



HAL
open science

Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, *Acyrtosiphon pisum*

Carole Smadja, P. Shi, R. Butlin, H. Robertson

► **To cite this version:**

Carole Smadja, P. Shi, R. Butlin, H. Robertson. Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, *Acyrtosiphon pisum*. *Molecular Biology and Evolution*, 2009, 26 (9), pp.2073-2086. 10.1093/molbev/msp116 . hal-01945562

HAL Id: hal-01945562

<https://hal.science/hal-01945562>

Submitted on 15 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Large Gene Family Expansions and Adaptive Evolution for Odorant and Gustatory Receptors in the Pea Aphid, *Acyrtosiphon pisum*

Carole Smadja,* Peng Shi,† Roger K. Butlin,* and Hugh M. Robertson‡

*Animal and Plant Sciences Department, University of Sheffield, Sheffield, United Kingdom; †State Key Laboratory of Evolutionary and Functional Genomics, Kunming Institute of Zoology, Chinese Academy of Sciences, China; and ‡Department of Entomology, University of Illinois at Urbana-Champaign

Gaining insight into the mechanisms of chemoreception in aphids is of primary importance for both integrative studies on the evolution of host plant specialization and applied research in pest control management because aphids rely on their sense of smell and taste to locate and assess their host plants. We made use of the recent genome sequence of the pea aphid, *Acyrtosiphon pisum*, to address the molecular characterization and evolution of key molecular components of chemoreception: the odorant (Or) and gustatory (Gr) receptor genes. We identified 79 Or and 77 Gr genes in the pea aphid genome and showed that most of them are aphid-specific genes that have undergone recent and rapid expansion in the genome. By addressing selection within sets of paralogous Or and Gr expansions, for the first time in an insect species, we show that the most recently duplicated loci have evolved under positive selection, which might be related to the high degree of ecological specialization of this species. Although more functional studies are still needed for insect chemoreceptors, we provide evidence that Grs and Ors have different sets of positively selected sites, suggesting the possibility that these two gene families might have different binding pockets and bind structurally distinct classes of ligand. The pea aphid is the most basal insect species with a completely sequenced genome to date. The identification of chemoreceptor genes in this species is a key step toward further exploring insect comparative genetics, the genomics of ecological specialization and speciation, and new pest control strategies.

Introduction

Aphids are model organisms in both evolutionary and applied biology. This status is mainly related to features of their biology that enable them to locate and exploit their host plants (Powell et al. 2006). Evolutionary biologists focus on host plant selection and use in phytophagous insects because species and host races that are specialized on different plants are model organisms for the study of population differentiation in sympatry and ecological speciation (Via 2001; Berlocher and Feder 2002). Ecologically based divergent (i.e., positive) selection is expected to drive the evolution of host plant specialization, and the joint evolution of host preference and host-associated performance is critical to speciation in phytophagous insects in the presence of gene flow (Drès and Mallet 2002). Particularly important is the possibility that the same recognition processes, and so genes, may underlie both preference and performance and incidentally generate assortative mating, overcoming the barrier to speciation caused by recombination (Jiggins et al. 2005; Servedio 2009).

More than 90% of the ~4,000 species of aphid are specialists on one or a few plant hosts and host races exist within many taxa (Bush and Butlin 2004). In particular, extensive studies on host races of the pea aphid (*Acyrtosiphon pisum*) have shown that these populations are ecologically specialized and genetically differentiated for host use and that reproductive isolation among races has evolved as a by-product of host plant specialization (the reduction of gene flow occurring by a combination of habitat choice, selection against migrants and ecologically based selection against hybrids) (Via 1999; Via et al. 2000; Hawthorne and Via 2001; Via and Hawthorne 2002; Ferrari et al. 2006, 2008). The availability of the

genome sequence of the pea aphid (The International Aphid Genomics Consortium, submitted) means that this species has enormous potential for understanding the genetic basis of host shifts and, more generally, of adaptive evolution and ecological speciation.

Many aphid species are also known as major crop pests: By damaging crops through direct impact on plant tissues and virus transmission while feeding on their host plants, aphids are responsible for crop losses estimated at hundreds of millions of dollars annually worldwide (e.g., Morrison and Peairs 1998). Therefore, applied research challenges reside in the development of pest control strategies to reduce the impact of aphids on crop plants (Pickett et al. 1997).

Host plant selection by aphids is not a random process; these insects employ a variety of sensory and behavioral mechanisms to locate and recognize their host plants, and chemoreception plays a fundamental role in this host selection process (Pickett et al. 1992; Park and Hardie 2003). Indeed, the feeding and reproduction of an aphid on a plant depends upon the successful completion of a series of behavioral acts based on olfaction and gustation: Firstly, the insect lands on a potential host, responding to visual and olfactory cues (Pickett et al. 1992; Powell and Hardie 2001; Pickett and Glinwood 2007); after settling, aphids probe the cuticle and the epidermis of the leaf with their mouthparts, this probing phase being the key phase in host acceptance as demonstrated in host races of the pea aphid (Caillaud and Via 2000). In addition, the choice of a plant which to land on can also be influenced by the recognition—through pheromone communication—of conspecifics present on a plant (Pope et al. 2007). Studies have already been started to decipher the neurophysiological pathways underlying such chemosensory processes in aphids (Hardie et al. 1995; Visser et al. 1996; Park et al. 2000; Park and Hardie 2004; Kristoffersen et al. 2008) and chemical ecologists have identified some semiochemicals and pheromones that mediate these behaviors (Birkett and Pickett 2003; Del Campo et al. 2003;

Key words: olfaction, gustation, adaptive evolution, gene family evolution, host plant specialization, speciation.

E-mail: c.smadja@sheffield.ac.uk.

Mol. Biol. Evol. 26(9):2073–2086. 2009

doi:10.1093/molbev/msp116

Advance Access publication June 19, 2009

Pickett and Glinwood 2007; Webster et al. 2008). However, although taste and smell appear to be the main senses underlying host selection in aphids, the genes underlying the detection of chemical compounds in the environment are still entirely unknown in this group of insects in which new insights into the mechanisms of host plant selection would be of great interest for a large community. Moreover, identifying the genes underlying chemoreception is a key step toward understanding the evolution of traits involved in host plant specialization and ecological speciation and the role of ecologically based positive selection in driving adaptive divergence among populations.

During the last decade, large families of genes encoding receptors responsible for detecting chemical stimuli in the environment have been revealed from genomes in various phyla (Matsunami and Amrein 2003; Ache and Young 2005; Bargmann 2006). The insect chemoreceptor (Cr) superfamily of 7TM ligand-gated ion channels (Sato et al. 2008; Wicher et al. 2008) consists of a basal gustatory receptor (Gr) family, of great protein diversity, and a more derived odorant receptor (Or) family of more limited diversity. Both families have been shown to have a primary role in stimulus detection and discrimination (Vosshall and Stocker 2007). The Gr and Or families can be recognized based on sequence homology and their expression and function primarily in gustatory and olfactory neurons, respectively (reviewed, e.g., in Rutzler and Zwiebel 2005; Benton 2006; Hallem et al. 2006; Vosshall and Stocker 2007). Insect Crs are characterized by high sequence divergence, gene duplication and gene loss, and low levels of expression in specific sensory neurons (e.g., Robertson and Wanner 2006; Nozawa and Nei 2007), these features making their discovery highly dependent on genome sequences (e.g., first characterization in *Drosophila melanogaster* Clyne et al. 1999; Gao and Chess 1999; Vosshall et al. 1999; Clyne et al. 2000; Robertson et al. 2003). As newly sequenced insect genomes have become available, structure and sequence similarity comparisons with already identified receptors have allowed the molecular characterization of the Cr repertoire in several *Drosophila* species (Guo and Kim 2007; McBride 2007; Gardiner et al. 2008), other Diptera (*Anopheles* and *Aedes* mosquitoes: Hill et al. 2002; Kent et al. 2008) and insects in several other orders (e.g., honey bee *Apis mellifera*: Robertson and Wanner 2006; red flour beetle *Tribolium castaneum*: Consortium T.G.S. 2008; Engsontia et al. 2008; silkworm moth *Bombyx mori*: Wanner and Robertson 2008). However, interspecific divergence, scarcity of orthologous groups, and species-specific expansions of particular genes (Krieger et al. 2003; Robertson and Wanner 2006) continue to challenge automated gene prediction software and the process of annotation still requires intensive manual effort, especially for groups like aphids that are distantly related to the taxa already characterized.

Only one protein putatively involved in olfaction has been identified in the entire hemipteroid assemblage (Dickens et al. 1998). However, the pea aphid (*A. pisum*) genome has recently been sequenced (The international Aphid Genomics Consortium, submitted), offering a unique opportunity to identify the chemoreceptor repertoire for the first time in an aphid species and to address the evolution of

this gene superfamily. If ecologically based selection drives host plant specialization and if host plant recognition relies on chemosensory processes, chemoreceptor genes are expected to show some signature of evolution under positive selection. Here, we report the manual annotation and characterization of the Or and Gr gene families in the genome of the pea aphid and we test the above prediction by testing for positive selection within sets of paralogous Or and Gr expansions. We provide evidence that the most recently duplicated Or and Gr loci have evolved under positive selection, which might be related to the high degree of ecological specialization of this species.

Material and Methods

Identification of Pea Aphid Ors and Grs by Bioinformatics

Known insect Ors and Grs whose sequences have been entered into GenBank (National Center for Biotechnology Information, NCBI) were used to search for similar genes in the pea aphid genome sequence (TBLASTN searches against pea aphid raw traces and the first genome Assembly 1.0 available at <http://www.hgsc.bcm.tmc.edu/>—Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; <http://genoweb1.irisa.fr/AphidBase/Blast/Blast.php>—AphidBase, INRA, Rennes, France, and in the Whole Genome Shotgun database—<http://blast.ncbi.nlm.nih.gov/Blast.cgi>—NCBI at the NIH, United States). Identified pea aphid Ors and Grs were in turn employed in searches to find more genes in an iterative process and PSI-BlastP searches were used to detect highly divergent sequences (Arthropod PSI-BLASTP server at <http://insects.eugenes.org/arthropods/blast/psiblast.html>). Genomic scaffold sequences were used to construct Or and Gr genes manually using known Or/Gr exons as templates and using programs to edit and manipulate sequences (PAUP* v4.0b10, Wilgenbusch and Swofford 2003 and DNA Dynamo, BlueTractorSoftware Ltd, 2007) and to predict exon–intron splice sites (SplicePredictor, <http://deepc2.psi.iastate.edu/cgi-bin/sp.cgi/>; BioEdit, Hall T., Ibis Biosciences, 2007). Pea aphid Or and Gr genes/proteins were named “ApGr” and “ApOr” followed by a number. Due to high divergence and/or discontinuity among contigs/scaffolds, not all detected Or/Gr genes could be entirely annotated: In these cases, deduced amino acid sequences shorter than 150 amino acids were discarded as probable gene fragments and for long-enough partial sequences a suffix N or C after the gene–protein name was used to indicate that the N or C terminus is missing (Kent et al. 2008). When frameshifts or stop codons were identified in the gene sequence, we defined these genes as putative pseudogenes (suffix P). When sequences looked very similar, we used pairwise identity measures to distinguish between paralogs or allelic variants, as well as their locations (if on a separate small contig they were considered to be likely allelic variants, but if in long tandem arrays in one contig, they were considered to be paralogs). A few apparent such alternative haplotypes were found (seven in the Grs and one in the Ors) and were excluded from the final list of Cr genes. That the intact genes belong to the chemoreceptor gene family was

supported by matches to other insect Crs (BlastP searches against the NCBI nonredundant protein database) and by the number and location of trans-membrane domains in the predicted proteins (TransMembrane prediction using Hidden Markov Models [TMHMM]; Krogh et al. 2001). Manually obtained gene models were checked on the final draft Assembly 1.0 and compared with automated predictions. Only a handful of manually annotated genes were perfectly automatically predicted, thus confirming that the annotation of chemoreceptor genes still requires intensive manual effort. Protein alignments were used to identify irregularities and refine the gene structures (ClustalX, Thompson et al. 1997). Two of these gene models were confirmed by the presence of some representative genes in the Expressed Sequence Tag (EST) database (ApGr24 and ApOr18), see Results for details. Although most of the Cr genes are likely to have been identified here, we cannot exclude having missed some highly divergent genes, despite using PSI-BLASTP searches, because if there is not at least a partial gene model in the automated gene set, PSI-BLASTP searches will not find them, and TBLASTN searches can miss genes with short exons encoding divergent amino acid sequences.

Phylogenetic Analyses

For each family, phylogenetic analyses were conducted on aligned aphid protein predictions (all identified Ors; all but four Grs—excluding two short and two divergent sequences) as well as other representative insect Crs showing evidence of “orthologous” relationships with aphid sequences in BlastP searches. Gr and Or sets of sequences (respectively, 101 and 83 sequences) were aligned by ClustalX, with manual adjustments. N-terminal, C-terminal and a few internal regions of poor alignment and large gaps were removed from the alignment. Phylogenetic analysis (distance and maximum likelihood trees) of these two large data sets was performed using corrected distance analysis in Tree-Puzzle v5.0 (Schmidt et al. 2002) and PAUP*v4.0b10 (see Hill et al. 2002; Robertson et al. 2003; Robertson and Wanner 2006; Kent et al. 2008). Bayesian analysis was performed using MrBayes v3.1 (Ronquist and Huelsenbeck 2003) with the Jones, Taylor, Thornton (JTT) substitution model (Jones et al. 1992), four chains, one million generations, and two runs. Trees were sampled every 100 generations, discarding a burn in of 250,000 generations. Support for major branches was calculated as percent of 1,000 uncorrected distance bootstrap replications, 10,000 quartet maximum likelihood puzzling steps, and Bayesian posterior probabilities (PPs).

Molecular Evolution Analysis

Tests for variation in selective pressures and for positive selection were performed using the codeml program of the PAML package (Yang 1997), which estimates by maximum likelihood method ω ratios of the normalized nonsynonymous substitution rate (d_N) to the normalized synonymous substitution rate (d_S). $\omega > 1$ is considered to be evidence of positive selection for amino acid replacements, whereas $\omega < 1$ indicates purifying selection. We

tested for patterns of positive selection in three groups of genes that underwent recent expansions with a mean synonymous distance less than 1 (these groups being called “Gr clade A,” “Or clade A,” and “Or clade B,” see details in the Results section and in figs. 1 and 2). We used only the putative intact genes, removing all the pseudogenes and partially annotated genes in this analysis (tables 1 and 2). We aligned the protein sequences using ClustalX (Thompson et al. 1997) and the nucleotide sequence alignment was guided by the protein sequence alignment. For each clade, the phylogenetic tree of intact sequences was rebuilt using the Neighbor Joining method (Saitou and Nei 1987) in MEGA 4 (Tamura et al. 2007). Patterns of variation in selective pressures and of positive selection among paralogous genes in each of the pea aphid-specific clades were tested using “branch-specific” and “site-specific” methods. The branch-specific method compares the following two models: The null model (one-ratio model) assumes all branches examined have the same ω ratio, whereas the alternative model (alternative model) allows the ω ratio to vary among branches in the phylogeny (Yang 1998; Yang and Nielsen 1998). Thus, a significantly higher likelihood of the alternative model than that of the null model indicates variation of selective pressures among branches, some of which might have undergone positive selection (Yang 1998; Yang and Nielsen 1998). For the analysis of the site-specific model, we first compared the recently implemented codon-substitution models M8a and M8. This test has been shown to be conservative and robust against violations of various model assumptions and to produce less false positives than the M7–M8 comparison (Swanson et al. 2003; Wong et al. 2004). The alternative model (M8) assumes a beta distribution to model variable selection pressure among sites and allows for an extra ω parameter that can be greater than 1. The null model (M8a) differs from the alternative model in that the extra ω is fixed at 1. The comparison between models was assessed using Likelihood-Ratio Tests (LRTs) for hierarchical models (Anisimova et al. 2001), a significantly higher likelihood of the alternative model than that of the null model indicating positive selection in the data set examined. Although M8 and M8a models were shown to be reliable and robust for detecting positive selection in computer simulations (Swanson et al. 2003; Wong et al. 2004), it remains possible that the evolutionary patterns of chemosensory receptors in the pea aphid genome could violate the assumptions of M8 and M8a models, which could lead to false detections of positive selection. To further confirm our results, we additionally compared LRT of M1a (Nearly Neutral) and M2a (Selection) models, which has been shown to be one of the most conservative tests for selection (e.g., Wong et al. 2004; Yang et al. 2005). For both M8a–M8 and M1a–M2a comparisons, we used degree of freedom, $df = 2$. For each analysis, correction for multiple testing (Bonferroni correction) was applied. Only in cases where LRT was significant, we used the Bayes empirical Bayes (BEB) procedure to calculate the PPs to identify sites under positive selection (Yang et al. 2005). We further examined the distribution of the inferred positively selected sites by mapping the amino acids under positive selection onto the chemoreceptor topology using TMHMM.

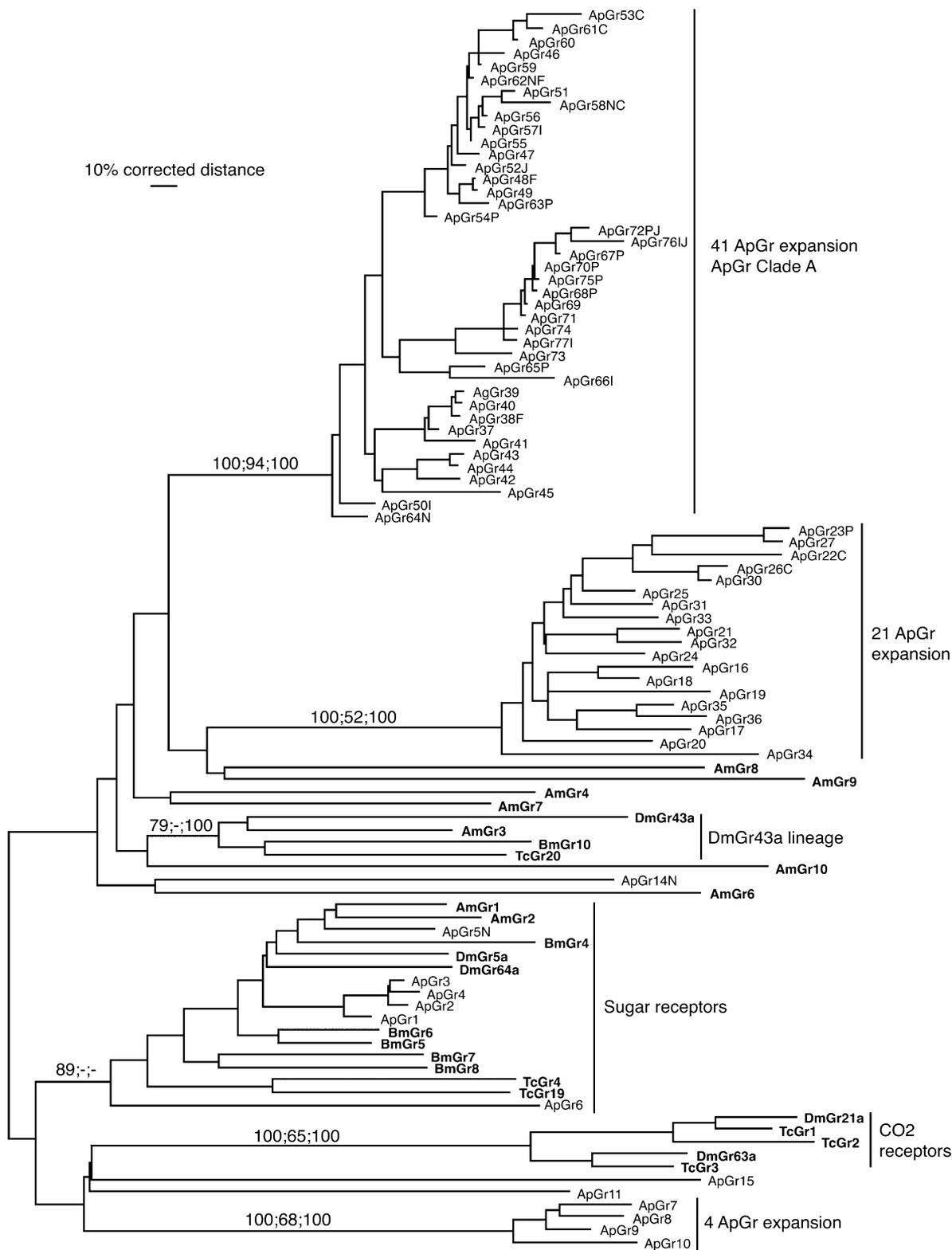


FIG. 1.—Phylogenetic relationships of the pea aphid ApGrS with representative GrS from other insects. This corrected distance tree was rooted at the midpoint. Major gene lineages discussed in the text are indicated by vertical bars on the right, including the clade examined for selection. Support for major branches is shown above them as percent of 1,000 uncorrected distance bootstrap replications, 10,000 quartet maximum likelihood puzzling steps, and Bayesian PPs, but only if they were above 50%. Suffixes after gene—protein names are: P—pseudogene, F—fixed gene model, J—joined gene model, N—N-terminal exon(s) or region missing, C—C-terminal exon(s) or region missing, I—internal exon(s) or region missing. Species abbreviations are Am—*Apis mellifera*, Ap—*Acyrtosiphon pisum*, Bm—*Bombyx mori*, Dm—*Drosophila melanogaster*, and Tc—*Tribolium castaneum*. Names of GrS from other species are in bold type.

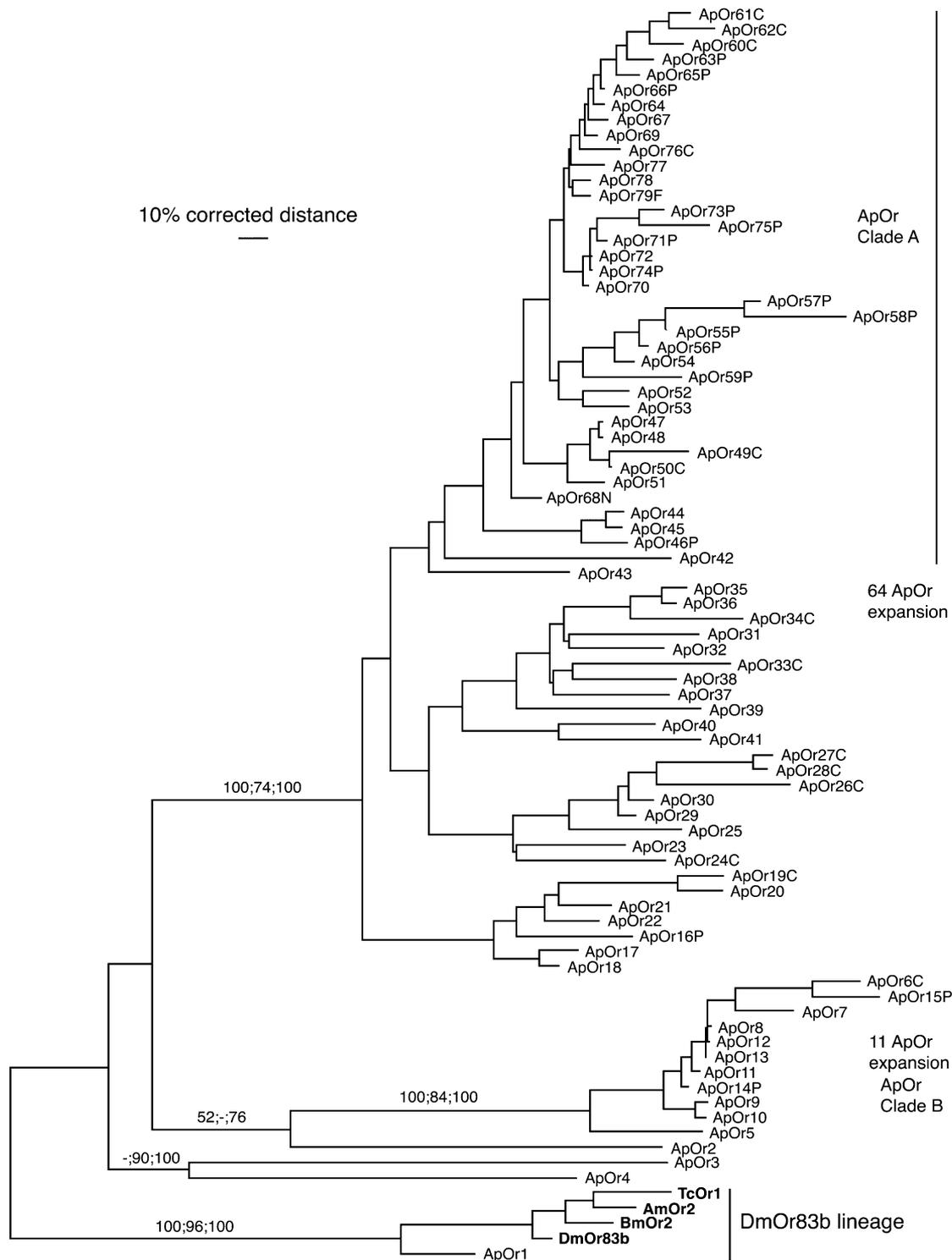


FIG. 2.—Phylogenetic relationships of the pea aphid ApOrs. This corrected distance tree was rooted by declaring the DmOr83b orthologs as the outgroup, based on the basal location of DmGr83b within the Or family and its similarity to the Grs (see Robertson et al. 2003). Other details are as in figure 1 legend.

Results

The Gustatory Receptor Family

We identified 77 ApGr genes in the draft genome sequence (supplementary table S1a, Supplementary Material online; deduced amino acid sequences in supplementary

table S2a, Supplementary Material online). Only three of these gene models are present in the REFSEQ set of gene models for aphid generated by NCBI, presumably because REFSEQ gene models require considerable comparative or experimental support, and only one corresponds perfectly

Table 1
LRTs of Variation in Selective Pressures on Branches within the Pea Aphid Gr and Or Receptor Gene Families

Clades	n^a	Likelihood Ratios		$2\Delta^b$ 1 versus Free	df ^c	<i>P</i> Value (without Bonferroni Correction)	<i>P</i> Value (after Bonferroni Correction)
		One Ratio	Free Ratio				
Gr genes							
Clade A: 41-ApGr expansion	23	-15,491.77	-15,447.41	88.71	43	5.13E-05	0.0003
Or genes							
Clade A: ApOr43-79 clade in 64-ApOr expansion	17	-12,742.14	-12,705.54	73.21	31	2.87E-05	0.0001
Clade B: 11-ApOr expansion	8	-5,230.68	-5,215.04	31.28	13	0.003	0.015

^a Number of sequences in the data set.

^b Twice the logarithm of likelihood ratio.

^c Df calculated as the difference in the number of free parameters between the two models under comparison.

with our annotation. As described further below, few of these ApGrS have significant sequence similarity to GrS in other insects, and there is EST support for just one gene model (see below). Automated gene models in the combined GLEAN gene set are present for an additional 48 genes; however, only four of these are correct. The remainder required changes ranging from addition of an exon, to splitting or fusion of existing gene models. Some of these remain as partial or truncated gene models with exons missing in assembly gaps, whereas others are in fact pseudogenes. Twenty-six new gene models are proposed, of which nine encode apparently full-length intact and potentially functional GrS. All of these gene models and proposed changes have been communicated to AphidBase and will be included in future releases of the genome annotation. Genes were numbered in a phylogenetic order as described below.

ESTs usually provide a source of experimental support for gene models, however the chemoreceptors in general (Vosshall et al. 1999), and the gustatory receptors in particular (e.g., Thorne and Amrein 2008), are usually ex-

pressed at such low levels that they seldom show up in EST sets (e.g., Patch et al. 2009). As expected then, when 168,186 pea aphid ESTs (e.g., Sabater-Munoz et al. 2006) were searched for matches to the Gr proteins encoded by our gene models using the TBLastN algorithm, there were several hits but most of these were overlapping Untranslated Regions from flanking genes or unspliced sequences of undetermined origin. Just one EST (GenBank accession EX606636.1) is from a spliced mRNA from ApGr24, which is the third Gr, in addition to two of the sugar GrS, for which there is a REFSEQ at GenBank. The splicing of three introns in this EST and gene model agrees with all the others in the set of 21 genes in an aphid Gr clade described below, providing additional confidence in our gene models. Finally, we note that no obvious instances of alternative splicing of long N-terminal exons into shared C-terminal exons, which are found in the GrS of flies (e.g., Hill et al. 2002; Robertson et al. 2003; Kent et al. 2008) and *Tribolium* (Consortium T.G.S. 2008), were detected.

Table 2
LRTs of Positive Selection on the Pea Aphid Gr and Or Gene Families

Clades	n^a	$2\Delta^b$		Parameters Estimated under M8	Positively Selected Sites ^c
		M1a versus M2a	M8 versus M8a		
Gr genes					
Clade A: 41-ApGr expansion	23	38.68**	32.93**	$p_0 = 0.903$; $p = 1.216$; $q = 2.060$; ($p_1 = 0.097$); $\omega = 1.839$	<u>41H</u> 63E 67T <u>109Q</u> 138Y 141W 144I <u>151I</u> <u>152L</u> 154F 256F 258S 259A 266F 370R 373Y 374R 375Q
Or genes					
Clade A: ApOr43-79 clade in 64-ApOr expansion	17	31.17**	26.88*	$p_0 = 0.892$; $p = 0.759$; $q = 0.866$; ($p_1 = 0.108$); $\omega = 1.960$	<u>28N</u> 30T 57P 58I 81S 111H 127I 138A 139F 143I 144F 152S 197S 198I 200Y <u>201V</u> 238T <u>240D</u> 276R <u>288S</u> 292F 296I <u>306N</u> 318V
Clade B: 11-ApOr expansion	8	5.34	5.49*	$p_0 = 0.933$; $p = 0.449$; $q = 0.235$; ($p_1 = 0.067$); $\omega = 2.654$	11N 12M 18H 19M <u>28A</u> 31L 34T 36S 39P 45T 46Q 47N 52L 59L 62A 63H 66A 67F 72I 78H 84H 86H 87G 88V 96R 99A 100T 112F 122S 123L 125W 126V 137I 145T 147V 168T 179V 184R 190N 192Q 197I 210K 215V 232M 237C 238G 245S 247F 252I 266V 269I 271C 276I 295A 310R 326K 330V 344F 356I 358R 364A

^a Number of sequences in the data set.

^b Twice the logarithm of likelihood ratio.

^c Positive selection sites estimated under M8 model by BEB approach with PPs > 50% are listed. The overlapping sites of M8 and M2a models are underlined and those with PP > 95% under M8 model in bold. The sites are indexed by their position in the alignment after having removed the gaps.

**Significant at the 0.1% level after Bonferroni correction; *significant at the 5% level after Bonferroni correction.

For phylogenetic analysis, we included 26 Grs from other species to provide context for the ApGrs, yielding a total of 101 proteins (ApGr12 and 13 were excluded from the analysis because they are so highly divergent they lead to alignment difficulties and exceptionally long branches, as were ApGr28 and 29 because they are short and severely damaged pseudogenes). The Grs from other species (see fig. 1 for species name abbreviations) are 5 carbon dioxide receptors (DmGr21a and 63a, and TcGr1–3), 11 sugar receptors (DmGr5a and 64a, BmGr4–8, TcGr4 and 19, and AmGr1/2), three orthologs of DmGr43a (BmGr10, TcGr20, and AmGr3), and AmGr4–10 representing the most basal published Grs (Robertson and Wanner 2006). The resultant phylogenetic tree based on corrected amino acid distances is shown in figure 1. Note that in the rest of the article, major Or and Gr clades will be sometimes referred to as subfamilies within each chemoreceptor family.

As was the case for *A. mellifera* (Robertson and Wanner 2006), no orthologs were discovered for the carbon dioxide heterodimer (*D. melanogaster*, Jones et al. 2007; Kwon et al. 2007) or heterotrimer (mosquitoes, *B. mori*, and *T. castaneum*, Lu et al. 2007; Robertson and Kent 2009). This gene lineage is also missing from all available more phylogenetically basal insect and related arthropod genomes, despite being highly conserved within the insects listed above, so it appears to have been lost independently from multiple insect lineages, or conserved only after the divergence of the Hymenoptera (Robertson and Kent 2009). However, it is still unknown whether aphids can detect carbon dioxide and whether this gas affects aphid oviposition (Peltonen et al. 2006; Guerenstein and Hildebrand 2008). There is also no apparent ortholog for the otherwise well-conserved DmGr43a protein, the ligand of which is unknown.

The pea aphid Gr family does contain several members of the sugar receptor subfamily that is present in all other studied insects (Kent and Robertson 2009 and references therein). This highly divergent Gr subfamily is not well resolved in the tree in figure 1 which should only be taken to show the presence of candidate sugar receptors in the pea aphid. More detailed analysis of these candidate aphid sugar receptors along with the data set of Kent and Robertson (2009) better reveals their relationships (data not shown). ApGr1–4 are a recent set of gene duplicates sharing at least 65% encoded amino acid identity in a tandem array in SCAFFOLD15735. They are most closely related to the AmGr1 and BmGr5–8 lineage in other insects, a relationship confirmed by the presence of a unique intron near the 5' end of these genes, intron "g" in Kent and Robertson (2009). ApGr5 is related to the AmGr2 and BmGr4 lineage in other insects, a relationship supported by the apparent presence of a unique intron in the middle of these genes, intron "l" in Kent and Robertson (2009), although the gene model is not completely unequivocal in this region (this gene model is also missing at least one exon at the N-terminus which might be in an upstream assembly gap). The presence of these two gene lineages in the pea aphid supports the prediction of Kent and Robertson (2009) that they represent an ancient and basal heterodimeric sugar receptor. In the pea aphid, like *B. mori*, the duplication of the

first lineage into ApGr1–4 suggests that each of these four Grs might form heterodimers with ApGr5, allowing detection and differentiation of various sugars. The pea aphid also has a third candidate sugar receptor lineage, the divergent protein ApGr6. It has no simple relationship to any of the other sugar receptors, but does appear to belong in the subfamily both phylogenetically and by virtue of having a glutamic acid immediately after the conserved TY pair in trans-membrane 7 domain, a defining feature of the sugar receptor subfamily (Kent and Robertson 2009), as well as many other distinctive amino acids and intron locations described therein. Detailed expression and functional studies will be required to determine whether it might form heterodimeric sugar receptors with ApGr1–5.

ApGr7–10 constitute another tandem array of four genes in 48-kb SCAFFOLD16347 and share at least 43% encoded amino acid identity. They have no simple phylogenetic relationship to any other known Grs. ApGr11–15 are five highly divergent proteins each on separate scaffolds (supplementary table S1a, Supplementary Material online), with no simple phylogenetic relationships with each other, other pea aphid Grs, or other insect Grs. Indeed, ApGr12 and 13 are so highly divergent they are hard to align with any of the other proteins and hence have very long branches in phylogenetic analysis and were excluded from figure 1. We are nonetheless confident they belong in the chemoreceptor superfamily because they have the conserved TYhhhhhhQF motif in TM7 domain that is characteristic of the Gr family (where h is any hydrophobic amino acid) (e.g., Wanner and Robertson 2008). ApGr16–36 form a monophyletic lineage of 21 genes sharing at least 25% amino acid identity, but dispersed around the genome in several sets of tandem duplications (supplementary table S1a, Supplementary Material online). This lineage includes two highly damaged pseudogenes that were not included in figure 1 (28 and 29), as well as two truncated gene models missing C-terminal exons in assembly gaps (22 and 26).

Finally, ApGr37–77 belong to a monophyletic lineage of 41 genes in several sets of recent tandem duplications (supplementary table S1a, Supplementary Material online). This lineage includes 11 incomplete gene models, all of which are assumed to be complete in the genome but are missing exons or parts of exons in assembly gaps (two of these gene models involve joining N- and C-terminal parts across interscaffold gaps that do not yet have experimental support for these joins, and another three are incomplete in the assembled genome but were repaired using raw traces from the Trace Archive at the NCBI). It also contains eight apparent pseudogenes containing various defects that should preclude their function, such as stop codons within alignable exons, intron splice defects, and frameshifts or large internal insertions or deletions. This lineage also contains the vast majority of the gene fragments encoding less than 50% of an average Gr, which were excluded from the numbering system. Most of these appear to be truly fragments in the genome rather than genes truncated by assembly gaps. This lineage therefore appears to be the most rapidly evolving Gr lineage in the pea aphid. The Grs in this lineage also have considerable divergence of the otherwise conserved TYhhhhhhQF motif in TM7, being (S/T)(G/A)hhThhQM in these proteins.

The Odorant Receptor Family

We identified 79 ApOr genes in the draft genome sequence (supplementary table S1b, Supplementary Material online; deduced amino acid sequences in supplementary table S2b, Supplementary Material online). Similar to the Grs, only three of these gene models are present in the RE-FSEQ set of gene models for aphid generated by NCBI. ApOrs, even more than ApGrs, are highly divergent from all other insect odorant receptors identified so far, making their identification by sequence similarity quite difficult. Although automated gene models in the combined GLEAN gene set are present for an additional 61 genes, only five of these are correct and the remainder required many changes. Among the 79 gene models proposed, 48 encode apparently full-length intact and potentially functional Ors. Ten other full-length genes showing frameshifts or stop codons are considered here as putative pseudogenes. When ESTs were searched for matches to the Or proteins encoded by our gene models, just one EST (GenBank accession CN586195) corresponded to a spliced mRNA from an ApOr gene (Or18). Nevertheless, this result provides additional support for the gene model characterizing the aphid Or subfamily to which ApOr18 belongs (described below). All of these gene models and proposed changes have been communicated to AphidBase and will be included in future releases of the genome annotation. Genes were numbered in a phylogenetic order as described below.

One Or, the DmOr83b protein from *D. melanogaster*, is unusual in that it is the heterodimeric partner for all the other Ors, at least in *D. melanogaster* (e.g., Larsson et al. 2004). Experimental evidence suggests the same in other insects such as moths (Kiely et al. 2007) and bees (Wanner, Nichols, et al. 2007). It is an unusually conserved protein encoded by a single copy gene in species studied to date. As expected, therefore, we found a single ortholog for DmOr83b and named it ApOr1. In our phylogenetic analysis, it clusters confidently with the single DmOr83b orthologs from other insects, and this protein lineage was declared the outgroup to root the tree (fig. 2). Unlike the Gr analysis, additional Ors from other insect species were not included in the tree analysis, because there are no hints of “orthologous” relationships between the remaining 78 ApOrs and other insect Ors. That is, all the other ApOrs form species-specific gene subfamily expansions or are highly divergent singletons. This observation is expected because even in comparisons between *Drosophila* and mosquito species there are few orthologous relationships (Hill et al. 2002), and the *B. mori*, *T. castaneum*, and *A. mellifera* Ors also all form lineage-specific subfamilies (Robertson and Wanner 2006; Wanner, Anderson, et al. 2007; Engstontia et al. 2008).

ApOrs form two large lineage-specific subfamily expansions, 1 of 11 genes (ApOr5–15) and one of 64 genes (ApOr16–79). These expansions include some tandem arrays of genes (i.e., ApOr20–22 on SCAFFOLD42, ApoOr23–24 on SCAFFOLD6001, ApOr40–41 on SCAFFOLD150003; see supplementary table S1b, Supplementary Material online) and most of the genes in these two main clades have apparently undergone relatively recent gene duplications. The 64-Or expansion actually consists

of two main clades: ApOr17–42 and ApOr43–79, the latter being the most recently expanded clade. The remaining three ApOrs (2–4) are singletons with different gene structures and no convincing relationships to the other ApOrs (although ApOr2 does consistently cluster with the ApOr5–15 subfamily), or other insect Ors.

Patterns of Positive Selection

To investigate the evolutionary forces shaping these newly duplicated paralogous genes, we compared the rates of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions on each branch of Or and Gr lineage-specific clades using the branch-specific model (Yang 1998; Yang and Nielsen 1998). To make our results conservative, we performed all PAML analyses on clades showing a moderate sequence divergence ($0.5 < d_S < 1$) because the LRT is reliable and powerful in detecting positive selection with computational simulations at this sequence divergence level and saturation may only be a problem at higher divergence levels (Yang 1998; Anisimova et al. 2001, 2002). Following this criterion, clade B in the Gr gene family (mean $d_S = 1.46$) and Or clade B (mean $d_S = 1.54$) were excluded from our analysis (although we do not exclude the possibility that some of the genes in these two clades are under the positive selection) and PAML analyses were only performed on the three remaining aphid-specific clades called Gr clade A, Or clade A, and Or clade B (tables 1 and 2 and figs. 1 and 2).

As shown in table 1, when the free-ratio model and one-ratio model were compared, the LRT was significant in all three clades examined, suggesting that selective pressure varies among branches in each targeted clade and that some of these proteins might have evolved under positive selection. However, this observation does not rule out the alternative possibility of a relaxation of selective constraint on some branches. To further test for positive selection, we evaluated the selective pressures on homologous sites among sequences in each clade using the site-specific models, a random distribution of amino acid substitutions being expected under a relaxation of selective pressure, whereas overrepresented amino acid changes in functional domains are expected under positive selection. Estimates of the parameter values of the ω distribution under M8 (table 2) indicate that a fraction of sites are under positive selection in each of the clades tested. Even after correction for multiple testing, LRTs (see Material and Methods) of M8 and M8a were significant and confirmed that the variation in selection pressure was due to the evolution of a subset of sites by positive Darwinian selection (table 2). Essentially similar results were obtained with the even more conservative M1a–M2a comparison, although the marginal significance for “Or clade B” under M8 was not detected under the M2a model (table 2).

We further examined the distribution of the inferred positively selected sites under the M8 model in the two clades showing some significant results for both tests (Gr clade A and Or clade A). We identified, at the PP > 50% level, 18 sites as potential targets of positive selection in Gr clade A and 24 sites in Or clade A (and, respectively,

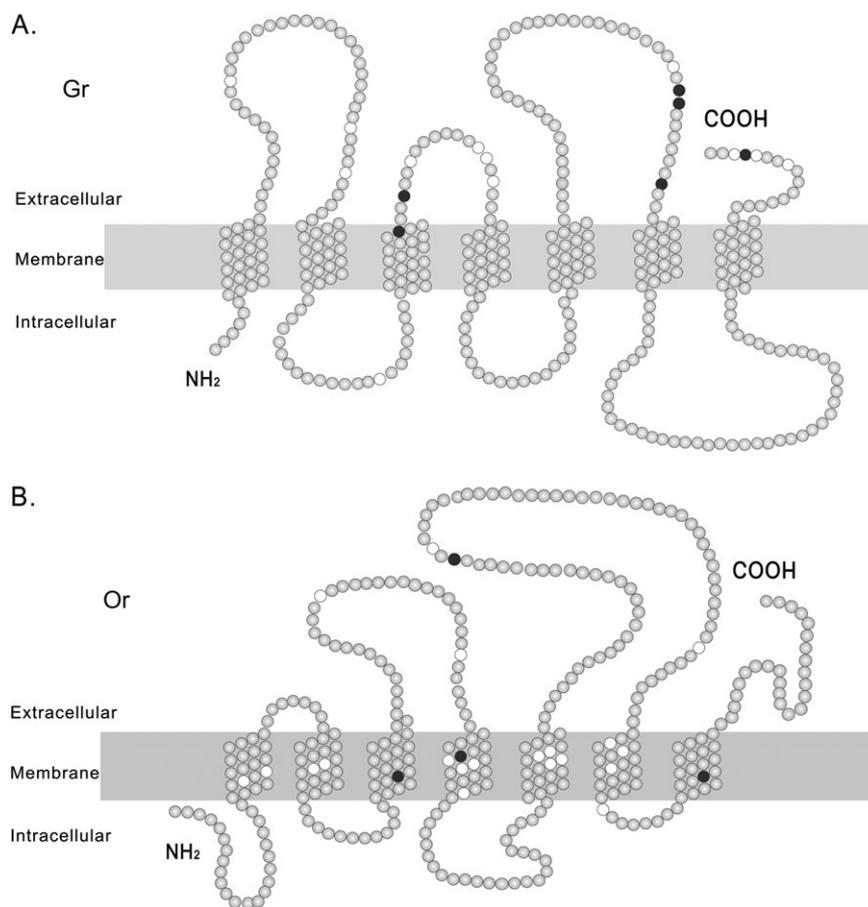


FIG. 3.—Positively selected sites on the structural topology of pea aphid (A) gustatory receptors (GRs) and (B) odorant receptors (ORs). Gr37 and Or43 separately represent the Gr and Or families and were used to predict the topology by TMHMM. Circles indicate the amino acids (gray: neutrally selected sites; white: positively selected sites with PPs > 50%; black: positively selected sites with PPs > 95%). The gray shaded areas denote the cell membrane.

12 and 11 overlapping sites between the M8 and M2a models) (table 2). For both Ors and Grs, TMHMM algorithms predicted an intracellular orientation of the NH₂-tails and an extracellular orientation of the COOH tails, supporting the “reverse” membrane topology of insect chemoreceptors (Sato et al. 2008; Smart et al. 2008; Wicher et al. 2008). When all the putatively positively selected sites under M8 were plotted onto the chemoreceptor topology, their distribution was clearly heterogeneous between the different protein regions (figs. 3 and 4). The proportion of positively selected sites in extracellular regions (ER) was 72% for Gr clade A, significantly greater than expected under

a homogeneous distribution model (χ^2 test, $P = 6.3E^{-5}$), as the ER only constitutes about 25% of the entire protein. By contrast, the percentage of positively selected sites in trans-membrane regions (TM) was 75%, significantly over-represented in Or clade A ($P = 4.9E^{-4}$). Therefore, we find that the distribution of candidate sites among regions differed between the receptor types ($G_4 = 26.16$, $P = 2.9E^{-5}$, fig. 4). If only the overlapping sites were considered, essentially similar results were obtained (table 2). In addition, this pattern is confirmed when we restrict our analysis to sites with high PP (Gr clade A, at PP > 95%, $P = 0.04$; Or clade A, at PP > 90%; $P = 0.03$). We caution that

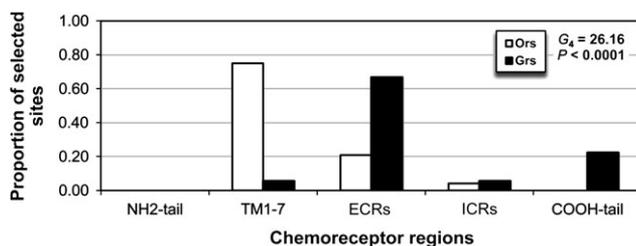


FIG. 4.—Distribution of the positively selected sites by protein regions (NH₂-tail, trans-membrane domains 1–7 (TM1–7), extracellular regions (ECRs), intracellular regions (ICRs) and COOH tail). Ors: odorant receptor (white bars); Grs: gustatory receptor (black bars). Statistical test: G test with four dfs.

the likelihood method is known to falsely detect positively selected sites under certain conditions (Suzuki and Nei 2002; Zhang 2004), so the positively selected sites reported here should be confirmed by other statistical methods and/or experiments.

Discussion

Our study addressing the molecular characterization and evolution of the chemoreceptor superfamily in the pea aphid genome provides the first molecular insights into the mechanisms of chemoreception in a hemipteran species.

Like in some other insect lineages (Robertson and Wanner 2006; Robertson and Kent 2009), no carbon dioxide receptors were identified. As far as we are aware, it is still unknown whether aphids can detect carbon dioxide (Peltonen et al. 2006; Guerenstein and Hildebrand 2008), but if they can, like honeybees they must be using other receptors for this purpose. Moreover, we show that the pea aphid has no simple ortholog of the DmGr43a gene lineage, whereas it is present in other insects as a single reasonably well-conserved ortholog in Diptera and *Apis* or as a duplication or an expansion of the lineage in *B. mori* (Wanner and Robertson 2008) and *T. castaneum* (Consortium T.G.S. 2008). This result might indicate that this gene was either lost at some point in the pea aphid lineage or conserved only in endopterygote insects.

Several members of the sugar receptor subfamily were identified in the pea aphid genome, which allows us to predict that aphids can detect sugars, as one might expect but which does not appear to have been demonstrated experimentally. Separate detailed analysis of these candidate sugar receptor genes and proteins compared with those from other insects examined by Kent and Robertson (2009) reveals that they represent three lineages, two of which are ancient within insects and are hypothesized to form a functional heterodimer. Independent expansion of one of these lineages in both *B. mori* and the pea aphid suggests that unlike *A. mellifera* and *Nasonia vitripennis*, which have single representatives of each lineage, aphids might be able to differentiate various sugars. The third lineage is a highly divergent protein of unclear role in sugar perception.

In addition to five highly divergent singletons (ApGr11–15), as expected from previous experience with other insects from different orders (*B. mori*, *T. castaneum*, and *A. mellifera*), there are three aphid-specific expansions of 4, 21, and 41 Grs. These expansions reveal typical features of recently expanded gene subfamilies, including commonly being in tandem duplications on single scaffolds. They also contain the few pseudogenes recognized in the Gr family. Although the functions of the four-gene subfamily (ApGr7–10) and the five highly divergent singletons (ApGr11–15) are quite obscure, we propose that the other two Gr subfamily expansions of 21 and 41 genes might represent bitter taste receptors involved in detecting the many plant defensive compounds aphids must utilize in finding suitable host plants (e.g., Pickett and Glinwood 2007; Webster et al. 2008). These candidate bitter receptors show no close relationship to the only fly receptor whose

bitter ligand has been identified, the DmGr66a receptor for caffeine (Moon et al. 2006), but this is not surprising as this receptor is only conserved in Diptera, with no simple relatives in other insects. The same is true for the other Grs in flies that have been implicated in perception of bitter compounds (Thorne et al. 2004; Wang et al. 2004).

As expected, the pea aphid genome encodes a single conserved ortholog of DmOr83b, the unusual member of the odorant receptor family that plays a distinct role as a heterodimeric partner for the specific Ors. In addition to three highly divergent singletons, the remainder of the pea aphid odorant receptor family consists of 2 aphid-specific subfamilies of 11 and 64 genes. We propose that these Ors are involved primarily in detection of host plant volatiles involved in host selection. However, a new family of chemoreceptor genes (ionotropic receptors, IRs), which has recently been discovered in *D. melanogaster* (Benton et al. 2009), may also play a role in the detection of odorants and tastants. As yet, nothing is known about these receptors in aphids.

Interestingly, the presence of aphid-specific expansions in the Or and Gr repertoires seems to be correlated with some signature of positive selection, with the clades experiencing the most recent duplications showing a high rate of nonsynonymous substitutions (figs. 1 and 2; table 1). Although previous studies on selection in insect Ors/Grs focused on *Drosophila* species orthologous comparisons (e.g., McBride 2007; Gardiner et al. 2008), our study for the first time addresses selection within sets of paralogous Or and Gr expansions in an insect species.

A relaxation of purifying selection could also explain the observed pattern (e.g., McBride 2007): If, for example, a protein under strong purifying selection is no longer in use in a new environment, the mutations that cause amino acid changes or premature stop codons in this superfluous receptor would no longer be deleterious and could become more frequent, leading to an overall ω closer to 1. However, one characteristic of the substitution rate that might help to rule out the possibility that a complete relaxation of purifying selection underlies ω close to 1 or slightly significantly greater than 1 is its spatial distribution along proteins. Although identifying the protein sites that could be the target of selection, we showed that these sites are not evenly distributed along the proteins (fig. 3), whereas we would expect new mutations to be fixed at random positions and the spatial distribution of amino acid substitutions not to mirror specific regions of the receptors if purifying selection was completely relaxed.

The amino acid residues we identified to be subject to positive selection may determine binding specificity and provide useful information on the diversity of odorants and tastants that aphids encounter when they explore new habitats and environments. An interesting observation from our analysis is that Grs and Ors have different sets of positively selected sites, suggesting the possibility that these two gene families might have different binding pockets and bind structurally distinct classes of ligand. Recently, Gardiner et al. (2009) analyzed the distribution of amino acid sites in Grs and Ors that show evidence for divergence under either positive selection or relaxed purifying constraints, in the genomes of 12 *Drosophila* species, and they also found significant differences between these two

receptor types (Gardiner et al. 2009). The two studies might not be directly comparable because the *Drosophila* study focused on orthologous comparisons and overall the distribution pattern of positively selected sites was quite different. Nevertheless, our result here is broadly consistent with their finding and suggests that insect Ors and Grs might have distinct molecular properties and mechanisms of ligand recognition and/or signal transduction.

It is also interesting to note that the majority of positively selected sites in vertebrate bitter taste receptors (T2Rs) are located in ERs that are thought to be involved in ligand binding (Shi et al. 2003; Soranzo et al. 2005; Shi and Zhang 2006), whereas positive selection has also been detected in trans-membrane regions, which are believed to form the binding pocket in vertebrate Ors (Emes et al. 2004). Therefore, our analysis might reflect the convergent evolution of insect and vertebrate chemoreceptors in binding of environmental chemicals. Functional studies on these receptors might benefit from focusing on the regions we identify as undergoing positive selection, and eventually a convergence of functional and evolutionary studies will hopefully contribute to an improved understanding of how this novel superfamily of chemoreceptors works.

The fact that aphid Ors and Grs in the most recent gene expansions underwent positive selection might indicate that gene duplication was important for aphids to acquire new chemosensory functions and may have been driven by a change of ligand (Nei et al. 2008). For example, in *Drosophila* species, lineage-specific gene duplication seems to have led to additional specialization in response to specific ecological conditions (Guo and Kim 2007; McBride 2007; McBride et al. 2007). Aphids are known to rely heavily on their senses of smell and taste to recognize stimuli in their environment, such as resources, natural enemies, and mates (Pickett et al. 1992). In particular, it has been shown that host plant acceptance, a key component of host plant specialization, relies mainly on chemosensory processes in the pea aphid (Caillaud and Via 2000) and probably in related species as well (Park et al. 2000; Park and Hardie 2003, 2004). In this context, the acquisition of a novel host may drive the adaptive divergence of sensory systems by positive selection, and the abandonment of an ancestral host may result in the deterioration of older sensory adaptations by genetic drift or positive selection. Therefore, the observation of signatures of positive selection in Or and Gr genes might be related to the striking ecological specialization observed in *A. pisum* (Via 1999; Ferrari et al. 2006, 2008), the chemosensory genes being subject to divergent evolutionary pressures when aphids enter new niches during host shifts or host specialization events. Future comparative studies among several insect species, and population analyses using polymorphism data within aphid species (like in *D. melanogaster*, Aguade 2009), will help to further explore the role of chemoreceptor genes in adaptive ecological specialization.

Conclusion

Emerging information on the molecular basis of aphid host plant recognition influences our understanding of the mechanisms underlying speciation and biodiversity in

aphids (Smadja et al. 2008; Via and West 2008) and the development of new pest control strategies (Pickett et al. 1997). Therefore, the identification of genes responsible for olfaction and gustation in the pea aphid will have significant positive impacts on fields as diverse as evolutionary biology, comparative insect genetics and agriculture. First, it offers new opportunities for comparative analyses across several insect orders. Indeed, although these chemoreceptors are being characterized in a growing number of insect species, the pea aphid chemoreceptors currently represent an outgroup to the other insect chemoreceptors. In this context, metanalyses of paralogous comparisons across several insect species is a promising route to identify Or and Gr candidate ligand-binding domains. Moreover, gaining insights into the molecular basis of olfaction and gustation in the pea aphid is a key step toward a better understanding of the biology of an insect relying mainly on chemical cues to locate and assess food, habitat, and conspecifics. Another application of our findings might be the definition of new pest control strategies based on the manipulation of the attraction of aphids to specific agricultural cultivars, this being potentially applied to pea aphids as well as other related species. Finally, the identification of genes involved in chemoreception and showing patterns of evolution under selection provides interesting candidate genes on which further investigation of the genetic basis of host plant specialization and speciation in aphids can be developed.

Supplementary Material

We provide supplementary tables S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Don Gilbert for his arthropod PSI-BlastP server, AphidBase staff, and especially Fabrice Legeai for entering manually annotated gene models into Apollo, the Baylor Human Genome Sequencing Center for access to the pea aphid genome assembly before publication, Anastasia Gardiner for helpful discussions on data analysis, and Hollis Woodard and Brielle Fischman for comments on the manuscript. C.S. was funded by a Marie Curie Fellowship (European Community); P.S. was funded by a start-up fund of “Hundreds-Talent Program” from Chinese Academy of Sciences; R.K.B. was funded by NERC; H.M.R. was funded by USDA grant 2008-35302-18815.

Literature Cited

- Ache BW, Young JM. 2005. Olfaction: diverse species, conserved principles. *Neuron*. 48:417–430.
- Aguade M. 2009. Nucleotide and copy-number polymorphism at the odorant receptor genes Or22a and Or22b in *Drosophila melanogaster*. *Mol Biol Evol*. 26:61–70.
- Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol*. 18:1585–1592.

- Anisimova M, Bielawski JP, Yang ZH. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Mol Biol Evol.* 19:950–958.
- Bargmann CI. 2006. Comparative chemosensation from receptors to ecology. *Nature.* 444:295–301.
- Benton R. 2006. On the ORigin of smell: odorant receptors in insects. *Cell Mol Life Sci.* 63:1579–1585.
- Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell.* 136:149–162.
- Berlacher SH, Feder JL. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? *Annu Rev Entomol.* 47:773–815.
- Birkett MA, Pickett JA. 2003. Aphid sex pheromones: from discovery to commercial production. *Phytochemistry.* 62:651–656.
- Bush GL, Butlin RK. 2004. Sympatric speciation in insects: an overview. In: Dieckmann MDU, Metz H, Tautz D, editors. *Adaptive speciation.* Cambridge (UK): Cambridge University Press. p. 229–248.
- Caillaud MC, Via S. 2000. Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. *Am Nat.* 156:606–621.
- Clyne PJ, Warr CG, Carlson JR. 2000. Candidate taste receptors in *Drosophila*. *Science.* 287:1830–1834.
- Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim JH, Carlson JR. 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron.* 22:327–338.
- Consortium The International Aphid Genomics. Genome sequence of the pea aphid *Acyrtosiphon pisum*. PLoS Biol. Submitted.
- Consortium *Tribolium* Genome Sequencing. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature.* 452:949–955.
- Del Campo ML, Via S, Caillaud MC. 2003. Recognition of host-specific chemical stimulants in two sympatric host races of the pea aphid *Acyrtosiphon pisum*. *Ecol Entomol.* 28:405–412.
- Dickens JC, Callahan FE, Wergin WP, Murphy CA, Vogt RG. 1998. Intergeneric distribution and immunolocalization of a putative odorant-binding protein in true bugs (Hemiptera, Heteroptera). *J Exp Biol.* 201:33–41.
- Drès M, Mallet J. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. *Phil Trans Roy Soc Lond Ser B-Biol Sci.* 357:471–492.
- Emes RD, Beatson SA, Ponting CP, Goodstadt L. 2004. Evolution and comparative genomics of odorant and pheromone-associated genes in rodents. *Genome Res.* 14:591–602.
- Engsontia P, Sanderson AP, Cobb M, Walden KKO, Robertson HM, Brown S. 2008. The red flour beetle's large nose: an expanded odorant receptor gene family in *Tribolium castaneum*. *Insect Biochem Mol Biol.* 38:387–397.
- Ferrari J, Godfray HCJ, Faulconbridge AS, Prior K, Via S. 2006. Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. *Evolution.* 60:1574–1584.
- Ferrari J, Via S, Godfray HCJ. 2008. Population differentiation and genetic variation in performance on eight hosts in the pea aphid complex. *Evolution.* 62:2508–2524.
- Gao Q, Chess A. 1999. Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics.* 60:31–39.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. *Drosophila* chemoreceptor gene evolution: selection, specialisation and genome size. *Mol Ecol.* 17:1648–1657.
- Gardiner A, Butlin RK, Jordan WC, Ritchie MG. 2009. Sites of evolutionary divergence differ between olfactory and gustatory receptors of *Drosophila*. *Biol Lett.* online publication 5:244–247.
- Guerenstein PG, Hildebrand JG. 2008. Roles and effects of environmental carbon dioxide in insect life. *Annu Rev Entomol.* 53:161–178.
- Guo S, Kim J. 2007. Molecular evolution of *Drosophila* odorant receptor genes. *Mol Biol Evol.* 24:1198–1207.
- Hallem EA, Dahanukar A, Carlson JR. 2006. Insect odor and taste receptors. *Annu Rev Entomol.* 51:113–135.
- Hardie J, Visser JH, Piron PGM. 1995. Peripheral odour perception by adult aphid forms with the same genotype but different host-plant preferences. *J Insect Physiol.* 41:91–97.
- Hawthorne DJ, Via S. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature.* 412:904–907.
- Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ. 2002. G protein coupled receptors in *Anopheles gambiae*. *Science.* 298:176–178.
- Jiggins CD, Emeliano I, Mallet J. 2005. Assortative mating and speciation as pleiotropic effects of ecological adaptation: examples in moths and butterflies. In: Fellowes M, Holloway G, Rolff J, editors. *Insect evolutionary ecology.* London: Royal Entomological Society. p. 451–473.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comp Appl Biosci.* 8:275–282.
- Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB. 2007. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature.* 445:86–90.
- Kent LB, Robertson HM. 2009. Molecular evolution of the sugar receptors in insects. *BMC Evol Biol.* 9:41.
- Kent LB, Walden KKO, Robertson HM. 2008. The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. *Chem Senses.* 33:79–93.
- Kiely A, Authier A, Kralicek AV, Warr CG, Newcomb RD. 2007. Functional analysis of a *Drosophila melanogaster* olfactory receptor expressed in Sf9 cells. *J Neurosci Meth.* 159:189–194.
- Krieger J, Klink O, Mohl C, Raming K, Breer H. 2003. A candidate olfactory receptor subtype highly conserved across different insect orders. *J Compar Physiol a-Neuroethol Sens Neural Behav Physiol.* 189:519–526.
- Kristoffersen L, Hansson BS, Anderbrant O, Larsson MC. 2008. Agglomerular Hemipteran antennal lobes-basic neuroanatomy of a small nose. *Chem Senses.* 33:771–778.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Evol.* 305:567–580.
- Kwon JY, Dahanukar A, Weiss LA, Carlson JR. 2007. The molecular basis of CO₂ reception in *Drosophila*. *Proc Natl Acad Sci USA.* 104:3574–3578.
- Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB. 2004. Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron.* 43:703–714.
- Lu T, Qiu YT, Wang G, et al. (11 co-authors). 2007. Odor coding in the maxillary palp of the malaria vector mosquito *Anopheles gambiae*. *Curr Biol.* 17: 1533–1544.
- Matsunami H, Amrein H. 2003. Taste and pheromone perception in mammals and flies. *Genome Biol.* 4:220.

- McBride CS. 2007. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci USA*. 104:4996–5001.
- McBride CS, Arguello JR, O'Meara BC. 2007. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*. 177:1395–1416.
- Moon SJ, Kottgen M, Jiao YC, Xu H, Montell C. 2006. A taste receptor required for the caffeine response in vivo. *Curr Biol*. 16:1812–1817.
- Morrison WP, Peairs FB. 1998. Response model concept and economic impact. In: Quisenberry SS, Peairs FB, editors. *Response model for an introduced pest—the Russian wheat aphid*. Lanham, MD: Entomological Society of America.
- Nei M, Niihura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat Rev Genet*. 9:951–963.
- Nozawa M, Nei M. 2007. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc Natl Acad Sci USA*. 104:7122–7127.
- Park KC, Elias D, Donato B, Hardie J. 2000. Electroantennogram and behavioural responses of different forms of the bird cherry-oat aphid, *Rhopalosiphum padi*, to sex pheromone and a plant volatile. *J Insect Physiol*. 46:597–604.
- Park KC, Hardie J. 2003. Electroantennogram responses of aphid nymphs to plant volatiles. *Physiol Entomol*. 28:215–220.
- Park KC, Hardie J. 2004. Electrophysiological characterisation of olfactory sensilla in the black bean aphid, *Aphis fabae*. *J Insect Physiol*. 50:647–655.
- Patch HM, Velarde RA, Walden KKO, Robertson HM. 2009. A candidate pheromone receptor and two odorant receptors of the hawkmoth *Manduca sexta*. *Chem Senses*. 34:305–316.
- Peltonen PA, Julkunen-Tiitto R, Vapaavuori E, Holopainen JK. 2006. Effects of elevated carbon dioxide and ozone on aphid oviposition preference and birch bud exudate phenolics. *Glob Change Biol*. 12:1670–1679.
- Pickett JA, Glinwood RT. 2007. Chemical ecology. In: van Emden H, Harrington R, editors. *Aphids as crop pests*. Oxford (United Kingdom): Oxford University Press. p. 235–260.
- Pickett JA, Wadhams LJ, Woodcock CM. 1997. Developing sustainable pest control from chemical ecology. *Agric Ecosyst Environ*. 64:149–156.
- Pickett JA, Wadhams LJ, Woodcock CM, Hardie J. 1992. The chemical ecology of aphids. *Annu Rev Entomol*. 37:67–90.
- Pope TW, Campbell CAM, Hardie J, Pickett JA, Wadhams LJ. 2007. Interactions between host-plant volatiles and the sex pheromones of the bird cherry-oat aphid, *Rhopalosiphum padi* and the damson-hop aphid, *Phorodon humuli*. *J Chem Ecol*. 33:157–165.
- Powell G, Hardie J. 2001. The chemical ecology of aphid host alternation: how do return migrants find the primary host plant? *Appl Entomol Zool*. 36:259–267.
- Powell G, Tosh CR, Hardie J. 2006. Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu Rev Entomol*. 51:309–330.
- Robertson HM, Kent LB. 2009. Evolution of the gene lineage encoding the carbon dioxide heterodimeric receptor in insects. *J Insect Sci*. 9:19.
- 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.
- Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*. 100:14537–14542.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
- Rutzler M, Zwiebel LJ. 2005. Molecular biology of insect olfaction: recent progress and conceptual models. *J Compar Physiol a-Neuroethol Sens Neural Behav Physiol*. 191:777–790.
- Sabater-Munoz B, Legeai F, Rispé C, et al. (22 co-authors). 2006. Large-scale gene discovery in the pea aphid *Acyrtosiphon pisum* (Hemiptera). *Genome Biol*. 7:R21.
- Saitou N, Nei M. 1987. The neighbor-joining method—a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 4:406–425.
- Sato K, Pellegrino M, Nakagawa T, Nakagawa T, Vosshall LB, Touhara K. 2008. Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*. 452:1002–1006.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. 18:502–504.
- Servedio MR. 2009. The role of linkage disequilibrium in the evolution of premating isolation. *Heredity*. 102:51–56.
- Shi P, Zhang J. 2006. Contrasting modes of evolution between vertebrate sweet/umami receptor genes and bitter receptor genes. *Mol Biol Evol*. 23:292–300.
- Shi P, Zhang J, Yang H, Zhang YP. 2003. Adaptive diversification of bitter taste receptor genes in mammalian evolution. *Mol Biol Evol*. 20:805–814.
- Smart R, Kiely A, Beale M, Vargas E, Carraher C, Kralicek AV, Christie DL, Chen C, Newcomb RD, Warr CG. 2008. *Drosophila* odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect Biochem Mol Biol*. 38:770–780.
- Smadja C, Galindo J, Butlin RK. 2008. Hitching a lift on the road to speciation. *Mol Ecol*. 17:4177–4180.
- Soranzo N, Bufe B, Sabeti PC, Wilson JF, Weale ME, Marguerie R, Meyerhof W, Goldstein DB. 2005. Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. *Current Biol*. 15:1257–1265.
- Suzuki Y, Nei M. 2002. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. *Mol Biol Evol*. 19:1865–1869.
- Swanson WJ, Nielsen R, Yang Q. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol*. 20:18–20.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*. 25:4876–4882.
- Thorne N, Amrein H. 2008. Atypical expression of *Drosophila* gustatory receptor genes in sensory and central neurons. *J Compar Neurol*. 506:548–568.
- Thorne N, Chromey C, Bray S, Amrein H. 2004. Taste perception and coding in *Drosophila*. *Current Biol*. 14:1065–1079.
- Via S. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*. 53:1446–1457.
- Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol*. 16:381–390.
- Via S, Bouck AC, Skillman S. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II.

- Selection against migrants and hybrids in the parental environments. *Evolution*. 54:1626–1637.
- Via S, Hawthorne DJ. 2002. The genetic architecture of ecological specialization: correlated gene effects on host use and habitat choice in pea aphids. *Am Nat*. 159:S76–S88.
- Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol Ecol*. 17: 4334–4345.
- Visser JH, Piron PGM, Hardie J. 1996. The aphids' peripheral perception of plant volatiles. *Entomol Exp Appl*. 80:35–38.
- Vosshall LB, Amrein H, Morozov PS, Rzhetsky A, Axel R. 1999. A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell*. 96:725–736.
- Vosshall LB, Stocker RE. 2007. Molecular architecture of smell and taste in *Drosophila*. *Annu Rev Neurosci*. 30: 505–533.
- Wang ZR, Singhvi A, Kong P, Scott K. 2004. Taste representations in the *Drosophila* brain. *Cell*. 117:981–991.
- Wanner KW, Anderson AR, Trowell SC, Theilmann DA, Robertson HM, Newcomb RD. 2007. Female-biased expression of odourant receptor genes in the adult antennae of the silkworm, *Bombyx mori*. *Insect Mol Biol*. 16:107–119.
- Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM. 2007. A honey bee odourant receptor for the queen substance 9-oxo-2-decenoic acid. *Proc Natl Acad Sci USA*. 104:14383–14388.
- Wanner KW, Robertson HM. 2008. The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Mol Biol*. 17:621–629.
- Webster B, Bruce T, Dufour S, Birkemeyer C, Birkett M, Hardie J, Pickett J. 2008. Identification of volatile compounds used in host location by the black bean aphid, *Aphis fabae*. *J Chem Ecol*. 34:1153–1161.
- Wicher D, Schafer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS. 2008. *Drosophila* odourant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature*. 452:1007–1011.
- Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinform*. Chapter 6:Unit 6.4.
- Wong WSW, Yang ZH, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*. 168:1041–1051.
- Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comp Appl Biosci*. 13:555–556.
- Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15:568–573.
- Yang ZH, Nielsen R. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. 46:409–418.
- Yang Z, Wong WS, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 22:1107–1118.
- Zhang J. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol Biol Evol*. 21:1332–1339.

Diethard Tautz, Associate Editor

Accepted June 3, 2009