

Neofunctionalization in Ligand Binding Sites of Ant Olfactory Receptors

Rana Saad¹, Amir B. Cohanim¹, Mickey Kosloff^{2,*}, and Eyal Privman^{1,*}

¹Department of Evolutionary and Environmental Biology, Institute of Evolution, University of Haifa, Israel

²Department of Human Biology, University of Haifa, Israel

*Corresponding authors: E-mails: eprivman@univ.haifa.ac.il; kosloff@sci.haifa.ac.il.

Accepted: June 22, 2018

Data deposition: De novo annotated OR gene sequences are available in supplementary information files S1-9, and also deposited in the Figshare database (DOI:10.6084/m9.figshare.5900275).

Abstract

Chemical communication is fundamental for the operation of insect societies. Their diverse vocabulary of chemical signals requires a correspondingly diverse set of chemosensory receptors. Insect olfactory receptors (ORs) are the largest family of chemosensory receptors. The OR family is characterized by frequent expansions of subfamilies, in which duplicated ORs may adapt to detect new signals through positive selection on their amino acid sequence. Ants are an extreme example with ~400 ORs per genome—the highest number in insects. Presumably, this reflects an increased complexity of chemical communication. Here, we examined gene duplications and positive selection on ant ORs. We reconstructed the hymenopteran OR gene tree, including five ant species, and inferred positive selection along every branch using the branch-site test, a total of 3326 tests. We find more positive selection in branches following species-specific duplications. We identified amino acid sites targeted by positive selection, and mapped them onto a structural model of insect ORs. Seventeen sites were under positive selection in six or more branches, forming two clusters on the extracellular side of the receptor, on either side of a cleft in the structure. This region was previously implicated in ligand activation, suggesting that the concentration of positively selected sites in this region is related to adaptive evolution of ligand binding sites or allosteric transmission of ligand activation. These results provide insights into the specific OR subfamilies and individual residues that facilitated adaptive evolution of olfactory functions, potentially explaining the elaboration of chemical signaling in ant societies.

Key words: gene duplication, positive selection, protein structure, olfactory receptors, social insects.

Introduction

Ants represent an extreme example of social behavior in animals and one of the most ecologically successful animal groups in the ecosystem. Chemical communication is central for the efficient operation and coordination of an ant colony, including distinction between reproductive (queens) and nonreproductive castes (workers), coordination and division of labor among worker groups. Ants also discriminate between nestmates and nonnestmates based on a shared, colony-specific odor, which is comprised of a complex mixture of very long chain cuticular hydrocarbons (Nowbahari et al. 1990; Soroker et al. 1995; Akino et al. 2004; Ozaki et al. 2005; Martin et al. 2008).

Insects detect odors and pheromones by chemosensory receptors. In ants, the olfactory receptor (OR) gene family stands out due to its dramatic expansion to ~300–400 OR genes in every ant genome (Zhou et al. 2012), the largest number in all studied insects. Although new OR genes may underlie various kinds of olfactory functions, it is commonly suggested that this dramatic expansion is primarily related to the evolution of complex chemical communication systems in ant societies (Smith CR et al. 2011). Ants use large numbers of pheromones, which may be composed of complex mixtures. For example, nestmate recognition is based on mixtures of hydrocarbons displayed on the ants' cuticle. In the desert ant *Cataglyphis niger*, a mixture of 55 different cuticular hydrocarbons were identified (Saar et al. 2014) and the composition

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

of this mixture changes rapidly among even closely related species (Martin and Drijfhout 2009). Olfactory perception of pheromones is vital for cooperative functions of the nest, and should therefore be an important fitness factor in ants. Hence, new OR genes with new olfactory functions are expected to evolve together with the evolution of ant social behavior.

Gene duplication plays a central role in evolutionary innovation—a gene is duplicated and the new paralogs evolve novel beneficial functions, possibly via positive selection (Ohno 1970). Gene duplication can be followed by neofunctionalization, where one of the paralogs evolves a new function while the other maintains the original function (Force et al. 1998). Alternatively, in subfunctionalization, a multifunctional gene is duplicated and the paralogs retain different subfunctions of the ancestral gene. In *Drosophila*, two thirds of gene duplications fit the neofunctionalization model (Assis and Bachtrog 2013). Thus, duplication of ant ORs may underlie the evolution of new olfactory functions. OR duplication may be followed by positive selection on the ligand binding site to recognize new odorants (neofunctionalization), or the duplicated ORs may become more specialized in detecting specific odors instead of one OR that responds to several similar odors (subfunctionalization). Different paralogs may also evolve differential regulation of gene expression or kinetics of ligand activation and deactivation.

ORs are seven-transmembrane domain proteins (Clyne et al. 1999; Vosshall et al. 1999), but in contrast to vertebrate ORs, insect ORs show an inverted membrane orientation—cytoplasmic N-terminus and extracellular C-terminus (Benton et al. 2006; Lundin et al. 2007; Smart et al. 2008). Insect ORs are heterodimers consisting of one of many specific ligand-binding ORs and one conserved coreceptor known as ORCO (Larsson et al. 2004). The heterodimer has been suggested to function as a ligand-gated ion channel, which upon activation leads to depolarization of olfactory sensory neurons (Sato et al. 2008; Wicher et al. 2008). The specific OR dictates ligand selectivity and may also transduce the signal by activating a G protein (Wicher et al. 2008). Relevantly, residues 146–150 in transmembrane helix 3 (TM3) and the adjacent extracellular domain 2 (EC2) of *Drosophila melanogaster* OR 85b (DmOR85b) have been shown to contribute to ligand binding (Nichols and Luetje 2010). Which additional residues contribute to ligand binding and receptor activation remains an open question.

Roux et al. (2014) and Engson et al. (2015) analyzed the gene phylogeny and inferred positive selection on ant ORs of two and five ant species, respectively. Engson et al. proposed a hypothesis where species-specific gene expansions evolve under positive selection, as they are associated with olfactory adaptation unique to each lineage. They tested for positive selection in these expansions using the site test of positive selection. However, this test does not

compare lineage-specific branches to the nonlineage-specific branches of the tree. Here, we apply the branch-site test, which allows this comparison, and thereby testing the hypothesis of more positive selection in species-specific expansions. We find a statistically significant difference between species specific ORs and nonspecies specific ORs, with slightly more positive selection on species-specific ORs. There is extensive evidence for positive selection in most OR subfamilies, with periods of multiple consecutive gene duplications and positive selection on multiple paralogs. Amino acids targeted by positive selection are primarily concentrated around the putative ligand binding site, which suggests adaptation for the perception of new ligands and elaboration and fine-tuning of the response to existing ligands.

Materials and Methods

OR Gene Sequences

OR sequences were collected from four previously annotated ant genomes (table 1) and we de novo annotated three more genomes: *Solenopsis invicta* (genome assembly version Si_gnH; NCBI accession AEAQ00000000), *S. fugax* (Sf_gnA; accession QKQZ00000000), and *Monomorium pharaonis* (Mp_gnA; accession QKWU00000000). De novo gene annotation was conducted using genBlast (She et al. 2009, 2011) (<http://genome.sfu.ca/genblast/>; last accessed July 2014), in which translated BLAST hits to exons were grouped to represent putative gene models, while stitching hits at predicted splice site junctions. The previously annotated ORs were used as queries for the BLAST searches. Between 239 and 461 ORs were annotated per species (table 1). Novel gene sequences are available in supplementary information files S1–S9, Supplementary Material online and also deposited in the Figshare database (DOI: 10.6084/m9.figshare.5900275). ORs of the wasp *Nasonia vitripennis* and the honeybee *Apis mellifera* were also included in the phylogenetic analysis as outgroups.

Phylogenetic Analysis

Nucleotide sequences were translated to amino acid sequences and aligned using MAFFT (version 7; accurate variant E-INS-i; Katoh et al. 2005). The OR gene tree was constructed using RAXML (version 8.1.15; Stamatakis 2006) with the PROTCATLG model, the LG rate matrix, and 100 bootstraps repeats. The phylogenetic tree included 2793 sequences and 5583 branches, which is too large for any dN/dS-based positive selection inference method. Thus, the tree was divided to 31 clades (subtrees) at branches with the highest possible bootstrap support, so that the largest clade contained 144 sequences. Clades were defined as monophyletic groups wherever possible, with the exception of clade 28, which is paraphyletic. In a few cases, subtrees were split at low

Table 1

Olfactory Receptors of the Species Included in the Study

	Species Names (abbreviation)	Number of ORs	References
Ants	<i>Linepithema humile</i> (LhOR)	320	Smith CD et al. (2011)
	<i>Pogonomyrmex barbatus</i> (PbOR)	291	Smith CR et al. (2011)
	<i>Harpegnathos saltator</i> (HsOR)	461	Zhou (2012)
	<i>Camponotus floridanus</i> (CfOR)	377	Zhou (2012)
	<i>Solenopsis invicta</i> (SiOR)	407	De novo annotation, here
	<i>Solenopsis fugax</i> (SfOR)	314	De novo annotation, here
	<i>Monomorium pharaonis</i> (MpOR)	239	De novo annotation, here
Bee	<i>Apis mellifera</i>	168	Robertson and Wanner (2006)
Wasp	<i>Nasonia vitripennis</i>	262	Werren et al. (2010)

bootstrap branches, which might also lead to subsets of sequences that are not monophyletic clades. However, the downstream analysis (the branch-site test) does not rely on an assumption of monophyly, so this should not lead to artifacts in our inference of positive selection. Due to the very long running time of the branch-site test (17 days for each branch in a tree of 200 sequences) we also restricted the positive selection analysis to the five more closely related ant species, including the four myrmicines: *P. barbatus* (PbOR), *S. invicta* (SiOR), *M. pharaonis* (MpOR), and *S. fugax* (SfOR); and one formicine: *C. floridanus* (CfOR). After this reduction, the sequences of each clade were realigned using *GUIDANCE* (version 2.01; Penn et al. 2010) with the aligner *PAGAN* (Löytynoja et al. 2012) as a codon alignment by reverse-translating the amino acid alignment. Unreliably aligned residues were masked at a *GUIDANCE* score cutoff of 0.8 by replacing low-scoring codons with “NNN.” Then, phylogenetic trees were built using *RAXML* for each clade as above, based on the amino acid alignment.

Positive Selection Inference

Each branch of the gene tree was tested for positive selection using the modified branch-site test (model A; Zhang et al. 2005), as implemented in *codeml* in the *PAML* package (version 3.15; Yang and Nielsen 2000). *Codeml* uses maximum likelihood to estimate the ω parameter, which represents the dN/dS ratio in the subset of sites under positive selection. Positive selection was inferred by comparing a model that allows positive selection ($dN/dS > 1$) only in one “foreground branch” with a null model that does not allow any positive selection ($dN/dS \leq 1$). Since the compared models are nested, likelihood ratio tests (LRT) were used to reject the null model. Thus, *P* values were obtained for each tested branch, and adjusted into *q* values to control the false discovery rate (FDR; Benjamini and Hochberg 1995) using the Bioconductor R package “*qvalue*” (Bass et al. 2015). For each branch where the LRT is successful at FDR of 10%, specific codons under positive selection

were identified based on posterior probability >0.9 for $dN/dS > 1$.

Mapping Selected Residues to the Protein Structure

Amino acid positions under positive selection in ant ORs were mapped onto a structural model of insect ORs, built by Hopf et al. (2015) using the evolutionary coupling method (Marks et al. 2012), which generated a 3D structure for the sequence of OR85b of *Drosophila melanogaster* (DmOR85b). To map the positions from the alignments of each of the 31 ant ORs clades to the DmOR85b structure, one PbOR representative from each clade and the DmOR85b sequence were aligned using *MAFFT* with the FFT-NS-i method. Positions in the ant alignments with posterior probability of 0.9 for positive selection were mapped to the homologous positions in the DmOR85b sequence using a perl script (included in [Supplementary Material](#)). Positions in insertions/deletions in either ant or fly sequences were mapped to the nearest amino acid position in the homologous sequence. Then, these amino acid positions were visualized in the 3D structure of DmOR85b using the molecular visualization software *Pymol* (version 1.3; DeLano 2002). The transmembrane regions of the structure were predicted based on amino acids with positive scores on the Kyle–Doolittle hydrophathy scale.

Results

OR Gene Tree

The OR gene tree was reconstructed based on the amino acid sequences of seven ant species: *Harpegnathos saltator*, *Linepithema humile*, *Camponotus floridanus*, *Pogonomyrmex barbatus*, *Monomorium pharaonis*, *Solenopsis fugax*, and *Solenopsis invicta*. The jewel wasp *Nasonia vitripennis* and honeybee *Apis mellifera* were included as outgroups. The tree includes 2793 sequences and 5583 branches. The ORCO subtree was used as an outgroup to root the tree. This tree is too large for positive selection inference, so we divided it to 31 clades, with no more than 200 sequences in each clade (fig. 1; the full tree

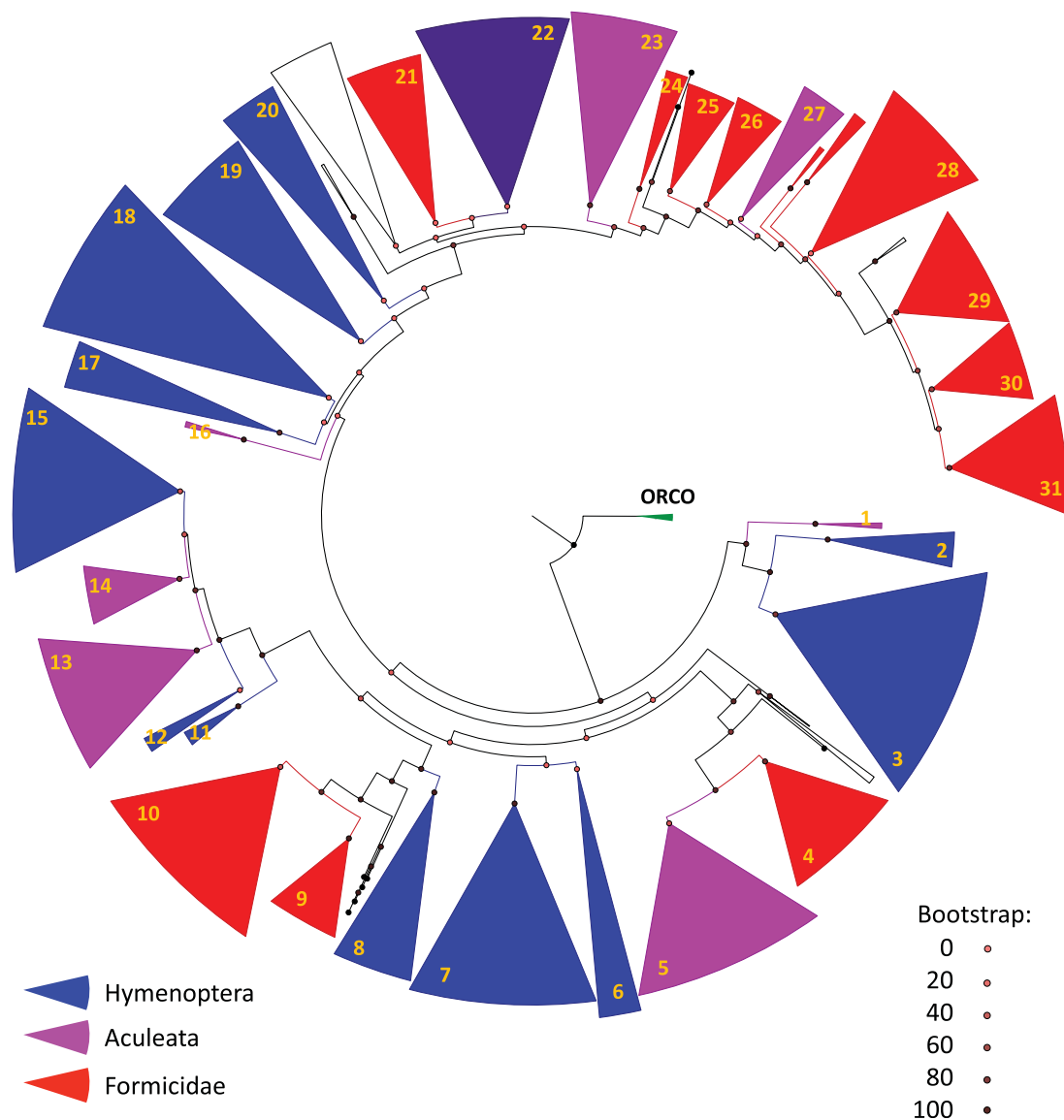


Fig. 1.—OR gene tree including 2,973 genes from seven ants, honeybee, and jewel wasp. The tree was reconstructed using *RAxML* based on a *MAFFT* amino acid alignment and divided into 31 clades for subsequent analyses. Formicidae-specific clades are colored red (only ant sequences), Aculeatan-specific clades are colored purple (include bee sequences), and Hymenopteran clades are colored blue (include wasp sequences). Clades colored white were not tested for positive selection.

is included in the [Supplementary Material](#) with distinct coloring of each clade). The hymenopteran OR gene tree consists of many subtrees that include ant, bee, and wasp genes, which represent paralogs that evolved in a common ancestor of these Hymenoptera. Each subtree contains more recently evolved paralogs, including gene duplications before the divergence of ants and bees and many more subsequent ant-specific duplications.

Extensive Positive Selection across OR Subfamilies

Due to computational run-time constraints of the branch-site test, positive selection was inferred in each of the 31

clades and only the five most closely related ant species were included in the analysis (see Materials and Methods). The reduced clades included 1706 sequences in total. A test for positive selection was applied to each branch of each clade, 3,326 branches in total. We also required that at least one codon site had a posterior probability >0.9 for $dN/dS > 1$. After correction for multiple testing (false discovery rate $q < 0.1$), this analysis identified 261 branches with evidence for positive selection ([table 2](#)). Virtually all clades contain some branches with positive selection and between 1 and 75 amino acid sites under positive selection.

A representative subfamily (clade 22) is shown in figure 2a (the other 30 are included as supplementary figs.,

Table 2

Branches and Sites Inferred to be Under Positive Selection Based on the Branch-Site Test

Clade	Number of Sequences	Branches with Positive Selection (%)	Number of Sites with $P > 0.9$ for $dN/dS > 1$
Clade1	5	0 (0.0%)	0
Clade2	21	1 (2.6%)	1
Clade3	144	22 (7.7%)	60
Clade4	78	14 (9.2%)	45
Clade5	123	12 (4.9%)	19
Clade6	19	2 (5.7%)	10
Clade7	124	21 (8.6%)	44
Clade8	39	3 (4.0%)	6
Clade9	35	4 (6.0%)	6
Clade10	141	21 (7.5%)	75
Clade11	10	1 (5.9%)	1
Clade12	8	0 (0.0%)	0
Clade13	72	13 (9.2%)	25
Clade14	36	4 (5.8%)	14
Clade15	84	12 (7.3%)	54
Clade16	6	0 (0.0%)	0
Clade17	15	5 (18.5%)	74
Clade18	74	9 (6.2%)	26
Clade19	47	8 (8.8%)	23
Clade20	27	4 (7.8%)	9
Clade21	54	16 (15.2%)	63
Clade22	98	17 (8.8%)	52
Clade23	56	11 (10.1%)	46
Clade24	19	2 (5.7%)	2
Clade25	34	3 (4.6%)	5
Clade26	36	7 (10.1%)	13
Clade27	14	1 (4.0%)	2
Clade28	90	10 (5.6%)	21
Clade29	81	15 (9.4%)	25
Clade30	48	7 (7.5%)	12
Clade31	68	16 (12.0%)	34
Total	1,706	261	

NOTE.—Positive results are those that passed the LRT (at false discovery rate $q < 0.1$) and had at least one site with posterior probability > 0.9 for $dN/dS > 1$.

Supplementary Material online). Fifteen out of 193 branches in this clade (7.8%) were inferred to be under positive selection (false discovery rate $q < 0.1$). Branch 121 had the largest number of sites under positive selection (13 codon sites). An inspection of the sequence alignment shows that these sites are located in reliable alignment blocks with no obvious alignment problems that could lead to false inference of positive selection (fig. 2c). Randomly chosen branches from additional clades were likewise manually inspected, leading us to estimate that no more than 10% of the branches inferred to be under positive selection might be affected by unreliable alignment.

Seven of the positively selected branches in clade 22 are immediately following a gene duplication event. For example, the root of this clade (branch 165) is a gene duplication event

that created two paralogs that were again duplicated, and each of these three duplications was followed by positive selection on at least one of the paralogs (branches 60 and 121). Most of these events preceded the divergence of the lineages represented by the five ant species included in this analysis, that is, before the divergence of Formicinae from Myrmicinae. Subsequently, there were many more duplications, including species-specific duplications (marked by brackets in fig. 2a), but these were generally not followed by positive selection on the species-specific paralogs.

Positive Selection following Species-Specific Expansions

The OR gene tree reveals dramatic expansions in many subfamilies. Many of the duplication events are species-specific, as highlighted for the example of clade 22 in figure 2a. Each of the studied species had numerous species-specific duplications in most of the clades, with the exception of the smallest clades (table 3). To test whether more paralogs are under positive selection after species-specific duplications compared with nonspecies-specific duplications, we counted positively selected branches following species-specific duplications. In total, 81 out of 905 (9.0%) species-specific branches were under positive selection and 171 out of 2602 (6.6%) nonspecies-specific branches were under positive selection, which is a statistically significant difference (two-tailed Fisher's exact test P value = 0.02).

Positively Selected Sites in the Context of Structural Domains

We mapped the amino acid positions under positive selection to the 3D structural model of insect ORs that was built using the evolutionary coupling method (Marks et al. 2012) by Hopf et al. (2015). The number of branches in which an amino acid site was inferred to be under positive selection was counted (table 4) and mapped to the 3D structure of the protein (fig. 3). Seventeen sites were under positive selection in six or more, and up to twelve branches, suggesting recurrent episodes of adaptive evolution in the same sites. Only two of these sites are located in the large intracellular region. The other 15 sites are located in the transmembrane region, with 11 sites located closer to the extracellular side of the receptor. The latter can be divided into two clusters (marked by ellipses in fig. 3b and d). The seventh transmembrane region (TM7) contains the largest concentration of positively selected sites (six sites). These sites are adjacent to four additional sites in TM5 and TM6, forming a dominant cluster in the transmembrane region that borders on the extracellular region (cluster 1). A second cluster of three positively selected sites (cluster 2) is located in the large second extracellular loop (EC2), in close proximity to TM2 and TM3. The sites of cluster 2 surround a region in TM3 (marked by an arrow in fig. 3b and d) that was shown by mutagenesis experiments to affect

Table 3
Positive Selection in Branches of Species-Specific Gene Duplications

Clade	<i>Camponotus floridanus</i>			<i>Pogonomymex barbatus</i>			<i>Monomorium pharaonis</i>			<i>Solenopsis fugax</i>			<i>Solenopsis invicta</i>			Nonspecies-Specific Branches		
	Br	PS	%PS	Br	PS	%PS	Br	PS	%PS	Br	PS	%PS	Br	PS	%PS	Br	PS	%PS
Clade 1	0	0	—	0	0	—	0	0	—	0	0	—	0	0	—	11	0	0
Clade 2	6	0	0	0	0	—	0	0	—	0	0	—	2	0	—	31	1	3
Clade 3	32	0	0	12	1	8	0	0	—	2	1	50	8	0	0	231	17	7
Clade 4	22	1	5	8	0	0	6	0	0	0	0	—	0	0	—	117	14	12
Clade 5	20	0	0	6	0	0	6	0	0	6	1	17	12	0	0	193	12	6
Clade 6	0	0	—	0	0	—	0	0	—	0	0	—	0	0	—	35	3	9
Clade 7	80	8	10	2	1	50	4	0	0	0	0	—	8	2	25	151	11	7
Clade 8	12	1	8	2	1	50	4	0	0	2	0	0	2	0	0	53	1	2
Clade 9	0	0	—	2	0	0	0	0	—	8	0	0	22	2	9	35	2	6
Clade 10	20	1	5	2	0	0	4	2	50	6	1	17	32	1	3	215	15	7
Clade 11	0	0	—	0	0	—	0	0	—	0	0	—	0	0	—	17	1	6
Clade 12	0	0	—	0	0	—	0	0	—	0	0	—	0	0	—	13	0	0
Clade 13	24	3	13	4	0	0	4	0	0	0	0	—	8	1	13	101	6	6
Clade 14	8	0	0	6	0	0	0	0	—	0	0	—	2	0	0	53	4	8
Clade 15	6	0	0	4	0	0	0	0	—	0	0	—	4	1	25	151	11	7
Clade 16	0	0	—	0	0	—	0	0	—	0	0	—	0	0	—	7	0	0
Clade 17	0	0	—	0	0	—	2	0	0	0	0	—	2	1	50	149	4	3
Clade 18	6	1	17	4	0	0	2	0	0	0	0	—	2	0	0	131	5	4
Clade 19	0	0	—	0	0	—	0	0	—	0	0	—	8	2	25	83	6	7
Clade 20	0	0	—	10	0	0	8	2	25	5	0	0	12	0	0	16	2	13
Clade 21	6	2	33	14	0	0	6	1	17	0	0	—	0	0	—	79	12	15
Clade 22	16	0	0	8	0	0	4	0	0	4	1	25	6	1	17	155	13	8
Clade 23	22	0	0	4	1	25	2	1	50	0	0	—	6	1	17	75	9	12
Clade 24	0	0	—	0	0	—	2	0	0	0	0	—	6	1	17	27	1	4
Clade 25	2	0	0	6	0	0	2	0	0	2	1	50	4	0	0	49	2	4
Clade 26	4	0	0	8	0	0	0	0	—	0	0	—	6	1	17	51	6	12
Clade 27	2	1	50	0	0	—	0	0	—	0	0	—	0	0	—	23	0	0
Clade 28	14	0	0	2	0	0	0	0	—	0	0	—	32	4	13	129	6	5
Clade 29	50	8	16	8	2	25	4	0	0	0	0	—	32	2	6	65	2	3
Clade 30	10	1	10	8	0	0	0	0	—	14	1	7	18	3	17	103	2	2
Clade 31	46	7	15	18	0	0	0	0	—	0	0	—	16	6	38	53	3	6
Total	408	34	8.3	138	6	4.3	60	6	10	49	6	12	250	29	12	2,602	171	6.6

Br, total number of the branches; PS, number of the branches with positive selection; %PS, percentage of branches with positive selection.

(nonfunctionalization); one paralog may retain the ancestral function while the other acquires new functions through adaptive evolution (neofunctionalization); or, the duplication may allow each paralog to evolve more specialized functions, so that the functions of each paralog are a subset of the functions of the ancestral gene (subfunctionalization). Neofunctionalization and subfunctionalization events can be distinguished by the signature of natural selection on paralogs after gene duplication. A duplication followed by an asymmetrical signature of positive selection in one paralog and conservation of the other suggests a neofunctionalization event, while symmetrical signature of positive selection on both paralogs suggests subfunctionalization and specialization. Most OR duplications fit the former scenario of neofunctionalization. For example, branch 9 in clade 22 appears to be a neofunctionalization event, since after duplication this

paralog experienced positive selection while the other paralog has a significantly shorter branch (leading to PbOR269 and SiOR449), which is in line with negative selection that maintains a conserved ancestral function (fig. 2a).

The extensive evidence for adaptive evolution in ant ORs parallels the evolutionary dynamics of chemosensory proteins (CSP) and odorant binding protein (OBP). These are two families of small sensillar lymph proteins that solubilize hydrophobic odorants. Kulmuni et al. (2013) showed that ants have a large number of species-specific duplications (especially *S. invicta*), as well as a higher dN/dS ratio in these species-specific CSP genes relative to nonspecies-specific genes. However, that study did not conduct a test for positive selection. McKenzie et al. (2014) tested for positive selection on some subfamilies of OBPs and CSPs using the site test for positive selection. Their analysis found adaptive evolution in

Table 4
Mapping of Positively Selected Sites to Amino Acid Positions in DmOR85b

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Count ^a	9	1	1	4	3	2	3	2	4	2	2	0	4	3	2	1	2	1	1	5	5	2	4	1	0
Region	IC1																								
Position	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
Count	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	1	2	0	0	2	1	2	5
Region	IC1												TM1												
Position	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Count	5	2	4	1	1	1	0	0	0	2	2	3	3	2	4	3	1	2	1	1	0	2	1	3	0
Region	TM1							EC1											TM2						
Position	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Count	2	2	0	0	3	2	1	3	0	2	2	0	1	2	4	7	1	2	1	1	1	1	2	1	2
Region	TM2																			IC2					
Position	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125
Count	1	0	0	2	4	2	3	1	3	0	0	1	1	1	3	0	1	0	1	2	2	3	2	5	0
Region	IC2																								
Position	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
Count	3	0	6	3	0	1	0	0	0	1	2	2	1	1	0	0	2	2	1	1	4	0	1	0	1
Region	TM3																								
Position	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175
Count	1	1	0	0	0	0	0	0	1	0	2	6	0	2	3	1	1	4	9	1	1	1	2	4	1
Region	TM3			EC2																					
Position	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
Count	1	1	1	2	1	1	1	1	3	2	1	0	5	4	6	1	2	2	1	0	1	0	4	2	2
Region	EC2				TM4																				
Position	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225
Count	2	2	4	5	5	3	1	2	1	1	3	3	1	2	2	5	0	0	2	3	0	5	2	2	0
Region	TM4												IC3												
Position	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250
Count	1	10	0	0	0	0	0	0	1	0	0	0	0	1	0	2	2	0	1	1	0	4	1	3	3
Region	IC3																								
Position	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275
Count	2	2	0	3	1	0	4	3	1	5	3	2	3	2	2	0	2	2	3	0	1	4	5	2	2
Region	IC3			TM5																					
Position	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300
Count	1	4	6	1	3	3	0	7	1	0	3	2	3	1	1	0	2	0	1	1	8	1	1	2	3
Region	TM5							EC3										TM6							
Position	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325
Count	1	11	4	1	0	2	2	1	2	5	5	2	2	2	4	3	1	1	4	5	10	2	5	0	2
Region	TM6												IC4												
Position	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350
Count	1	4	2	2	2	2	3	1	0	2	4	4	2	0	3	4	3	3	0	1	2	1	0	2	2
Region	IC4																								
Position	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375
Count	4	1	3	2	3	4	1	2	2	1	3	0	4	4	3	6	2	4	5	1	12	4	2	7	4
Region	IC4												TM7												
Position	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392								
Count	3	7	1	7	1	1	1	3	3	1	6	3	2	2	1	0	0								
Region	TM7																								
Color scale	0	1	2	3	4	5	6	7	8	9	10	11	12												

^aCounts of branches with positive selection for each amino acid site, after mapping positions in ant sequences to the homologous positions on the structural model of DmOR85b. Predicted extracellular (EC), intracellular (IC), and transmembrane (TM) regions are marked by shaded boxes.

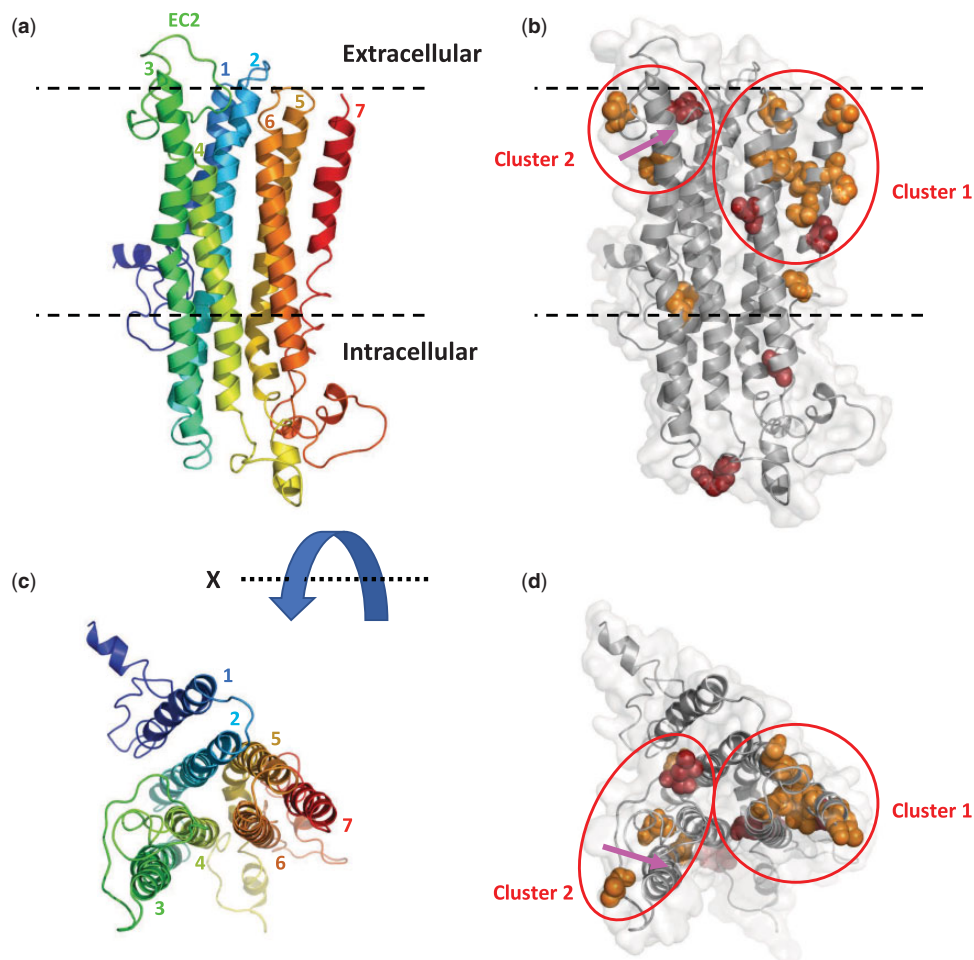


FIG. 3.—Mapping of positively selected sites to a 3D structural model of insect ORs. (a) Ribbon diagram of DmOR85b is rainbow colored from the amino terminus (blue) to the carboxy terminus (red), with the seven transmembrane helices numbered. (b) Amino acids under selection in ≥ 6 branches of the gene tree are shown as spheres and colored orange (between 6 and 8 branches) or red (between 9 and 12 branches). Two clusters of positively selected sites are marked by ellipses. A region in TM3 that affects ligand activation based on the mutagenesis experiments in *Drosophila melanogaster* is marked by an arrow. (c and d) Top view (from the extracellular side) of (a) and (b), rotated 90° about the X axis.

both CSPs and OBPs, suggesting they evolved hand in hand with ORs to produce novel olfactory functions such as response to new pheromones.

The OR gene tree allows dating of gene duplications relative to the speciation of taxonomic groups: recent gene duplication events that created species-specific genes, less recent duplications of genes found in multiple ant species (e.g., all Formicinae), and ancient duplications that preceded the divergence of all ant species included in our analysis. ORs were duplicated throughout evolution of Hymenoptera, yet gene duplications were even more rapid in ants. Engsontia et al. (2015) hypothesized that species-specific duplications are followed by more positive selection pressures than non-species-specific duplications. However, that study only applied the site test for positive selection to the subtrees of species-specific expansions, and the branch-site test only to a few selected branches. Thus, their analyses do not provide a

comprehensive comparison between branches in the species-specific expansions and the rest of the tree. The analyses in the present study applied the branch-site test to the majority of branches of the full gene tree. Our comparison shows a statistically significant difference in the proportion of branches with positive selection, with more positive selection in the species-specific expansions relative to nonspecies-specific branches. However, the magnitude of this difference is not large—9.0% versus 6.6% (table 3). Ergo, neofunctionalization of ORs proceeded both before the divergences of ant subfamilies/genera and after, in specific lineages.

Conversely, our findings are in agreement and complementary to the gene births and deaths analysis by Engsontia et al. Their results show that clades J (equivalent to clade 17 in our tree), L (clades 12–15), U (clades 9 and 10), V (clades 4 and 5), and 9-exon (clades 21–31) underwent substantial expansion. In our branch-site test analyses these are among the clades

with the highest number of positively selected branches. In other clades that had little or no expansion, very few or no branches show evidence for positive selection, for example: clade B (equivalent to clade 16 in our tree), C (clade 1), and K (clade 11). In this respect, both types of analyses highlight specific ant OR subfamilies that underwent adaptive radiation.

Our results also shed light on the functional relevance of positive selection in ORs. Mapping sites under positive selection to the 3D protein structure showed that positive selection targets mainly the extracellular-facing portions of the receptor. Two clusters of positive selection stand out. One is located around a TM3 region that was shown to affect ligand activation in a mutagenesis study of a *D. melanogaster* OR, indicating its proximity to the ligand binding site (Nichols and Luetje 2010). The second and larger cluster of positively selected sites is located across a cleft in the extracellular face of the receptor that separates TM2, TM3, and EC2 from TM5 and TM6. According to the structural model this cleft is accessible from the extracellular side of the receptor. Thus, our results are consistent with a binding site located between TM2/3 and TM5/6, and suggest that the most positively selected regions of ORs are related to ligand binding or receptor activation following ligand binding. These sites may play roles in ligand-binding specificity of ant ORs or in receptor activation in response to ligand binding. Following the extensive duplications of ant ORs, adaptive evolution of these sites may have modified receptor specificities toward new odorants and pheromones and/or modulated olfactory response in different paralogs. Multiple paralogs may have also allowed the elaboration of a richer repertoire of responses to existing pheromones, such as the complex mixtures of cuticular hydrocarbons that display variation in the relative proportion of many structurally related compounds.

In conclusion, our analyses revealed extensive evidence for positive selection on the ant ORs following many gene duplication events, typically fitting the signature of neofunctionalization. The distribution of selected sites on the protein structure suggests that ligand binding is the main target of positive selection. Thus, this study supports the hypothesis that the dramatic expansion of the OR superfamily in ants and the adaptive evolution of novel olfactory functions allowed recognition of an expanded collection of olfactory cues, potentially facilitating more complex chemical communication using a larger vocabulary. This work lays the ground for further detailed investigations both into specific subfamilies and periods of adaptive evolution in ants as well as experimental investigation of OR functional sites. Such investigations can be guided by the sites identified here as targets of positive selection.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

All computations were conducted on the Hive computer cluster of the Faculty of Natural Science at the University of Haifa. E.P. was supported by grants from the Israel Science Foundation (grant numbers 646/15, 2140/15, 2155/15) and the US-Israel Binational Science Foundation (grant 2013408). M.K. was supported by grants from the Israel Science Foundation (grant numbers 1454/13, 1959/13, 2155/15)

Literature Cited

- Akino T, Yamamura K, Wakamura S, Yamaoka R. 2004. Direct behavioral evidence for hydrocarbons as nestmate recognition cues in *Formica japonica* (Hymenoptera: formicidae). *Applied Entomology and Zoology* 39(3):381–387.
- Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc Natl Acad Sci USA*. 110(43):17409–17414.
- Attrill H, et al. 2016. FlyBase: establishing a Gene Group resource for *Drosophila melanogaster*. *Nucleic Acid Res.* 44(D1):D786–D792.
- Bass JD, Dabney A, Robinson D. (2015). qvalue: Q-value estimation for false discovery rate control R package version 2.10.0. Available from: <http://github.com/jdstorey/qvalue>.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 57:289–300.
- Benton R, Sachse S, Michnick SW, Vosshall LB. 2006. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS Biol.* 4(2):e20.
- Clyne PJ, et al. 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* 22(2):327–338.
- DeLano WL. (2002). The PyMOL molecular graphics system. Available from: <http://pymol.org>.
- Engsontia P, Sangket U, Robertson HM, Satasook C. 2015. Diversification of the ant odorant receptor gene family and positive selection on candidate cuticular hydrocarbon receptors. *BMC Res Notes* 8:380.
- Force A, et al. 1998. Preservation of duplicate genes by complementary. Degenerative mutations. *Genetics* 151:1531–1545.
- Hopf TA, et al. 2015. Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors. *Nat Commun.* 6(1):6077.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33(2):511–518.
- Kulmuni J, Wurm Y, Pamilo P. 2013. Comparative genomics of chemosensory protein genes reveals rapid evolution and positive selection in ant-specific duplicates. *Heredity* 110(6):538–547.
- Larsson MC, et al. 2004. *Or83b* encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* 43(5):703–714.
- Löytynoja A, Vilella AJ, Goldman N. 2012. Accurate extension of multiple sequence alignments using a phylogeny-aware graph algorithm. *Bioinformatics* 28(13):1684–1691.
- Lundin C, et al. 2007. Membrane topology of the *Drosophila* OR83b odorant receptor. *FEBS letters* 581:5601–5604.
- Marks DS, Hopf TA, Sander C. 2012. Protein structure prediction from sequence variation. *Nature biotechnology* 30:1072–1080.
- Martin SJ, Drijfhout FP. 2009. A review of ant cuticular hydrocarbons. *J Chem Ecol.* 35(10):1151–1161.
- Martin SJ, Vitikainen E, Helanterä H, Drijfhout FP. 2008. Chemical basis of nest-mate discrimination in the ant *Formica exsecta*. *Proc R Soc B* 275(1640):1271–1278.

- McKenzie SK, Oxley PR, Kronauer DJC. 2014. Comparative genomics and transcriptomics in ants provide new insights into the evolution and function of odorant binding and chemosensory proteins. *BMC Genomics* 15:718.
- Nichols AS, Luetje CW. 2010. Transmembrane segment 3 of *Drosophila melanogaster* odorant receptor subunit 85b contributes to ligand-receptor interaction. *J Biol Chem* 285(16):11854–11862.
- Nowbahari E, et al. 1990. Individual, geographical and experimental variation of cuticular hydrocarbons of the ant *Cataglyphis cursor* (Hymenoptera, Formicidae) – their use in nest and subspecies recognition. *Biochem Syst Ecol*. 18(1):63–73.
- Ohno S. 1970. Evolution by gene duplication. New York: Springer Science & Business Media.
- Ozaki M, et al. 2005. Ant nestmate and non-nestmate discrimination by a chemosensory sensillum. *Science* 309(5732):311–314.
- Penn O, Privman E, Landan G, Graur D, Pupko T. 2010. An alignment confidence score capturing robustness to guide tree uncertainty. *Mol Biol Evol*. 27(8):1759–1767.
- 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*. 16:1395–1403.
- Roux J, et al. 2014. Patterns of positive selection in seven ant genomes. *Mol Biol Evol*. 31(7):1661–1685.
- Saar M, Leniaud L, Aron S, Hefetz A. 2014. At the brink of supercoloniality: genetic, behavioral, and chemical assessments of population structure of the desert ant *Cataglyphis niger*. *Front Ecol Evol*. 2:13.
- Sato K, et al. 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature* 452(7190):1002–1006.
- She R, Chu JS, Wang K, Pei J, Chen N. 2009. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res* 19:143–149.
- She R, et al. 2011. genBlastG: using BLAST searches to build homologous gene models. *Bioinformatics* 27:2141–2143.
- Smart R, et al. 2008. Drosophila odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect Biochem Mol Biol*. 38(8):770–780.
- Smith CD, et al. 2011. Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proc Natl Acad Sci USA*. 108(14):5667–5678.
- Smith CR, et al. 2011. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc Natl Acad Sci USA*. 108(14):5667–5672.
- Soroker V, Vienne C, Hefetz A. 1995. Hydrocarbon dynamics within and between nestmates in *Cataglyphis niger* (hymenoptera: formicidae). *J Chem Ecol*. 21(3):365–378.
- Stamatakis A. 2006. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.
- Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. 1999. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* 96(5):725–736.
- Werren JH, et al. 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* 327:343–348.
- Wicher D, et al. 2008. *Drosophila* odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature* 452(7190):1007–1011.
- Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 17(1):32–43.
- Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22(12):2472–2479.
- Zhou X, et al. 2012. Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genet*. 8(8):e1002930.

Associate editor: Belinda Chang