bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title:

The evolutionary dynamics of the odorant receptor gene family in corbiculate bees

Authors:
Philipp Brand[1,2] and Santiago R. Ramírez[1]

Author affiliations:
[1]Department for Evolution and Ecology, Center for Population Biology, University of California Davis, One Shields Avenue, 95616 Davis, USA
[2]Population Biology Graduate Group, University of California Davis, One Shields Avenue, 95616 Davis, USA

Corresponding author:
Philipp Brand, Department for Evolution and Ecology, Center for Population Biology, University of California Davis, One Shields Avenue, 95616 Davis, USA, +1 530-752-7614, pbrand@ucdavis.edu

Keywords:
Odorant receptors, Chemosensory genes, birth-death process, Genomic architecture, Positive selection, Corbiculate bees, Honey bees, Bumble bees, Stingless bees, Orchid bees

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Insects rely on chemical information to locate food, choose mates, and detect potential predators. It has been hypothesized that adaptive changes of the olfactory system facilitated the specialization and diversification of numerous insect lineages. In fact, modifications of life history strategies often occur in parallel with evolutionary changes in Odorant Receptor (OR) genes. However, the molecular mechanisms underlying the evolution of insect OR genes remain poorly understood. Corbiculate bees include economically and ecologically important bee pollinators that exhibit varying degrees of social organization, diverse sex pheromone systems, and manifold associations with floral resources. We investigated the molecular mechanisms governing the evolution of the OR gene family in corbiculate bees. Our results indicate that the genomic organization of the OR gene family has remained highly conserved for approximately 80 million years, despite exhibiting major changes in repertoire size among bee lineages. Moreover, the evolution of OR genes appears to be driven by lineage-specific gene duplications in few genomic regions harboring large numbers of OR genes. A selection analysis revealed that OR genes evolve under positive selection, with the strongest signals detected in recently duplicated copies. Our results indicate that chromosomal translocations had a minimal impact on OR evolution, and instead show that local molecular mechanisms drive OR repertoire size. Our results provide empirical support to the longstanding hypothesis that positive selection shaped the diversification of the OR gene family. Our results shed new light on the molecular mechanisms underlying the evolution of olfaction in insects and gene families in general.

Significance Statement

Despite playing a central role in the diversification of insect lineages, the evolution of insect olfactory systems remains poorly understood. Here, we analyze the Odorant Receptor (OR) repertoire across the corbiculate bees—a lineage of important insect pollinators—to test the hypothesis that genomic re-organization and positive selection shaped the evolution of the OR gene family. We show that locally acting molecular mechanisms are the driving force of OR gene family evolution in corbiculate bees. Moreover, we provide empirical support for the longstanding hypothesis that positive selection impacts the evolution of recently duplicated insect OR genes. Our results augment our understanding of the molecular mechanisms that govern the evolution of olfactory systems in insects and gene families in general.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Animals have evolved sophisticated sensory systems that can detect and discriminate airborne volatile chemicals and provide precise information about food, enemies, and mating partners (1). Insects detect olfactory signals and cues via olfactory sensory neurons located in the antenna. Each olfactory sensory neuron expresses Odorant Receptor (OR) proteins on the cell membrane (2-4). It is the interaction of volatile odorant molecules with these OR proteins that initiates the signal transduction and transmission toward the olfactory centers in the insect brain (2, 4). Insect genomes are endowed with a diverse array of functional OR genes, each of which encodes a unique odorant receptor protein tuned to a specific set of odors (5-7). The entire repertoire of OR genes expressed in the olfactory organs determines the spectrum of chemical volatiles that an insect species can detect.

Comparative analyses of insect genomes have revealed that the number of OR genes and divergence among them can vary widely between species. Insect genomes may contain as few as 10 OR genes (head lice, (8)) and as many as 300 OR genes (ants, (9, 10)). In fact, the expansion and contraction of the OR gene family has been linked to shifts in the sensory ability of some insect lineages. This supports the idea that adaptation to novel food resources (11-13) or modifications in the pheromone communication system (14) can be mediated by peripheral changes in the OR gene repertoire. In addition, changes in the amino acid sequence of existing OR genes have been shown to correspond to adaptive shifts in sensory tuning (15, 16). However, we lack a clear understanding of the specific genetic and molecular mechanisms that drive the diversification of the insect OR gene family (17, 18).

Gene family evolution is determined by multiple molecular mechanisms including genomic drift, natural selection, and chromosomal rearrangements. Birth-death processes can directly impact the number of genes within a gene family (19). For example, genomic drift may generate new gene copies through gene duplication (gene 'birth') while existing copies can be purged via pseudogenization or deletion (gene 'death'). Novel mutations resulting in gene duplication or gene loss may be subsequently fixed through neutral genetic drift or positive selection (17, 19, 20). Although OR gene repertoire sizes have been analyzed extensively in various insect lineages, the relative contribution of positive selection and neutral processes in a birth-death evolutionary framework remains uncertain (17, 18).

The genomic organization and architecture of gene families may contain signatures of past evolutionary forces that shaped their diversification. For instance birth-death processes often produce tandem arrays of OR genes, which are groups of ancestrally duplicated genes located in close physical proximity along the genome (19, 21). Chromosomal rearrangements and transposition, on the other hand, can lead to inter-chromosomal translocations of OR genes (22, 23), and it has been hypothesized that changes in the location of OR genes along the genome can influence their evolution (18, 23, 24). Although both tandem duplications and

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

chromosomal rearrangements may influence the evolution of the OR gene family, the relative contribution of each process remains unknown.

Like most insects, bees rely on olfactory information to regulate a wide array of behaviors, including the location of food sources and nesting materials, the identification of mating partners, and social interactions with other colony members in social species. Corbiculate bees encompass a group of ecologically and economically important bee lineages, including honey bees, bumble bees, stingless bees, and orchid bees. Honey bees are estimated to pollinate one-third of agricultural crops (25), and stingless bees and orchid bees are the major pollinators of numerous tropical flowering plant species (26, 27). Corbiculate bees have evolved a variety of phenotypic and behavioral traits that require specialized olfactory functions. These include the evolution of an obligate cooperative social lifestyle (eusociality), which has resulted in the evolution of specialized pheromone communication systems in honey bees, bumble bees and stingless bees (28-31). In addition, orchid bees evolved a unique pheromone communication system in which male bees concoct species-specific perfume bouquets from scents collected from flowers and other sources to subsequently use during courtship display (32-34). However, despite the central role that olfactory encoding plays in the biology of corbiculate bees, we lack a clear understanding of the evolution of the OR gene family in this diverse group of bees.

Here we analyzed the evolution of the OR gene family in a set of species including the four corbiculate bee clades, spanning 80 million years (my) of evolution. We tested whether the evolution of the OR gene family has been influenced by changes in the genomic organization of genes, as hypothesized based on patterns observed in *Drosophila* and mammalian genomes (18, 23, 24). Second, we tested the hypothesis that birth-death processes, in combination with positive selective pressures, promoted the origin of novel OR genes. Finally, we estimated the relative importance of positive selection in shaping OR sequence divergence, a mechanism that has been hypothesized to drive sensory adaptation in insect lineages (15, 35, 36).

Results

The Evolution of Corbiculate Bee Odorant Receptors is Highly Dynamic. To analyze the evolutionary dynamics of odorant receptor (OR) gene repertoires in corbiculate bees, we reconstructed the phylogenetic relationships among the entire OR gene family based on both whole-genome sequences and transcriptome data obtained from ten bee species (Fig. 1). An initial comparison of OR annotation based on genomic and transcriptomic data indicated no differences in annotation accuracy, but as expected, the number of genes tended to be lower for transcriptome data (Supplemental Text). Our analysis included published whole-genome sequence data from at least one species per corbiculate bee tribe (21, 37-39). In addition, we sequenced the genome of one orchid bee (*Euglossa dilemma;* (40)) and the antennal transcriptomes of five additional orchid bee species. The

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

resulting dataset, which included three orchid bee species pairs, allowed us to analyze the evolution of the OR gene family at multiple taxonomic levels over a broad range of divergence times (0.15 to 80mya; Fig. 1). We compared the evolution of the OR gene family, the largest and most dynamic chemosensory gene family in insects, to the evolution of four additional chemosensory gene families, including the Gustatory Receptors (GRs), Ionotropic Receptors (IRs), Odorant Binding Proteins (OBPs), and Chemosensory Proteins (CSPs). We used genomic data only to derive evolutionary patterns of gene family size since transcriptome data often yield incomplete gene sets (Fig. 1).

Gene repertoire size varied substantially among bee lineages, ranging from 142 OR genes in the orchid bee *Eufriesea mexicana* to 196 OR genes in the stingless bee *Melipona quadrifasciata* (Fig. 1). The identified ORs were organized in 115 homologous OR ortho-groups consisting of orthologs and/or inparalogs (*i.e.* lineage-specific expansions, Fig. 2, Fig. S1). Of these ortho-groups, 24 did not contain duplications but instead consisted of simple 1:1 orthologous genes that were present in three or more bee species (Fig. 2b). The remaining 91 ortho-groups included species-specific, genus-specific, and tribe-specific duplications or larger expansions (Fig. 2c). In contrast to the other four gene families analyzed, the repertoire size of ORs is highly variable among corbiculate bee lineages. The majority of genes in the non-OR gene families revealed simple 1:1 orthology with few lineage-specific duplications resulting in similar repertoire sizes among bee lineages (Fig. 1; Fig. S2). An exception to this was an observed increase in gene family size in honey bee OBPs as well as bumble bee GRs (Supplemental Text). Among all chemosensory genes, we identified 32 gene duplications and large expansions in the honey bees, 36 in the bumble bees, 26 in the stingless bees, and 18 in the orchid bees, most of which were observed in the OR gene family (89%, Tab. S1). These results show that the evolution of the corbiculate bee OR gene family repertoire is characterized by highly dynamic lineage-specific gene expansions.

The Genomic Organization of the OR Gene Family is Highly Conserved in Corbiculate Bees. Next, we investigated the genomic organization of OR genes across the corbiculate bees. Our analysis revealed that OR genes are widely distributed throughout the five bee genomes we analyzed. This supports previous studies where OR genes were detected on scaffolds assigned to almost all chromosomes in those species with available linkage groups (*i.e.* honey bees and bumble bees; (21, 37, 38)). Congruent with a birth-death process for OR gene evolution, the majority of genes were clustered in large tandem arrays and only few ORs were found in isolated genomic regions as singletons. We tested whether the chromosomal location of OR genes is conserved among corbiculate bees by comparing the 200 kb flanking regions of all OR singletons and tandem arrays among species using an “all-against-all” reciprocal blastn approach.

With the exception of a small number of ancestral translocations (Supplementary Text), we found that the flanking regions and thus the physical chromosomal locations of most OR genes were strictly homologous in all corbiculate bee species

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Tab. S2). Furthermore, all orthologous OR genes found in two or more species were located in homologous chromosomal regions. Paralogous genes within species-specific, genus-specific, and tribe-specific duplications and larger OR sublineage expansions were always located in the same tandem array. Similarly, ORs located within the same genomic region were generally more closely related to each other than ORs located in different genomic regions that formed highly supported monophyletic gene clades (Fig. 2a). On the basis of this strong correlation between phylogenetic relationship and genomic location, we defined a total of 25 well-supported subfamilies in the OR gene family that largely corresponded to groups previously identified in ants (Fig. 2a; (41, 42)). In addition to the high conservation of genomic locations of OR genes and tandem arrays, we found that the organization of genes within tandem arrays (*i.e.* gene synteny) is also highly conserved (Supplementary Text; Fig. S3). Our analysis identified similar patterns of conservation in the genomic organization of other chemosensory gene families, including Gustatory Receptors, Ionotropic Receptors, Odorant Binding Proteins, and Chemosensory Genes (Dataset S1).

The Evolution of OR Repertoire Sizes is Governed by Dynamic Evolutionary Changes in Large Subfamilies. Our analysis revealed that although the genomic organization of OR tandem arrays is highly conserved across corbiculate bee lineages, OR evolution is highly dynamic, with multiple taxon-specific OR subfamily expansions. In order to investigate potential localized molecular mechanisms that shape the evolution of the OR gene family, we analyzed the divergence of the 25 OR subfamilies (see above). We found that variation in subfamily size among species was positively correlated with mean subfamily size (Fig. 3a, Pearson coefficient $r=0.85$, $p=0$; Orchid bees only: $r=0.6$, $p=0.0013$), a common pattern in gene family evolution (Dennis et al. 2017). This result suggests that larger OR subfamilies have a higher gene turnover than smaller gene subfamilies. Accordingly, all subfamilies with one or two genes per species consisted of simple 1:1 orthologous genes across the corbiculate bee genomes and transcriptomes (Fig. S1; Tab. S3).

Three of the OR subfamilies that exhibited high interspecific variation in OR repertoire size had an expansion in a single bee species. The G04A subfamily was expanded in the honey bee *A. mellifera* (Fig. 2c), the G09A subfamily was expanded in the orchid bee *Ef. mexicana,* and the G09C subfamily was expanded in the stingless bee *M. quadrifasciata* (Fig. 3b, Tab. S3). Several subfamily expansions were observed in two or more species. Although most expansions are not shared by multiple bee species with similar life history traits, the G02A subfamily was expanded in all obligate eusocial bee species we included in our analysis (honey bee, bumble bee and stingless bee; Fig. 3b). Conversely, ORs in the subfamily GUnA and G12A were expanded in orchid bees (Fig. 3b, Tab. S3), which are known for their unique perfume collecting behavior. The expansion we observed in the subfamilies G02A and GUnA correspond to well-supported lineage-specific clusters of paralogs in our phylogenetic analysis (Fig. S1), suggesting that these subfamilies are independent expansions leading to elevated OR numbers. In contrast, more than half of the orchid bee OR expansion in subfamily G12A was shared between the

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

orchid bee species included in our analysis (*i.e.* orthologous copies were present), suggesting that the expansion was already present in their common ancestor.

Positive Selective Pressures Drive the Evolution of Both Paralogous and Orthologous OR Genes. To investigate the selective pressures that shaped the evolution of the corbiculate bee OR gene family, we directly compared the evolution of ORs with the evolutionary patterns of the remaining major chemosensory gene families, including the GRs, IRs, OBPs, and CSPs. We classified all annotated chemosensory genes into orthologous groups consisting of at least four orthologs and/or inparalogs based on our phylogenetic analysis (see Fig. 2b-c for examples of orthologous groups). The resulting 159 orthologous groups (115 OR, 13 GR, 9 IR, 15 OBP, and 7 CSP orthologous groups) contained a total of 1588 chemosensory genes (Dataset S1). We tested each of the orthologous groups for the presence of episodic positive selection using branch-site model based dN/dS estimates (Smith et al. 2015). We identified 54 chemosensory genes (2.1% of all terminal branches) and 57 deeper branches (6.1% of 941 deep branches) with signatures of positive selection under a 5% false discovery rate (FDR). This corresponds to 4.4% of all branches tested. We found no correlation between divergence time and positive selective pressures (Supplemental Text; Tab. S4). Furthermore, there was no difference in the number of branches under positive selection between the four corbiculate bee tribes (Tab. 1). However, a comparison across gene families revealed a higher proportion of branches under positive selection in the GRs relative to any other gene family in all bee lineages, supporting the hypothesis that GRs exhibit elevated signatures of positive selection across many insect groups ($p<0.0001$; Fisher's Exact Test; Supplementary Text) (12).

Those branches that exhibited signatures of positive selection were distributed throughout the entire OR gene tree (Fig. S1). Interestingly, only 32% of all branches that we identified to evolve under positive selection belonged to lineage-specific duplications or larger gene expansion events (Tab. S5). Thus, the majority of branches under positive selection correspond to orthologous rather than paralogous gene copies. This pattern held true when we considered only those five species with genomic data. However, when normalized by the total number of the two branch classes (duplication and non-duplication branches, respectively), we found that a higher proportion of all tested duplication branches were under positive selection in comparison to non-duplicated branches (6% vs. 3.9%, Fisher's Exact Test, $p< 0.05$, Tab. S5). Together, these results suggest that positive selective pressures play an important role in the evolution and divergence of duplicated OR genes.

In addition, we observed that 10 of the 11 OR subfamilies without any branches under positive selection included simple (single-copy) 1:1 orthologs only (Subfamilies G01A, G01B, G02B, G05A, G09D, G14B, G15A, G15B, GUnB, ORco; Fig. S1). This observation suggests the existence of a core set of OR genes with conserved chemosensory functions across all corbiculate bees.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Our analysis focused on documenting signatures of positive selection as well as changes in the genomic organization of genes in the OR gene family analyzing 1394 ORs spanning multiple evolutionary time scales across the corbiculate bees. Our analysis revealed that the corbiculate bee OR gene family is highly dynamic (Fig. 1, Fig. 2, Fig. S1), a pattern consistent with that observed in other insect lineages. However, the genomic organization we identified in corbiculate bees differs from that previously reported in other insect groups. First, we found that the genomic organization of OR genes is highly conserved between corbiculate bee lineages, despite sharing a common ancestor 80 million years ago (Tab. S2). Second, OR gene repertoire sizes vary significantly between bee species due to lineage-specific expansions in a few large tandem arrays, some of which correlate with the evolution of obligate eusociality and pheromone communication systems (Fig. 3; Fig. S2). These findings suggest that the highly dynamic evolutionary history of the OR gene family is independent of changes in genomic organization and instead more likely influenced by locally acting molecular mechanisms. Furthermore, we found evidence for elevated rates of positive selection on duplicate ORs (Tab. 1; Tab S6), supporting the main prediction from the birth-death model of gene family evolution. Finally, we identified a fraction of bee ORs under positive selection that have not undergone previous duplication, highlighting the importance of positive selective pressures in OR sequence divergence, possibly in response to ecological adaptation.

Locally Acting Molecular Mechanisms are the Driving Force of Chemosensory Evolution. Our phylogenetic and genome annotation analysis revealed a high degree of genomic landscape conservation in the OR gene family across the corbiculate bees. This pattern was accompanied by a fast diversification rate of OR genes, indicating that genomic rearrangements had a relatively weak influence relatively to the stronger influence of local molecular mechanisms. The high degree of conservation in the overall genomic organization holds true even between corbiculate bee lineages that shared a common ancestor 80 million years ago. This is an unprecedented level of genomic conservatism in OR organization, and is particularly surprising given the diverse factors known to affect genome composition and genome architecture. In fact, the bee lineages we examined exhibit more than two-fold variation in chromosome number (9 in stingless bees to 20 in orchid bees; (43)), an order of magnitude of variation in genome size (ranging from 0.25GB in honey bees to 3GB in orchid bees; (37-40)), as well as exceptionally high recombination rates (44-46). Moreover, the high level of conservation stands in sharp contrast to known evolutionary mechanisms that affect genomic organization of OR genes in drosophilid flies, the only other insect group for which the evolution of chemosensory genes has been examined in comparable detail. The genomes of drosophilid flies exhibit complex genome rearrangements leading to a high degree of ortholog OR gene translocation within evolutionary time scales similar to those spanning the bee species we analyzed (22, 47, 48). A similarly dynamic pattern of OR tandem array architecture was previously observed in mammalian genomes (49). These observations lead to the hypothesis that genomic rearrangements are

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

important in the evolution of chemosensory receptors across vertebrate and invertebrate lineages (18, 23, 24). In contrast, our results show that the evolution of the OR gene family in corbiculate bees has resulted in highly divergent gene repertoires without undergoing genome rearrangements.

Our results revealed no correlation between the evolution of genomic organization and genome size. In fact, genome size of most corbiculate bees resembles those reported for drosophilid flies (37, 50). Instead, a recent comparison of the genomes of honey bees and bumble bees revealed high conservation of genome synteny between these species (44), indicating that conserved genomic architecture of OR tandem arrays may result from a low precedence of genome rearrangements in corbiculate bees. Indeed, our analysis revealed that synteny among OR genes in orchid bees is almost perfectly conserved even within highly dynamic genomic areas that are dominated by gene duplication (Supplementary Text). Despite the high level of conservation in the genomic organization reported here, we find a high degree of OR gene divergence and turnover among corbiculate bees. This result suggests that chemosensory gene evolution in corbiculate bees, and perhaps insects in general, is shaped by local rather than global mechanisms of genome evolution.

Large Tandem Arrays are Hotspots for the Evolution of Novel ORs. The high rate of variation we observed in OR repertoire size among lineages, along with the prevalence of lineage-specific gene expansions in large OR tandem arrays, suggests that large gene subfamilies play an important role in the evolution of novel OR genes in corbiculate bees. Insect OR repertoires are highly dynamic between lineages in comparison to other chemosensory gene families (18, 51). Our results confirm this pattern for corbiculate bees. In contrast to previous studies, however, we find that these expansions are driven by lineage-specific expansions within large tandem arrays harboring entire OR subfamilies. This result is congruent with a recent phylogenetic analysis that showed that gene duplication and pseudogenization events of OR genes in Hymenoptera are often found in large subfamilies (52). The increased gene turnover rate could be influenced by elevated rates of genome replication errors resulting from non-allelic homology of a large number of closely related paralogous loci in close proximity within large tandem arrays (53-55). Nevertheless, other factors such as transposable element activity and the timing of DNA replication can influence segmental duplication rates (53, 56). Independent of the underlying genomic mechanism, our results suggest that large OR tandem arrays are a major source of novelty in chemosensory receptors in corbiculate bees and potentially other Hymenoptera. Accordingly, large tandem arrays are likely an important source of functional novelty associated with the detection of novel molecules or blends of chemical compounds. Hence, it is possible that the gene expansions we identified coincide with key evolutionary innovations such as complex social chemical communication in the obligate eusocial honey bees, bumble bees and stingless bees or perfume-collecting behavior in orchid bees.

It has been hypothesized that OR gene family size is correlated with the degree of complexity in the chemical communication system displayed by eusocial

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Hymenoptera (57). Although our data do not support this general trend, we find a positive correlation between OR gene subfamily size and the presence of obligate eusociality. For instance, the OR subfamily G02A is expanded in all three obligate eusocial bee species analyzed, relatively to the smaller size exhibited by the solitary to weakly social orchid bees. Moreover, the OR subfamily G02A is expanded in ants but not in the facultative social halictid bee *Lasioglossum albipes*, which exhibits subfamily G02A gene numbers comparable to orchid bees (Subfamily L in (42, 52)). Thus, it is possible that ORs in subfamily G02A play a central role in the olfactory detection and processing of pheromones involved in social behavior. Similarly, an expansion of the '9-exon' subfamily (here subfamilies G04A, G05B, G11A, G11B, and G12A; Fig. 2a; Fig. 3b; Fig S1) was hypothesized to be important in the evolution of eusociality due to a putative role of '9-exon' ORs in the detection of social pheromones (namely cuticular hydrocarbons, (9, 10, 58)). We find an expansion of the '9-exon' subfamily in the eusocial honey bees and bumble bees but not eusocial stingless bees. This observation contrasts with the previous suggestion that the expansion of this OR subfamily was concomitant with the evolution of eusociality in Hymenoptera. However, it is possible that ORs involved in social pheromone communication evolved independently in different Hymenoptera lineages. Future studies should investigate the possibility that olfactory adaptations to social behavior evolved through multiple analogous routes in social Hymenoptera (39, 59).

In addition, we found two OR subfamilies that exhibit an expansion in the orchid bee lineage (namely GUnA and G12A). While large expansions were in general species-specific clusters of paralogous ORs (*e.g.* Fig. 2c; Fig. S1), more than half of the ORs in the orchid bee expansion of subfamily G12A were orthologous ORs found in all the seven species included in our analysis. This shared expansion may correspond to an orchid bee specific olfactory specialization. Since the only known olfactory behavior clearly distinguishing orchid bees from all other corbiculate bees is male perfume-collecting behavior (34, 60), it is possible that the G12A subfamily is involved in the detection and encoding of these exogenous compounds. Functional analyses are required to investigate the olfactory coding mediated by members of this OR subfamily.

Selective Patterns of Corbiculate Bee Chemosensory Genes Support the Birth-Death Model of Gene Family Evolution. It has been proposed that chemosensory gene families evolve under a birth-death process where duplications of existing genes give rise to new gene copies (gene 'birth') whereas pseudogenization purges genes from the genome (gene 'death'; (19)). The birth-death model predicts a relaxation of purifying selective pressures after duplication allowing subsequent diversification of one or both copies through positive selection for newly accumulated beneficial mutations. Accordingly, the evolution of highly dynamic chemosensory gene families is likely driven by the action of positive selection. Nevertheless, while previous studies detected a high impact of gene duplication and loss on chemosensory gene family evolution, there is very little evidence for

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

diversification through positive selection (12, 22, 52, 61-63), even in large lineage-specific gene family expansions (42).

Our analysis provides empirical evidence congruent with the selective pressures predicted by the birth-death model of chemosensory gene family evolution. We show that positive selective pressures are stronger on duplicated chemosensory genes than in orthologous chemosensory genes across corbiculate bees. The observed higher proportion of branches under positive selection is in contrast to previous analyses of insect chemosensory genes and might be partly explained by the former use of less powerful algorithms, reliance on a priori expectations, incomplete testing, or differences in taxon sampling (64). It is thus unclear whether the disparity between studies reflects technical differences or differences in chemosensory gene evolution between different groups of insects. Regardless, our findings suggest a non-negligible impact of positive selective pressures on chemosensory gene evolution in corbiculate bees.

Diversification of Orthologous ORs through Positive Selection Drives Divergence in Chemosensory Systems. Our selection analysis indicates that positive selection is involved in the divergence of orthologous ORs, providing an evolutionary mechanism for the diversification of chemosensory genes independent of gene duplication. We found 54 ortholog chemosensory genes and deeper branches that exhibit signatures of positive selection. Moreover, our analysis using three pairs of sympatric species revealed the presence of orthologous chemosensory genes evolving under positive selection irrespective of divergence times (0.15 - 13 my). This result suggests that positive selection is an important evolutionary force driving the diversification of orthologous chemosensory genes in corbiculate bees. Similarly, Leary et al. (2012) (15) showed that positive selection of specific amino acid substitutions in OR orthologs led to novel receptor functions through changes in ligand binding affinities in a lineage of moths. Under this model, orthologous OR genes undergo functional divergence driven by positive selection in the absence of a duplication event. While OR gene duplications are more common in large subfamilies in corbiculate bees, this mechanism might be important for the evolution of interspecific functional divergence in smaller subfamilies. Thus, our results suggest that positive selection is not only important in driving divergence between OR duplicates within taxa, but also orthologous ORs between taxa. It is also possible that a combination of these two mechanisms drives the evolution of chemosensory genes in other insect lineages.

Evidence for ORs with Core Chemosensory Function in Corbiculate Bees. We detected the existence of multiple highly conserved OR lineages that appear to be subject to strong purifying selection. This observation supports the existence of a core set of OR genes with conserved functions across all corbiculate bees and perhaps other more distantly related bee lineages. We detected 10 OR subfamilies and 14 orthologous groups within larger subfamilies that solely consisted of 1:1 orthologs. While most OR subfamilies and orthologous groups were characterized by taxon-specific duplications and divergence driven by positive selective pressures,

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

our results suggest that these OR lineages are highly conserved throughout at least 80 million years of evolution. This observation implies that corbiculate bee genomes contain a core set of olfactory genes with conserved chemosensory functions. In fact, high conservation of OR genes is typical for some species with highly similar natural history (12, 47, 65). Detailed functional analysis of these core OR genes may elucidate the role of conserved olfactory functions in corbiculate bees of ecological importance.

Conclusion

Odorant receptors (ORs) have been hypothesized to play an important role in chemosensory adaptation, for example by enabling the evolution of novel pheromone detection systems associated with social communication or mate attraction (1, 66). However, little is known about the molecular mechanisms driving the evolution of insect OR gene families. Our analysis of the OR gene family in an important lineage of insect pollinators, the corbiculate bees, elucidated the role of positive selective pressures and genomic organization on the evolution of the OR gene family.

We show that OR gene family size is highly dynamic between corbiculate bee species. In contrast, the underlying genomic organization of OR genes is highly conserved among lineages over at least 80 million years. Consistent with these findings, ORs are arranged in tandem arrays located in close proximity along the genome. Moreover, the genomic location correlates with the phylogenetic relationships among genes that fall in 25 well-supported OR subfamilies. We find that a few large subfamilies contain a considerable amount of lineage-specific gene duplications, explaining most of the variation in gene family size. Further, we show that ORs evolve under strong positive selective pressures, with a higher impact on duplicated genes. Our study suggests that locally acting molecular mechanisms are the driving force of OR gene family evolution in corbiculate bees. Moreover, we provide empirical support for the longstanding hypothesis that positive selection is an important mechanism in the evolution of duplicated OR genes in insects. These results elucidate the OR gene family evolution in corbiculate bees and extend the current understanding of OR gene family evolution to other insect groups.

Methods

Sampling and Sequencing

Sampling and sequencing procedures followed Brand et al. 2015. Briefly, males of the four orchid bee species *Euglossa flammea*, *Eg. imperialis*, *Eulaema bombiformis* and *El. meriana* were sampled at the La Gamba Field Station in Costa Rica using different chemical baits (67). Antennal transcriptomes for each species were based on pools of the antennae of 20 individuals for the *Euglossa* species and 5 individuals for the *Eulaema* species. Bees were chilled on ice and the antennae of each torpid male were dissected using sterile forceps and transported either on liquid nitrogen

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

or RNA-later (Thermo Fisher Scientific, Waltham, MA) and kept frozen until RNA-extraction.

Total-RNA was extracted and quantified as described (35). Barcoded cDNA libraries were prepared using the NEBnext Ultra RNA Library Prep kit for Illumina (New England Biolabs, Ipswich, MA) following the manufacturers protocol and sequenced in two pools of two species on one HiSeq 2500 (*Euglossa*) and one HiSeq 4000 (*Eulaema*) lane for 2x100bp cycles. (Raw sequence reads are available at the NCBI Sequence Read Archive XXX).

De novo *transcriptome assembly and transcript recovery.*

Identical raw reads were merged using fastuniq v1.1 (68) and subsequently trimmed by trim_galore v0.3.7 (Babraham Bioinformatics) if sequencing primers or low-quality bases (Phred-score <20) were detected. The pre-processed reads of each species were used for *de novo* transcriptome assembly following the meta-assembly approach developed in Brand et al. 2015 using the Trinity (69, 70) and SOAPdenovo-trans v2 (71) assemblers. Briefly, each of the two assemblers was run with multiple parameter sets controlling assembly stringency resulting in 25 assemblies per species. Trinity settings were set as described (35), whereas SOAPdenovo-trans parameters –e and –d were set to 0, 2, 3, or 5 and 1, 3, 5, or 7, respectively and run in all possible combinations. While Trinity was run with the default single-k-mer mode, SOAPdenovo-trans was run in multi-k-mer mode with nine different k-values (25, 31, 37, 43, 49, 59, 69, 79 and 89).

Annotation of chemosensory genes.

Chemosensory genes of five gene families (ORs, IRs, GRs, OBPs, and CSPs) were annotated in the newly sequenced antennal transcriptomes of the orchid bee species *Eg. flammea*, *Eg. imperialis*, *El. bombiformis*, and *El. meriana*, as well as the publicly available genomes of the orchid bee species *Eg. dilemma* (40) and *Eufriesea mexicana* (39), and the stingless bee *Melipona quadrifasciata* (39). Additionally, the CSP gene family was annotated for the bumble bee *Bombus terrestris* (38). All four *de novo* transcriptomes were annotated using a homology-based iterative tBLASTn approach as previously described (35). In order to annotate the genomes, tBLASTn was used to detect scaffolds harboring chemosensory genes with an e-value cutoff of 10e-6 based on gene family specific query libraries comprised of published bee, wasp, ant and dipteran OR, IR, GR, CSP and OBP sequences (21, 35, 38, 41, 51, 72, 73). Subsequently, the queries were used to predict exon-intron boundaries with exonerate (74) in a scaffold specific manner. All annotations were curated manually and corrected if needed. All annotations can be found in Dataset S1.

Phylogenetic reconstruction of chemosensory gene families.

For each chemosensory gene family we estimated a maximum likelihood (ML) gene-tree using RaxML (75) to infer the genealogical histories of the candidate corbiculate gene family members. On that account, gene family specific protein sequence alignments were constructed using MAFFT (76, 77) applying the L-INS-I algorithm with the --maxiterate option set to 1000 (78). Based on the resulting

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

alignments ML trees were inferred under a JTT + G substitution model for each family with 20 independent ML searches and 100 bootstrap replicates.

Analysis of the genomic organization of chemosensory gene families.

Based on genomic annotations of *Eg. dilemma*, *Ef. mexicana*, *M. quadrifasciata* (this study), *A. mellifera* (21), and *B. terrestris* (38), we analyzed the evolutionary history of genomic structure of chemosensory gene families within the corbiculate bees. Therefore, we applied a reciprocal blast approach that compares the genomic location of all chemosensory gene family tandem arrays and singletons to the genomes of all other species. We required both 200kb flanking regions to be the reciprocal best blast hit (79) in order to call homology of the genomic regions. In case annotations were located less than 200kb away from the end of a scaffold, we used the maximum flanking region length possible, which never corresponded to less than 100kb. The location of ORs in the resulting genomic regions in combination with the gene family phylogeny was used to classify all ORs into subfamilies. Subfamilies were named after the linkage group they were located on in the honey bee genome. Letters were used to allow for the unique indication of subfamilies located on the same linkage group (*e.g.* G12B is the second subfamily found to be located on linkage group 12 of the honey bee). In case a gene family was located on a honey bee scaffold that could not be assigned to a linkage group, it was named 'Un', in short for 'unassigned'.

In a second approach, we used the OR gene family phylogeny to infer the evolutionary histories of micro-genomic synteny within tandem arrays. Orthologs and paralogs were directly inferred from the ML phylogenies and subsequently mapped to respective genomic coordinates. Since more distantly related species share less orthologous than paralogous gene copies (*e.g.* (21, 52)), we focused on the two orchid bees *Eg. dilemma* and *Ef. mexicana* for this analysis.

Selection analysis

Based on the phylogenetic reconstructions, we defined monophyletic orthologous groups for each gene family consisting of four or more ortholog copies including in-paralogs that were used for all subsequent selection analyses. In order to identify genes under positive selection, we applied the adaptive branch-site relative effects-likelihood (aBSREL) algorithm as implemented in HYPHY (80) to each subgroup to identify branches with dN/dS ratios significantly larger than 1, indicating positive selection along these branches (64). P-values for each aBSREL run were corrected for multiple testing using a false-discovery-rate (FDR) of 5%. In order to analyze patterns of recent gene family divergence in the orchid bees, we conducted pairwise dN/dS analyses for the *Eulaema* species pair and the two *Euglossa* species pairs. Therefore, we extracted orthologous copies from the gene family phylogenies and performed pairwise nucleotide sequence alignments for each orthologous pair. Each alignment was subsequently used for ML estimation of the pairwise dN/dS ratio (model M1) in codeml (81, 82). dN/dS ratios for each pairwise comparison were tested for significant deviations from the neutral expectation (dN/dS=1; model M0) using likelihood-ratio tests.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgements

We thank Michelle Stitzer, Judy Wexler, Mascha Koenen and the Ramírez lab for valuable comments on the manuscript. We thank Hugh Robertson for sharing chemosensory gene sequences of the honey bee and bumble bee. P.B. was supported by a fellowship of the German academic exchange service (Deutscher Akademischer Austauschdienst, DAAD) for parts of the project. SRR received support from the David & Lucile Packard Foundation and the National Science Foundation (DEB-1457753). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668, S10RR027303, and OD018174.

References

1. Hansson BS, Stensmyr MC (2011) Evolution of insect olfaction. *Neuron* 72(5):698–711.

2. Vosshall LB, Wong AM, Axel R (2000) An olfactory sensory map in the fly brain. *Cell* 102(2):147–159.

3. Elmore T, Smith DP (2001) Putative Drosophila odor receptor OR43b localizes to dendrites of olfactory neurons. *Insect Biochem Mol Biol* 31(8):791–798.

4. Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR (2003) Integrating the Molecular and Cellular Basis of Odor Coding in the Drosophila Antenna. *Neuron* 37(5):827–841.

5. Hallem EA, Ho MG, Carlson JR (2004) The molecular basis of odor coding in the Drosophila antenna. *Cell* 117(7):965–979.

6. Hallem EA, Carlson JR (2006) Coding of odors by a receptor repertoire. *Cell* 125(1):143–160.

7. Wang G, Carey AF, Carlson JR, Zwiebel LJ (2010) Molecular basis of odor coding in the malaria vector mosquito Anopheles gambiae. *Proc Natl Acad Sci USA* 107(9):4418–4423.

8. Kirkness EF, et al. (2010) Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci USA* 107(27):12168–12173.

9. Smith CD, et al. (2011) Draft genome of the globally widespread and invasive Argentine ant (Linepithema humile). *Proc Natl Acad Sci USA* 108(14):5673–5678.

10. Smith CR, et al. (2011) Draft genome of the red harvester ant Pogonomyrmex barbatus. *Proc Natl Acad Sci USA* 108(14):5667–5672.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11. McBride CS (2007) Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. *Proc Natl Acad Sci USA* 104(12):4996–5001.

12. McBride CS, Arguello JR (2007) Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177(3):1395–1416.

13. Goldman-Huertas B, et al. (2015) Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci USA* 112(10):3026–3031.

14. Gould F, et al. (2010) Sexual isolation of male moths explained by a single pheromone response QTL containing four receptor genes. *Proc Natl Acad Sci USA* 107(19):8660–8665.

15. Leary GP, et al. (2012) Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proc Natl Acad Sci USA* 109(35):14081–14086.

16. Pellegrino M, Steinbach N, Stensmyr MC, Hansson BS, Vosshall LB (2011) A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. *Nature* 478(7370):511–514.

17. Nei M, Niimura Y, Nozawa M (2008) The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet* 9(12):951–963.

18. Sánchez-Gracia A, Vieira FG, Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103(3):208–216.

19. Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39(1):121–152.

20. Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 11(2):97–108.

21. Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family. *Genome Res* 16(11):1395–1403.

22. Guo S, Kim J (2007) Molecular evolution of Drosophila odorant receptor genes. *Mol Biol Evol* 24(5):1198–1207.

23. Conceição IC, Aguadé M (2008) High Incidence of Interchromosomal Transpositions in the Evolutionary History of a Subset of Or Genes in Drosophila. *J Mol Evol* 66(4):325–332.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

716 24. Kratz E, Dugas JC, Ngai J (2002) Odorant receptor gene regulation: implications
717 from genomic organization. *Trends in Genetics* 18(1):29–34.

718 25. Klein A-M, et al. (2007) Importance of pollinators in changing landscapes for
719 world crops. *Proceedings of the Royal Society B: Biological Sciences*
720 274(1608):303–313.

721 26. Slaa EJ, Chaves LAS, Malagodi-Braga KS, Hofstede FE (2006) Stingless bees in
722 applied pollination: practice and perspectives. *Apidologie* 37(2):293–315.

723 27. Ramírez S, Dressler RL, Ospina M (2002) Abejas euglosinas (Hymenoptera:
724 Apidae) de la Región Neotropical: Listado de especies con notas sobre su biología.
725 3:7.

726 28. Grüter C, Keller L (2016) Inter-caste communication in social insects. *Current*
727 *Opinion in Neurobiology* 38:6–11.

728 29. Kocher SD, Grozinger CM (2011) Cooperation, Conflict, and the Evolution of
729 Queen Pheromones. *J Chem Ecol* 37(11):1263–1275.

730 30. Leonhardt SD, Menzel F, Nehring V, Schmitt T (2016) Ecology and Evolution of
731 Communication in Social Insects. *Cell* 164(6):1277–1287.

732 31. Leonhardt SD (2017) Chemical Ecology of Stingless Bees. *J Chem Ecol* 2(2):1–
733 18.

734 32. Vogel S (1966) Parfümsammelnde Bienen als Bestäuber von Orchidaceen und
735 Gloxinia. *Österr bot Z* 113(3-4):302–361.

736 33. Eltz T, Whitten WM, Roubik DW, Linsenmair KE (1999) Fragrance Collection,
737 Storage, and Accumulation by Individual Male Orchid Bees. *J Chem Ecol*
738 25(1):157–176.

739 34. Roubik DW, Hanson PE (2004) *Orchid bees of tropical America: biology and field*
740 *guide* (Instituto Nacional de Biodiversidad (INBio), Santo Domingo De Heredia).

741 35. Brand P, et al. (2015) Rapid evolution of chemosensory receptor genes in a pair of
742 sibling species of orchid bees (Apidae: Euglossini). *BMC Evol Biol* 15(1):176.

743 36. McBride CS, et al. (2014) Evolution of mosquito preference for humans linked to
744 an odorant receptor. *Nature* 515(7526):222–227.

745 37. Weinstock GM, et al. (2006) Insights into social insects from the genome of the
746 honeybee Apis mellifera. *Nature* 443(7114):931–949.

747 38. Sadd BM, et al. (2015) The genomes of two key bumblebee species with primitive
748 eusocial organization. *Genome Biology* 16(1):76.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

749 39. Kapheim KM, et al. (2015) Genomic signatures of evolutionary transitions from
750 solitary to group living. *Science*:aaa4788.

751 40. Brand P, et al. (2017) The Nuclear And Mitochondrial Genomes Of The
752 Facultatively Eusocial Orchid Bee Euglossa dilemma. *bioRxiv*:123687.

753 41. Zhou X, et al. (2012) Phylogenetic and transcriptomic analysis of chemosensory
754 receptors in a pair of divergent ant species reveals sex-specific signatures of odor
755 coding. *PLoS Genet* 8(8):e1002930.

756 42. Engsontia P, Sangket U, Robertson HM, Satasook C (2015) Diversification of the
757 ant odorant receptor gene family and positive selection on candidate cuticular
758 hydrocarbon receptors. *BMC Research Notes* 8(1):380.

759 43. Ross L, Blackmon H, Lorite P, Gokhman VE, Hardy NB (2015) Recombination,
760 chromosome number and eusociality in the Hymenoptera. *Journal of Evolutionary*
761 *Biology* 28(1):105–116.

762 44. Stolle E, et al. (2011) A second generation genetic map of the bumblebee Bombus
763 terrestris (Linnaeus, 1758) reveals slow genome and chromosome evolution in the
764 Apidae. *BMC Genomics* 12(1):48.

765 45. Rueppell O, et al. (2016) A New Metazoan Recombination Rate Record and
766 Consistently High Recombination Rates in the Honey Bee Genus Apis
767 Accompanied by Frequent Inversions but not Translocations. *Genome Biol*
768 *Evol*:evw269.

769 46. Wilfert L, Gadau J, Schmid-Hempel P (2007) Variation in genomic recombination
770 rates among animal taxa and the case of social insects. *Heredity* 98(4):189–197.

771 47. Nozawa M, Nei M (2007) Evolutionary dynamics of olfactory receptor genes in
772 Drosophila species. *Proc Natl Acad Sci USA* 104(17):7122–7127.

773 48. Obbard DJ, et al. (2012) Estimating divergence dates and substitution rates in the
774 Drosophila phylogeny. *Mol Biol Evol* 29(11):3459–3473.

775 49. Niimura Y, Nei M (2005) Comparative evolutionary analysis of olfactory receptor
776 gene clusters between humans and mice. *Gene* 346:13–21.

777 50. Bosco G, Campbell P, Leiva-Neto JT, Markow TA (2007) Analysis of drosophila
778 species genome size and satellite DNA content reveals significant differences
779 among strains as well as between species. *Genetics* 177(3):1277–1290.

780 51. Croset V, et al. (2010) Ancient protostome origin of chemosensory ionotropic
781 glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet*
782 6(8):e1001064.

783 52. Zhou X, et al. (2015) Chemoreceptor evolution in Hymenoptera and its

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

784 implications for the evolution of eusociality. *Genome Biol Evol*:evv149.

785 53. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in
786 gene copy number. *Nat Rev Genet* 10(8):551–564.

787 54. Perry GH, et al. (2008) Copy number variation and evolution in humans and
788 chimpanzees. *Genome Res* 18(11):1698–1710.

789 55. Perry GH, et al. (2006) Hotspots for copy number variation in chimpanzees and
790 humans. *Proc Natl Acad Sci USA* 103(21):8006–8011.

791 56. Cardoso-Moreira M, Emerson JJ, Clark AG, Long M (2011) Drosophila
792 Duplication Hotspots Are Associated with Late-Replicating Regions of the
793 Genome. *PLoS Genet* 7(11):e1002340.

794 57. LeBoeuf AC, Benton R, Keller L (2013) The molecular basis of social behavior:
795 models, methods and advances. *Current Opinion in Neurobiology* 23(1):3–10.

796 58. McKenzie SK, Fetter-Pruneda I, Ruta V, Kronauer DJC (2016) Transcriptomics
797 and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant
798 receptors in chemical communication. *Proc Natl Acad Sci USA*:201610800.

799 59. Woodard SH, et al. (2011) Genes involved in convergent evolution of eusociality
800 in bees. *Proc Natl Acad Sci USA* 108(18):7472–7477.

801 60. Michener CD (2007) *The Bees of the World. 2nd* (Ed. Johns Hopkins).

802 61. Tunstall NE, Sirey T, Newcomb RD, Warr CG (2007) Selective pressures on
803 Drosophila chemosensory receptor genes. *J Mol Evol* 64(6):628–636.

804 62. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG (2008) Drosophila
805 chemoreceptor gene evolution: selection, specialization and genome size.
806 *Molecular Ecology* 17(7):1648–1657.

807 63. Engsontia P, Sangket U, Chotigeat W, Satasook C (2014) Molecular evolution of
808 the odorant and gustatory receptor genes in lepidopteran insects: implications for
809 their adaptation and speciation. *J Mol Evol* 79(1-2):21–39.

810 64. Smith MD, et al. (2015) Less is more: an adaptive branch-site random effects
811 model for efficient detection of episodic diversifying selection. *Mol Biol Evol*
812 32(5):1342–1353.

813 65. Neafsey DE, et al. (2015) Highly evolvable malaria vectors: The genomes of 16
814 Anopheles mosquitoes. *Science* 347(6217):1258522–1258522.

815 66. Smadja C, Butlin RK (2009) On the scent of speciation: the chemosensory system
816 and its role in premating isolation. *Heredity* 102(1):77–97.

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

817 67. Pokorny T, et al. (2013) Acquisition of species-specific perfume blends: influence
818 of habitat-dependent compound availability on odour choices of male orchid bees
819 (Euglossa spp.). *Oecologia* 172(2):417–425.

820 68. Xu H, et al. (2012) FastUniq: a fast de novo duplicates removal tool for paired
821 short reads. *PLoS ONE* 7(12):e52249.

822 69. Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-Seq
823 data without a reference genome. *Nat Biotechnol* 29(7):644–652.

824 70. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq
825 using the Trinity platform for reference generation and analysis. *Nat Protoc*
826 8(8):1494–1512.

827 71. Xie Y, et al. (2014) SOAPdenovo-Trans: de novo transcriptome assembly with
828 short RNA-Seq reads. *Bioinformatics* 30(12):1660–1666.

829 72. Forêt S, Maleszka R (2006) Function and evolution of a gene family encoding
830 odorant binding-like proteins in a social insect, the honey bee (Apis mellifera).
831 *Genome Res* 16(11):1404–1413.

832 73. Vieira FG, Rozas J (2011) Comparative genomics of the odorant-binding and
833 chemosensory protein gene families across the Arthropoda: origin and
834 evolutionary history of the chemosensory system. *Genome Biol Evol* 3(0):476–
835 490.

836 74. Slater GSC, Birney E (2005) Automated generation of heuristics for biological
837 sequence comparison. *BMC Bioinformatics* 6(1):31.

838 75. Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for
839 maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*
840 21(4):456–463.

841 76. Katoh K, Misawa K, Kuma K-I, Miyata T (2002) MAFFT: a novel method for
842 rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids*
843 *Res* 30(14):3059–3066.

844 77. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software
845 version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–
846 780.

847 78. Katoh K, Kuma K-I, Toh H, Miyata T (2005) MAFFT version 5: improvement in
848 accuracy of multiple sequence alignment. *Nucleic Acids Res* 33(2):511–518.

849 79. Rivera MC, Jain R, Moore JE, Lake JA (1998) Genomic evidence for two
850 functionally distinct gene classes. *Proc Natl Acad Sci USA* 95(11):6239–6244.

851 80. Pond SLK, Frost SDW, Muse SV (2005) HyPhy: hypothesis testing using

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

phylogenies. *Bioinformatics* 21(5):676–679.

81. Bielawski JP, Yang Z (2004) A Maximum Likelihood Method for Detecting Functional Divergence at Individual Codon Sites, with Application to Gene Family Evolution. *J Mol Evol* 59(1):121–132.

82. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591.

Figures

Fig. 1 Gene family size dynamics in corbiculate bees. Numbers of previously (*A. mellifera, B. terrestris, Eg. viridissima*) and newly annotated (all others) genes for each gene family are indicated. Annotations derived from whole genomes and antennal transcriptomes are labeled. Phylogenetic relationships and divergence times of the ten corbiculate bee species following Ramirez et al. 2010 and Romiguier et al. 2015 are shown. Annotated nodes indicate MRCA of bumble bees and stingless bees (A), honey bees + A (B) and all corbiculate bees (C).

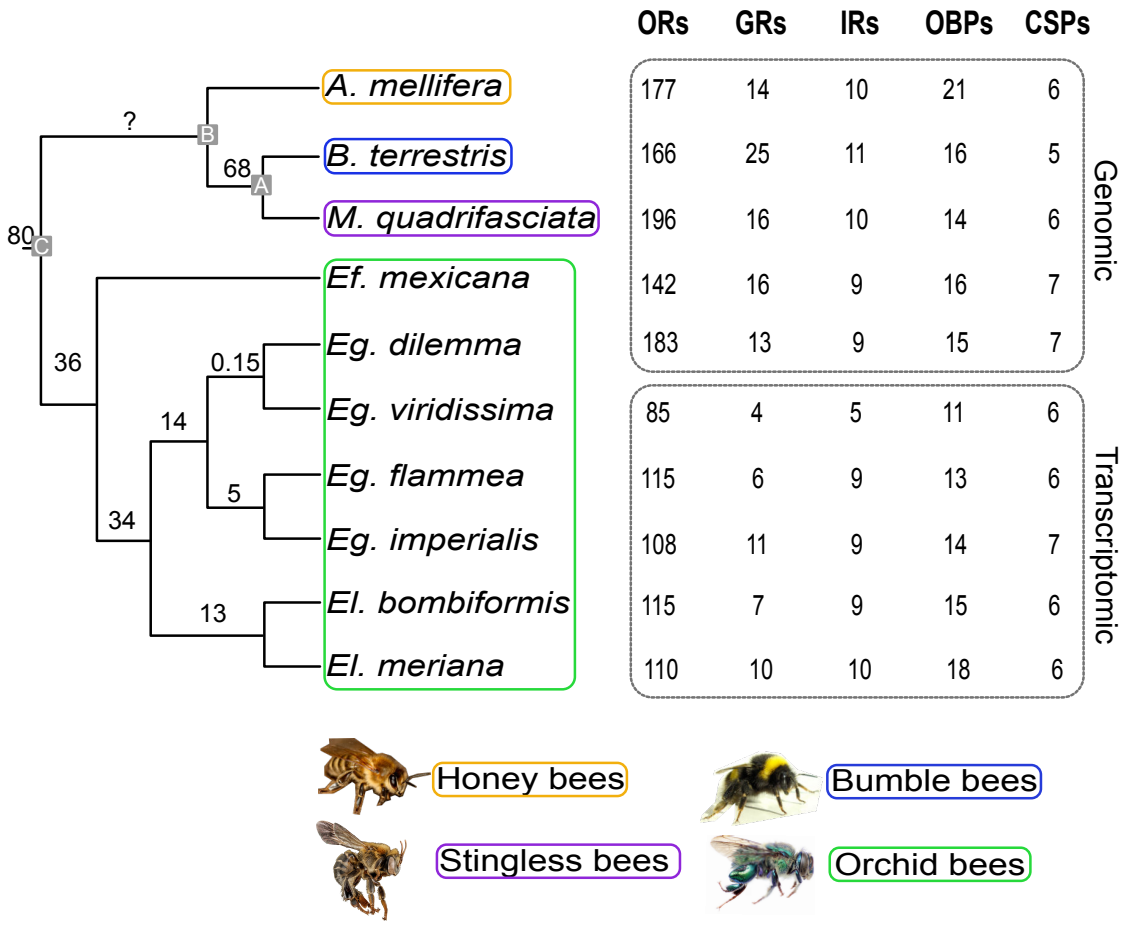
Fig. 2 Odorant receptor subfamily dynamics in corbiculate bees. Relationships between subfamilies (**A**) are based on all annotated ORs of 10 corbiculate bee species. Observed differences of within subfamily dynamics are shown with subfamily G13A and subfamily G04A as examples (**B, C**). (**A**) OR subfamilies have high bootstrap support while deeper nodes are of low support. All genes within a subfamily are located in a single tandem array with highly conserved genomic landscapes throughout corbiculate bees. The few exceptions are indicated with *. (**B, C**) ORs in subfamily G13A are mainly simple 1:1 single copy orthologs found in all or a subset of the analyzed species (marked '1:1') in contrast to subfamily G04A, where lineage specific duplications and larger expansions in *A. mellifera*, *B. terrestris*, *Eg. dilemma* and *Ef. mexicana* dominate (marked 'E'). Branches evolving under positive selective pressures are indicated by a red star. Both, duplicate branches as well as branches in 1:1 ortholog groups were under positive selection. The branch indicated with a * in (**B**) indicates the sistergroup to all other ORs in subfamily G13A which is found at a genomic location different from the rest of the subfamily in all corbiculate bee species representing an ancient translocation event (see text for details).

Fig. 3 Evolution of OR subfamily size. (**A**) OR subfamily size variation among corbiculates correlates with mean subfamily size. The pattern holds true for all corbiculates (white dots, black dotted line) and the two orchid bee species alone (black dots, grey dotted line). **(B)** OR subfamily repertoire numbers among corbiculate bees. Numbers of ORs in all 25 subfamilies for the five species with whole-genome data are indicated. Pseudogenes are excluded. *Eg. dilemma* and *Ef. mexicana* represent the two orchid bee species. OR subfamily sizes are highly variable between species, while increased subfamily sizes are shared by multiple

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

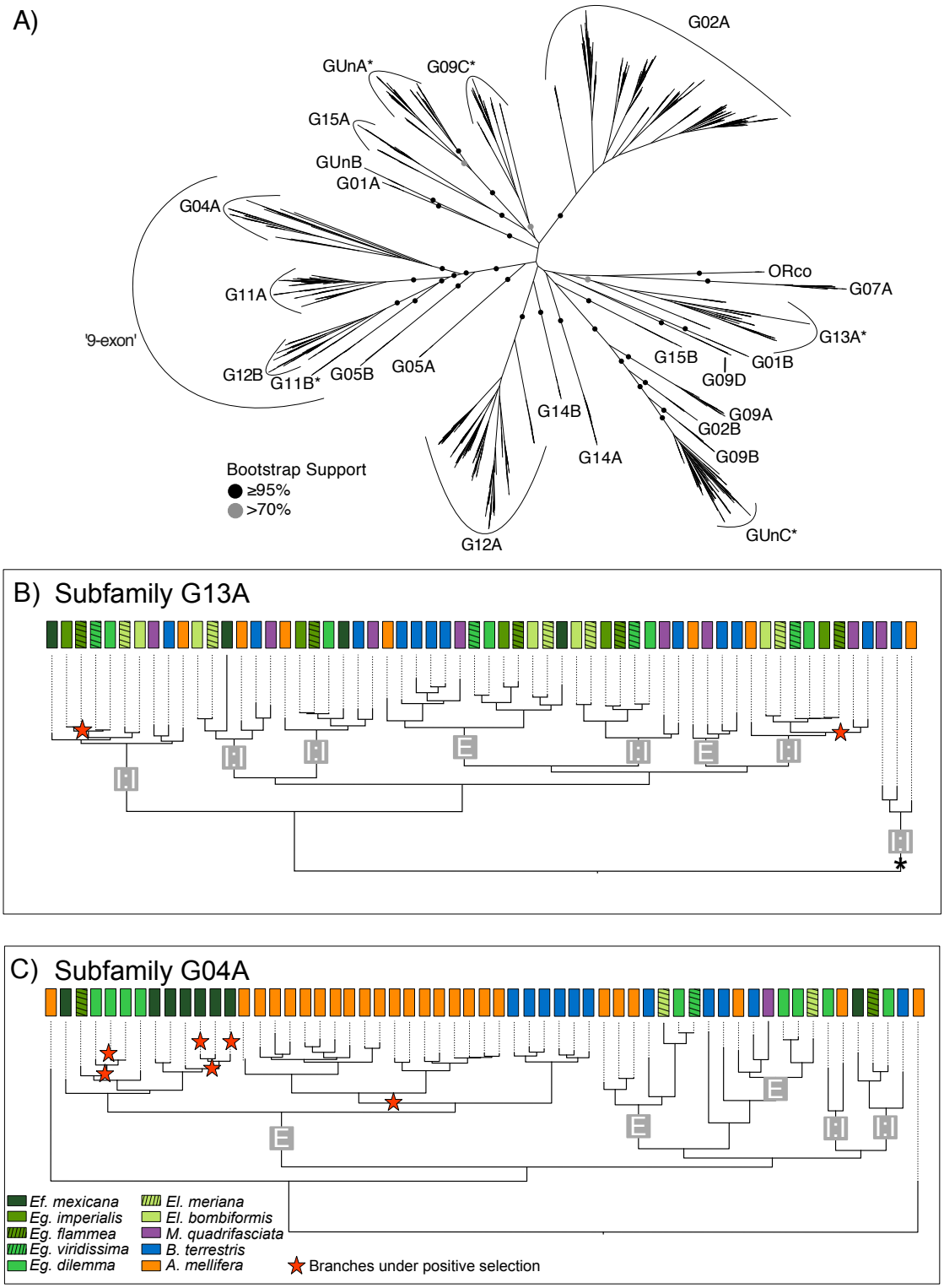
lineages or occur in single lineages only. 9Exon contains the combined repertoire of subfamilies G04A, G05B, G11A, G11B, and G12A.

Fig. 1



bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

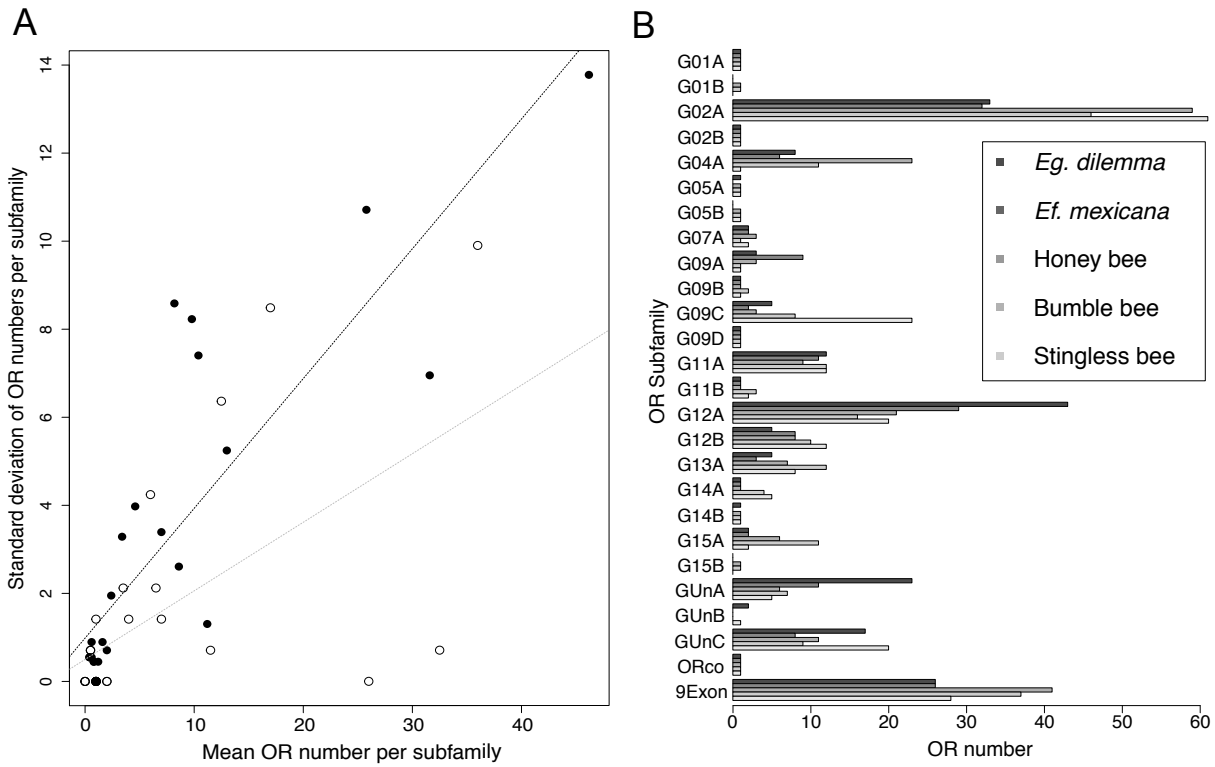
902 **Fig. 2**



903
904

bioRxiv preprint doi: https://doi.org/10.1101/130781; this version posted April 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

905 **Fig. 3**



906
907
908

909 ## Tables

910

911 **Tab. 1** Numbers of branches under positive selection for each chemosensory gene
912 family among the four corbiculate bee tribes.

913

914 **Tab. 1**

		Orchid bees	Honey bees	Bumble bees	Stingless bees	Total
ORs	Tested	1234	216	194	237	1881
	P ≤ 0.05	42 (3.4%)	5 (2.3%)	3 (1.6%)	14 (5.9%)	64 (3.4%)
GRs	Tested	116	11	32	15	174
	P ≤ 0.05	17 (14.7%)	2 (18.2%)	5 (15.6%)	4 (26.7%)	28 (16.1%)
IRs	Tested	101	9	9	9	128
	P ≤ 0.05	5 (5.0%)	0 (0%)	0 (0%)	0 (0%)	5 (3.9%)
OBPs	Tested	178	29	18	12	237
	P ≤ 0.05	3 (1.7%)	4 (13.8%)	0 (0%)	0 (0%)	7 (3.0%)
CSPs	Tested	76	6	5	6	93
	P ≤ 0.05	1 (1.3%)	0 (0%)	0 (0%)	0 (0%)	1 (1.1%)

915 All branches tested as well as the numbers and percentages of branches under positive selection after 5% FDR
916 correction are represented.