

RESEARCH ARTICLE

Open Access



Positive selection at sites of chemosensory genes is associated with the recent divergence and local ecological adaptation in cactophilic *Drosophila*

Fernando Diaz¹, Carson W. Allan¹ and Luciano M. Matzkin^{1,2,3*} 

Abstract

Background: Adaptation to new hosts in phytophagous insects often involves mechanisms of host recognition by genes of sensory pathways. Most often the molecular evolution of sensory genes has been explained in the context of the birth-and-death model. The role of positive selection is less understood, especially associated with host adaptation and specialization. Here we aim to contribute evidence for this latter hypothesis by considering the case of *Drosophila mojavensis*, a species with an evolutionary history shaped by multiple host shifts in a relatively short time scale, and its generalist sister species, *D. arizonae*.

Results: We used a phylogenetic and population genetic analysis framework to test for positive selection in a subset of four chemoreceptor genes, one gustatory receptor (*Gr*) and three odorant receptors (*Or*), for which their expression has been previously associated with host shifts. We found strong evidence of positive selection at several amino acid sites in all genes investigated, most of which exhibited changes predicted to cause functional effects in these transmembrane proteins. A significant portion of the sites identified as evolving positively were largely found in the cytoplasmic region, although a few were also present in the extracellular domains.

Conclusions: The pattern of substitution observed suggests that some of these changes likely had an effect on signal transduction as well as odorant recognition and protein-protein interactions. These findings support the role of positive selection in shaping the pattern of variation at chemosensory receptors, both during the specialization onto one or a few related hosts, but as well as during the evolution and adaptation of generalist species into utilizing several hosts.

Keywords: Odorant receptor, Gustatory receptor, Population genetics, Molecular evolution, Adaptation, Cactophilic *Drosophila*

Background

Studying host adaptation is critical to understanding the ecological and evolutionary history of phytophagous insects [1–5]. Host shifts represent a distinct and sometimes novel set of complex challenges since new host plants can differ in nutritional content, chemical composition, microorganisms and defensive quality [6–8]. The use of novel resources when

shifting to a new host can therefore affect survival of larva and adult stages, fecundity of feeding females as well as behavioral patterns of oviposition and mating, causing ultimately substantial impact on fitness. This has led to the hypothesis that in addition to genes involved in performance relative to a new environment, adaptation to new hosts often involves mechanisms of host preference [8–10]. Given their role in odorant and gustatory sensory and ultimately in the discrimination of suitable hosts, chemoreceptors have been the focus of several studies examining the evolutionary mechanisms underlying host adaptation [11–17].

* Correspondence: lmatzkin@email.arizona.edu

¹Department of Entomology, University of Arizona, Tucson, AZ 85721, USA

²Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

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



Chemical stimuli recognition occurs through odorant (*Or*) and gustatory receptor (*Gr*) proteins, converting volatile and soluble chemicals from the environment into electrical outputs through nerve impulses [18, 19]. In insects, these proteins have seven trans-membrane domains located on the dendritic membranes of sensory neurons housed in sensilla of receptor organs, with a reverse orientation when compared to mammalian topology (COOH-tail extracellular) [20, 21]. While the number of genes in these families seems to be largely stable across *Drosophila* evolution, there have been constant gene loss and gain events through the birth-and-death model in insects [22–28] where divergence arises from the selection of beneficial mutations between paralogous genes [29]. This mode of evolution, as well as the clustered organization of these genes has provided strong support to the hypothesis that these loci play a critical role in ecological shifts and specialization [9, 11–17, 22, 23]. In particular, the ecological specialization observed within *Drosophila* has been suggested to have contributed to the higher pseudogenization levels observed in sensory genes [14, 15]. However, the analysis of 12 *Drosophila* species genomes [30], has shown that genome size or endemism could also predict pseudogenization rates, but the few specialist species compared and the lack of evolutionary data over short time scales has prevented efforts to disentangle these causes [31].

The dramatic heterogeneity in family size and divergence among chemosensory gene paralogs suggest that these gene families have evolved rapidly. However, the pattern of variation and divergence at a majority of genes are more consistent with purifying selection and only a small portion of genes have shown evidence of positive selection, some of which have been linked to specialization [14, 15, 29, 31]. Thus, specialist species such as *D. sechellia* and *D. erecta* have shown higher K_a/K_s ratios when compared with their closest generalist species [15]. The majority of amino acid sites showing signatures of positive selection in these families have been found in the cytoplasmic domain, suggesting that selection changes are associated with signal transduction rather than the detection of odorant molecules [32]. To date most comparisons have involved a few highly divergent species; here we examine the case of the cactophilic local specialist *D. mojavensis* and its generalist sister species *D. arizonae* to investigate the role of positive selection in four candidate chemoreceptor genes (one *Gr* and three *Or*). These species have adapted to develop and feed on necrotic cacti, showing partially overlapping distributions in xeric regions of southwestern United States and northwestern Mexico [33, 34]. The four candidate loci examined here (*Or67c*, *Or83c1*, *Or83c2* and *Gr63a*) have shown a pattern of differential expression associated with host shifts suggesting their role in host

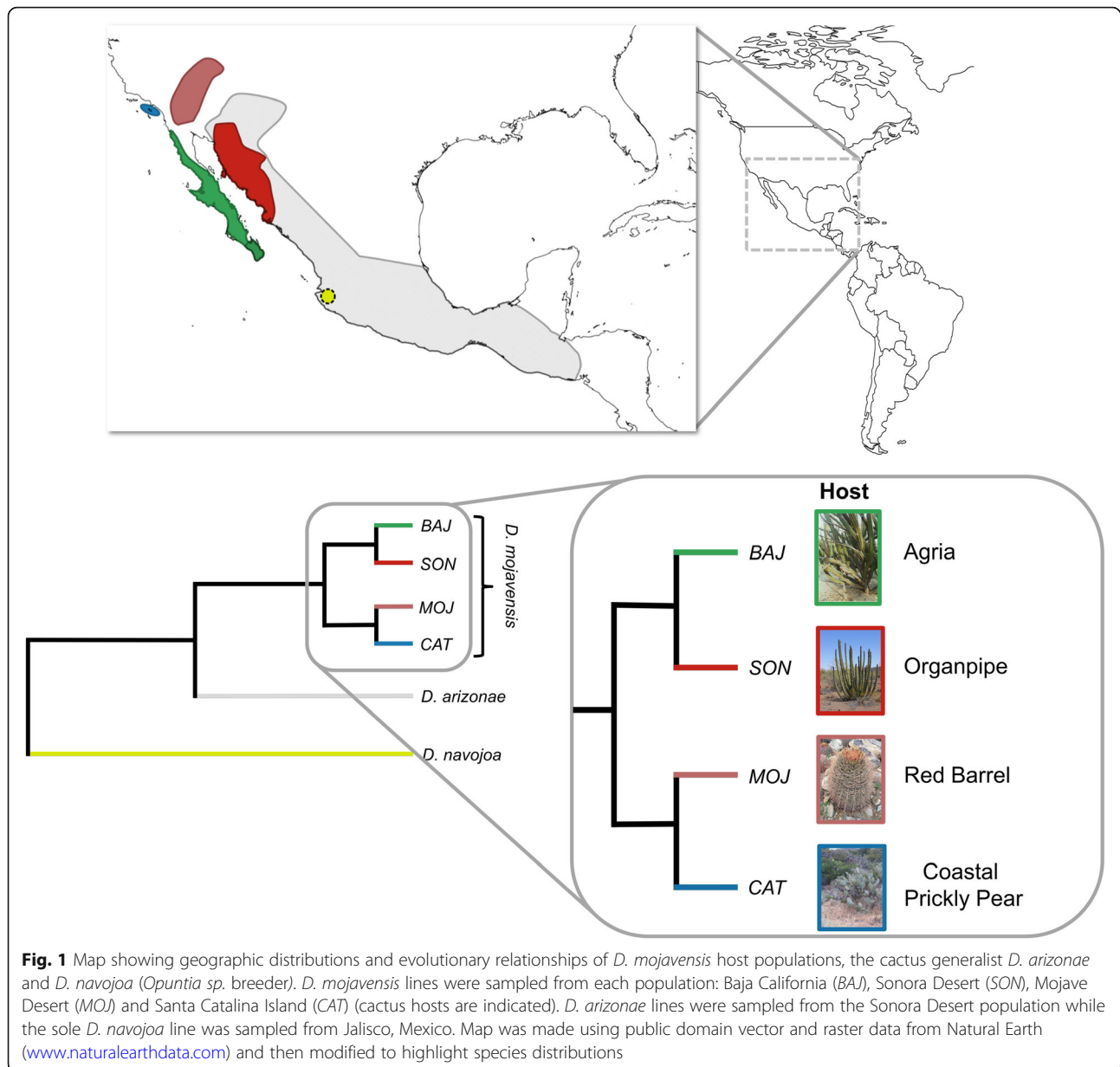
preference, presumably by detecting volatiles associated with either the presence of nutrients or toxins produced in rotting cacti [35–37]. *Or67c* and *Gr63a* are known to play a role in detecting the by-products of alcoholic fermentation as well as CO₂ [38, 39], while *Or83c1* and *Or83c2* have been suggested to play role in food detection [38, 40].

Drosophila mojavensis offers a unique case to study host adaptation since its biogeographical history has been shaped by multiple host shifts in a relatively short period of time [41]. Since its divergence from its sister species *D. arizonae*, the *D. mojavensis* lineage has expanded through a rapid specialization process (<0.5 Mya; [42, 43]) producing four host populations adapted to different cactus species (Fig. 1): agria (*Stenocereus gummosus*) in Baja California; coastal prickly pear (*Opuntia littoralis*) in Santa Catalina Island; red barrel (*Ferocactus cylindraceus*) in the Mojave Desert and organpipe (*S. thurberi*) in the Sonora Desert population [33, 34]. We investigated patterns of molecular evolution in four chemoreceptor genes across the four host populations of *D. mojavensis* and the generalist *D. arizonae* to test for positive selection using a multiple approach of gene-wide, polymorphic-based, codon-based and functional tests. Then, we analyzed the distribution of sites showing evidence of positive selection on the structure of these trans-membrane proteins in order to test whether selection has been biased towards specific domains and discussed the implication of these results in terms of the role of positive selection in host preference adaptation.

Results

Genetic diversity, neutrality tests and divergence

For each locus examined, between 7 and 15 sequences were generated for each of the four *D. mojavensis* populations (Sonora Desert, SON; Santa Catalina Island, CAT; Mojave Desert, MOJ and Baja California, BAJ), as well as between 8 and 12 sequences of *D. arizonae*. After excluding introns, these amplicons had between 1170 bp for *Or83c2* to 1479 bp for *Gr63a* (Table 1). Given the population structure in *D. mojavensis* (Additional file 1: Table S2) [44, 45], we sampled multiple individuals per population. Here we report average values per population of *D. mojavensis* to make them comparable with those in *D. arizonae* (Table 1). Between 12 and 35 polymorphic sites and 6 and 12 different haplotypes were detected across the loci and species examined (Table 1). Diversity estimates in *D. arizonae* often fall between the estimates for individual population in *D. mojavensis* (Additional file 1: Table S2). However, most genetic diversity estimates were higher in *D. arizonae* when compared with average estimates across *D. mojavensis* populations. Thus, haplotype diversity was higher in *D.*



arizonae for all genes, genetic diversity for three genes and nucleotide diversity for two of genes analyzed (Table 1). Tajima's D and Fu and Li's D and F neutrality tests were not significant for most genes (Table 1). Fu's F_S statistic was significantly negative for the majority of genes in both species (Table 1). For one of the loci (*Gr63a*) and in one *D. mojavenensis* population (MOJ) these tests were significantly positive (Additional file 1: Table S3), indicating a recent bottleneck causing a heterogeneous pattern in the demographic history of this species.

The genetic divergence assessed by comparing the number of synonymous (K_s) and nonsynonymous (K_a) substitutions ratio (K_a/K_s) with respect to the *D. navojoa* outgroup was often larger in *D. arizonae* for all loci with

the exception of *Or83c1* (Table 1). The genetic structure as assessed by Φ_{ST} indicated the existence of significant population structure in *D. mojavenensis*, with Φ_{ST} values ranging between 0.26 and 0.59 (Table 1). Population structure was highly significant for all genes as well as the majority of pairwise comparisons between *D. mojavenensis* populations (Additional file 1: Table S4).

Evidence of positive selection

All genes showed significant signatures of positive selection in at least one of the gene-wide (Table 2), codon-based and polymorphism-based (Table 3) tests performed. *Codeml* consistently showed all genes evolving under positive selection, ie. LRT favored the M8 selection model (Table 2). The

Table 1 Descriptive parameters for genetic diversity, neutrality tests and divergence. Results are shown for each gene in *D. mojavensis* and *D. arizonae*. In the case of *D. mojavensis*, genetic diversity and divergence values correspond to the average across the four populations (See [Methods](#) for abbreviations)

Gene/ species	Gr63a		Or67c		Or83c1		Or83c2	
	<i>D. moj</i>	<i>D. ari</i>	<i>D. moj</i>	<i>D. ari</i>	<i>D. moj</i>	<i>D. ari</i>	<i>D. moj</i>	<i>D. ari</i>
<i>N_s</i>	1479	1479	1215	1215	1182	1182	1170	1170
<i>N</i>	8.5	8	12.5	12	12.2	12	13.7	12
<i>h</i>	6.2	8	9	12	8.5	12	8	12
<i>H_d</i>	0.929	1.00	0.866	1.00	0.741	1.00	0.831	1.00
<i>S</i>	18.2	12	20.5	27	20.7	35	12.2	20
π	0.0049	0.0023	0.0052	0.0048	0.0049	0.0073	0.0029	0.0061
π_s	0.0180	0.0069	0.0192	0.0159	0.0121	0.0207	0.0094	0.0237
π_{ns}	0.0007	0.0008	0.0010	0.0014	0.0026	0.0031	0.0010	0.0009
θ_W	0.0046	0.0031	0.0055	0.0074	0.0059	0.0098	0.0033	0.0057
θ_{W_s}	0.0168	0.0107	0.0207	0.0259	0.0152	0.0257	0.0116	0.0210
$\theta_{W_{ns}}$	0.0006	0.0007	0.0010	0.0018	0.0031	0.0048	0.0010	0.0011
<i>D_T</i>	0.37	-1.38	-0.41	-1.59	-0.27	-1.15	-0.04	0.36
<i>D</i>	0.43	-1.93	-0.01	-2.60	-0.65	-1.79	-0.37	0.37
<i>F</i>	0.50	-2.21	-0.15	-2.77	-0.70	-1.96	-0.35	0.46
<i>F_S^a</i>	(+)	-5.21	(-)	-7.41	(-)	-5.52	(-)	-6.35
Φ_{ST}	0.26	na	0.35	na	0.34	na	0.59	na
<i>K_a/K_s</i>	0.036	0.038	0.059	0.066	0.38	0.30	0.27	0.36

Significant and marginally significant values after *FDR* correction are in bold

^aIn the case of *D. mojavensis* the significance (negative or positive) is indicated when at least one population was significant following Additional file 1: Table S3

D. moj: *D. mojavensis*

D. ari: *D. arizonae*

π_s : Nucleotide diversity for synonymous sites

π_{ns} : Nucleotide diversity for nonsynonymous sites

θ_W : Total theta Watterson

θ_{W_s} : Theta Watterson for synonymous sites

$\theta_{W_{ns}}$: Theta Watterson for nonsynonymous sites

Table 2 Gene-wide tests of positive selection and McDonald-Kreitman test. A summary of *P*-values obtained for selection tests is shown for each gene. Results for McDonald-Kreitman test are shown for the whole gene as well as for each domain (Cytoplasmic, transmembrane or extracellular) (See [Methods](#) for abbreviations)

Gene	<i>Codeml_{MB}</i>	<i>PARRIS</i>	<i>BUSTED</i>	<i>BSR</i>	<i>aBSREL</i>	<i>MK_{Total}</i>	<i>MK_{Cyto}</i>	<i>MK_{Trans}</i>	<i>MK_{Extr}</i>
<i>Gr63a</i>	< 0.001	0.131	1.000	> 0.05	> 0.05	0.502	1.000	na	0.350
<i>Or67c</i>	< 0.001	0.000	0.629	> 0.05	> 0.05	0.015^b	0.084	1.000	na
<i>Or83c1</i>	< 0.001	0.280	0.996	> 0.05	> 0.05	0.885	1.000	0.662	1.000
<i>Or83c2</i>	< 0.001	0.017	0.053	0.014^a	0.040	0.001^c	0.011	0.661	0.109

Significant and marginally significant values after *FDR* correction are in bold

^a This test showed positive selection specific to *D. mojavensis* lineage

^b This test was also significant for pairwise comparisons using *D. mojavensis* populations (*SON* and *BAJ*) and *D. arizonae* but not with ancestral *D. navojoa* outgroup (Additional file 1: Table S5)

^c This test was also significant for pairwise comparisons using *D. mojavensis* populations (*SON*, *BAJ* and *CAT*) and *D. arizonae* as well as ancestral *D. navojoa* outgroup (Additional file 1: Table S5)

MK_{Total}: McDonald-Kreitman test for whole gene

MK_{Cyto}: McDonald-Kreitman test for cytoplasmic domain

MK_{Trans}: McDonald-Kreitman test for transmembrane domain

MK_{Extr}: McDonald-Kreitman test for extracellular domain

Table 3 Codon-based tests of positive selection. A summary of *P*-values and posterior probabilities obtained for selection tests are shown for those codons where at least one test resulted with significant value after FDR correction (global $\alpha = 0.05$). The codon location in domains (*Cyto*, *trans* or *extr*) is indicated for each codon under positive selection. Biochemical properties significantly associated with amino acid sites under selection are indicated for *PRIME* test (See [Methods](#) for abbreviations)

Gene/Codon	Domain	codeml _{BEB}	SLAC	FEL	IFEL	REL	MEME	FUBAR	PRIME
<i>Gr63a</i>									
73	<i>Cyto</i>	0.571	0.471	0.369	> 0.050	0.980	0.227	0.732	> 0.050
93	<i>Cyto</i>	< 0.500	0.748	0.642	> 0.050	0.955	0.487	0.650	> 0.050
482	<i>Extr</i>	0.977	0.189	0.184	> 0.050	0.990	0.078	0.941	V
485	<i>Extr</i>	0.906	0.428	0.434	> 0.050	0.986	0.139	0.906	H
<i>Or67c</i>									
100	<i>Trans</i>	< 0.500	0.132	0.092	> 0.050	0.933	0.141	0.648	> 0.050
116	<i>Cyto</i>	< 0.500	0.982	0.815	> 0.050	0.807	0.583	0.401	> 0.050
125	<i>Cyto</i>	< 0.500	0.752	0.529	> 0.050	0.866	0.449	0.467	> 0.050
236	<i>Cyto</i>	< 0.500	0.199	0.348	> 0.050	0.901	0.023	0.653	S
253	<i>Cyto</i>	0.998	0.237	0.029	> 0.050	0.976	0.058	0.906	C, VR
274	<i>Cyto</i>	< 0.500	0.434	0.424	> 0.050	0.885	0.388	0.623	> 0.050
335	<i>Cyto</i>	< 0.500	0.850	0.965	> 0.050	0.778	0.664	0.451	> 0.050
391	<i>Trans</i>	1.000	0.025	0.000	> 0.050	1.000	0.000	1.000	S
404	<i>Extr</i>	1.000	0.434	0.268	> 0.050	0.911	0.271	0.621	> 0.050
<i>Or83c1</i>									
9	<i>Cyto</i>	0.764	0.411	0.155	> 0.050	0.975	0.133	0.712	> 0.050
99	<i>Cyto</i>	1.000	0.075	0.004	0.026	1.000	0.001	0.999	> 0.050
230	<i>Cyto</i>	1.000	0.088	0.145	> 0.050	0.999	0.121	0.973	P, R, I
236	<i>Cyto</i>	1.000	0.136	0.105	> 0.050	1.000	0.131	0.972	H
362	<i>Trans</i>	0.536	0.451	0.136	> 0.050	0.976	0.151	0.727	> 0.050
386	<i>Extr</i>	0.999	0.330	0.051	> 0.050	0.996	0.033	0.971	H
387	<i>Extr</i>	1.000	0.089	0.024	> 0.050	1.000	0.036	0.992	> 0.050
391	<i>Extr</i>	0.965	0.386	0.070	> 0.050	0.986	0.066	0.926	> 0.050
<i>Or83c2</i>									
7	<i>Cyto</i>	0.712	0.459	0.203	> 0.050	0.972	0.229	0.828	> 0.050
21	<i>Cyto</i>	< 0.500	0.459	0.285	> 0.050	0.959	0.064	0.794	> 0.050
43	<i>Trans</i>	1.000	0.007	0.003	0.018	1.000	0.005	0.999	C
92	<i>Cyto</i>	0.773	0.572	0.518	> 0.050	0.855	0.000	0.644	C, RP
178	<i>Extr</i>	< 0.500	0.478	0.271	> 0.050	0.720	0.029	0.701	> 0.050
187	<i>Trans</i>	0.957	0.228	0.064	0.019	0.998	0.085	0.947	> 0.050
238	<i>Cyto</i>	1.000	0.496	0.205	> 0.050	0.986	0.047	0.930	> 0.050
284	<i>Trans</i>	0.845	0.176	0.063	> 0.050	0.992	0.07	0.984	> 0.050

Significant and marginally significant values after FDR correction are in bold

genes *Or67c* and *Or83c2* were also significant in one (*PAR-RIS*) and three (*PARRIS*, *BSR*, *aBSREL*) of the additional gene-wide selection tests performed respectively, as well as the McDonald-Kreitman test (Table 2). The BSR test identified selection in *Or83c2* for specific branches belonging to the *SON* and *CAT* populations of *D. mojavensis* as well as branches belonging to *D. arizonae* lineage (Table 2). McDonald-Kreitman test was also performed for pairwise

comparisons between host populations of *D. mojavensis* and *D. arizonae*, showing signatures of positive selection for *SON* and *BAJ* in *Or67c* and *SON*, *BAJ* and *CAT* in *Or83c2* (Table 2 and Additional file 1: Table S5). Only the gene *Or83c2* showed signatures of positive selection in both the *D. mojavensis* and *D. arizonae* lineage as determined by the McDonald-Kreitman tests and polarizing the fixed sites utilizing the outgroup species *D. navojoa* (Table 2 and

Additional file 1: Table S5). Furthermore, all genes showed evidence of positive selection at individual sites in at least three of the codon-based selection tests performed and all tests provided significant values in at least two genes (Table 3). A combined analysis of these tests led us to identify four sites evolving under positive selection for the gene *Gr63a* (using the *codeml_{BEB}*, *REL*, *FUBAR* and *PRIME* tests); nine sites in *Or67c* (using the *codeml_{BEB}*, *SLAC*, *FEL*, *REL*, *MEME*, *FUBAR* and *PRIME* tests); eight sites in *Or83c1* from (using the *codeml_{BEB}*, *FEL*, *IFEL*, *REL*, *MEME*, *FUBAR* and *PRIME* tests); and eight sites positively selected in the gene *Or83c2* (using the *codeml_{BEB}*, *SLAC*, *FEL*, *IFEL*, *REL*, *MEME*, *FUBAR* and *PRIME* tests) (Table 3, Fig. 2). Though we identified several sites evolving under positive selection, we also investigated whether these changes led to changes in amino acid characteristics that might affect protein structure and hence function. Results of the *PRIME* analysis showed that 38% of the candidate sites across the four loci indicated possible substantial changes to protein structure (Table 3).

Since the performance of selection tests based on phylogenetic comparisons could be affected by recombination

increasing the number of false positively selected sites [46–48], we tested for the presence of recombination breakpoints and found evidence for such events in *Or83c1* and *Or83c2*. Therefore, we also performed the test of selection on these two loci within each linkage block. After controlling for recombination, two sites in *Or83c1* and four sites in *Or83c2* continued to indicate a pattern of positive selection. The remaining sites did not pass the threshold of the Bayes factors (BF > 100) in the *REL* test, but posterior probabilities remained over 0.8. Suggesting that, although the power of the test is decreased since blocks are analyzed instead of genes, a similar selective history affected the pattern of variation at these sites as well. For this reason and because the phylogenetic relationships among the three species did not change between partitions, suggesting that the assumptions of selection tests are not violated, only the results for the gene as a whole without partitions were further analyzed.

Signatures of selection among gene domains

Given the pattern observed at these loci suggesting positive selection between *D. mojavenensis* and *D. arizonae*, we

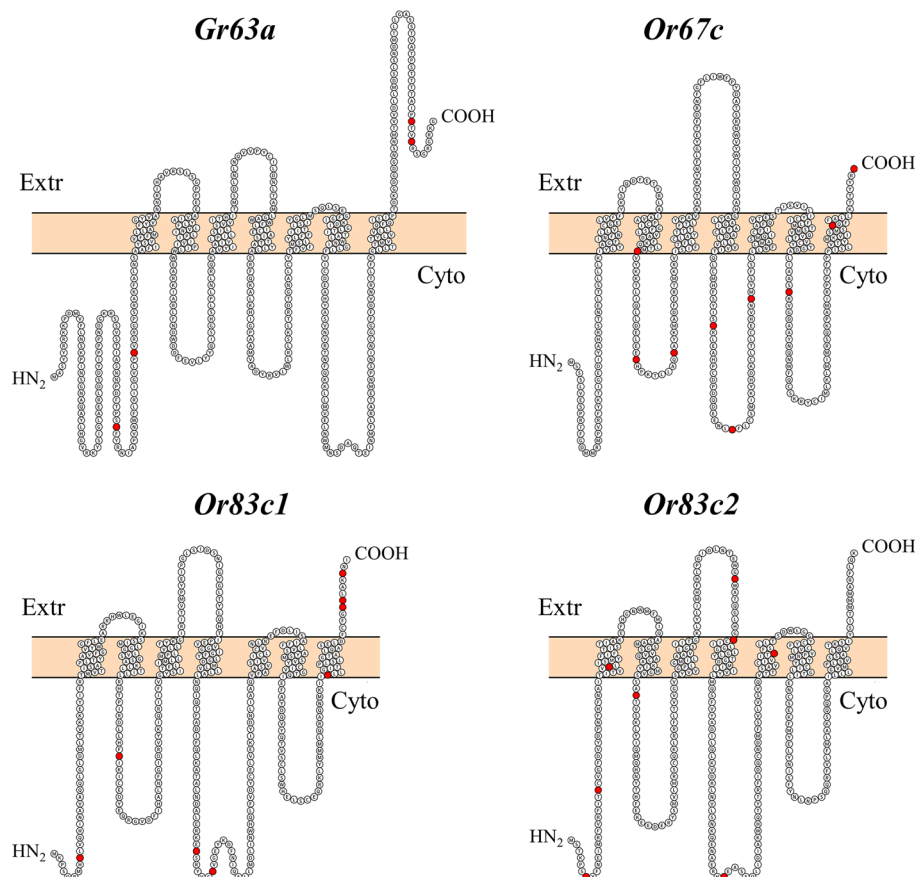


Fig. 2 Predicted topology of chemoreceptor proteins showing seven transmembrane domains and a cytoplasmic NH₂-tail. Positively selected sites are highlighted in red. Topology was predicted by the TMHMM software

further tested whether selection had occurred within specific domains of these transmembrane proteins. Approximately 55% of the sites under positive selection are located in the cytoplasmic domain while only 24 and 21% are in the extracellular and transmembrane domains, respectively (Fig. 2). We assessed the distribution of selected sites across the protein domains of all four loci through a GLM model (Table 4). We found no significant gene effect, but a significant domain effect (Table 4), with the cytoplasmic domain enriched for positively selected sites when compared with the transmembrane and extracellular domains according to multiple comparisons (Fig. 3). We additionally tested for such structural heterogeneity in the level of selection by running the McDonald-Kreitman independently for each domain of each gene. These tests showed significant evidence of positive selection in the cytoplasmic domain for *Or83c2* and marginally for *Or67c*, while no significant effects for the transmembrane and extracellular domains (Table 2). The proportion of sites under positive selection was just marginally significant when accounting for the domain size in the GLM model for *Or83c2* and *Or67c* (Table 4), which points to selection also being influenced by domain size. This is expected since, larger domains have a higher number of sites available for selection to act on.

Discussion

It has been suggested that specialist species not only lose a subset of sensory pathway genes during host specialization, but some of those that are retained undergo periods of positive selection [14, 15]. Here we investigated the role of positive selection in the molecular evolution of four chemoreceptor genes (one *Gr* and three *Or*) during the evolution of *D. mojavensis*, a species that has experienced recent bouts of specialization resulting in a set of four cactus host populations, and its generalist sister *D. arizonae*. We used multiple approaches to test for positive selection including codon-based, gene-wide, polymorphism and functional analyses. All genes investigated showed evidence of positive selection at codon and gene-wide approaches while

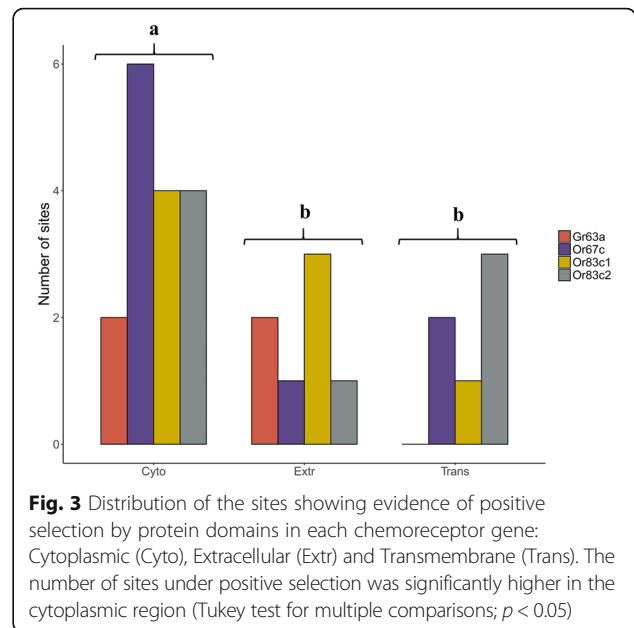


Fig. 3 Distribution of the sites showing evidence of positive selection by protein domains in each chemoreceptor gene: Cytoplasmic (Cyto), Extracellular (Extr) and Transmembrane (Trans). The number of sites under positive selection was significantly higher in the cytoplasmic region (Tukey test for multiple comparisons; $p < 0.05$)

two loci indicated positive selection when polymorphism information was analyzed. Although most sites at these genes are likely to be evolving under purifying selection, our approach suggests evidence of positive selection shaping the evolution of several amino acid sites involved in molecular and physiological functions, which allowed us to investigate the distribution of candidate sites under positive selection across the domains of these chemosensory receptor proteins.

The level of genetic diversity was higher in *D. arizonae* when compared with the average value across the *D. mojavensis* populations. This pattern is in agreement with previous studies of other nuclear genes in these species [45], which is likely a consequence of a higher overall effective population size in *D. arizonae* [49, 50]. This pattern of variation is however dissimilar to what has been observed from mitochondrial loci [44], which suggests that recent evolutionary forces are shaping nuclear and mitochondrial genes differently [45]. Likewise, the Fu's F_S was significantly negative at most loci, which is evidence of excess number of alleles [51]. Since we

Table 4 ANOVA of the GLM analysis for the number of sites showing evidence of positive selection and the proportion of sites under positive selection (controlling for the number of amino acids within a domain) (See Methods for abbreviations)

Effect	Df	Deviance Resid.	Df Resid.	Dev.	P
<i>Number of sites under selection</i>					
Gene	3	8.8x10 ⁻⁵	8	0.0005	0.4366
Domain	2	2.7x10 ⁻⁴	6	0.0002	0.0152
<i>Proportion of sites under selection</i>					
Gene	3	0.0102	8	0.0232	0.1913
Domain	2	0.0103	6	0.0129	0.0925

Significant values are in bold

controlled for the detected population structure performing tests per population, this pattern is more likely indicating that both species are undergoing population expansions. Although we did not find such evidence from all neutrality tests, most values were consistently negative and the Fu's F_S statistic is particularly sensitive to recent changes in population sizes [51]. This expansion in both species has been previously suggested to be associated with a shared biogeographical history in some of the *D. mojavensis* populations (*SON* and *BAJ*) and *D. arizonae*. As suggested by Machado et al. [45], these populations have experienced expansion and contraction events due to Pleistocene glaciation cycles that affected a great part of the biota [52–54]. This is consistent with our results per population since the sampled population of *D. arizonae* is from the Sonoran Desert and only the populations from Sonora and Baja California showed signatures of expansion in *D. mojavensis*. The sole significantly positive value indicates a deficiency of alleles in the Mojave Desert population of *D. mojavensis* at a single locus (*Gr63a*) [51]. Given that demographic changes would be expected to affect the majority of loci, this particular case could possibly be a result of balancing selection. Further evidence is needed to obtain a more complete understanding of what is shaping variation at this locus in this population.

Most of the evidence for the role of positive selection in *Or* and *Gr* genes during host specialization comes from the genome-wide analysis in *Drosophila* species, particularly the case of *D. sechellia* and *D. erecta* [14, 15, 31]. McBride and Arguello [15] found higher *Ka/Ks* ratios in chemoreceptors of these specialist species when compared with their closest generalists, although such ratios could be alternatively explained by relaxed purifying and not necessarily positive selection. Such comparisons have been based on genome-wide studies, where the lack of polymorphic variation and specialization events in short time scales have made the efforts to disentangle these two hypotheses difficult. Gardiner et al. [32] found 20 genes with evidence of functional divergence in a genome-wide study of 12 *Drosophila* genomes, but only six were considered to be under positive selection. This suggests that the number of genes positively selected in these families is often low in insects, however, as suggested by Gardiner et al. [32], the low power of the phylogenetic-based tests should caution one against making broad generalizations. Despite the low *Ka/Ks* values of these genes, suggesting that most sites are evolving under purifying selection, we were able to identify a number of candidate codons under selection, some of which might contribute to significant effects on protein structure and function. These changes may be involved in detection of new odor molecules as well as protein-protein interactions during signal transduction. These findings suggest that the low number of sites under

positive selection previously reported in sensory genes (as recognized by Gardiner et al. [32]) may be at least in part due to the low power of the particular selection tests previously utilized and the absence population level polymorphism data.

According to the specialization hypothesis [14], beneficial mutation associated with amino acid changes that improve the recognition capacity of smell and taste receptors should be favored in specialists, resulting in a narrowly tuned receptor [11]. This proposed mechanism of evolution at chemosensory loci might also have shaped the variation of chemosensory loci during the specialization process of the four cactus host populations of *D. mojavensis* [33, 34, 55]. The necroses of each cactus differ in their chemical composition as well as the yeast and bacterial communities on which *D. mojavensis* feeds [56–58]. It has been shown that these populations not only differ in terms of their performance while utilizing different necrotic cacti [59], but also exhibit different behavioral and electrophysiological responses of olfactory organs, which suggests that their peripheral nervous system has been shaped by local ecological differences [59, 60].

Although we provide evidence supporting the hypothesis of positive selection shaping the evolution of amino acid sites involved in molecular functions of sensory genes during the host specialization in *D. mojavensis*, not all the analyses performed allow for the possibility to detect lineage specific changes. Nevertheless, for those analyses addressing specific lineages, the evidence supports the role of positive selection during the evolution of *D. mojavensis*. This was evident from the *BSR* and McDonald-Kreitman tests in *Or83c2*. While *Or67c* was not significant for the McDonald-Kreitman test when compared with *D. navojoa*, both species showed similar fixed differences. In addition, the changes specific to the *D. mojavensis* lineage are supported by strong genetic differentiation among *D. mojavensis* populations as estimated by Φ_{ST} which are often larger than what has been previously reported for neutral *SSR* markers [61], suggesting divergent selection may be increasing Φ_{ST} values in coding sequences among *D. mojavensis* populations.

We also found selective changes exclusive to the *D. arizonae* lineage. This was evident from the *Ka/Ks* value using *D. navojoa* as an outgroup, which was often larger in *D. arizonae* as well as the McDonald-Kreitman tests for *Or83c2* and *Or67c* (Table 2). These findings in *D. arizonae* pose an interesting question about the role of positive selection in the evolution of a generalist that has diverged from a specialized species. Cytological and molecular evidence suggests that the *D. mojavensis/D. arizonae* lineage diverged from the lineage of the *Opuntia* specialist *D. navojoa* approximately 4 Mya (reviewed in [41]). Hence *D. arizonae* has acquired the ability to use

columnar cacti, as well as *Opuntia*, among other hosts. The potential role of positive selection during the acquisition of new hosts in the evolutionary transition between specialist to generalist species may favor mutations associated with amino acid changes involved in the chemical recognition of a wider repertoire of odors, a broadly tuned receptor [11]. Alternatively, the changes observed could be associated with the modulation of odorant binding signals during the early stages of the signal transduction cascade [62, 63].

The majority of the sites under positive selection in the *Or* and *Gr* genes map to the cytoplasmic domain of these transmembrane proteins (Tables 2, 3 and 4). In general, this pattern was unexpected since this domain interacts with secondary messengers involved in signal transduction and was therefore expected to be conserved. Interestingly, this pattern seems to be more common than previously thought, since there has been a previous report of the cytoplasmic domain of chemosensory receptors evolving rapidly [32]. We only analyzed one *Gr*, which prevents us from including gene family as a factor in the *GLM* model. Despite such limitations of the model, the sole *Gr* gene analyzed here showed clear differences in the distribution of candidate sites when compared with *Or* genes. For example, *Gr63a* had the lowest number of candidate sites and was the only gene with the same number of candidate sites in extracellular and cytoplasmic regions, with no evidence of positive selection in the transmembrane domain. Such differences still need to be confirmed by comparing more *Gr* genes, but again, this result agrees with those patterns reported elsewhere [32]. As suggested by Gardiner et al. [32], this probably reflects the functional and molecular differences between *Or* and *Gr* receptors. *Or* genes are associated with volatile recognition while *Gr* recognize soluble ligands, pheromones and particularly for *Gr63a*, levels of CO₂ [11, 20, 21, 38, 64]. In addition, the olfactory receptors requires the interaction with the universal receptor *Or83b* [11, 20, 21], while *Gr63a* requires interaction with *Gr21a* for its functionality [39, 65]. Such functional and molecular differences are likely to lead to distinct regions of constraint as reflected by the observed differences in the distribution of positively selected sites.

The majority of positively selected sites mapped to IL1 and IL2 loops, which would be expected since these loops are not involved in the dimerization with *Or83b*, the universal co-receptor [32]. *Or67c* was the only gene showing a candidate site in the IL3 domain. This loop interacts with IL3 of *Or83b* during dimerization [64], which make it likely to be under constraint. We also found candidate sites in the COOH-tail for *Or83c1* and *Or67c*, which contrasts with the results of Gardiner et al. [32] since they found no candidate sites in this region for the ten *Or* genes under positive selection among the 12 *Drosophila* species studied. This terminal region is also important for

coupling with *Or83b*, being conserved even in highly divergent genes [64].

The higher number of positively selected sites in the cytoplasmic domain suggest that selection has been acting primarily on the transduction of signal captured by chemoreceptors and not only its detection. Nevertheless, after controlling for the number of amino acids per domain, the *GLM* analysis was just marginally significant. Additionally, *Or83c1* and *Or67c* also show divergent selection in the extracellular region (COOH-tail) essential for signal detection [64], suggesting that both signal detection and transduction have been shaped by positive selection during the molecular evolution of these chemoreceptors.

Conclusions

Local ecological adaptation can shape the evolutionary trajectories of myriads of traits and their respective underlying genetic architecture. Chemosensory pathways, especially those involved in host preference in insects, can be shaped by the selective forces associated with changes in host utilization. We have provided evidence supporting the hypothesis of host specialization as an evolutionary scenario that favors positive selection at specific amino acid sites in genes of sensory pathways. Given that selection has been particularly biased towards cytoplasmic domains at amino acid sites involved in structural and physiological functions, these selective forces appear to be largely shaping the post-odorant binding role of these chemosensory receptors, i.e. signal transduction. Our approach highlights the advantage of using a set of recently diverged species and populations locally adapted to distinct ecological conditions in the understanding of the mechanisms of chemosensory evolution. Furthermore, the incorporation of population level polymorphism data which allowed multiple and powerful approaches for detecting positive selection in an ecological context proved to be highly effective.

Methods

Samples

Isofemale lines from all four *D. mojavensis* cactus host populations were utilized in this study: Baja California (*BAJ*), Sonora Desert (*SON*), Mojave Desert (*MOJ*) and Santa Catalina Island (*CAT*) (Fig. 1). Additionally, sequences from a *D. arizonae* population from the Sonora Desert and one line of *D. navojoa* (*Drosophila* Species Stock Center 15081–1374.11), a cactophilic (*Opuntia* spp) species, from Jalisco, Mexico was utilized for comparisons requiring an outgroup. For both *D. mojavensis* and *D. arizonae* approximately 7–15 independent isofemale lines per population/species were sampled. These isofemale lines have been maintained in the laboratory for 60–120 generations and are highly inbred (Fig. 1). All stocks were

maintained in 8-dram vials with banana-molasses media [66] in a 25°C incubator on a 14:10 h light:dark cycle.

Molecular procedures and alignments

DNA samples for each species were amplified via PCR using Apex Taq DNA Polymerase (Genesee Scientific) following manufacturer's recommendations in a 25 µl final volume, involving 1 µl of DNA and 0.2 µM of each primer. Amplifications were performed using a PCR program of 35 cycles involving 94 °C for 30 s followed by the corresponding annealing temperature for each gene for 30 s (Additional file 1: Table S1), 72 °C for 1 min, and then a final extension at 72 °C for 10 min. PCR products were purified using QIAquick PCR Purification columns (Qiagen) following manufacturer's recommended procedures and sanger sequenced in both directions (Operon Eurofins). Sequences were visually inspected for quality using Geneious [67] and aligned based on their amino acid sequences using ClustalW and manual verification. Haplotype phase was inferred for each sample using the software PHASE [68–70].

Genetic diversity, neutrality tests and divergence

The genetic diversity at nucleotide and haplotype levels was estimated for each species through a set of descriptive parameters using the software DNAsp v5 [71], including the average number of sequences (N), total number of sites (N_S), number of variable sites (S), number of haplotypes (h), haplotype diversity (H_d), nucleotide diversity (π) and the average number of nucleotide differences (θ_w). In order to infer any deviation from neutral expectations at the population level, genetic structure or drastic changes in population sizes (recent bottlenecks or population expansion), a set of neutrality tests were performed, including Tajima's D_T [72], D and F tests [73] and Fu's F_S statistic [74] using the software DNAsp. All genes were assessed for recombination events using the *GARD* test, a genetic algorithm for recombination detection [75, 76] implemented in the package HyPhy v2 [77].

Divergence between *D. mojavensis* populations and *D. arizonae* was estimated by the number of synonymous (K_s) and nonsynonymous (K_a) substitutions ratio (K_a/K_s) with respect to the *D. navojoa* outgroup using the software DNAsp v5 [71]. The genetic structure among *D. mojavensis* populations was estimated through pairwise Φ_{ST} , an analogue version of the Wright's fixation index F_{ST} [78, 79] which is estimated by an Analysis of Molecular Variance (AMOVA), taking into account information on the genetic distances among haplotypes as well as their frequencies following Excoffier et al. [80]. Significance was based on 10,000 permutations in the software Arlequin version 3.5.2 [81].

Phylogenetic inference

Because several of the selection tests we performed require a phylogenetic tree of aligned sequences, each gene was individually analyzed using Bayesian inference in MrBayes [82] after testing for the best evolutionary model inferred in the software jModelTest v0.1.1 [83]. Two independent runs were used from different starting points by a Metropolis-coupled Markov Chain Monte Carlo analysis with four chains, one cold and three incrementally heated (heating parameter = 0.1) for 10 million generations, sampling every 5,000th tree. Both runs were checked for convergence of the Markov chains with the standard deviation of split frequencies being less than 0.001. Parameter estimates were then analyzed in Tracer [84] to ensure that these had reached stable values with adequate mixing and ESSs above 200.

Assessing positive selection

Signatures of positive selection were assessed based on the nonsynonymous/synonymous substitution ratio ($d_N/d_S = \omega$). When there is no selection, both classes of substitutions are expected to become fixed with the same probability ($\omega = 1$). On the other hand, selection can increase fixation probabilities for nonsynonymous mutations because of selective advantages (Positive selection, $\omega > 1$), or decrease these probabilities due to selective constraints (Purifying / negative selection, $\omega < 1$). A number of methods have been developed for detecting positive selection based on ω ratios at different levels, whole alignment, phylogenetic branches, codons and combinations of those. Therefore, we tested for selection using multiple tests and four approaches: *i*) gene-wide; *ii*) polymorphic-based; *iii*) codon-based; *iv*) functional implications of candidate sites.

i) Gene-wide selection tests were performed in order to detect general signatures of selection at the gene level in the alignments or trees among sequences of the three species, without making any assumption about foreground branches. First, the software *codeml*, implemented in the package *PAML* v4 [85] was used to compare a null model (M7) in which ω is assumed to be beta-distributed among sites and a selection model (M8), where codons are allowed to have an extra category of positively selected sites $\omega > 1$. The significance of these comparisons was assessed using a likelihood-ratio test [86]. We also used the package HyPhy v2 [77] in order to perform a set of gene-wide tests for positive selection to complement and provide a comparison with the LRTs in *codeml*. Thus, we used the *PARRIS* test (a partitioning approach for robust inference of selection), which makes maximum likelihood inference of positive selection robust to the presence of recombination. The *BUSTED* test (Branch-site Unrestricted Statistical Test for Episodic Diversification) [87] was used to specifically assesses whether a gene has experienced positive selection on at least one site and one branch given

a phylogeny. The *BSR* (Branch-Site Random effects likelihood test) [88] was used to test for episodic diversifying selection detecting lineages with $\omega > 1$. The *aBSREL* test (adaptive Branch-Site Random Effects Likelihood), the improved version of the commonly-used “branch-site” models, was used to test if positive selection had occurred on a proportion of branches. *aBSREL* models both site-level and branch-level ω heterogeneity, testing whether a proportion of sites have evolved under positive selection. An FDR correction was performed to account for multiple comparisons across genes for each test.

ii) A polymorphism-based approach was used to test for positive selection comparing the ω ratios within species with those between species using the McDonald-Kreitman test [89]. This test takes advantage of the intraspecific polymorphic variation, since the substitution ω ratios within species are expected to equal those ratios between species under a neutral scenario but differ under selective scenarios. An FDR correction was performed to account for multiple comparisons across genes for each test.

iii) We performed a set of codon-based selection tests for identifying specific sites within genes showing signatures of selection. Bayes Empirical Bayes (*BEB*) [90] was used to estimate the posterior probability of sites under positive selection following the M8 model from *codeml* results. Again, these results were compared with codon-based selection tests available in the package HyPhy, including *SLAC* (Single-Likelihood Ancestor Counting method) [91], *FEL* (Fixed Effects Likelihood method), *IFEL* (Internal Fixed Effects Likelihood), *REL* (Random Effects Likelihood method), *MEME* (Mixed Effects Model of Evolution) [92] and *FUBAR* tests (Fast Unconstrained Bayesian Approximation for inferring selection) [93].

iv) Finally, we investigated whether changes at the amino acid level were associated with altered biochemical properties using the PRIME test (PRoperty Informed Model of Evolution) following Conant-Stadler [94] and Atchley categories [95]. The five predefined amino acid properties proposed by Conant et al. [94] are: chemical composition of the side chain (C), residue polarity (RP), volume of the residue side chain (VR), isoelectric point of the side chain (IC), and hydrophathy (H). The five composite properties proposed by Atchley et al. [95] are: polarity index (P), secondary structure factor (S), volume (V), refractivity (R), and isoelectric point (I). Although the method performs multiple tests on each site, *p*-values are reported after Bonferroni correction to control for the number of false positives.

Sites with evidence of positive selection were mapped onto the receptor’s topology predicted with the software *TMHMM* Server v2 (<http://www.cbs.dtu.dk/services/TMHMM/>) and then diagrams of the 2D receptors structure were created using *PROTTER* [96].

Signatures of selection among gene domains

We used two approaches in order to investigate whether selection is acting on specific domains in transmembrane receptor genes (cytoplasmic, transmembrane or extracellular). First, the number of sites under positive selection was compared among genes and domains through a Generalized linear model using the R package *GLM* after data normalization using the logarithmic transformation. This analysis was done twice, for raw counts of sites under selection and after accounting for the size of each domain in order to test for the potential effect of domain size on the number of sites showing signatures of selection, since longer domains have higher chance to show sites under positive selection. Additionally, the McDonald-Kreitman test was performed for each domain independently.

Additional file

Additional file 1: Table S1. Primer sequences (Forward and Reverse) for each gene and species. **Table S2.** Descriptive parameters for haplotype, nucleotide diversity and neutrality tests. Results are shown for each gene in each *D. mojavensis* population and *D. arizonae* (See [Methods](#) for abbreviations). **Table S3.** Neutrality tests for each gene in each *D. mojavensis* population and *D. arizonae* (See [Methods](#) for abbreviations). **Table S4.** Divergence between *D. mojavensis* and *D. arizonae* species and populations within *D. mojavensis*. *Ka/Ks* ratios and genetic structure estimated by Φ_{ST} are given for pairwise comparisons between species and populations for each gene (See [Methods](#) for abbreviations). **Table S5.** McDonald-Kreitman test for each *D. mojavensis* and *D. arizonae*. A summary of *P*-values is shown for each gene using *D. arizonae* and *D. navojoa* as an outgroup (See [Methods](#) for abbreviations). (DOCX 50 kb)

Abbreviations

aBSREL: adaptive Branch-Site random effects likelihood; *BAJ*: Baja California; *BSR*: Branch-Site random effects likelihood test; *BUSTED*: Branch-site unrestricted statistical test for episodic diversification; C: Chemical composition of the side chain; *CAT*: Santa Catalina Island; *Codeml_{M8}*: Codeml comparison between M7 and M8 models; *Cyto*: Cytoplasmic; *D*: Fu and Li’s *D* test; *D. ari*: *D. arizonae*; *D. moj*: *D. mojavensis*; *D. nav*: *D. navojoa*; *D_T*: Tajima’s *D* test; *Extr*: Extracellular; *F*: Fu and Li’s *F* test; *F₅*: Fu’s *F₅* statistic; *H*: Hydrophathy; *h*: Number of haplotypes; *H_d*: Haplotype diversity; *I*: Isoelectric point; *IC*: Isoelectric point of the side chain; *Ka/Ks*: Nonsynonymous (*Ka*)/synonymous (*Ks*) ratio; *MK_{Cyto}*: McDonald-Kreitman test for cytoplasmic domain; *MK_{Extr}*: McDonald-Kreitman test for extracellular domain; *MK_{Total}*: McDonald-Kreitman test for whole gene; *MK_{Trans}*: McDonald-Kreitman test for transmembrane domain; *MOJ*: Mojave Desert; *N*: Average number of sequences; *N_s*: Total number of sites; *P*: Polarity index; *PARRIS*: Partitioning approach for robust inference of selection; *PRIME*: PRoperty Informed Model of Evolution; *R*: Refractivity; *RP*: Residue polarity; *S*: Number of variable sites; *S*: Secondary structure factor; *SON*: Sonora Desert; *Trans*: Transmembrane; *V*: Volume; *VR*: Volume of the residue side chain; $\theta_{W, ns}$: Theta Watterson for nonsynonymous sites; $\theta_{W, s}$: Theta Watterson for synonymous sites; θ_W : Total theta Watterson; π : Total nucleotide diversity; π_{ns} : Nucleotide diversity for nonsynonymous sites; π_s : Nucleotide diversity for synonymous sites; Φ_{ST} : Genetic structure index

Acknowledgements

We would like to thank Laurel Brandsmeier for the laboratory work she performed on this project. We also would like to thank two anonymous reviewers whose comments improved this manuscript.

Funding

This project was initially supported by a National Science Foundation grant DEB-1219387 to LMM and later by a National Science Foundation grant IOS-1557697 to LMM.

Availability of data and materials

All nucleotide data are available for each gene in GenBank (Gr63a: MH750445 - MH750487, Or67c: MH750488 - MH750550, Or83c: MH750551 - MH750612 and Or83c2: MH750613 - MH750680).

Authors' contributions

The project was conceived and designed by LMM. CWA performed the majority of the laboratory work. FD performed most of the statistical analyses. All authors participated in the data analysis and the writing of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This section is not applicable to the present study.

Consent for publication

This section is not applicable to the present study.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Entomology, University of Arizona, Tucson, AZ 85721, USA. ²Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA. ³BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA.

Received: 20 April 2018 Accepted: 20 August 2018

Published online: 20 September 2018

References

- Bernays EA, Chapman R. In: Bernays EA, Chapman RF, editors. Host-plant selection by phytophagous insects. New York: Chapman & Hall; 1994.
- Jaenike J. Host specialization in phytophagous insect. *Annu Rev Ecol Syst.* 1990;21:243–73.
- Bruce TJA, Wadhams LJ, Woodcock CM. Insect host location: a volatile situation. *Trends Plant Sci.* 2005;10:269–74.
- Matsubayashi KW, Ohshima I, Nosil P. Ecological speciation in phytophagous insects. *Entomol Exp Appl.* 2010;134:1–27.
- Hardy NB, Otto SP. Specialization and generalization in the diversification of phytophagous insects: tests of the musical chairs and oscillation hypotheses. *Proc R Soc B Biol Sci.* 2014;281:20132960.
- Agosta SJ. On ecological fitting, lant-insect associations, herbivore host shifts, and host plant selection. *Oikos.* 2006;114:556–65.
- Forbes AA, Devine SN, Hippee AC, Tvedte ES, Ward AKG, Widmayer HA, et al. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution (NY).* 2016;71:1126–37.
- Thompson JN. Evolutionary ecology of the relationship between oviposition preference and performance of offspring in phytophagous insects. *Entomol Exp Appl.* 1988;47:3–14.
- Vertacnik KL, Linnen CR. Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics. *Ann NY Acad Sci.* 2017;1389: 186–212.
- Dworkin I, Jones CD. Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. *Genetics.* 2009;181:721–36.
- Andersson MN, Lofstedt C, Newcomb RD. Insect olfaction and the evolution of receptor tuning. *Front Ecol Evol.* 2015;3:1–14.
- Bohbot JD, Pitts RJ. The narrowing olfactory landscape of insect odorant receptors. *Front Ecol Evol.* 2015;3:1–10.
- Matsuo T. Genes for host-plant selection in *Drosophila*. *J Neurogenet.* 2008; 22:195–210.
- McBride CS. Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc Natl Acad Sci U S A.* 2007;104: 4996–5001.
- McBride CS, Arguello JR. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics.* 2007;177:1395–416.
- Simon JC, D'alençon E, Guy E, Jacquin-Joly E, Jaquiéry J, Nouhau P, et al. Genomics of adaptation to host-plants in herbivorous insects. *Brief Funct Genomics.* 2015;14:413–23.
- Smadja CM, Canbäck B, Vitalis R, Gautier M, Ferrari J, Zhou JJ, et al. Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution (NY).* 2012;66: 2723–38.
- Amrein H, Thorne N. Gustatory perception and behavior in *Drosophila melanogaster*. *Curr Biol.* 2005;15:673–84.
- Grabe V, Sachse S. Fundamental principles of the olfactory code. *BioSystems.* 2018;164:94–101.
- Touhara K, Vosshall LB. Sensing odorants and pheromones with chemosensory receptors. *Annu Rev Physiol.* 2009;71:307–32.
- Vosshall LB, Stocker RF. Molecular architecture of smell and taste in *Drosophila*. *Annu Rev Neurosci.* 2007;30:505–33.
- Sánchez-Gracia A, Vieira FG, Rozas J. Molecular evolution of the major chemosensory gene families in insects. *Heredity (Edinb).* 2009;103:208–16.
- Sánchez-Gracia A, Vieira FG, Almeida FC, Rozas J. Comparative genomics of the major chemosensory gene families in arthropods. *Encyclopedia of Life Sciences (ELS).* Wiley, Ltd: Chichester. 2011. <https://doi.org/10.1002/9780470015902.a0022848>.
- Benton R. Multigene family evolution: perspectives from insect chemoreceptors. *Trends Ecol Evol.* 2015;30:590–600.
- Guo S, Kim J. Molecular Evolution of *Drosophila* Odorant Receptor Genes. *Mol Biol Evol.* 2007;24:1198–207.
- Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. Family size evolution in *Drosophila* chemosensory gene families: a comparative analysis with a critical appraisal of methods. *Genome Biol Evol.* 2014;6:1669–82.
- Nozawa M, Nei M. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc Natl Acad Sci U S A.* 2007;104:7122–7.
- Vieira FG, Rozas J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the arthropoda: Origin and evolutionary history of the chemosensory system. *Genome Biol Evol.* 2011;3: 476–90.
- Vieira FG, Sánchez-Gracia A, Rozas J. Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: Purifying selection and birth-and-death evolution. *Genome Biol.* 2007;8:R235.
- Consortium (Drosophila 12 Genomes consortium). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007;450:203–18.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. *Drosophila* chemoreceptor gene evolution: Selection, specialization and genome size. *Mol Ecol.* 2008;17:1648–57.
- Gardiner A, Butlin RK, Jordan WC, Ritchie MG. Sites of evolutionary divergence differ between olfactory and gustatory receptors of *Drosophila*. *Biol Lett.* 2009;5:244–7.
- Heed WB. Ecology and genetics of Sonoran Desert *Drosophila*. In: Brussard PF, editor. *Ecol. Genet. Interface.* New York: Springer-Verlag; 1978. p. 109–26.
- Ruiz A, Heed WB, Wasserman M. Evolution of the *mojavensis* cluster of cactophilic *Drosophila* with descriptions of two new species. *J Hered.* 1990;81:30–42.
- Matzkin LM, Markow TA. Transcriptional differentiation across the four subspecies of *Drosophila mojavensis*. In: Michalak P, editor. *Speciat. Nat. Process. Genet. Biodivers: Nova Science Publishers, Inc;* 2013. p. 119–35.
- Matzkin LM. Population transcriptomics of cactus host shifts in *Drosophila mojavensis*. *Mol Ecol.* 2012;21:2428–39.
- Matzkin LM, Watts TD, Bitler BG, Machado CA, Markow TA. Functional genomics of cactus host shifts in *Drosophila mojavensis*. *Mol Ecol.* 2006;15:4635–43.
- Mansourian S, Stensmyr MC. The chemical ecology of the fly. *Curr Opin Neurobiol.* 2015;34:95–102.
- Turner SL, Ray A. Modification of CO2 avoidance behaviour in *Drosophila* by inhibitory odorants. *Nature.* 2009;461:277–81.
- Ronderos DS, Lin C-C, Potter CJ, Smith DP. Farnesol-detecting olfactory neurons in *Drosophila*. *J Neurosci.* 2014;34:3959–68.
- Matzkin LM. Ecological genomics of host shifts in *Drosophila mojavensis*. In: Landry C, Aubin-Horth N, editors. *Adv. Exp. Med. Biol.* vol. 781. New York: Springer; 2014. p. 233–47.

42. Matzkin LM. The molecular basis of host adaptation in cactophilic drosophila: Molecular evolution of a glutathione S-transferase gene (GstD1) in *Drosophila mojavensis*. *Genetics*. 2008;178:1073–83.
43. Smith G, Lohse K, Etges WJ, Ritchie MG. Model-based comparisons of phylogeographic scenarios resolve the intraspecific divergence of cactophilic *Drosophila mojavensis*. *Mol Ecol*. 2012;21:3293–307.
44. Reed LK, Nyboer M, Markow TA. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Mol Ecol*. 2007;16:1007–22.
45. Machado CA, Matzkin LM, Reed LK, Markow TA. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. *Mol Ecol*. 2007;16:3009–24.
46. Anisimova M, Nielsen R, Yang Z. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics*. 2003;164:1229–36.
47. Arenas M, Posada D. Coalescent simulation of intracodon recombination. *Genetics*. 2010;184:429–37.
48. Arenas M, Posada D. The influence of recombination on the estimation of selection from coding sequence alignments. In: Fares MA, editor. *Nat. Sel. Methods Appl*. Boca Raton: CRC Press/Taylor & Francis; 2014. p. 112–25.
49. Matzkin LM. Population genetics and geographic variation of Alcohol dehydrogenase (Adh) paralogs and Glucose-6-Phosphate dehydrogenase (G6pd) in *Drosophila mojavensis*. *Mol Biol Evol*. 2004;21:276–85.
50. Matzkin LM, Eanes WF. Sequence variation of alcohol dehydrogenase (Adh) paralogs in cactophilic *Drosophila*. *Genetics*. 2003;163:181–94.
51. Ramírez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*. 2008;179:555–67.
52. Nason JD, Hamrick JL, Fleming TH. Historical vicariance and postglacial colonization effects on the evolution of genetic structure in *Lophocereus*, a Sonoran desert columnar cactus. *Evolution (N Y)*. 2002;56:2214–26.
53. Dyer RJ, Nason JD. Population Graphs: the graph theoretic shape of genetic structure. *Mol Ecol*. 2004;13:1713–27.
54. Smith CI, Farrell BD. Range expansions in the flightless longhorn cactus beetles, *Moneilema gigas* and *Moneilema armatum*, in response to Pleistocene climate changes. *Mol Ecol*. 2005;14:1025–44.
55. Fellows DP, Heed WB. Factors affecting host plant selection in Desert-adapted cactophilic *Drosophila*. *Ecology*. 1972;53:850–8.
56. Kircher HW. Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*. In: Barker JSF, Starmer WT, editors. *Ecol. Genet. Evol. Cactus-Yeast-Drosophila Syst*. New York: Academic Press; 1982. p. 143–58.
57. Fogleman JC, Starmer WT, Heed WB. Larval selectivity for yeast species by *Drosophila mojavensis* in natural substrates. *Proc Natl Acad Sci*. 1981;78:4435–9.
58. Fogleman JC, Danielson PB. Chemical interactions in the cactus-microorganism- *Drosophila* model system of the Sonoran Desert. *Am Zool*. 2001;41:877–89.
59. Date P, Crowley-Gall A, Diefendorf AF, Rollmann SM. Population differences in host plant preference and the importance of yeast and plant substrate to volatile composition. *Ecol Evol*. 2017;7:3815–25.
60. Date P, Dweck HKM, Stensmyr MC, Shann J, Hansson BS, Rollmann SM. Divergence in olfactory host plant preference in *D. mojavensis* in response to cactus host use. *PLoS One*. 2013;8:1–10.
61. Ross CL, Markow TA. Microsatellite variation among diverging populations of *Drosophila mojavensis*. *J Evol Biol*. 2006;19:1691–700.
62. Koshland D, Goldbeter A, Stock J. Amplification and adaptation in regulatory and sensory systems. *Science*. 1982;217:220–5.
63. Mobashir M, Schraven B, Beyer T. Simulated evolution of signal transduction networks. *PLoS One*. 2012;7:1–10.
64. Benton R, Sachse S, Michnick SW, Vosshall LB. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS Biol*. 2006;4:240–57.
65. Isono K. Molecular and cellular designs of insect taste receptor system. *Front Cell Neurosci*. 2010;4:1–16.
66. Coleman JM, Benowitz KM, Jost AG, Matzkin LM. Behavioural evolution accompanying host shifts in cactophilic *Drosophila* larvae. *Ecol Evol*. 2018; <https://doi.org/10.1002/ece3.4209>.
67. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9.
68. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68:978–89.
69. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*. 2003;73:1162–9.
70. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet*. 2005; 76:449–62.
71. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–2.
72. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–95.
73. Fu Y-X, Li W-H. Statistical tests of neutrality of mutations. *Genetics*. 1993;133:693–709.
74. Fu Y-X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. 1997;147:915–25.
75. Kosakovsky-Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. GARD: a genetic algorithm for recombination detection. *Bioinformatics*. 2006;22:3096–8.
76. Kosakovsky-Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol*. 2006;23:1891–901.
77. Kosakovsky-Pond SL, Frost SDW, Muse SV. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*. 2005;21:676–9.
78. Wright S. The genetic structure of populations. *Ann Eugen*. 1951;15:323–54.
79. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution (N. Y)*. 1984;38:1358–70.
80. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 1992;131:479–91.
81. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564–7.
82. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 2001;17:754–5.
83. Posada D. jModelTest: Phylogenetic model averaging. *Mol Biol Evol*. 2008;25:1253–6.
84. Drummond A, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:1–8.
85. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
86. Zhang J, Nielsen R, Yang Z. Evaluation of an improved Branch-Site Likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 2005;22:2472–9.
87. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. *Mol Biol Evol*. 2015;32:1365–71.
88. Kosakovsky-Pond SL, Murrell B, Fourment M, Frost SDW, Delpont W, Scheffler K. A random effects Branch-Site model for detecting episodic diversifying selection. *Mol Biol Evol*. 2011;28:3033–43.
89. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. 1991;351:652–4.
90. Yang Z, Wong WSW, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*. 2005;22:1107–18.
91. Kosakovsky-Pond SL, Frost SDW. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 2005;22:1208–22.
92. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky-Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 2012;8:e1002764.
93. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky-Pond SL, et al. FUBAR: A fast, unconstrained bayesian Approximation for inferring selection. *Mol Biol Evol*. 2013;30:1196–205.
94. Conant GC, Wagner GP, Stadler PF. Modeling amino acid substitution patterns in orthologous and paralogous genes. *Mol Phylogenet Evol*. 2007;42:298–307.
95. Atchley WR, Zhao J, Fernandes AD, Druke T. Solving the protein sequence metric problem. *Proc Natl Acad Sci*. 2005;102:6395–400.
96. Omasits U, Ahrens CH, Müller S, Wollscheid B, Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*. 2014;30:884–6.