of plastic phenotypes in two
1019 eusocial insect species with simple societies. *Proc Natl Acad Sci USA*. 112:13970–13975.

1020 Pearson WR. 2013. Selecting the right similarity-scoring matrix. *Curr Protoc Bioinformatics.*
1021 43:3.5.1-3.5.9.

1022 Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A,
1023 Podsiadlowski L, Petersen M, Lanfear R, et al. 2017. Evolutionary History of the Hymenoptera.
1024 *Curr Biol.* 27:1013-1018.

1025 Post DC, Jeanne RL. 1983. Male reproductive behavior of the social wasp *Polistes fuscatus*
1026 (Hymenoptera: Vespidae). *Z Tierpsychol.* 62:157-171.

1027 Post DC, Jeanne RL. 1984. Recognition of conspecifics and sex by territorial males of the social
1028 wasp *Polistes fuscatus* (Hymenoptera: Vespidae). *Behaviour.* 91:78-92.

1029 R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for
1030 Statistical Computing, Vienna, Austria. https://www.R-project.org/.

1031 Raguso RA. 2008. Wake up and smell the roses: the ecology and evolution of floral scent. *Annu*
1032 *Rev Ecol Evol Syst.* 39:549-569.

1033 Ramdya P, Benton R. 2010. Evolving olfactory systems on the fly. *Trends Genet.* 26:307-316.

1034 Reed HC, Landolt PJ. 1990. Sex attraction in paper wasp, *Polistes exclamans* Viereck
1035 (Hymenoptera: Vespidae), in a wind tunnel. *J Chem Ecol.* 6(4):1277-1287.

1036 Reeve HK. 1991. Polistes. In: Ross KG, Matthews RD, editors. The social biology of wasps.
1037 Cornell University Press, Ithaca, NY. p. 99-148.

1038 Rehan SM, Toth AL. 2015. Climbing the social ladder: the molecular evolution of sociality.
1039 *Trends Ecol Evol.* 30:426-433.

1040 Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor
1041 gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci USA.* 100:14537-14542.

1042 Robertson HM, Wanner KW. 2006. The chemoreceptor superfamily in the honeybee *Apis*
1043 *mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res.* 16:1395-
1044 1403.

1045 Robertson HM, Gadau J, Wanner KW. 2010. The insect chemoreceptor superfamily of the
1046 parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol Biol.* 19:121-136.

1047 Ross KG, Matthews RW. 1991. The social biology of wasps. Cornell University Press, Ithaca,
1048 NY.

1049 Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J,
1050 Grimmelikhuijzen CJP, Hasselmann M, Lozier JD, et al. 2015. The genomes of two key
1051 bumblebee species with primitive eusocial organization. *Genome Biol.* 16:76

1052 Sánchez-Gracia A, FG Vieira, J Rozas. 2009. Molecular evolution of the major chemosensory
1053 gene families in insects. *Heredity.* 103:208-216.

1054 Sawyer SA. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol.* 6:526-538.

1055 Seeholzer LF, Seppo M, Stern DL, Ruta V. 2018. Evolution of a central neural circuit underlies
1056 *Drosophila* mate preferences. *Nature.* 559:564-569.

1057 Sheehan MJ, Botero CA, Hendry TA, Sedio BE, Jandt JM, Weiner S, Toth AL, Tibbetts EA.
1058 2015. Different axes of environmental variation explain the presence vs. extent of cooperative
1059 nest founding associations in *Polistes* paper wasps. *Ecol Lett.* 18:1057-1067.

1060 Singer TL, Camann MA, Espelie KE. 1992. Discriminant analysis of cuticular hydrocarbons of
1061 social wasp *Polistes exclamans* Viereck and surface hydrocarbons of its nest paper and pedicel. *J*
1062 *Chem Ecol.* 18:785-797.

1063 Singer TL. 1998. Roles of hydrocarbons in the recognition systems of insects. *Amer Zool.*
1064 38:394-405.

1065 Sledge MF, Boscaro F, Turillazzi S. 2001a. Cuticular hydrocarbons and reproductive status in
1066 the social wasp *Polistes dominulus*. *Behav Ecol Sociobiol.* 49:401-409.

1067 Sledge MF, Dani FR, Cervo R, Dapporto L, Turillazzi S. 2001b. Recognition of social parasites
1068 as nest-mates: adoption of colony-specific host cuticular odours by the paper wasp parasite
1069 *Polistes sulcifer*. *Proc R Soc B.* 268:2253-2260.

1070 Sledge MF, Trinca I, Massolo A, Boscaro F, Turillazzi S. 2004. Variation in cuticular
1071 hydrocarbon signatures, hormonal correlates and establishment of reproductive dominance in a
1072 polistine wasp. *J Insect Physiol.* 50:73-83.

1073 Slone JD, Pask GM, Ferguson ST, Millar JG, Berger SL, Reinberg D, Liebig J, Ray A, Zwiebel
1074 LJ. 2017. Functional characterization of odorant receptors in the ponerine ant, *Harpegnathos*
1075 *saltator*. *Proc Natl Acad Sci USA*. 114:8586-8591.

1076 Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E,
1077 Elsik CG, et al. 2011. Draft genome of the globally widespread and invasive Argentine ant
1078 (*Linepithema humile*). *Proc Natl Acad Sci USA.* 108:5673-5678.

1079 Smith CR, Smith CD, Robertson HM, Helmkampf M, Zimin A, Yandell M, Holt C, Hu H,
1080 Abouheif E, Benton R, et al. 2011. Draft genome of the red harvester ant *Pogonomyrmex*
1081 *barbatus*. *Proc Natl Acad Sci USA.* 108:5667-5672.

1082 Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. 2015. Less is more: an
1083 adaptive branch-site random effects model for efficient detection of episodic diversifying
1084 selection. *Genome Biol Evol.* 32:1342-1353.

1085 Sonnhammer ELL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting
1086 transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol.* 6:175-182.

1087 Stamatakis A. 2006. Phylogenetic models of rate heterogeneity: a high performance computing
1088 perspective. *Proceedings 20th IEEE International Parallel and Distributed Processing*
1089 *Symposium*, Rhodes Island, doi: 10.1109/IPDPS.2006.1639535.

1090 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1091 large phylogenies. *Bioinformatics.* 30:1312–1313.

1092 Standage DS, Berens AJ, Glastad KM, Severin AJ, Brendel VP, Toth AL. 2016. Genome,
1093 transcriptome and methylome sequencing of a primitively eusocial wasp reveal a greatly reduced
1094 DNA methylation system in a social insect. *Mol Ecol.* 25:1769-1784.

1095 Stensmyr MC, Dweck HKM, Farhan A, Ibba I, Strutz A, Mukunda L, Linz J, Grabe V, Steck K,
1096 Lavista-Llanos S, et al. 2012. A conserved dedicated olfactory circuit for detecting harmful
1097 microbes in *Drosophila*. *Cell.* 151:1345-1357.

1098 Sturgis SJ, Gordon DM. 2012. Nestmate recognition in ants (Hymenoptera: Formicidae): a
1099 review. *Myrmecol News.* 16:101-110.

1100 Toth AL, Tooker JF, Radhakrishnan S, Minard R, Henshaw MT, Grozinger CM. 2015. Shared
1101 genes related to aggression, rather than chemical communication, are associated with
1102 reproductive dominance in paper wasps (*Polistes metricus*). *BMC Genomics*. 15:75.

1103 Touhara K, Vosshall L. 2009. Sensing odorants and pheromones with chemosensory receptors.
1104 *Annu Rev Physiol.* 71:307-332.

1105 Trible W, Olivos-Cisneros L, McKenzie SK, Saragosti J, Chang N-C, Matthews BJ, Oxley PR,
1106 Kronauer DJ. 2017. Orco mutagenesis causes loss of antennal lobe glomeruli and impaired social
1107 behavior in ants. *Cell.* 170:727-735.

1108 Tsutsui ND. 2013. Dissecting ant recognition systems in the age of genomics. *Biol Lett.*
1109 9:20130416.

1110 van Zweden JS, d'Ettorre P. 2010. Nestmate recognition in social insects and the role of
1111 hydrocarbons. In: Blomquist GJ, Bagnères A-G, editors. Insect Hydrocarbons. Cambridge
1112 University Press. p. 222-243.

1113 Vernot B, Stolzer M, Goldman A, Durand D. 2008. Reconciliation with non-binary species trees.
1114 *J Comput Biol.* 15:981-1006.

1115 Vogt RG, Prestwich GD, Lerner MR. 1991. Odorant-binding-protein subfamilies associate with
1116 distinct classes of olfactory receptor neurons in insects. *J Neurobiol.* 22:74-84.

1117 Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM. 2007. A
1118 honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid. *Proc Natl Acad Sci*
1119 *USA.* 104:14383-14388.

1120 Wicher D. 2015. Olfactory signaling in insects. In: Glatz R, editor. Molecular Basis of Olfaction.
1121 Elsevier. p. 37-54.

1122 Yan H, Opachaloemphan C, Mancini G, Yang H, Gallitto M, Mlejnek J, Leibholz A, Haight K,
1123 Ghaninia M, Huo L. 2017. An engineered orco mutation produces aberrant social behavior and
1124 defective neural development in ants. *Cell.* 170:736-747.

1125 Yan H, Jafari S, Pask G, Zhou X, Reinberg D, Desplan C. 2020. Evolution, developmental
1126 expression and function of odorant receptors in insects. *J Exp Biol.* 223:jeb208215.

1127 Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under
1128 realistic evolutionary models. *Mol Biol Evol.* 17:32-43.

1129 Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586-
1130 1591.

1131 Yapici N, Zimmer M, Domingos AI. 2014. Cellular and molecular basis of decision-making.
1132 *EMBO Rep.* 15:1023-1035.

1133 Yu G, Smith D, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an R package for visualization and
1134 annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol.*
1135 *Evol.* 8:28-36.

1136 Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, Ray A, Reinberg D, Zwiebel LJ. 2012.
1137 Phylogenetic and Transcriptomic Analysis of Chemosensory Receptors in a Pair of Divergent
1138 Ant Species Reveals Sex-Specific Signatures of Odor Coding. *PLoS Genet.* 8, e1002930.

1139 Zhou X, Rokas A, Berger SL, Liebig J, Ray A, Zwiebel LJ. 2015. Chemoreceptor Evolution in
1140 Hymenoptera and Its Implications for the Evolution of Eusociality. *Genome Biol Evol.* 7:2407-
1141 2416.

1142 Zube C, Rössler W. 2008. Caste- and sex-specific adaptations within the olfactory pathway in
1143 the brain of the ant *Camponotus floridanus*. *Arthropod Struct Dev.* 37:469-479.