

# UC Davis

## UC Davis Previously Published Works

### Title

Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini)

### Permalink

<https://escholarship.org/uc/item/23p5k8jn>

### Journal

BMC Ecology and Evolution, 15(1)

### ISSN

2730-7182

### Authors

Brand, Philipp  
Ramírez, Santiago R  
Leese, Florian  
et al.

### Publication Date

2015-12-01

### DOI

10.1186/s12862-015-0451-9

Peer reviewed

RESEARCH ARTICLE

Open Access



# Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini)

Philipp Brand<sup>1,2\*</sup>, Santiago R. Ramírez<sup>2</sup>, Florian Leese<sup>1,4</sup>, J. Javier G. Quezada-Euan<sup>3</sup>, Ralph Tollrian<sup>1</sup> and Thomas Eltz<sup>1</sup>

## Abstract

**Background:** Insects rely more on chemical signals (semiochemicals) than on any other sensory modality to find, identify, and choose mates. In most insects, pheromone production is typically regulated through biosynthetic pathways, whereas pheromone sensory detection is controlled by the olfactory system. Orchid bees are exceptional in that their semiochemicals are not produced metabolically, but instead male bees collect odoriferous compounds (perfumes) from the environment and store them in specialized hind-leg pockets to subsequently expose during courtship display. Thus, the olfactory sensory system of orchid bees simultaneously controls male perfume traits (sender components) and female preferences (receiver components). This functional linkage increases the opportunities for parallel evolution of male traits and female preferences, particularly in response to genetic changes of chemosensory detection (e.g. Odorant Receptor genes). To identify whether shifts in pheromone composition among related lineages of orchid bees are associated with divergence in chemosensory genes of the olfactory periphery, we searched for patterns of divergent selection across the antennal transcriptomes of two recently diverged sibling species *Euglossa dilemma* and *E. viridissima*.

**Results:** We identified 3185 orthologous genes including 94 chemosensory loci from five different gene families (Odorant Receptors, Ionotropic Receptors, Gustatory Receptors, Odorant Binding Proteins, and Chemosensory Proteins). Our results revealed that orthologs with signatures of divergent selection between *E. dilemma* and *E. viridissima* were significantly enriched for chemosensory genes. Notably, elevated signals of divergent selection were almost exclusively observed among chemosensory receptors (i.e. Odorant Receptors).

**Conclusions:** Our results suggest that rapid changes in the chemosensory gene family occurred among closely related species of orchid bees. These findings are consistent with the hypothesis that strong divergent selection acting on chemosensory receptor genes plays an important role in the evolution and diversification of insect pheromone systems.

## Background

Olfaction allows animals to perceive volatile chemicals from the environment and is therefore essential for the detection and discrimination of food resources, predators, and conspecifics in a diverse array of taxa [1, 2]. In insects, intraspecific olfactory communication is predominantly based on the recognition of endogenous pheromones that are used to trigger a plethora of behaviors, including social

interaction, mate choice, and mate identification [3, 4]. Chemosensory genes expressed in the peripheral sensory neurons of the insect antennae enable the detection of pheromone compounds, and thus are crucially important for the detection of olfactory cues with diverse ecological functions [5, 6]. Closely related species of insects often exhibit pheromone signals with minute quantitative and qualitative differences [7]. Thus, highly specialized and sensitive signal recognition systems are necessary for discrimination of conspecific individuals and co-occurring (sympatric) species. However, despite the relative importance of pheromone detection in the evolution of insect communication and speciation, the genetic mechanisms

\* Correspondence: pbrand@ucdavis.edu

<sup>1</sup>Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Universitätsstrasse 150, D-44801 Bochum, Germany

<sup>2</sup>Department for Evolution and Ecology, Center for Population Biology, University of California Davis, One Shields Avenue, 95616 Davis, USA

Full list of author information is available at the end of the article



underlying the differentiation of pheromone recognition systems remain poorly understood [7, 8].

Recent work on reproductively isolated sympatric races of the European corn borer, *Ostrinia nubilalis*, has shown that divergence in pheromone recognition might be best explained by nucleotide substitutions in pheromone receptor genes [9]. Although this study lacked a direct test on candidate genes, it suggests that molecular divergence of chemosensory genes could have promoted the early differentiation of pheromone sensory tuning in these two species of moth. This mechanism has been put forward to explain the rapid evolution of pheromone communication systems in other sympatric sibling species of Lepidoptera [10, 11]. In fact, a single receptor mutation was shown to drive the early divergence in pheromone detection in *Ostrinia furnacalis*, a close relative to *O. nubilalis* [12]. However, it remains unclear whether these findings are broadly applicable to other insect taxa that rely on pheromone reception for sexual communication, mainly because few studies have examined the evolution of peripheral olfactory systems in closely related species. Additionally, other molecular mechanisms could lead to changes in the odor perception ability of a species, including changes in expression rates of genes involved in olfaction [13–15] and the evolution of the gene repertoire through gene duplication and gene loss [16, 17]. However, the relative impact of these two mechanisms on pheromone recognition of closely related species is less well understood.

The process of differentiation of communication channels requires concomitant shifts in both signal emission and signal perception [18]. Thus, a change in the pheromone detection pathway is expected to take place along with a shift in pheromone composition. Pheromone signals are usually synthesized metabolically *de novo* from relatively simple precursor building blocks [19]. Therefore, a shift in the chemical communication system of an insect lineage requires the correlated modifications of two independent pathways: the biosynthesis of pheromone compounds and the olfactory detection of such pheromone compounds [3, 12, 20]. Here we introduce an insect communication system in which a single pathway (olfactory) is responsible for changes in both signal production and signal detection.

Orchid bees (Apidae; Euglossini) are some of the most important pollinators in the neotropical region, where they pollinate thousands of plant species from numerous angiosperm families [21]. Unlike other insects, male orchid bees utilize unmodified exogenous volatiles to communicate species affiliation [22–27] that are further hypothesized to address females in the context of mating (e.g. [26]). Male orchid bees collect chemical substances from various floral and non-floral sources to concoct a species-specific perfume blend [28–32]. The perfume is

stored in specialized pouches in the bee's hind-tibiae and is eventually released in a ritualized courtship display behavior at sites where females arrive for mating [30, 33–35]. Accordingly, it is expected that both male and female orchid bees use overlapping gene sets to detect perfume compounds (e.g. [27]). As a result, orchid bees rely on their olfactory sensory system to produce and detect species-specific chemical signals.

Although the precise physiological mechanisms of perfume discrimination remain unknown, previous studies indicate that the olfactory periphery (*i.e.* the antennal processes involved in translating chemical odor signals into neurophysiological responses) plays a critical role in compound discrimination. Preferences for conspecific perfume blends are accompanied by increased antennal responses in comparison to responses to perfumes of closely and distantly related species [25, 27]. Furthermore, single compounds that exclusively occur in the bouquet of a given species can elicit higher antennal responses of conspecifics compared to individuals of closely related species [27, 36]. This suggests the presence of chemosensory genes that are tuned towards key perfume compounds, similar to what has been described for lepidopteran pheromone receptors (e.g. [12]).

All insect pheromone receptor genes characterized to date, including the honeybee queen pheromone receptor, belong to the Odorant Receptor (OR) gene family, the largest of three chemosensory receptor multi-gene families involved in insect odor detection [6, 37–40]. Unlike ORs, only a subset of Ionotropic Receptors (IRs) and few members of the Gustatory Receptor (GR) gene family are associated with olfaction [41–43]. In addition to these three receptor gene families, Odorant Binding Proteins (OBPs) and Chemosensory Proteins (CSPs) play a crucial role in peripheral olfactory recognition [6, 44]. These two non-receptor multi-gene families encode soluble globular proteins that presumably help transport hydrophobic odorant molecules through the hydrophilic sensillum lymph [44]. While CSPs were shown to be involved in nest mate recognition in ants [45], OBPs are necessary for pheromone reception in different insect species [46–48]. The involvement of these five gene families in insect olfaction makes them potential targets of selection with cascading effects on intraspecific communication channels.

In this study we provide an analysis of the chemosensory gene families of two recently diverged (~0.15–0.11 mya) sibling species of orchid bees, namely *Euglossa dilemma* Bembé & Eltz and *E. viridissima* Friese from the Yucatán Peninsula of Mexico [49]. The perfume profiles of these two morphologically [49] and ecologically [50] similar species differ mainly in the presence of a single compound (2-hydroxy-6-nona-1,3-dienyl-

benzaldehyde, hereafter HNDB) in *E. dilemma* and its complete absence in *E. viridissima* perfumes [27, 49]. This compound of unknown origin is found in the perfume of only one other orchid bee species, namely *Euglossa mixta*, which is distantly related to *E. dilemma*, suggesting possible multiple independent origins of HNDB collection in orchid bees [32]. In fact, HNDB comprises on average more than 60 % of the *E. dilemma* perfume blend. Moreover, HNDB attracts volatile-seeking males of *E. dilemma* but not *E. viridissima* when presented as a single compound in chemical bioassays in the field [27]. Concordantly, antennae of both male and female *E. dilemma* are more sensitive to HNDB than those of *E. viridissima* suggesting that behavioral differences might be based on divergence in the antennal periphery ([27]; T. Eltz, unpublished). The marked recent divergence between these sibling species provides unique opportunities to study the evolutionary genetic mechanisms that shaped peripheral olfactory recognition systems. For this purpose, we first identified the repertoires of all five focal chemosensory gene families in *E. dilemma* and *E. viridissima*. Due to the large genome sizes (~4Gb, each; Ramírez et al. unpublished data) we employed an antennal transcriptome sequencing approach coupled with a highly conservative *de novo* meta-assembly strategy. We analyzed the orthologous chemosensory gene sets found in both species and screened for patterns of divergent evolution. Our findings demonstrate that, despite the overall low divergence between these sibling species, divergent evolution of key chemosensory genes is accelerated, possibly due to divergent selection on the OR gene family.

## Results

### Candidate gene detection

To identify chemosensory genes of the OR, IR, GR, OBP, and CSP gene families in *Euglossa dilemma* and *E. viridissima*, we reconstructed the antennal transcriptomes for each species using a conservative meta-assembly approach (see Methods). In order to validate assembly quality, we annotated the transcriptomes by BLAT comparisons to 10,602 honeybee reference protein sequences that are not members of the focal chemosensory gene families (non-chemosensory (NC) gene set; RefSeq database accessed 10/22/12). This resulted in 8710 unique annotations of which more than 90 % were detected independently in both species and >70 % of all shared annotations showed  $\geq 95$  % completeness and contiguity of open reading frames (ORFs) (see Additional file 1: Table S1). Overall, 3091 full-length ortholog NC genes passed further conservative filter criteria and were later used for divergence analyses.

Chemosensory gene discovery using iterative tBLASTn searches on the antennal transcriptomes revealed 117 *Euglossa* loci homologous to known members of the five

targeted chemosensory gene families of bees and wasps (Additional file 1: Table S4) of which 95 (81 %) full-length orthologs were shared between *E. dilemma* and *E. viridissima* and thus were supported by independent discovery in the two sibling species (Table 1).

With 90 candidate genes, the largest number of loci we identified among chemosensory genes belonged to the OR gene family, corresponding to 77 % of all candidate chemosensory loci detected. Of all OR genes, 75 (83 %) were present in full length in the transcriptomes of both species. Additionally, of five candidate GRs and five candidate IRs, one and four orthologs were shared between species, respectively. The non-receptor chemosensory gene families were represented by 17 unique candidate genes, of which only two could not be detected in both species' antennal transcriptomes. This resulted in 10 and five shared full-length ortholog OBPs and CSPs, respectively.

We note that it is unlikely that the detected candidate genes represent the complete repertoire of the *E. dilemma* and *E. viridissima* chemosensory gene families, because detection is not possible if expression levels of target genes are too low, or specific to unexamined sexes, life stages or tissues. For example, it has been established that there is a typical 1:1 relationship of ORs and the number of glomeruli in the antennal lobes across insect lineages [51, 52]. Thus, based on ~160 glomeruli in the antennal lobes of each analyzed *Euglossa* species (Ramírez and Eltz unpublished), we estimate that we detected ~50 % of the functional ORs. However, the detected genes likely cover a large fraction of chemosensory genes important in intra-specific olfactory communication, as these are typically among the highest expressed chemosensory genes in insects and thus very likely to be detected in antennal transcriptome analyses [11, 40, 53].

### Chemosensory gene family dynamics

Phylogenetic inferences of the candidate chemosensory genes validated the homology of all the loci detected in relation to their respective gene families. Each *Euglossa* locus clustered with known honeybee representatives of the assigned gene family (Fig. 1; Additional file 2: Figures S1-S4). In contrast to the OR gene family, we identified simple 1:1 orthologous relationships for all but two non-OR chemosensory genes to known chemosensory genes of the honeybee, the closest relative with a completely known chemosensory gene family set [54]. This included all five IRs, all of which were orthologous to genes of the olfactory 'antennal IR' subfamily ([42, 55]; Additional file 2: Figure S1). Of five GRs, four had simple orthologous relationships to known honeybee GRs including an ortholog of a candidate sugar receptor (EvirGR04). Interestingly, the GR without a simple ortholog formed the outgroup of a cluster of three honeybee GR

**Table 1** Chemosensory genes detected in the antennal transcriptomes of *E. dilemma* and *E. viridissima*

	ORs	GRs	IRs	OBP	CSP	Total
<i>E. dilemma</i> <sup>a</sup>	86 (5 3)	2 (1 1)	4 (0 0)	10 (0 0)	5 (0 0)	107 (6 4)
<i>E. viridissima</i> <sup>a</sup>	85 (4 3)	4 (3 1)	5 (1 0)	11 (1 0)	6 (1 0)	111 (10 4)
Unique genes	90	5	5	11	6	117
Full-length orthologs <sup>b</sup>	75 (0.83)	1 (0.20)	4 (0.80)	10 (0.91)	5 (0.83)	95 (0.81)

<sup>a</sup>In brackets: unique genes found in only one of the two species | number of genes with missing N or C terminus

<sup>b</sup>Amount of full-length orthologs per gene family present in the antennal transcriptomes of both sibling species given in brackets

pseudogenes (Additional file 2: Figure S2). Simple orthologs were also identified for all CSPs. All but one OBP (Additional file 2: Figure S3-S4) were also assigned to honeybee orthologs including one ortholog of OBP3 of the mason bee *Osmia cornuta* for which binding affinities to several odors have been established [56]. All the *Euglossa* OBPs we identified were found to be orthologous to OBPs of the classical subfamily and/or exhibited six conserved cysteines typical for that subfamily (Additional file 2: Figures S3 and S5). Interestingly, two candidate OBPs were orthologous to honeybee OBPs that are not expressed in antennal tissue of either sex or caste, and were also absent in antennal transcriptomes of *O. cornuta* [56, 57].

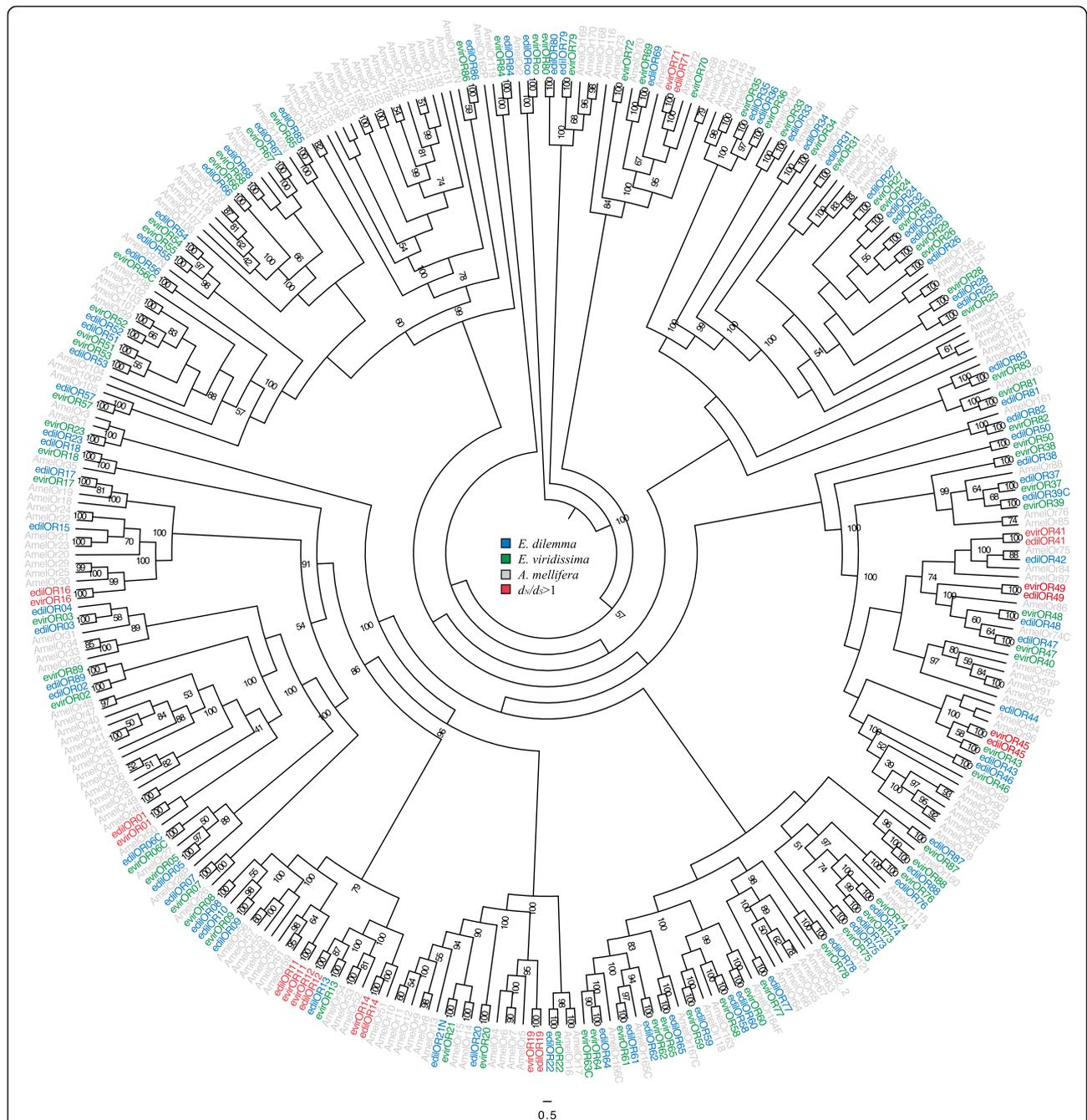
The relationships between the ORs of *Euglossa* and the honeybee were less unambiguous in relation to the dynamics of the other chemosensory families. Only 28 of the 90 detected *Euglossa* ORs (31 %) showed simple 1:1 orthology to known honeybee ORs (22 of these with  $\geq 95$  % bootstrap support, Fig. 1), indicating increased gene family divergence. Although the *Euglossa* homolog of the highly conserved OR co-receptor [58, 59] could be identified, none of the *Euglossa* ORs were orthologous to any of the three honeybee ORs that have been functionally characterized to date [38, 60]. Thus, potential ligands of *Euglossa* ORs remain unidentified. The lack of orthology of several *Euglossa*-specific ORs to honeybee ORs, indicated 10 *Euglossa*-specific OR duplications signified by clusters of one honeybee OR as outgroup to two different *Euglossa* ORs (Fig. 1; e.g. AmelOR114 to evir/edilOR73-74). We furthermore found two subfamily expansions of three or more *Euglossa* ORs (Fig. 1; e.g. evir/edilOR08-10). On the other hand, several honeybee ORs had no direct orthologs that could be detected in *Euglossa*. This is not surprising given the likely incompleteness of OR sets in the analyzed species of *Euglossa* (see above).

#### Patterns of nucleotide polymorphism and diversifying selection

We mapped antennal reads from a pool of 40 haploid males for each species against the ORFs of all the detected full-length chemosensory and NC genes. We excluded ORFs that lacked a mean per-base coverage  $\geq 10$ -fold in both orthologs, but retained 3091 NC genes as well as 74

ORs, four IRs, one GR, 10 OBPs and five CSPs (Table 2). Overall, 42 chemosensory loci and 387 NC loci were variable (polymorphic) between *E. dilemma* and *E. viridissima*, corresponding to 45 % and 13 % of the reconstructed candidate genes, respectively. With 36 loci, the majority (86 %) of variable chemosensory genes belonged to the OR gene family, while for each remaining chemosensory gene family only one or two loci showed polymorphisms between the two species (Table 2). This is not surprising given the number of reconstructed loci in each gene family. In total, we identified 1207 variable sites, of which 218 were found in chemosensory genes and 989 in NC genes, with 101 (46 %) and 376 (38 %) sites, respectively, representing fixed differences between *E. dilemma* and *E. viridissima*. Altogether, 24 chemosensory loci (21 ORs, two IRs and one OBP) and 157 NC loci contained sites fixed for different nucleotides in the two species (Table 2). Overall, the ratio of fixed to polymorphic sites was higher in the chemosensory gene family sets (0.86 and 0.61 for chemosensory and NC genes, respectively; Fisher's exact test;  $p = 0.02637$ ). Furthermore, the chemosensory gene families showed a significantly elevated ratio of non-synonymous to synonymous fixed sites in comparison to the NC gene set (Table 2; Fisher's exact test;  $p < 2.2e-16$ ).

Patterns of nucleotide polymorphisms enriched for fixed non-synonymous in comparison to synonymous substitutions as observed for the combined chemosensory gene set are expected in the presence of diversifying selection. To test for diversifying selection on the 181 candidate genes with fixed interspecific differences, we used the non-synonymous and synonymous substitution rates between *E. dilemma* and *E. viridissima* to calculate pairwise replacement to silent substitution rate ratios ( $d_N/d_S$ ). Therefore, we only took fixed sites into consideration because polymorphic sites are known to inflate  $d_N/d_S$  estimates, especially between species with comparatively low divergence times (see Methods; [61]). Consistent with the observed patterns of nucleotide polymorphisms,  $d_N$  was significantly higher for chemosensory loci than for NC loci (Mann-Whitney U = 2933;  $p$ -value =  $8.462e-7$ ) while  $d_S$  showed similar values for both sets of genes (Mann-Whitney U = 1624;  $p$ -value = 0.2768).



**Fig. 1** Phylogenetic relationships of the 86 candidate *E. dilemma* and 85 *E. viridissima* Odorant Receptors (ORs) to *Apis mellifera*. The maximum likelihood tree was rooted by the OR co-receptor orthologs of all three species. Bootstrap values for branches with  $\geq 50\%$  support are indicated. ORs marked in red correspond to the 10 orthologs with  $d_N/d_S > 1$  (Table 2; see main text for details). Symbols after the OR descriptors: C: C-terminus is missing, N: N-terminus is missing, F: gene model was manually assembled, P: pseudogene (after [54])

This resulted in an elevated mean  $d_N/d_S$  for the chemosensory loci (0.91 vs. 0.12 for the NC gene set; Fig. 2a; Table 2) indicating relaxed purifying and/or increased diversifying selection on chemosensory genes. Although mean  $d_N/d_S$  was smaller than 1 for both gene sets, we detected 12 chemosensory receptors (10 ORs and 2 IRs; Table 3) and 23 NC genes with  $d_N/d_S > 1$  (Fig. 2b; Table 2;

Additional file 1: Table S2), indicating that positive selection may have driven divergence in these candidate genes. The set of genes corresponds to 12.8 % of the chemosensory loci identified and 0.7 % of the NC gene set. This observation suggests that divergent selective pressures are increased in genes of the olfactory periphery, in particular in chemosensory receptors

**Table 2** Nucleotide polymorphisms and patterns of selection between orthologous genes

	ORs	GRs	IRs	OBPs	CSPs	Total	NC
Unique genes <sup>a</sup>	74	1	4	10	5	94	3091
Variable genes <sup>b</sup>	36 (0.49)	1 (1.0)	2 (0.5)	2 (0.2)	1 (0.2)	42 (0.45)	387 (0.13)
Total variable sites	194	3	13	10	1	218	989
Total polymorphic <sup>c</sup>							
synonymous	55	0	2	3	1	61	479
non-synonymous	50	3	2	4	0	56	134
Total fixed <sup>c</sup>							
synonymous	23	0	1	1	0	25	277
non-synonymous	66	0	8	2	0	76	99
Genes fixed <sup>d</sup>	21	0	2	1	0	24	157
Mean $d_N^e$	0.0034857	0	0.0026	0.006000	0	0.0035167	0.0005127
Mean $d_S^e$	0.0037524	0	0.00105	0.011600	0	0.0038542	0.0042777
Mean $d_N/d_S^e$	0.9289340	-	2.4761905	0.5172414	-	0.9124324	0.1198630
$d_N/d_S > 1^{ef}$	10 (1)	-	2 (0)	0 (0)	-	12 (1)	23 (0)
S ratio [%] <sup>g</sup>	13.51	0	50	0	0	12.77	0.74

<sup>a</sup>Homologous genes identified independently in the antennal transcriptomes of *E. dilemma* and *E. viridissima* with  $\geq 10$ -fold mean per-base coverage

<sup>b</sup>Genes with fixed differences between the two species. In brackets: relative amount of all unique genes

<sup>c</sup>Total polymorphic/ fixed sites of all variable sites

<sup>d</sup>Genes with fixed differences between the two species

<sup>e</sup> $d_N/d_S$  calculations are based on genes containing fixed differences

<sup>f</sup>Number of pairwise  $d_N/d_S$  estimates significantly higher than 1 in brackets

<sup>g</sup>Ratio of orthologous genes with fixed differences among all orthologous genes detected per gene set

(Fisher's exact test;  $p = 3.843e-10$ ). Accordingly, the set of genes with  $d_N/d_S > 1$  was significantly enriched for chemosensory receptors compared to those with  $d_N/d_S < 1$  (Fisher's exact test;  $p = 3.92e-10$ ).

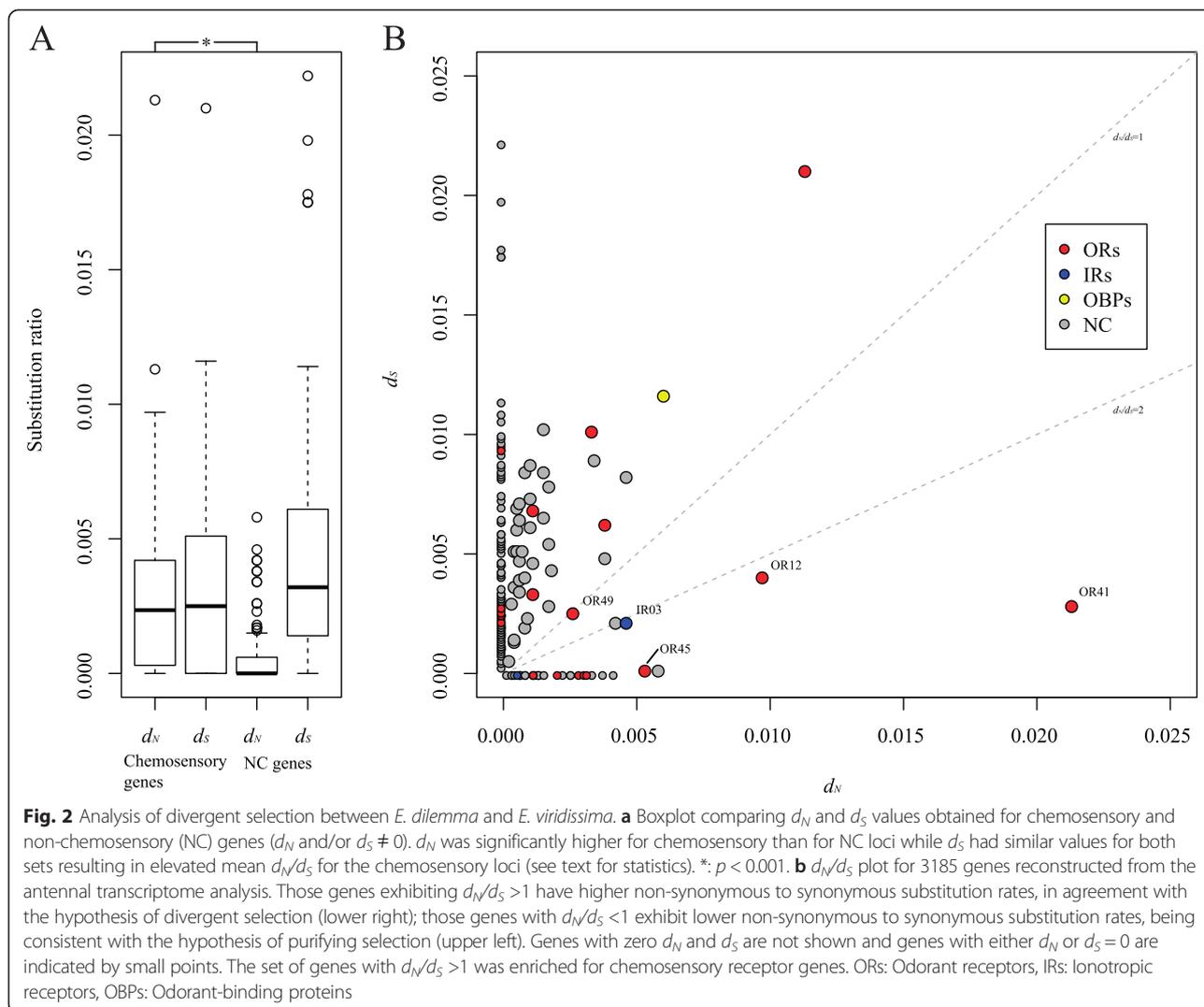
We estimated a mean of 2.9 fixed substitutions per gene among the 35 genes with  $d_N/d_S > 1$  (101 fixed substitutions in 35 genes; median: 1). In fact, only 12 of these genes contained at least three fixed substitutions (seven out of 12 chemosensory genes (58 %), five out of 23 NC genes (22 %)). In addition, 20 of the 23 remaining genes had only 1 fixed substitution, of which the majority (85 %) belonged to the NC gene set (three chemosensory genes, 17 NC genes; Table 3; Additional file 1: Table S2). Thus, most of the genes with  $d_N/d_S > 1$  were fixed for just one non-synonymous substitution between *E. dilemma* and *E. viridissima*.

To test whether the  $d_N/d_S$  values differ from a null model of neutral evolution, we applied likelihood ratio tests to all orthologous pairs of chemosensory and NC genes. Interestingly, of the 181 gene pairs with fixed interspecific differences, only two  $d_N/d_S$  estimates were significantly different from the null model. Of these OR41 showed a  $d_N/d_S$  significantly higher than 1 ( $d_N/d_S = 7.73$ ; Likelihood-ratio test  $\Delta = 7.09$ ;  $p < 0.01$ ; Table 2 & 3) and OR06 showed a  $d_N/d_S$  significantly lower than 1 ( $d_N/d_S = 0.001$ ; Likelihood-ratio test  $\Delta = 5.62$ ;  $p < 0.05$ ; Additional file 1: Table S5). This low number of genes diverging from the neutral null model is likely to reflect the generally low power of

pairwise  $d_N/d_S$  estimates [62] as well as the low number of fixed differences between the two species.

#### Spatial distribution of non-synonymous substitutions

Non-synonymous changes in ligand binding domains of receptor proteins can alter affinities towards ligands, modifying ligand interaction patterns [12, 63–67]. We determined the spatial distribution of non-synonymous substitutions along OR and IR protein sequences with  $d_N/d_S > 1$  to examine potential effects on ligand binding domains. Therefore, we predicted transmembrane domains (Additional file 1: Table S6; Additional file 2: Figure S6), the regions of OR proteins most sensitive to non-synonymous substitutions with regard to ligand binding [12, 64, 66]. Moreover, we used homology to known *Drosophila* IRs and the closely related ionotropic glutamate receptors (iGluRs) to infer ligand-binding domains (see Methods). In total, 24 (51 %) of the 47 non-synonymous substitutions fixed in the 10 ORs having  $d_N/d_S > 1$  between *E. dilemma* and *E. viridissima* were located in one of the seven transmembrane domains (Table 3; Fig. 3) which covered between 19.3 % and 35.8 % of the OR amino acid sequence (Mean: 30.5 %; Additional file 1: Table S6). Additionally, 3 (38 %) of 8 replacement substitutions were located in the IR ligand binding domains that covered 16.5 % and 14.7 % of the IR03 and IR11 amino acid sequence, respectively (Mean: 15.6). Interestingly, only three of the



12 chemosensory receptors (two ORs and one IR) did not reveal any change in the amino acid sequence of respective ligand binding domains. In order to test whether fixed non-synonymous substitutions are randomly distributed among ORs and IRs, we applied a goodness-of-fit test on the observed number of substitutions by estimating the mean proportion of receptor proteins that span ligand-binding domains. These tests revealed that the observed number of fixed non-synonymous substitutions were non-randomly distributed among ORs and were significantly enriched for transmembrane domains (Goodness-of-fit  $\chi^2 = 9.38$ ;  $p < 0.01$ ; IRs:  $\chi^2 = 2.91$ ;  $p < 0.1$ ). Furthermore, non-synonymous substitutions in ligand-binding domains were positively correlated with the number of non-synonymous substitutions (Pearson's correlation coefficient;  $r = 0.91$ ;  $p < 0.001$ ). Concomitantly, the four chemosensory receptors that exhibited at least five fixed substitutions had the most replacement substitutions in ligand binding domains (up to

9 in OR41; Table 3), thus increasing the likelihood that such non-synonymous substitutions lead to changes in ligand-binding affinities.

Graphical analysis of the spatial distribution along the ORs revealed that the majority of non-synonymous substitutions are located towards the N-terminus rather than towards the C-terminus of the receptor (Fig. 3), matching known patterns of higher OR sequence conservation towards the C-terminus [68–70]. This pattern is also retained for a subset of the three ORs with at least five fixed substitutions between *E. dilemma* and *E. viridissima*.

## Discussion

The differentiation of intraspecific chemical communication systems depends on correlated shifts of signal production and signal detection [7]. In the case of chemical signaling of insect species, the emitter and receiver components are typically controlled by independent genetic

**Table 3** Fixed and polymorphic non-synonymous and synonymous substitutions of orthologous chemosensory genes with  $d_N/d_S > 1$ 

Gene	Non-synonymous <sup>a</sup>		Synonymous <sup>a</sup>		$d_N$	$d_S$	$d_N/d_S^b$	ln LBD <sup>c</sup>
	Fixed	Polymorphic	Fixed	Polymorphic				
OR41	18	1	1	5	0.0213	0.0028	7.7278*	9
OR12	10	6	1	2	0.0097	0.0040	2.4235	3
OR45	5	1	0	0	0.0053	0.0000	∞	5
OR14	3	2	0	4	0.0021	0.0000	∞	1
OR01	3	1	0	3	0.0031	0.0000	∞	2
OR16	2	5	0	1	0.0032	0.0000	∞	1
OR71	2	5	0	0	0.0029	0	∞	1
OR49	2	2	1	1	0.0026	0.0025	1.0536	2
OR11	1	0	0	1	0.0012	0	∞	0
OR19	1	0	0	0	0.0012	0	∞	0
IR03	7	2	1	1	0.0046	0.0021	2.2076	3
IR11	1	0	0	1	0.0006	0	∞	0

Genes with less than 3 fixed substitutions between the two species are highlighted in grey

<sup>a</sup>Fixed and polymorphic non-synonymous and synonymous substitutions between orthologs of given genes of *E. dilemma* and *E. viridissima*

<sup>b</sup>Receptors with  $d_N/d_S$  significantly higher than 1 are indicated by \*

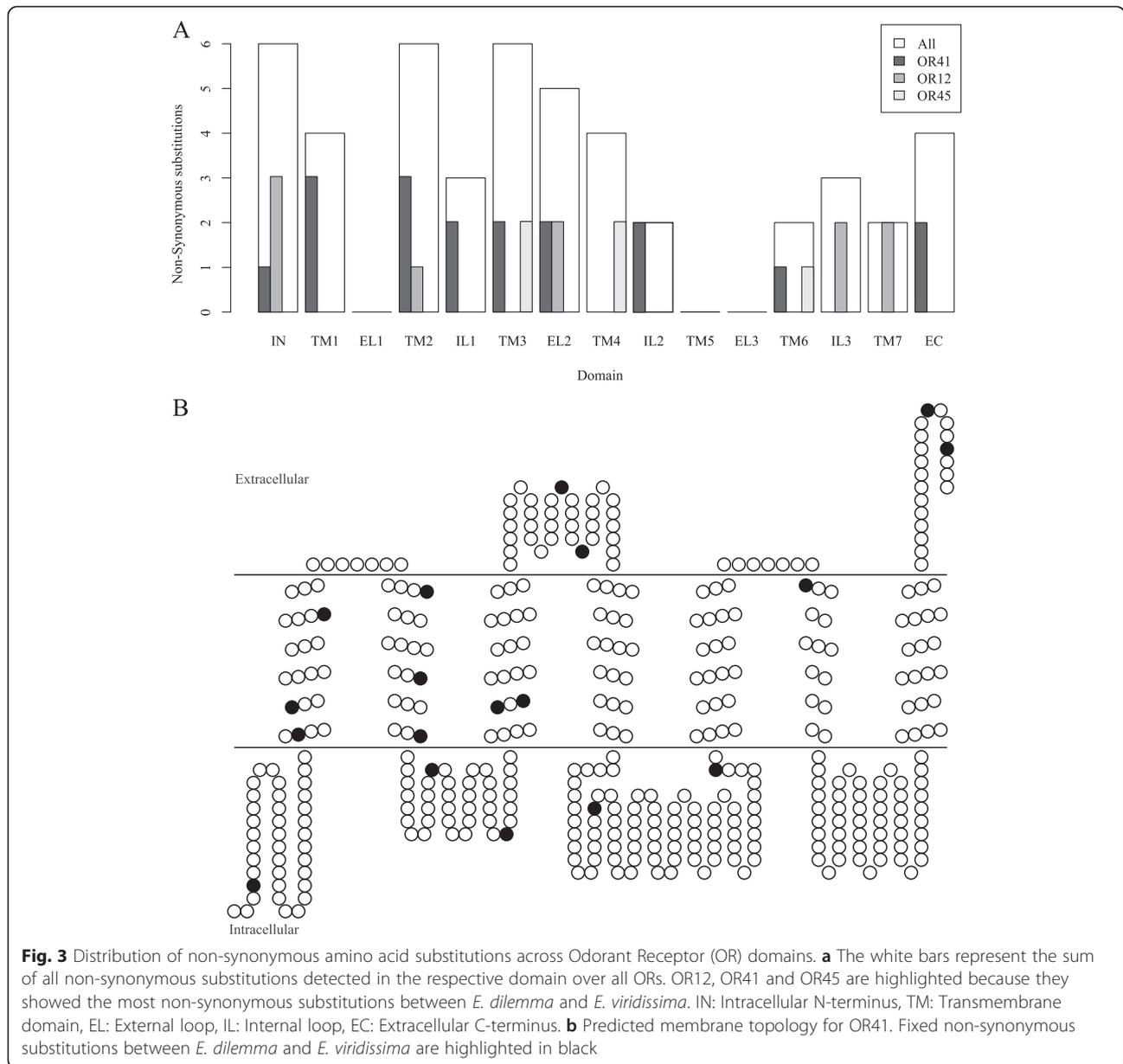
<sup>c</sup>Fixed substitutions in ligand binding domains (LBD) of ORs (transmembrane regions) and IRs (S1 and S2 LBD)

pathways [18]. In the present study we focused on an exceptional communication system, where a single pathway regulates both signal production and signal detection. Orchid bees rely on their peripheral olfactory system to detect compounds for the composition of their so-called perfumes. These perfume phenotypes are thought to function as intraspecific recognition signals and have been shown to be highly differentiated even among closely related species [25–27, 32]. As a result, chemosensory genes are likely the targets of selection during the evolutionary process of signal differentiation. We tested this hypothesis in a pair of recently diverged sibling species of orchid bees, *Euglossa dilemma* and *E. viridissima* [49] whose perfumes differ mainly in one key chemical component. We compared the evolutionary patterns of differentiation of 94 orthologous chemosensory genes and 3091 genes that are not involved in chemoreception (NC genes) derived from a *de novo* antennal transcriptome analysis.

#### Accelerated evolution of the olfactory periphery

Comparisons of the estimated  $d_N$  and  $d_S$  patterns based on orthologous chemosensory and NC genes indicate faster evolution of the chemosensory gene families driven in part by strong divergent selective pressures. Our analysis identified fixed nucleotide differences between the two sister species in 13 % of all analyzed genes. This low percentage probably reflects the short time span since the divergence of *E. dilemma* and *E. viridissima* [49]. However, chemosensory genes exhibited an elevated mean non-synonymous substitution rate ( $d_N$ ) that produced a mean  $d_N/d_S$  ratio ~7.5 times

higher than those estimated for NC genes. While this pattern is consistent with accelerated diversification rates in the chemosensory gene families, several underlying mechanisms could be at play. First, this pattern could be explained by relaxed purifying selection on chemosensory genes. Relaxation of purifying selection in chemosensory genes is usually expected following gene duplication events [71, 72]. However, analyses of the OR gene families from a number of *Drosophila* species have indicated that lineage-specific duplications among closely related sibling species (<10 mya divergence time) are rare [73, 74]. Although the OR gene family is the most dynamic among all chemosensory gene families, previous studies have identified very few lineage-specific duplication events, even among species evolving under highly divergent ecological conditions. For instance a maximum of four duplication events was estimated between the host specialist *D. sechellia* and the host generalist *D. simulans*, which have a divergence time of ~0.5 mya [73, 74]. In contrast, the elevated  $d_N/d_S$  ratio that we observed between *E. dilemma* and *E. viridissima* is based on 42 variable chemosensory genes including 36 variable ORs. Hence, it is unlikely that the observed patterns are greatly influenced by lineage specific duplication events. Consistently, we could not find any evidence for lineage specific duplications in either species. However, we note that recent duplication events may have gone undetected in our dataset for several reasons (e.g. low expression levels) and that relaxation of purifying selective pressures might occur even in the absence of duplication events.



We therefore cannot exclude relaxed purifying selection as a mechanism that shaped the evolution of chemosensory genes in orchid bees.

Alternatively, the observed patterns of  $d_N/d_S$  ratios could reflect signatures of positive selection. We identified 12 chemosensory genes with  $d_N/d_S > 1$ , which is consistent with the hypothesis of positive selection in one or both *Euglossa* lineages. Nevertheless, due to the overall low variability between species, the calculated  $d_N/d_S$  ratios for several loci were based on few fixed differences alone. While previous studies have reported relatively low variability in chemosensory genes under divergent selective pressures in closely related insect species [75], we detected genes with both  $d_N/d_S > 1$

and comparably high interspecific diversity of which one (OR41) was significantly different from a neutral null model of sequence divergence (Table 2). The fact that only a single gene exhibited a significantly elevated  $d_N/d_S$  ratio is not surprising given the low amount of fixed interspecific differences between *E. dilemma* and *E. viridissima* and the low power of pairwise  $d_N/d_S$  tests in detecting genes under divergent selection [62, 76]. Accordingly, we expect that the test produced several false negatives that may be revealed by more sensitive phylogeny-based  $d_N/d_S$  tests, as well as more comprehensive lineage sampling [62]. In addition, the comparatively short divergence time between *E. dilemma* and *E. viridissima* bears the potential of biasing  $d_N/d_S$  estimates, that could lead to false

positives [61, 77]. To account for this, we chose an approach that enabled the detection and exclusion of the main source of  $d_N/d_S$  inflation, namely segregating polymorphisms [61, 77]. The resulting patterns of  $d_N$  and  $d_S$  that we observed between chemosensory genes and the NC gene set (see above) using fixed differences only are similar to that observed in other insects that exhibit much greater divergence times [16], thus indicating that our approach is suited for the detection of selective pressures on the chemosensory gene sets of *E. dilemma* and *E. viridissima*.

Our results are consistent with the hypothesis that genes of the olfactory peripheral system of *E. dilemma* and *E. viridissima* have evolved under strong divergent selective pressures. Together, these observations support a significant trend of increased divergent selective pressures that may have shaped the recent evolution of chemosensory genes in *E. dilemma* and *E. viridissima*.

#### Are the observed patterns of diversifying selection related to divergence in chemical signaling?

The observed sequence divergence in orthologous olfactory receptor genes of *E. dilemma* and *E. viridissima* is likely linked to differences in the sensory tuning of each species, and possibly represents a response to divergent selection on chemosensory traits [7]. Various ecological factors may have promoted such differentiation, including host shifts [78, 79] and changes in mating ecology [12]. In solitary bees, several ecological factors may impose selective pressures on sensory detection, including foraging on different food resources (generalist vs. specialist), the use of different nesting materials, and the detection of suitable mating places and partners. However, both *E. dilemma* and *E. viridissima* are pollen generalists that are known to use very similar food resources [50], and the two species lack any noticeable differences in nesting biology (T. Eltz and S. Ramírez, pers. obs.). Thus, it is unlikely that the divergence we observed in olfactory receptor genes was due to selective pressures acting on foraging specialization. The most pronounced difference that has been documented between the two sympatric sibling species is on the chemical composition of male perfumes [27, 49].

The males of all species of orchid bees collect and accumulate species-specific perfumes [32] from a variety of sources. Likewise, *E. dilemma* and *E. viridissima* show distinct species-specific perfume phenotypes that are qualitatively and quantitatively differentiated. In particular, these two species differ in the presence of HNDB in *E. dilemma* perfumes and the complete absence in *E. viridissima* [27, 49]. This difference corresponds to pronounced behavioral and physiological responses. Males of *E. dilemma* are strongly attracted to HNDB, whereas males of *E. viridissima* are never attracted to this

compound [27]. Correspondingly, the antennae of *E. dilemma* exhibit a significantly stronger neurophysiological response and sensitivity to HNDB, compared to a weaker response in *E. viridissima* [27]. Consequently, the observed divergence of the chemosensory gene families on the molecular level suggest that the chemical, physiological, and behavioral differences between this pair of species might be mediated—at least partially—by some of the genetic differences we identified in the chemosensory gene families.

Previous studies have hypothesized that perfume bouquets in orchid bees function as species-specific signals that are addressed to conspecific females in the context of mating (see Discussion in [26]). Thus, because olfaction determines both signal production and signal detection, evolutionary shifts in the olfactory pathway may lead to concomitant divergence in sexual communication in orchid bees. Genomic studies on closely related species of drosophilid flies have revealed that the primary targets of diversifying selection are, in fact, sex-related genes [80]. Moreover, previous studies on Lepidoptera and drosophilid flies have determined that those OR genes with elevated  $d_N/d_S$  ratios tend to be involved in the perception of sex pheromones [12, 13]. Our results lend support to the hypothesis that candidate chemosensory receptors evolved under strong divergent selection in the *E. dilemma* and *E. viridissima* lineage, and thus selective forces may have contributed to shifts in the detection of odors that mediate sexual communication in one or both species.

#### Chemosensory receptor driven evolution of the peripheral olfactory pathway

The observed pattern of divergent selection in chemosensory receptors suggests that the olfactory peripheral system plays a major role in the evolution of olfactory specialization in *E. dilemma* and *E. viridissima*. Among all the five families of chemosensory genes we studied, signatures of divergent selection were only present in OR genes and antennal IR genes. This result is partially consistent with an earlier hypothesis that a shift in peripheral olfaction between *E. dilemma* and *E. viridissima* might be driven by molecular divergence in OR genes [27], a common mechanism of olfactory diversification in insects [12, 64, 65].

Currently we lack information on the functional properties of the detected divergent chemosensory receptors. To gain some insight into potential impact of the observed substitutions, we inferred the transmembrane (TM) and ligand-binding domains in both ORs and IRs. Based on these topological predictions, the majority of the molecular differences between the sister species could affect the biochemical features of the receptors. For most of the ORs that exhibited signatures of diversifying selection,

amino acid substitutions were detected in TM regions, where a single replacement can be sufficient to elicit shifts in ligand binding properties, as previously demonstrated in several insect taxa [12, 64–67]. Predicted amino acid substitutions in the S1 and S2 binding domains of a candidate IR might have similar consequences [42, 81, 82]. Accordingly, diversifying selection acting on *Euglossa* ORs and IRs might have led to divergent ligand binding properties of chemosensory receptor orthologs of the two species and in turn could account for the observed differences in antennal responses.

The neurophysiological difference in the response of *E. dilemma* and *E. viridissima* to volatile compounds could also be determined by higher-level integration of the central nervous system, e.g. neuronal networks in the antennal lobe or mushroom body. Nevertheless, the strong differentiation observed in antennal responses supports a scenario where the olfactory peripheral system profoundly affects the functioning of this chemical sexual communication system [27]. In orchid bees, differentiation of olfactory tuning through divergence of chemosensory receptors could have cascading effects on the species-specific perfumes that male bees acquire. Similarly, studies on moths have identified single non-synonymous substitutions in OR genes that lead to strong differentiation in pheromone detection abilities [9, 12, 14]. Comparative analyses have shown that the perfume phenotypes of orchid bees evolve exceptionally fast, even among closely related species [32]. Because non-synonymous substitutions in the olfactory periphery may simultaneously affect both male traits and female preference, diversifying selection acting on chemosensory receptors could serve as a mechanism to account for the fast evolutionary rates observed in perfume phenotypes among orchid bees. In fact, physical genetic linkage between sender and receiver genes could potentially accelerate the evolution of assortative mating and rapid modification of sexual communication channels (e.g. [83]). This, in turn, could help explain the high species richness of orchid bees throughout the neotropical region [84], and could lead to an understanding of the speciation mechanisms in orchid bees.

## Conclusions

While the vast majority of insect species relies on chemical communication to find mates in order to successfully reproduce, the genetic mechanisms underlying the evolution of pheromone recognition systems remain poorly understood, especially for non-model organisms. In this study we show that gene families in the olfactory periphery of *Euglossa dilemma* and *E. viridissima*, two recently diverged orchid bee species, likely evolved under divergent selection. Because signal production and signal detection are genetically linked in orchid bees, our findings support the hypothesis that divergent evolution of

OR genes likely played a role in shaping both olfactory perception and divergence of chemical mating signals. Our results are consistent with previous studies on lepidopterans and indicate the general significance of selection acting on chemosensory receptors as a driver mechanism of diversification in insect pheromones.

## Methods

### Sampling and sequencing

Males of the two orchid bee species *Euglossa dilemma* and *E. viridissima* were sampled in the Yucatán Peninsula, Mexico in October 2011 near the city of Xmatkuil and between Muna and Uxmal (distance to Xmatkuil: ~50 km) using different chemical baits [85]. Sampling was performed with the necessary permits issued by the Secretaría de Medio Ambiente y Recursos Naturales to J.J.G. Quezada-Euan. Bees were kept in small cages in a greenhouse (temperature 20–24 °C). During the two to eight days in captivity a 1:3 mixture of honey and tap water was provided as food source. To produce antennal transcriptomes representing Yucatán populations, the antennae of 40 male specimens of each species were pooled for RNA-extraction. Bees were chilled on ice and the antennae of each torpid male were dissected by sterile forceps and immediately shock-frozen on liquid nitrogen. Antennae were kept on liquid nitrogen/dry ice until RNA-extraction.

Total-RNA was extracted using the TRIzol extraction method (Invitrogen) following the manufacturers tissue preparation protocol, except for an extended incubation time of 15 min in the phase separation step to maximize RNA yield. Extracted RNA was resuspended in 30 µl of RNase free water. All optional steps were skipped. RNA pools were treated with DNaseI to purge potential DNA contamination and subsequently quantified on the Experion Automated Electrophoresis System (Bio-Rad) with the Experion StdSens Analysis Kit (Bio-Rad) according to the standard protocol. Afterwards, 4 µg and 2 µg of total-RNA of the *E. dilemma* and *E. viridissima* pool, respectively were sent to GATC-Biotech (Constance, Germany) for barcoded cDNA library preparation using the TruSeq mRNA kit (Illumina) and subsequent 100-bp single-end sequencing on an Illumina HiSeq 2000 lane (Raw sequence reads are available at the NCBI Sequence Read Archive [SRA: SRX765918, SRA: SRX765888]).

### Pre-processing

Identical raw reads were merged using Fulcrum 0.4.2 [86] to improve assembly quality and computing efficiency [87]. The merged read sets were quality checked and reads were trimmed on both sides if sequencing primers or low-quality bases (Phred-score ≤ 20) were detected, applying a sliding window approach in SeqtrimNext [88] with a window size of 3. Furthermore, homopolymeric reads and reads < 21 bp were discarded using an inhouse Perl-script.

### Assembly and transcript recovery

Due to a lack of comprehensive sequence data for orchid bees, the antennal transcriptomes had to be assembled *de novo*. As the scope of this study required accurate reconstruction of candidate olfaction related ORFs, a thorough validation process was utilized. To minimize the probability of annotating misassembled transcripts, a meta-assembly-like approach was chosen (*cf.* [89, 90]).

### *De novo* transcriptome assembly

The pre-processed reads of each species were assembled using the two *de novo* transcriptome assemblers Trinity release 2012-03-17 [91, 92] and Velvet v1.2.04/Oases v1.2.03 [93, 94]. Since different assembler settings show different performances in transcript reconstruction (Additional file 1: Table S1; P. Brand, pers. obs.), the assemblers were run with nine different combinations of two different parameters controlling the threshold for contig elongation in respect to overall contig-coverage. The chosen combinations represent a range from relaxed to conservative settings for both assemblers. In these combinations Parameter 1 was set to 3, 5 or 7 and Parameter 2 to 0.05, 0.10 or 0.33, where Parameter 1 refers to the `-cov_cutoff` and `-min_glue` parameters, and Parameter 2 to the `-edgeFractionCutoff` and `-min_iso_ratio` parameters of Oases and Trinity, respectively. Applying all possible combinations of the two parameters for each assembler on both species' read sets resulted in 18 assemblies per species. Trinity was always run with the default k-value of 25 while Oases was used in multiple-k-mer mode [95] with k-values of 21, 23, 25, 31, 37, 43, 49, 59, 69, 79 and 89 that were merged with a k-value of 27 using the merge mode of Oases. Minimum contig length was set to 100 bp for both assemblers.

### Detection of candidate transcripts

We applied standalone BLAT (BLAT-score threshold: 100, minimum sequence identity: 75 %; [96]) to annotate all 36 transcriptome assemblies independently using 10,602 unique *Apis mellifera* Refseq proteins not involved in olfaction as reference (accessed 10/22/12; [97]). For each species, all annotations with a completeness and contiguity  $\geq 95$  % in at least one assembly of each assembler (see Additional file 1: Table S1 for differences in assemblies) were extracted, translated to the corresponding amino acid sequences and validated via BLASTp homology searches against the Refseq proteins. To find orthologous sequences present in both species sets, we used a reciprocal BLAST approach [98]. Two annotated open reading frames (ORFs) were considered orthologous when showing identical length and at least 95 % sequence identity. In fact, the number of orthologs did not increase when decreasing the minimum required sequence identity levels to as low as 50 % in any gene

family (data not shown). Candidate orthologs were discarded when frameshifts or preliminary stop codons were present. All orthologs passing our filter settings were combined in the non-chemosensory gene set (NC genes).

To detect the maximum number of chemosensory genes we applied an iterative tBLASTn approach (*cf.* [53]). Homology searches were based on gene family specific query libraries comprised of published hymenopteran OBP, CSP, GR, IR and OR protein sequences [54–57, 99–101] (Additional file 1: Table S3). Transcripts with BLAST hits  $\leq 1e^{-06}$  were searched for all possible ORFs  $\geq 300$  bp for OBPs and CSPs and  $\geq 900$  bp for the chemosensory receptors. Detected ORFs were validated via BLASTp homology searches against the respective query library and subsequently reused as queries to search for potentially undetected ORFs. This loop was repeated until no further ORFs of sufficient length and/or e-value were detected. The pipeline to detect chemosensory genes was implemented in Perl. Scripts are available on github (<https://github.com/pbrec/CSGanalysis>).

Only those ORFs that were reconstructed by both assemblers for a given species were included in preliminary candidate ORF sets for each gene family. Orthologs present in both species transcriptomes were identified using a reciprocal BLAST approach (see above). In order to prevent annotations of non gene family members and assembly artifacts, we mapped all sequence reads back to all loci in the preliminary candidate gene sets using Bowtie2 v2.0.4 [102] in the default global alignment mode. In this way all detected putative members of the chemosensory gene family sets were curated manually in Geneious v6.0.5 [103] taking sequence coverage and mapping accuracy into account.

In rare cases, high similarity between sequences detected in the *Euglossa* transcriptomes and the reference sequences allowed for subsequent manual identifications of homologs of the other species only assembled by one assembler. The resulting chemosensory gene family sets constituted the basis of all subsequent analyses.

### Phylogenetic analyses

For each chemosensory gene family we calculated a maximum likelihood (ML) phylogenetic gene-tree using RaxML v7.2.7 [104] to infer the potential genealogical histories of the candidate euglossine gene family members. On that account, gene family specific alignments of the protein sequences of all candidate euglossine and other known hymenopteran proteins of the same gene family (see Additional file 1: Table S3 for details) were produced using MAFFT v7.031b [105, 106] applying the L-INS-I algorithm with the `-maxiterate` option set to 1000 [107].

In order to find the model for amino acid sequence evolution that fits the data best, Prottest v3.2 [108] was applied on each alignment testing for all 120 models available. The proposed model (JTT + G for ORs; CpREV + I + G for GRs; LG + I + G for IRs, OBPs and CSPs) was used to infer an unrooted ML tree in RaxML. Ten independent ML searches on ten randomized parsimony trees were conducted to find the tree with the highest likelihood. Of these, the tree with the highest likelihood was chosen and bootstrap analyses with 1000 replicates were conducted. Then, the tree was rooted by a known outgroup or, for gene families lacking a known outgroup, by mid-point rooting (See Fig. 1 and Additional file 2: Figures S1-S4).

### Signatures of selection

We used the pooled read sets representing 40 males per species to infer sites fixed for differences between all detected orthologous NC and chemosensory genes with  $\geq 10$ -fold mean per-base coverage in *Euglossa dilemma* and *E. viridissima*. Polymorphic sites were discarded since the nature of non-barcoded pooled RNA-Seq data prevents the reconstruction of individual haplotypes. Pre-processed reads were mapped onto the nucleotide sequences of the detected genes and the complete ORF together with 100 bp upstream and downstream the ORF was used as reference to allow inclusion of reads spanning the ORF and adjacent untranslated regions. For the mapping step global alignments with Bowtie2 were performed in the highly sensitive mode with -L set to 21 to adjust for the minimum read length.

A site was considered fixed if a minimum of 95 % of all reads spanning the site exhibited an identical nucleotide character at the specific site in the sequence of one species absent at the homologous site in the homologous sequence of the other species. Nucleotide characters were deemed present at a minimum of five reads, or five percent of all reads when higher than five, showing the respective nucleotide character at the site of interest. These restrictions were applied to counteract bias through probable sequencing errors. All pairs of orthologs showing fixed differences were saved as pairwise nucleotide sequence alignments and subsequently used for the maximum-likelihood estimation of pairwise  $d_N/d_S$  ratios [109] using estimated transition to transversion ratios (model M1) in codeml of the PAML package v4.6 [110]. The  $d_N/d_S$  ratios for each pairwise comparison were tested for significant deviations from a neutral model ( $d_N/d_S = 1$ ). Therefore, we conducted likelihood-ratio tests of the likelihood estimates for the M1 model and the likelihood estimated with  $d_N/d_S$  fixed to one (model M0;  $\Delta = 2(\ln(M1) - \ln(M0))$  with  $\Delta$  approximating a chi-square distribution with one degree of freedom).

### Prediction of ligand binding domains

Since non-synonymous mutations in the TM regions of ORs and in the S1 and S2 ligand binding domains of the IR related iGluRs can lead to differences in ligand binding affinities [12, 63–65, 67, 81, 82] we predicted TM topology and S1 and S2 domains for ORs and IRs, respectively, revealing patterns of positive selection.

For TM prediction we used online versions of TMPred ([http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html); [111]), TMHMM (<http://www.cbs.dtu.dk/services/TMHMM/>; [112]) and Topcons (<http://topcons.net>; [113]). The respective highest-ranking prediction of topologies was selected for each candidate OR and prediction tool. Non-synonymous substitutions were mapped onto the predicted receptor topologies and those mapping within TM regions supported by at least two of the three prediction tools were called TM substitutions. This step was necessary, since bioinformatic tools for TM prediction can deviate in their outputs for the same genes (e.g. [114]).

We aligned candidate *Euglossa* IRs and the conserved *Drosophila melanogaster* IR8a (DmelIR8a) receptor using MAFFT as described earlier to predict the S1 and S2 ligand binding domains. On that account, the known S1 and S2 ligand binding sites of DmelIR8a [42] were used to transfer annotations to the *Euglossa* IRs.

### Availability of supporting data

The data sets supporting the results of this article are included within the article and its additional files. Raw sequence reads are available at the NCBI Sequence Read Archive [SRA: SRX765918, SRA: SRX765888].

### Additional files

**Additional file 1: Supplementary Tables S1-S6.** (XLSX 3214 kb)

**Additional file 2: Supplementary Figures S1-S5.** (PDF 3452 kb)

### Abbreviations

CSP: Chemosensory protein; Gb: Gigabases; GR: Gustatory receptor; HNDB: 2-hydroxy-6-nona-1,3-dienyl-benzaldehyde; iGluR: Ionotropic glutamate receptor; IR: Ionotropic receptor; ML: Maximum likelihood; NC: Non-chemosensory; OBP: Odorant-binding protein; OR: Odorant receptor; ORF: Open reading frame; TM: Transmembrane.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

TE, SR, FL, PB conceived and designed the experiments. PB performed all experiments, analyzed the data and drafted the manuscript. SR, TE, FL helped writing the paper and discussed the data. TE, FL, RT, JQE contributed reagents and materials. All authors read and approved the final version of the manuscript.

### Acknowledgements

The authors thank Tamara Pokorny for providing bee samples, Anna Eckart for technical assistance and Andrey Rozenberg for discussions on the bioinformatic methods of the project. We also thank Klaus Lunau and Martin Beye for providing laboratory resources for antennal preparations at the

Heinrich-Heine University, Düsseldorf. We thank Julie Cridland for discussions and two anonymous reviewers for comments on the manuscript. This work was supported by the Germany Scholarship and a fellowship of the Deutscher Akademischer Austauschdienst to PB, a German Science Foundation (El 249/6) grant to TE, the Ruhr University Bochum, and in part by a grant of the Dinter Foundation within the Deutsches Stiftungszentrum (Essen) to FL and RT. SR received support from The David and Lucile Packard Foundation.

#### Author details

<sup>1</sup>Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Universitätsstrasse 150, D-44801 Bochum, Germany. <sup>2</sup>Department for Evolution and Ecology, Center for Population Biology, University of California Davis, One Shields Avenue, 95616 Davis, USA. <sup>3</sup>Departamento de Apicultura, Universidad Autónoma de Yucatán, Mérida, Mexico. <sup>4</sup>Present address: Faculty of Biology, Aquatic Ecosystems Research, University of Duisburg and Essen, Universitätsstrasse 5, D-45141 Essen, Germany.

Received: 22 March 2015 Accepted: 10 August 2015

Published online: 28 August 2015

#### References

- Hansson BS. *Insect Olfaction*. Heidelberg: Springer Verlag; 1999.
- Mombaerts P. Genes and ligands for odorant, vomeronasal and taste receptors. *Nat Rev Neurosci*. 2004;5:263–78.
- Birch MC, Haynes KF. *Insect Pheromones*. London: Hodder Arnold; 1982.
- Wyatt TD. Fifty years of pheromones. *Nature*. 2009;457:262–3.
- Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR. Integrating the molecular and cellular basis of odor coding in the drosophila antenna. *Neuron*. 2003;37:827–41.
- Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annu Rev Entomol*. 2013;58:373–91.
- Smadja C, Butlin RK. On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity*. 2009;102:77–97.
- Hansson BS, Stensmyr MC. Evolution of insect olfaction. *Neuron*. 2011;72:698–711.
- Lassance J-M, Bogdanowicz SM, Wanner KW, Löfstedt C, Harrison RG. Gene genealogies reveal differentiation at sex pheromone olfactory receptor loci in pheromone strains of the European corn borer, *Ostrinia nubilalis*. *Evolution*. 2011;65:1583–93.
- Vásquez GM, Syed Z, Estes PA, Leal WS, Gould F. Specificity of the receptor for the major sex pheromone component in *Heliothis virescens*. *J Insect Sci*. 2013;13:160–12.
- Jiang X-J, Guo H, Di C, Yu S, Zhu L, Huang L-Q, et al. Sequence similarity and functional comparisons of pheromone receptor orthologs in two closely related *Helicoverpa* species. *Insect Biochem Mol Biol*. 2014;48:63–74.
- Leary GP, Allen JE, Bunger PL, Luginbill JB, Linn CE, Macallister IE, et al. Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proc Natl Acad Sci U S A*. 2012;109:14081–6.
- Kopp A, Barmina O, Hamilton AM, Higgins L, McIntyre LM, Jones CD. Evolution of gene expression in the *Drosophila* olfactory system. *Mol Biol Evol*. 2008;25:1081–92.
- Vásquez GM, Fischer P, Grozinger CM, Gould F. Differential expression of odorant receptor genes involved in the sexual isolation of two *Heliothis* moths. *Insect Mol Biol*. 2011;20:115–24.
- Wang G, Vásquez GM, Schal C, Zwiebel LJ, Gould F. Functional characterization of pheromone receptors in the tobacco budworm *Heliothis virescens*. *Insect Mol Biol*. 2011;20:125–33.
- McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci U S A*. 2007;104:4996–5001.
- Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. Evolution of herbivory in *Drosophilidae* linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci U S A*. 2015;112:3026–31 (2015).
- Boake CRB. Coevolution of senders and receivers of sexual signals: genetic coupling and genetic correlations. *Trends Ecol Evol*. 1991;6:225–7.
- Tillman JA, Seybold SJ, Jurenka RA, Blomquist GJ. Insect pheromones—an overview of biosynthesis and endocrine regulation. *Insect Biochem Mol Biol*. 1999;29:481–514.
- Niehuis O, Buellesbach J, Gibson JD, Pothmann D, Hanner C, Mutti NS, et al. Behavioural and genetic analyses of *Nasonia* shed light on the evolution of sex pheromones. *Nature*. 2013;494:345–8.
- Ramírez S, Dressler RL, Ospina M. Abejas euglosinas (Hymenoptera: Apidae) de la región Neotropical: listado de especies con notas sobre su biología. *Biota Colombiana*. 2002;3:7–118.
- Kimsey LS. The behaviour of male orchid bees (Apidae, Hymenoptera, Insecta) and the question of leks. *Anim Behav*. 1980;28:996–1004.
- Williams NH. The biology of orchids and euglossine bees in J. Arditti (ed.). *Orchid Biology: Reviews and Perspectives*, II. Cornell University Press, Ithaca, NY. 1982;119–71.
- Cameron SA. Phylogeny and biology of neotropical orchid bees (Euglossini). *Annu Rev Entomol*. 2004;49:377–404.
- Eltz T, Ayasse M, Lunau K. Species-specific antennal responses to tibial fragrances by male orchid bees. *J Chem Ecol*. 2006;32:71–9.
- Zimmermann Y, Roubik DW, Eltz T. Species-specific attraction to pheromonal analogues in orchid bees. *Behav Ecol Sociobiol*. 2006;60:833–43.
- Eltz T, Zimmermann Y, Pfeiffer C, Pech JR, Twele R, Francke W, et al. An olfactory shift is associated with male perfume differentiation and species divergence in orchid bees. *Curr Biol*. 2008;18:1844–8.
- Ackerman JD. Specificity and mutual dependency of the orchid-euglossine bee interaction. *Biol J Linn Soc*. 1983;20:301–14.
- Whitten WM, Young AM, Stern DL. Nonfloral sources of chemicals that attract male euglossine bees (Apidae: Euglossini). *J Chem Ecol*. 1993;19:3017–27.
- Eltz T, Whitten WM, Roubik DW, Linsenmair KE. Fragrance collection, storage, and accumulation by individual male orchid bees. *J Chem Ecol*. 1999;25:157–76.
- Pemberton RW, Wheeler GS. Orchid bees don't need orchids: evidence from the naturalization of an orchid bee in Florida. *Ecology*. 2006;87:1995–2001.
- Zimmermann Y, Ramírez SR, Eltz T. Chemical niche differentiation among sympatric species of orchid bees. *Ecology*. 2009;90:2994–3008.
- Kimsey LS. The behavioural and structural aspects of grooming and related activities in euglossine bees (Hymenoptera: Apidae). *J Zool*. 1984;204:541–50.
- Bembé B. Functional morphology in male euglossine bees and their ability to spray fragrances (Hymenoptera, Apidae, Euglossini). *Apidologie*. 2004;283–291.
- Eltz T, Sager A, Lunau K. Juggling with volatiles: exposure of perfumes by displaying male orchid bees. *J Comp Physiol A*. 2005;191:575–81.
- Schorkopf DLP, Mitko L, Eltz T. Enantioselective preference and high antennal sensitivity for (–)-lpsdienol in scent-collecting male orchid bees, *Euglossa cyanura*. *J Chem Ecol*. 2011;37:953–60.
- Sakurai T, Nakagawa T, Mitsuno H, Mori H, Endo Y, Tanoue S, et al. Identification and functional characterization of a sex pheromone receptor in the silkworm *Bombyx mori*. *Proc Natl Acad Sci U S A*. 2004;101:16653–8.
- Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM. A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid. *Proc Natl Acad Sci U S A*. 2007;104:14383–8.
- Miura N, Nakagawa T, Touhara K, Ishikawa Y. Broadly and narrowly tuned odorant receptors are involved in female sex pheromone reception in *Ostrinia* moths. *Insect Biochem Mol Biol*. 2010;40:64–73.
- Wanner KW, Nichols AS, Allen JE, Bunger PL, Garczynski SF, Linn CE, et al. Sex pheromone receptor specificity in the European corn borer moth, *Ostrinia nubilalis*. *PLoS One*. 2010;5:e8685.
- Jones WD, Cayirlioglu P, Kadow IG, Vossball LB. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature*. 2007;445:86–90.
- Benton R, Vannice KS, Gomez-Diaz C, Vossball LB. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell*. 2009;136:149–62.
- Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GSXE, et al. Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *J Neurosci*. 2011;31:13357–75.
- Pelosi P, Calvillo M, Ban L. Diversity of odorant-binding proteins and chemosensory proteins in insects. *Chem Senses*. 2005;30 Suppl 1:i291–2.

45. Ozaki M, Wada-Katsumata A, Fujikawa K, Iwasaki M, Yokohari F, Satoji Y, et al. Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. *Science*. 2005;309:311–4.
46. Grosse-Wilde E, Svatoš A, Krieger J. A pheromone-binding protein mediates the bombykol-induced activation of a pheromone receptor *in vitro*. *Chem Senses*. 2006;31:547–55.
47. Grosse-Wilde E, Gohl T, Bouché E, Breer H, Krieger J. Candidate pheromone receptors provide the basis for the response of distinct antennal neurons to pheromonal compounds. *Eur J Neurosci*. 2007;25:2364–73.
48. Laughlin JD, Ha TS, Jones DNM, Smith DP. Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein. *Cell*. 2008;133:1255–65.
49. Eltz T, Fritzsche F, Pech JR, Zimmermann Y, Ramírez S, Quezada-Euan JGG, et al. Characterization of the orchid bee *Euglossa viridissima* (Apidae: Euglossini) and a novel cryptic sibling species, by morphological, chemical, and genetic characters. *Zool J Linn Soc*. 2011;163:1064–76.
50. Villanueva-Gutiérrez R, Quezada-Euan J, Eltz T. Pollen diets of two sibling orchid bee species, *Euglossa*, in Yucatán, southern Mexico. *Apidologie*. 2013;44:440–6.
51. Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell*. 1999;96:725–36.
52. Vosshall LB, Wong AM, Axel R. An olfactory sensory map in the fly brain. *Cell*. 2000;102:147–59.
53. Mitchell RF, Hughes DT, Luetje CW, Millar JG, Soriano-Agatón F, Hanks LM, et al. Sequencing and characterizing odorant receptors of the cerambycid beetle *Megacyllene caryae*. *Insect Biochem Mol Biol*. 2012;42:499–505.
54. Robertson HM, Wanner KW. The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. *Genome Res*. 2006;16:1395–403.
55. Crosset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet*. 2010;6:e1001064.
56. Yin X-W, Iovinella I, Marangoni R, Cattonaro F, Flamini G, Sagona S, et al. Odorant-binding proteins and olfactory coding in the solitary bee *Osmia cornuta*. *Cell Mol Life Sci*. 2013;70:3029–39.
57. Forêt S, Maleszka R. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res*. 2006;16:1404–13.
58. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron*. 2004;43:703–14.
59. Jones WD, Nguyen T-AT, Kloss B, Lee KJ, Vosshall LB. Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr Biol*. 2005;15:R119–21.
60. Claudianos C, Lim J, Young M, Yan S, Cristino AS, Newcomb RD, et al. Odor memories regulate olfactory receptor expression in the sensory periphery. *Eur J Neurosci*. 2014;39:1642–54.
61. Mugal CF, Wolf JBW, Kaj I. Why time matters: codon evolution and the temporal dynamics of dN/dS. *Mol Biol Evol*. 2014;31:212–31.
62. Bielawski JP, Yang Z. Maximum likelihood methods for detecting adaptive protein evolution. In: *Statistical Methods in Molecular Evolution*. New York: Springer New York; 2005. p. 103–24 [Statistics for Biology and Health].
63. Inomata N, Goto H, Itoh M, Isono K. A single-amino-acid change of the gustatory receptor gene, Gr5a, has a major effect on trehalose sensitivity in a natural population of *Drosophila melanogaster*. *Genetics*. 2004;167:1749–58.
64. Pellegrino M, Steinbach N, Stensmyr MC, Hansson BS, Vosshall LB. A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. *Nature*. 2011;478:511–4.
65. Richgels PK, Rollmann SM. Genetic variation in odorant receptors contributes to variation in olfactory behavior in a natural population of *Drosophila melanogaster*. *Chem Senses*. 2012;37:229–40.
66. Hughes DT, Wang G, Zwiebel LJ, Luetje CW. A determinant of odorant specificity is located at the extracellular loop 2-transmembrane domain 4 interface of an *Anopheles gambiae* odorant receptor subunit. *Chem Senses*. 2014;39:761–9.
67. Xu P, Leal WS. Probing insect odorant receptors with their cognate ligands: insights into structural features. *Biochem Biophys Res Commun*. 2013;435:477–82.
68. Tunstall NE, Sirey T, Newcomb RD, Warr CG. Selective pressures on *Drosophila* chemosensory receptor genes. *J Mol Evol*. 2007;64:628–36.
69. Carraher C, Authier A, Steinwender B, Newcomb RD. Sequence comparisons of odorant receptors among tortricid moths reveal different rates of molecular evolution among family members. *PLoS One*. 2012;7:e38391.
70. Miller R, Tu Z. Odorant receptor c-terminal motifs in divergent insect species. *J Insect Sci*. 2008;8:1–10.
71. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol Ecol*. 2008;17:1648–57.
72. Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. Family size evolution in *Drosophila* chemosensory gene families: a comparative analysis with a critical appraisal of methods. *Genome Biol Evol*. 2014;6:1669–82.
73. Guo S, Kim J. Molecular evolution of *Drosophila* odorant receptor genes. *Mol Biol Evol*. 2007;24:1198–207.
74. McBride CS, Arguello JR, O'Meara BC. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*. 2007;177:1395–416.
75. Willett CS. Evidence for directional selection acting on pheromone-binding proteins in the genus *Choristoneura*. *Mol Biol Evol*. 2000;17:553–62.
76. Endo T, Ikeo K, Gojobori T. Large-scale search for genes on which positive selection may operate. *Mol Biol Evol*. 1996;13:685–90.
77. Peterson GI, Masel J. Quantitative prediction of molecular clock and ka/ks at short timescales. *Mol Biol Evol*. 2009;26:2595–603.
78. Smadja C, Shi P, Butlin RK, Robertson HM. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrtosiphon pisum*. *Mol Biol Evol*. 2009;26:2073–86.
79. Smadja CM, Canbäck B, Vitalis R, Gautier M, Ferrari J, Zhou J-J, et al. Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*. 2012;66:2723–38.
80. Civetta A, Singh RS. Sex-related genes, directional sexual selection, and speciation. *Mol Biol Evol*. 1998;15:901–9.
81. Mayer ML. Glutamate receptors at atomic resolution. *Nature*. 2006;440:456–62.
82. Naur P, Hansen KB, Kristensen AS, Dravid SM, Pickering DS, Olsen L, et al. Ionotropic glutamate-like receptor delta2 binds D-serine and glycine. *Proc Natl Acad Sci U S A*. 2007;104:14116–21.
83. Wiley C, Ellison CK, Shaw KL. Widespread genetic linkage of mating signals and preferences in the Hawaiian cricket *Laupala*. *Proc Biol Sci*. 2012;279:1203–9.
84. Roubik DW, Hanson PE. *Orchid Bees of Tropical America: Biology and Field Guide*. Santo Domingo De Heredia: Instituto Nacional de Biodiversidad (INBio); 2004.
85. Pokorny T, Hannibal M, Quezada-Euan JGG, Hedenström E, Sjöberg N, Bång J, et al. Acquisition of species-specific perfume blends: influence of habitat-dependent compound availability on odour choices of male orchid bees (*Euglossa* spp.). *Oecologia*. 2013;172:417–25.
86. Burriesci MS, Lehnert EM, Pringle JR. Fulcrum: condensing redundant reads from high-throughput sequencing studies. *Bioinformatics*. 2012;28:1324–7.
87. Martin J, Bruno VM, Fang Z, Meng X, Blow M, Zhang T, et al. Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads. *BMC Genomics*. 2010;11:663.
88. Falgueras J, Lara AJ, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, Claros MG. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics*. 2010;11:38.
89. Kumar S, Blaxter ML. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics*. 2010;11:571.
90. Feldmeyer B, Wheat CW, Krezdom N, Rotter B, Pfenninger M. Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, Basommatophora, Pulmonata), and a comparison of assembler performance. *BMC Genomics*. 2011;12:317.
91. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
92. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8:1494–512.
93. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18:821–9.

94. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*. 2012;28:1086–92.
95. Surget-Groba Y, Montoya-Burgos JI. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res*. 2010;20:1432–40.
96. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res*. 2002;12:656–64.
97. Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI reference sequences: current status, policy and new initiatives. *Nucleic Acids Res*. 2009;37(Database issue):D32–6.
98. Rivera MC, Jain R, Moore JE, Lake JA. Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A*. 1998;95:6239–44.
99. Robertson HM, Gadau J, Wanner KW. The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*. *Insect Mol Biol*. 2010;19 Suppl 1:121–36.
100. Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol Evol*. 2011;3:476–90.
101. Vieira FG, Forêt S, He X, Rozas J, Field LM, Zhou J-J. Unique features of odorant-binding proteins of the parasitoid wasp *Nasonia vitripennis* revealed by genome annotation and comparative analyses. *PLoS One*. 2012;7:e43034.
102. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
103. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
104. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22:2688–90.
105. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059–66.
106. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
107. Katoh K, Kuma K-I, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 2005;33:511–8.
108. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27:1164–5.
109. Bielawski JP, Yang Z. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol*. 2004;59:121–32.
110. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
111. Hoffmann K. TMBASE-A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler*. 1993;374:166.
112. Krogh A, Larsson B, Heijne von G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567–80.
113. Bernal A, Viklund H, Hønnedal A, Elofsson A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res*. 2009;37(Web Server issue):W465–8.
114. Möller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*. 2001;17:646–53.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

