

# Weak Polygenic Selection Drives the Rapid Adaptation of the Chemosensory System: Lessons from the Upstream Regions of the Major Gene Families

Pablo Librado and Julio Rozas\*

Departament de Genètica, Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

\*Corresponding author: E-mail: jroz@ub.edu.

Accepted: July 18, 2016

## Abstract

The animal chemosensory system is involved in essential biological processes, most of them mediated by proteins encoded in multigene families. These multigene families have been fundamental for the adaptation to new environments, significantly contributing to phenotypic variation. This adaptive potential contrasts, however, with the lack of studies at their upstream regions, especially taking into account the evidence linking their transcriptional changes to certain phenotypic effects. Here, we explicitly characterize the contribution of the upstream sequences of the major chemosensory gene families to rapid adaptive processes. For that, we analyze the genome sequences of 158 lines from a population of *Drosophila melanogaster* that recently colonized North America, and integrate functional and transcriptional data available for this species. We find that both, strong negative and strong positive selection, shape transcriptional evolution at the genome-wide level. The chemosensory upstream regions, however, exhibit a distinctive adaptive landscape, including multiple mutations of small beneficial effect and a reduced number of *cis*-regulatory elements. Together, our results suggest that the promiscuous and partially redundant transcription and function of the chemosensory genes provide evolutionarily opportunities for rapid adaptive episodes through weak polygenic selection.

**Key words:** *Drosophila*, chemosensory, transcription, upstream regions, natural selection.

## Introduction

The chemosensory system plays a major role in essential processes such as nutrition, reproduction and social communication (Krieger and Ross 2002; Matsuo et al. 2007; Whiteman and Pierce 2008; Smadja and Butlin 2009). In insects, the specificity and sensibility of this system highly rely on peripheral processes occurring in sensilla (Pelosi 1996; Hildebrand and Shepherd 1997), which involve the uptake of chemical stimuli, its transport through the sensilla lymph, and the interaction with the chemoreceptor that will activate the signal transduction cascade. Most of these peripheral processes are mediated by proteins encoded in multigene families (Sanchez-Gracia et al. 2009; Depetris-Chauvin et al. 2015; Joseph and Carlson 2015), including extracellular ligand-binding proteins, such as Odorant Binding Proteins (OBPs) and Chemosensory Proteins (CSPs), as well as membrane receptor proteins, such as Odorant Receptors (ORs), Gustatory Receptors (GRs) and Ionotropic Receptors (IRs). As the expression of such genes

determines the ability to discriminate external chemical cues (and thus the individual fitness), the chemosensory system constitutes an excellent candidate to study the role of natural selection in driving transcriptional changes at the molecular level.

Several studies have revealed that the chemosensory protein-coding regions mainly evolve under purifying selection (Sanchez-Gracia et al. 2009; Croset et al. 2010; Almeida et al. 2014; Torres-Oliva et al. 2016). Yet, the selective constraints are typically lower for the “odorant binding” Gene Ontology (GO) category than for the genome average (Clark et al. 2007; Arguello et al. 2016), and a few positions show the molecular hallmark of positive selection (Foret and Maleszka 2006; McBride et al. 2007; Sánchez-Gracia and Rozas 2008). This relatively well-defined mode of protein-coding gene evolution contrasts with the lack of studies surveying their upstream regions (hereafter, chemosensory upstream regions). Indeed, there is compelling evidence

linking their transcriptional changes with particular phenotypic effects. For example: (1) expression changes of the *Or10* (Rollmann et al. 2010), *Or43a* (Stortkuhl et al. 2005), and *Obp99a* (Wang et al. 2010) genes have been associated with behavioral responses to benzaldehyde, an odorant involved in the *Drosophila melanogaster* feeding behavior; (2) several OBP and OR genes are upregulated in specialist species, such as *Obp50a* in *D. sechellia*, and *Or22a* in *D. sechellia* and *D. erecta* (Kopp et al. 2008; Linz et al. 2013; Shiao et al. 2015); (3) OBP genes are often located in chromosome regions of elevated transcriptional noise, as it may be advantageous in unstable external environments (Librado and Rozas 2013). Overall, these findings support that the transcriptional changes of the chemosensory genes significantly contribute to the phenotypic variation, both within (polymorphism) and between (divergence) *Drosophila* species.

Comparison of polymorphism and divergence DNA variability patterns between 5' upstream and neutrally evolving regions can be instrumental for understanding the action of natural selection at regulatory sequences (Jenkins et al. 1995; Egea et al. 2008). This powerful approach allows—in addition—estimating the fraction of nucleotide substitutions fixed by positive selection (i.e., the so-called  $\alpha$  parameter; Fay et al. 2001). Noticeably,  $\alpha$  estimates in noncoding regions widely vary among species, ranging from ~0% in humans (Keightley et al. 2005; Eyre-Walker and Keightley 2009) to ~50% in the UTRs of *Drosophila* (Andolfatto 2005; Hadrill et al. 2008). Nevertheless, this variation does not necessarily represent uneven adaptive pressures, but differences in the effective population sizes ( $N_e$ ) (Jensen and Bachtrog 2011; Gossmann et al. 2012). To disentangle these two confounding factors, (Schneider et al. 2011) developed a two-step approach that first accounts for  $N_e$  variations over time, and then estimates relevant selective parameters, such as the distribution of fitness effects of new deleterious mutations (DFE), the proportion of new adaptive mutations ( $p_a$ ) with a certain selective strength ( $s_a$ ), the rate of adaptive substitution ( $\alpha$ ), and the ratio of adaptive-to-neutral substitution rates ( $\omega_a$ ).

Here, we assess the relative contribution of the upstream regions to rapid adaptive processes, especially focusing on those of the major chemosensory gene families. For that, we applied the demographic-aware method developed by Schneider and colleagues (Schneider et al. 2011) to analyze the genome sequences of 158 *D. melanogaster* lines that recently colonized North America (Mackay et al. 2012). We investigated whether the selective parameters inferred for the chemosensory upstream regions are indeed distinctive, or associated with peculiarities of the chemosensory system, as characterized by their Gene Ontology annotations, and the functional (Roy et al. 2010; Kharchenko et al. 2011) and transcriptional (Ayroles et al. 2009) information available for this species.

## Materials and Methods

### Genomic Alignments

We downloaded the genome-wide multiple sequence alignment (MSA) of 158 *D. melanogaster* lines, with *D. simulans* and *D. yakuba* as outgroup species, from the PopDrowser database (Ramia et al. 2012). For each protein-coding gene ( $n_G=14,380$ , after excluding mitochondrial and Y-linked genes in the *D. melanogaster* release 5.42), we extracted the MSA of the following: (1) its largest transcript isoform, (2) its 4-fold degenerate positions, (3) its intronic regions and, (4) its upstream sequence (the 2 kb 5' upstream region, provided it does not overlap with any neighbor protein-coding region).

### Handling Residual Heterozygosity and Missing Data

DGRP freeze 1 includes a suite of 158 *D. melanogaster* lines from a natural population of Raleigh, inbred to near homozygosity (Mackay et al. 2012). We treated residual heterozygosity as missing data. As ~40% of the MSA positions include missing data (residual heterozygosity and/or "N"s), removing these alignment columns (complete deletion) would yield an important loss of information. To elude such undesirable situation, we circumscribed our study to MSA positions having at least 150 valid nucleotides ( $n_P=105,630,385$  positions), a trade-off that optimizes the number of positions and samples analyzed (supplementary fig. S1, Supplementary Material online).

Analyzing positions with a minimum number of 150 valid nucleotides (i.e., neither missing nor heterozygous) implies an uneven sample size along the MSA (e.g., some positions have 158 valid alleles, while others 150). To deal with positions having >150 valid alleles, we used a probabilistic approach. Suppose an alignment position with 152 valid alleles (e.g., 150 adenines, two derived thymines, and six missing variants), from which we have to sample only 150 valid nucleotides. Three different samples could be extracted: 148 adenines and two thymines, 149 adenines and one thymine, or 150 adenines and zero thymines. The probability of each configuration can be calculated by means of the hypergeometric distribution. In this example, we would consider a doubleton, singleton or monomorphic mutations/position with probabilities 0.9738, 0.0261 and  $8.7138 \times 10^{-05}$ , respectively.

### Handling Intra-Locus-Linked Positions

In population genomics, most neutrality tests based on the frequency spectrum assume that nucleotide positions evolve independently. In low recombining regions, however, natural selection can increase (e.g., via selective sweeps; Smith and Haigh 1974; Stephan et al. 1992) or reduce (e.g., via background selection; Charlesworth et al. 1993) the frequency of nearby-linked neutral alleles (supplementary fig. S2, Supplementary Material online). To deal with such

confounding effects, we applied two different approaches. First, as recombination can break up linked variants, we only analyzed the nucleotide diversity patterns at high-recombining regions (>2 cM/Mb; Mackay et al. 2012), which were identified using the Recombination Rate Calculator v2.2 (Fiston-Lavier et al. 2010). Second, for each group of linked mutations (haplotype block), we only analyzed the most central SNP as representative (tagSNP) of the allele frequencies within the haplotype block. To conduct such pruning step, we first detected all significant pairs of linked positions, by means of the Fisher exact test (FET) with multiple testing correction (Benjamini and Hochberg 1995). We further clustered these pairs into larger groups (i.e., the haplotype blocks) using mcl v10-201 program (Van Dongen 2008).

### Nucleotide Diversity across Different Site Categories

We used VariScan (Vilella et al. 2005; Hutter et al. 2006) to estimate the nucleotide diversity ( $\pi$ ) at the 2-kb upstream regions and 4d-fold sites of each protein-coding gene (supplementary tables S1 and S2, Supplementary Material online). We also computed  $\pi$  at positions 8–30 of short introns ( $\leq 65$  bp) (supplementary table S2, Supplementary Material online), as they likely represent the most neutrally evolving class of sites (Parsch et al. 2010). For each site category, we computed the mean and the confidence intervals of  $\pi$ , using the wtd.mean and wtd.quantile functions of the [R] programming language (R Core Team 2015).

### Unfolded Site Frequency Spectrum

To estimate the unfolded site-frequency spectrum (uSFS) we first reconstructed the nucleotide sequence of the most recent *D. melanogaster*-*D. simulans* ancestor. For that, we applied the joint ancestral reconstruction method implemented in the “baseml” program (Yang 2007), using as input the consensus DNA sequence of the 158 *D. melanogaster* lines (selecting the most frequent nucleotide variant at each position), and the sequences of *D. simulans* and *D. yakuba* (to polarize mutations). Given this ancestral sequence, we computed the uSFS using a parsimony criterion (at biallelic alignment positions with at least 150 valid nucleotides).

### Distribution of Fitness Effects, and Rates of Adaptive Evolution

We used the DFE v2.03 program (Keightley and Eyre-Walker 2007; Schneider et al. 2011) to estimate the impact of natural selection at the 5' upstream regions that are located in high-recombining regions (9,279 in total, 170 of which belong to chemosensory genes; supplementary tables S1 and S2, Supplementary Material online). As demography may mimic the molecular hallmark yield by natural selection, we first estimated the underlying demographic history of the *D. melanogaster* Raleigh population, from the SIP frequency spectrum (used as a reference for neutrally evolving sites). In particular,

**Table 1**

Mathematical Definition of the Main Parameters of Positive and Negative Selection

	Adaptive	Nonadaptive
<b>Proportion of mutations</b>	$p_a$	$1-p_a$
<b>Fitness effect of mutations</b>	$s_a (>0)$	$s (\leq 0)$
<b>Mean fixation probability</b>	$U_a = \frac{1-e^{-s_a}}{1-e^{-2N_e s_a}}$	$U_d = \frac{1-e^{-s}}{1-e^{-2N_e s}}$
<b>Adaptive subst. rate</b>	$\alpha = \frac{p_a U_a}{p_a U_a + (1-p_a) U_d}$	NA
<b>Adaptive-to-neutral subst. rate</b>	$\omega_a = a/d_s = 2N_e p_a U_a$	NA

NOTE.—The proportion and fitness effects of new mutations are directly estimated by the DFE v2.03 program. In particular, the distribution of fitness effects of the new nonadaptive mutations (DFE) is modeled using a Gamma distribution. The shape of such Gamma distribution determines the proportion of nonadaptive mutations that is nearly neutral ( $0 < -N_e s < 1$ ), slightly deleterious ( $1 < -N_e s < 10$ ), mildly deleterious ( $10 < -N_e s < 100$ ) and strongly deleterious ( $100 < -N_e s$ ). Given these parameters, the mean fixation probability,  $\alpha$ , and  $\omega_a$  are calculated by applying the corresponding formulas.

we investigated if the population has experienced: (1) a constant effective population size ( $N_e$ ), (2) a single  $N_e$  change event (either an expansion or a contraction) some generations ago (two-epoch model), or (3) two  $N_e$  changes (expansions or contractions) in the past (three-epoch model, which can account for a bottleneck scenario). The weighted Akaike Information Criterion (wAIC) indicated that the best-fit model is the last one (three-epoch; see “Results” section).

Over this three-epoch demographic baseline, we evaluated the fit of three selection regimes to the uSFS of the upstream regions: (1) the “no selection model”, which considers that the change in the allele frequencies is only governed by the underlying demographic history (i.e., the three-epoch model), (2) “negative selection model”, which contemplates the three-epoch demographic scenario plus the fitness effects of new deleterious mutations (DFE, which is modeled through a Gamma distribution) (table 1), and (3) the “complex selection model”, which includes the three-epoch demographic scenario, the DFE, and a proportion of new mutations ( $p_a$ ) with adaptive effects ( $s_a$ ) (table 1). We determined the statistical significance of positive selection at the chemosensory upstream regions (table 2), comparing the fit of the “negative selection model” and the “complex selection model” (an explicit test for positive selection) via the likelihood ratio test (using a chi-squared distribution with 2 d.f.).

### Functional and Expression Data

To gain insights into the evolutionary impact of the *cis*-regulatory elements (CREs) that are located within the 5' upstream regions, we gathered different sources of functional genomics data available for *D. melanogaster*. We collected information of the transcription factor binding sites (Marygold et al. 2013) by mining the “TF\_binding\_sites”, “enhancer”, “silencer” and “insulator” keywords in the “feature” field of the FlyBase GFF file (*D. melanogaster* release 5.54). This data

**Table 2**

Rate and Strength of Natural Selection in the Concatenated Chemosensory Upstream Sequences

	<i>P</i> value	$-N_e s$	$P_a$ (SE)	$S_a$ (SE)	$\alpha$ (SE)	$\omega_a$ (SE)
CSP	$2.30 \times 10^{-7}$	119.02 (1.07)	0.006 ( $3.23 \times 10^{-4}$ )	0.386 ( $3.10 \times 10^{-3}$ )	0.93 ( $4.03 \times 10^{-4}$ )	0.32 ( $4.33 \times 10^{-4}$ )
OBP	$1.57 \times 10^{-49}$	29.68 (0.04)	0.034 ( $1.12 \times 10^{-4}$ )	0.068 ( $6.51 \times 10^{-4}$ )	0.94 ( $3.11 \times 10^{-4}$ )	0.35 ( $1.91 \times 10^{-4}$ )
OR	$7.57 \times 10^{-84}$	27.77 (0.03)	0.038 ( $1.10 \times 10^{-4}$ )	0.071 ( $3.61 \times 10^{-4}$ )	0.98 ( $1.58 \times 10^{-4}$ )	0.42 ( $1.53 \times 10^{-4}$ )
GR	$8.75 \times 10^{-81}$	26.05 (0.03)	0.037 ( $1.77 \times 10^{-4}$ )	0.077 ( $9.32 \times 10^{-4}$ )	0.97 ( $2.89 \times 10^{-4}$ )	0.44 ( $1.92 \times 10^{-4}$ )
alR	$1.43 \times 10^{-06}$	43.91 (0.15)	0.005 ( $3.28 \times 10^{-4}$ )	0.380 ( $2.27 \times 10^{-3}$ )	0.77 ( $1.22 \times 10^{-3}$ )	0.27 ( $5.67 \times 10^{-4}$ )
dIR	$2.43 \times 10^{-81}$	29.27 (0.05)	0.063 ( $1.39 \times 10^{-4}$ )	0.045 ( $1.33 \times 10^{-4}$ )	0.99 ( $6.97 \times 10^{-5}$ )	0.44 ( $1.70 \times 10^{-3}$ )

NOTE.—The “*P* value” column shows the *P* value of the “positive selection” test (LRT, using a chi-squared distribution with 2 d.f.). As this test is significant for all six chemosensory families, the following columns summarize the impact of negative and positive selection, with their corresponding SEs.

represents a valuable resource summarizing many functional assays currently conducted in *D. melanogaster*, mostly including CRE candidates inferred by chromatin immunoprecipitation techniques. We moreover downloaded the 30 chromatin states model, which classifies each *D. melanogaster* nucleotide position according to the combinatorial profile of 18 histone marks (Kharchenko et al. 2011). We utilized the chromatin model from the BG3 cell line, because as derives directly from neurons may provide a reasonable approximation of the high-order regulatory mechanisms acting at the chemosensory upstream regions.

We used the 160 microarrays obtained by (Ayroles et al. 2009) to quantify the intra-specific transcriptional variability of the chemosensory genes. This data set encompasses transcriptional information for 40 DGRP lines, including two technical replicates for each male and female (ArrayExpress accession number: E-MEXP-1). We parsed and normalized the raw signal CEL files using the *affy*, *affyPLM*, *drosophila2cdf*, *org.Dm.eg.db* and *drosophila2.db* [R] packages of Bioconductor (Gentleman et al. 2004). For each biallelic position, we classified the 40 DGRP lines as carriers of the ancestral or the derived allele, and computed the expression intensity ratio (EIR) between these two groups. The biallelic positions with an EIR significantly deviating from the genome-wide pattern (top-1% or bottom-1% outliers) were considered eQTLs.

## Results

### Demographic History of the *D. melanogaster* Raleigh Population

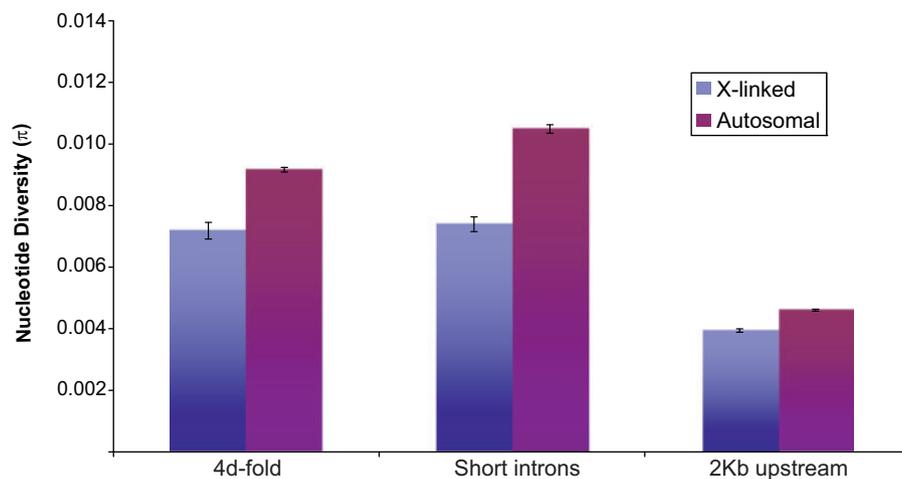
Quantifying the impact of natural selection on the 5' chemosensory upstream regions is challenging, as the bottleneck experienced by the *D. melanogaster* Raleigh population (Keller 2007; Duchon et al. 2012; Garud et al. 2015) can similarly yield a reduction of nucleotide diversity, as well as increased levels of linkage disequilibrium. To discern between demography and natural selection, we analyzed selectively neutral evolving sites in high-recombining regions (>2 cM/Mb). In particular, we assessed the neutrality of three classes of sites: the 5' upstream regions (2 kb, or shorter if it overlaps

with a neighbor upstream gene), the 4-fold degenerate sites, and the positions 8–30 of short introns ( $\leq 65$  bp) (short intron positions, or SIP). In agreement with previous findings (Parsch et al. 2010), our genome-wide analysis shows that the SIPs exhibit the highest levels of interspecific divergence and intra-specific polymorphism. This levels of nucleotide diversity are higher than those estimated at 4-fold degenerate positions (supplementary table S2, Supplementary Material online), which may reflect some codon usage bias for translational efficiency or accuracy (Nielsen et al. 2007; Poh et al. 2012; Lawrie et al. 2013). Moreover, the X-to-autosomes polymorphism ratio at SIPs approaches to the neutral expectations of 0.75 (fig. 1 and supplementary table S2, Supplementary Material online). Taken together, these results support that, in the North America population of *D. melanogaster*, SIPs are negligibly affected by any form of natural selection and, therefore, validate their utility as neutral markers to estimate the timing and severity of the bottleneck event.

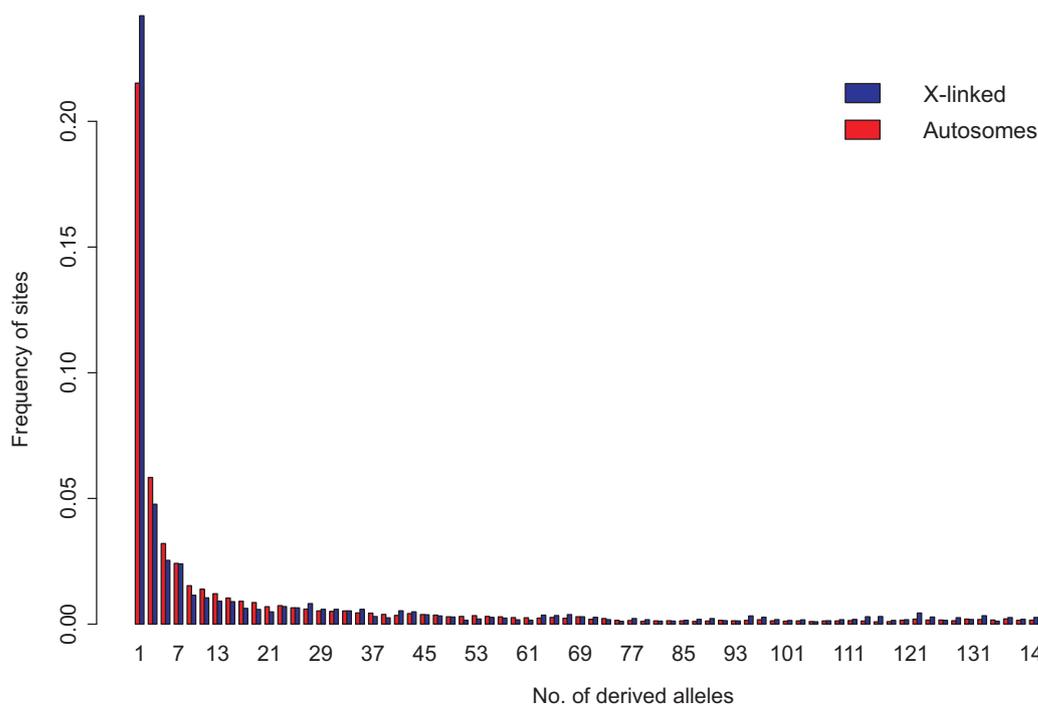
Demographic processes may unevenly shape autosomal and X-linked SIPs if, for example, males and females differentially contributed to the North America colonization. Using the *dfe* program (see “Materials and Methods” section), we separately fit different demographic scenarios to the uSFS of the autosomal and X-linked markers (fig. 2). In agreement with previous studies (Keller 2007; Duchon et al. 2012; Garud et al. 2015), we consistently found that the best-fit scenario is the 3-epoch model ( $wAIC_{\text{auto}}=0.99$  and  $wAIC_x=0.99$ , respectively), with ML estimates indicating a population bottleneck with a very recent recovery (supplementary table S3, Supplementary Material online). Due to its temporal proximity and severity, current allele frequencies are expected to be mostly influenced by the bottleneck. Therefore, explicitly accounting for this bottleneck is essential to accurately disentangle rate and strength of natural selection.

### Impact of Natural Selection on the Chemosensory Upstream Regions

We assessed the impact of natural selection by concatenating the 5' upstream regions of each chemosensory gene family, separating the two major IR subfamilies, the antennal (alR) and



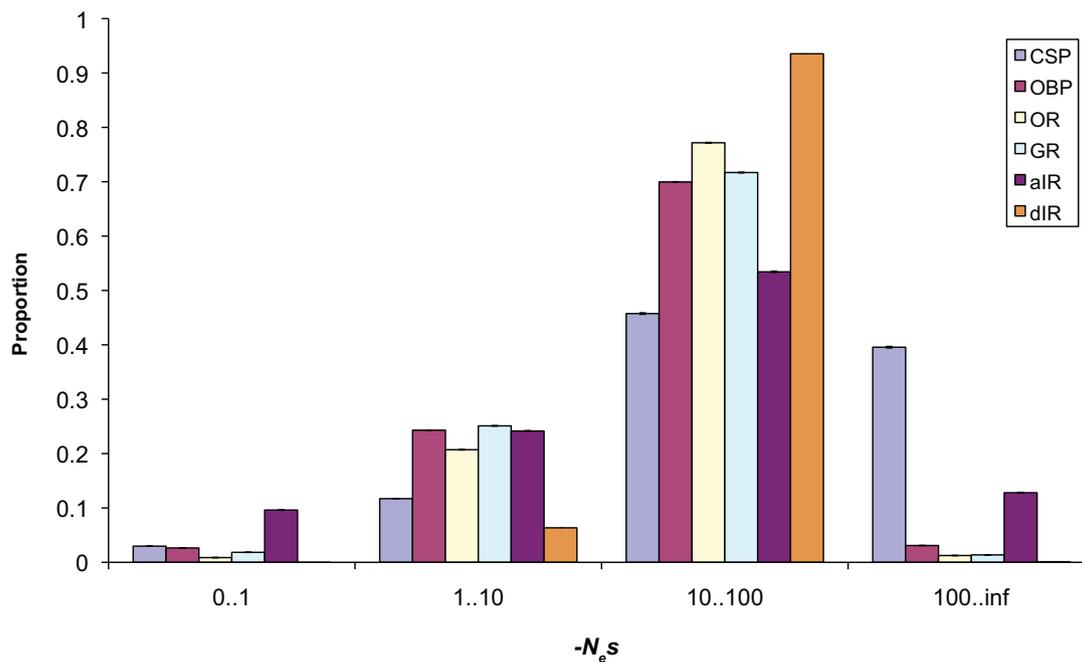
**FIG. 1.**—Nucleotide diversity levels ( $\pi$ ) at three putatively neutral classes of sites: the 4-fold degenerate sites, the positions 8–30 of short introns ( $\leq 65$  bp), and 2-kb upstream regions.  $\pi$  was separately computed for all the autosomal and X-linked regions.



**FIG. 2.**—The unfolded site frequency spectrum summarizing the proportion of biallelic sites with a given number of derived alleles. It was calculated from the positions 8–30 of all short introns ( $\leq 65$  bp) across the autosomal and X chromosomes.

the divergent (dIR) Ionotropic Receptors. Briefly, we compared the fit of two nested models to the unfolded Site Frequency Spectrum (uSFS) of the chemosensory upstream concatenates, including a null one considering the demographic model parametrized above plus the action of negative selection versus an alternative model additionally including the action of positive selection (see “Materials and Methods” section).

We found that the latter fits the uSFSs significantly better ( $P < 1.43 \times 10^{-6}$ ), especially for the upstream regions of the OR, GR and dIR genes ( $P = 7.57 \times 10^{-84}$ ,  $8.75 \times 10^{-81}$  and  $2.43 \times 10^{-81}$ , respectively) (table 2). These results indicate that positive selection has unequivocally played a significant role in the molecular evolution of the chemosensory upstream regions. Under this alternative model, which also considers the



**FIG. 3.**—Distribution of fitness effects of new deleterious mutations in autosomal chromosomes, partitioned into four categories: nearly neutral ( $0 < -N_e s < 1$ ), slightly deleterious ( $1 < -N_e s < 10$ ), mildly deleterious ( $10 < -N_e s < 100$ ) and strongly deleterious ( $100 < -N_e s$ ).

action of negative selection, we inferred that most new mutations are detrimental, with a distribution of fitness effects (DFE) skewed towards mildly ( $10 < -N_e s < 100$ ) or strongly deleterious ( $-N_e s > 100$ ) mutations, particularly at the CSP and alR upstream regions (fig. 3). Complementarily, a reduced but statistically significant proportion of new mutations ( $p_a$ ) have advantageous fitness effects ( $s_a$ ). For example, only the 3.8% ( $p_a = 0.038$ ) of mutations at the OR upstream regions are beneficial, with a mean advantageous fitness effects of  $s_a = 0.071$  (table 2).

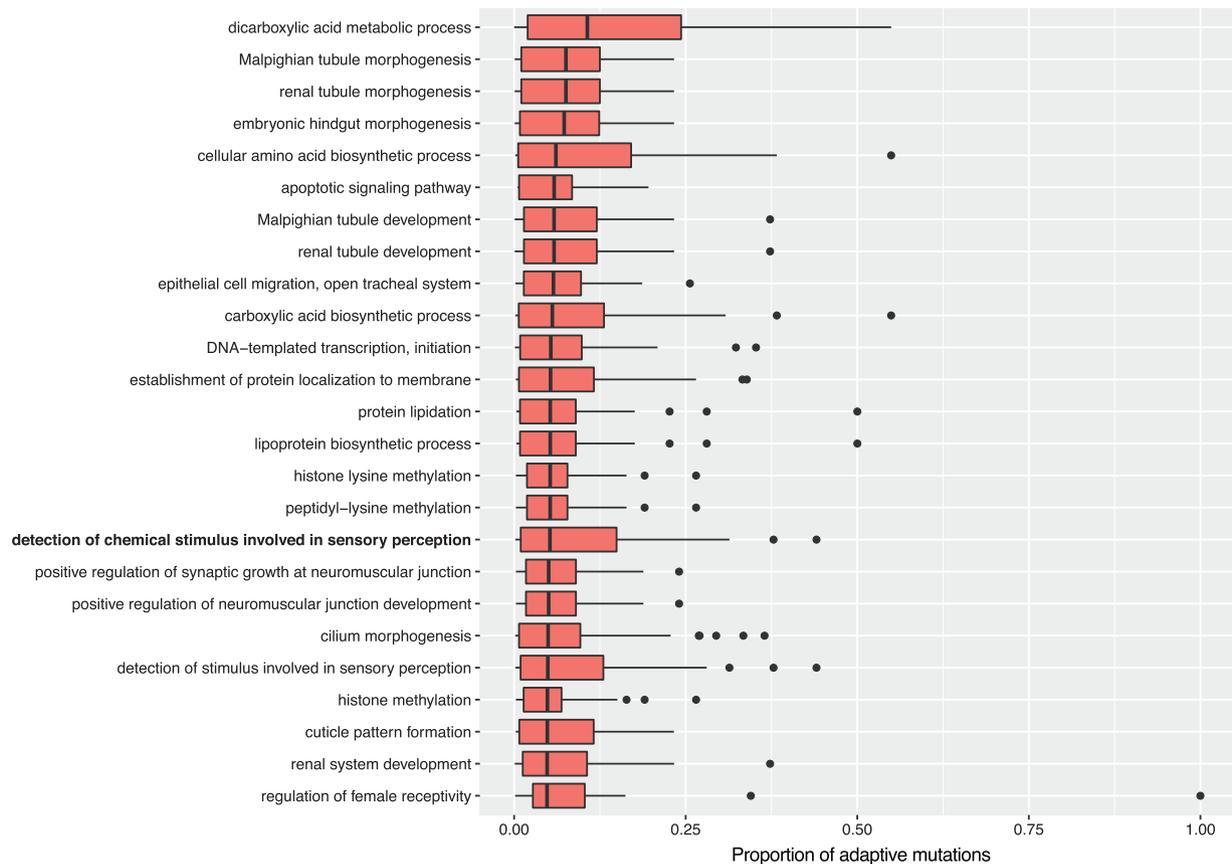
To contextualize these  $p_a$  and  $s_a$  estimates, we conducted the same analysis on the 9,279 upstream sequences located in high-recombining regions. We found a clear inverse relationship between these two parameters (Spearman rank correlation;  $\rho = -0.9294$ ;  $P = 0$ ), indicating that  $p_a$  and  $s_a$  are rarely high at the same time. This anticorrelation delineates, thus, two competing modes of adaptation, one including multiple changes of weak fitness effects, whereas another involving a few but strongly selected alleles. To further explore what drives the mode of selection, we grouped genes according to their Gene Ontology (GO) terms. Remarkably, we observed a complex nonmonotonic relationship with the GO term size, whereby functional categories encompassing a low number of genes ( $< 20$  members) tend to exhibit extreme  $p_a$  and  $s_a$  values (supplementary fig. S3, Supplementary Material online;  $F$ -test;  $P < 2 \times 10^{-16}$ ). It is worth noting that this significance mostly remains regardless of the cut-off value used to classify GO terms as small and large, and of the GO level considered (not shown). This likely reflects that biological

processes controlled by a large number of genes offer extra sources of potential DNA variation, and thus that positive selection can accommodate a phenotype that is adaptive for a particular biological process through multiple changes of small fitness effect. This trend is also recapitulated at the upstream regions of chemosensory genes, where families with high number of members (i.e., OBPs, ORs, GRs, and dIRs) concomitantly harbor a high proportion of adaptive sites with a low beneficial effect (table 2).

We then contrasted the  $p_a$  and  $s_a$  estimates for genes annotated as “detection of chemical stimulus involved in sensory perception” (which encompasses all the surveyed chemosensory gene families) against comparable GO categories; i.e., at the same level in the GO hierarchical structure and with minimum size of 20 members. We found that “detection of chemical stimulus involved in sensory perception” ranks amongst the categories with lowest  $s_a$  values (711th out of 727 GO terms), but top- $p_a$  estimates (17th out of 727; fig. 4). This collectively supports that weak polygenic selection has been especially prevalent at the chemosensory upstream regions, possibly because the combinatorial nature of the olfactory and gustatory code is complex, and involve multiple actors, which enables rapidly shaping a phenotype through diverse genetic changes.

### Strong versus Weak Selection in Transcriptional Evolution

To assess the relative contribution of strong and weak polygenic selection to transcriptional evolution, we conducted a



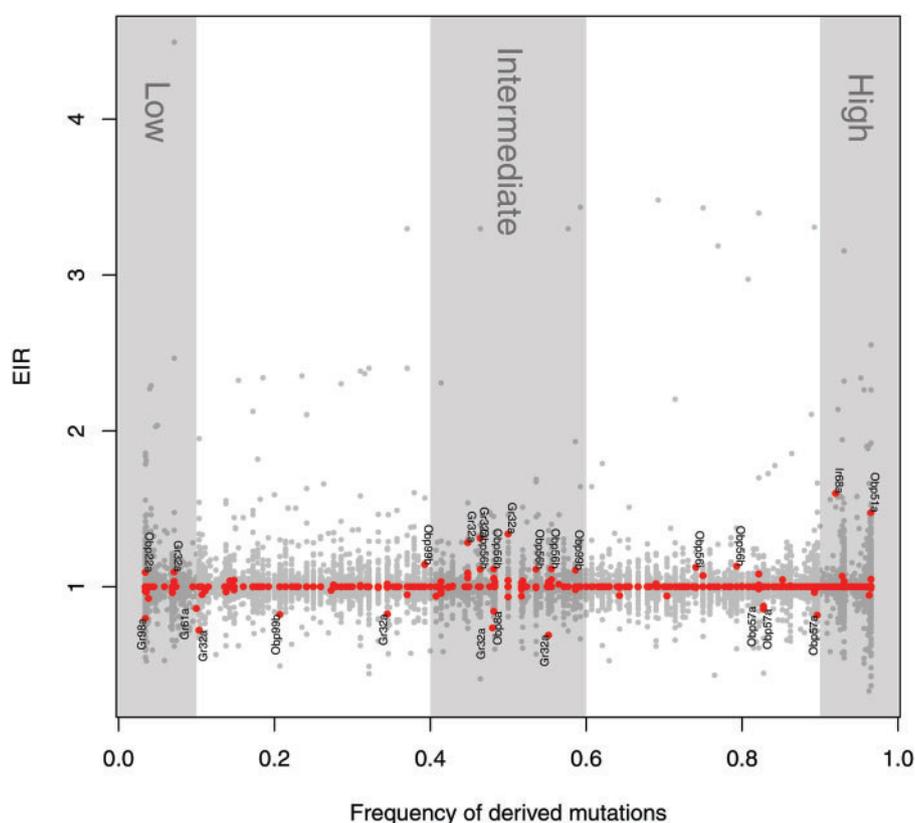
**Fig. 4.**—Top 25 GO categories according to their median  $p_a$  values. In bold, the category “detection of chemical stimulus involved in sensory perception”, which comprises all surveyed chemosensory families. It ranks in position 17 (out of 727 GO terms with >20 members), approximately representing the top-2%.

finer-scale analysis, by associating single nucleotide variants with expression intensity changes. In particular, we explored the relationship between the frequency of derived mutations and the expression-fold change they induce (i.e., ratio of transcriptional intensity between individuals with derived and ancestral alleles; see “Materials and Methods” section).

We observed many mutations significantly associated with gene transcriptional changes (expression QTLs, or eQTLs; fig. 5). The genes harboring such eQTLs at their upstream regions are mainly targeted to the “extracellular region” (Hypergeometric test, performed with the DAVID tool (Huang et al. 2008); adjusted  $P$  value =  $1.0 \times 10^{-10}$ ), and involved in “defence response” (adjusted  $P$  value =  $2.9 \times 10^{-6}$ ). These eQTLs are remarkably skewed towards extreme allele frequencies, revealing a strong action of negative and positive selection in shaping transcriptional evolution (FET; two-tailed  $P=0$ ; fig. 5). Unlike this genome-wide pattern, the eQTLs at the upstream regions of the chemosensory genes are slightly enriched at intermediate frequencies (fig. 5; two-tailed FET;  $P=0.0182$ ). For example, the upstream region of *Gr32a* ranks in the top 1% of the genome according to its Tajima’s  $D$  value

( $D=1.201$ ) (Tajima 1989), a DNA variability pattern that may be potentially exacerbated by its pleiotropic function, including DEET detection (Lee et al. 2010) and modulation of the male courtship behavior (Fan et al. 2013; Andrews et al. 2014). Among all eQTLs inferred to be targeted by strong Darwinian selection (derived allele frequency > 0.9; fig. 5), only two are located at the upstream regions of chemosensory genes (*Ir68a* and *Obp51a*). Notably, both genes are abundantly expressed in male accessory glands (Chintapalli et al. 2007) and known to be regulated by multiple CREs, as defined in FlyBase (Marygold et al. 2013) (supplementary fig. S4, Supplementary Material online).

To further investigate the relationship between the mode of selection and the number of CREs, we estimated  $p_a$  and  $s_a$  inside and outside the DNA regions known to be bound by regulatory proteins, and found that the proportion of adaptive mutations is not higher inside ( $p_a=0.0267$ ) than outside ( $p_a=0.0402$ ) CREs, but just the opposite (Wilcoxon test,  $P=9.38 \times 10^{-6}$ ). Beneficial mutations within known CREs have, however, larger fitness effects ( $s_a=0.0889$  vs.  $0.0662$ ; Wilcoxon test,  $P=0.0308$ ). Remarkably, the distinctive



**Fig. 5.**—Association between the frequencies of derived mutations and the transcriptional fold change they induce (EIR; see “Materials and Methods” section). Red and grey points depict SNPs in chemosensory and nonchemosensory upstream regions, respectively. eQTLs at chemosensory upstream regions (red) producing a significant transcriptional change are labeled with the corresponding gene name. The background shaded areas define the categories for extreme ( $>0.9$ , or  $<0.1$ ) and intermediate (from 0.4 to 0.6) derived allele frequencies.

selective pressure outside CREs, characterized by low  $s_a$  and high  $p_a$  estimates, resembles that estimated for the chemosensory upstream regions (fig. 4). We thus compared the CRE number at the chemosensory and rest of upstream regions, and with the recurrent exception of the aIR and CSP, we indeed found a lower amount of regulatory elements at the OBP, OR, GR and dIR upstream sequences (supplementary table S4, Supplementary Material online; Wilcoxon test;  $P < 0.05$ ). The lack of CREs is also supported by the chromatin state at the chemosensory upstream regions, which collectively shows a deficit of nucleotide positions annotated as “Promoter and TSS” and “Regulatory Regions” (supplementary table S5, Supplementary Material online; two-tailed Fisher Exact Test;  $P < 2.2 \times 10^{-16}$ ), while being enriched in nucleotides annotated as “Transcriptionally Silent, Intergenic” (two-tailed FET;  $P < 2.2 \times 10^{-16}$ ).

## Discussion

It is well-known that mutations affecting upstream gene regions are a major force driving the evolution of transcriptional regulation, significantly contributing to the phenotypic

variation within and between species (Wittkopp and Kalay 2012). We hypothesized that the role of positive selection may be particularly noteworthy at the upstream regions of the major chemosensory gene families, reflecting the adaptive pressures imposed by changing external environments. To infer the phenotypic effects of positively selected alleles, we comprehensively exploited diverse sources of relevant information, including population genomics analyses, as well as transcriptional and functional data available for this species.

### The Demographic History of *D. melanogaster*

Although already described elsewhere (Keller 2007; Duchen et al. 2012; Garud et al. 2015), we re-estimated the underlying *D. melanogaster* demography to ensure an homogeneous treatment of residual heterozygosity and missing data, which has been proven to be essential in large genome panels, including for the detection of selection (Ferretti et al. 2012). Indeed, the North America colonization offered opportunities to natural selection to act in response to the new climate environment pressures. If adaptation has occurred, the colonization is so recent ( $\sim 5000$ – $7000$  generations) that the

molecular footprint of positive selection may remain in the *D. melanogaster* genome and, particularly, in some chemosensory genes (which may play a relevant role in the adaptive process). We actually inferred that positive selection plays a major role in driving the nucleotide divergence at the chemosensory upstream regions (table 2), with rates of adaptive substitution much higher than those previously estimated at the *D. simulans* noncoding regions (Andolfatto 2005). Nevertheless, it does not necessarily reflect a huge adaptive pressure, but merely the recent demographic history of this population; i.e., the recovery from the bottleneck is so recent that hardly the strongly favored mutations have had time enough to reach fixation (supplementary fig. S5, Supplementary Material online). Therefore, and to avoid misleading interpretations about the impact of positive selection, we focused the analyses on the proportion of new adaptive mutations ( $p_a$ ) and their coefficient of selection ( $s_a$ ).

#### Differential Pace of Natural Selection at the Chemosensory Upstream Regions

We have detected a significant and inverse relationship between  $p_a$  and  $s_a$  parameters. This has profound implications to understand the pace of selection at the chemosensory upstream regions. Positive selection scans have traditionally focused on identifying strong departures of the expected levels and patterns of nucleotide variability, typically hard or soft selective sweep signatures (Stephan 2016). Garud and colleagues (Garud et al. 2015), for example, have recently applied a powerful statistical test to detect selective sweeps in the DGRP panel, and only identified a few chemosensory genes among the top-10% (strongly selected) candidate regions.

The phenotype of most quantitative traits is likely to be simultaneously controlled by multiple loci (Pritchard et al. 2010; Huang et al. 2012). Although it depends on the specific genetic architecture, the action of natural selection on complex traits is known to yield subtle molecular footprints, whereby the beneficial alleles rarely reach fixation, but remain segregating, trapped in the complexity of the rugged adaptive landscape (Pavlidis et al. 2012; Stephan 2016). This implies that alleles underlying complex selected traits may be counter-intuitively found at intermediate frequencies. Unlike the genome-wide pattern, shaped by both, strong negative and positive selection (fig. 5), we interestingly found that the eQTLs located at the chemosensory upstream regions are slightly but significantly enriched at intermediate frequencies. More conclusively, the GO term “detection of chemical stimulus involved in sensory perception” exhibits one of the largest proportions of adaptive mutations, but with small fitness effects. Our results are well in line with a recent study showing that the adaptive substitution rate is not particularly prevalent at the chemosensory protein families, and that chemosensory adaptation involves multiple but subtle changes from standing

variation (Arguello et al. 2016). Altogether, these results suggest that weak polygenic selection at the chemosensory upstream region contributed to the recent colonization of North America.

The fact that the chemosensory upstream regions belong to moderate-sized multigene families likely conditioned the pace of selection. It is well known, for example, that different members of the same chemosensory family recognize partially overlapping sets of cues (Laissue and Vosshall 2008). This redundancy may enable accumulating slightly or mildly deleterious variation that, eventually, can become beneficial in front of new external environments. Tolerating or even promoting moderate levels of phenotypic plasticity and/or segregating variation may be adaptive in varying environments, enabling to quickly buffer and survive sudden habitat alterations.

#### Functional Data and Phenotypic Assays

The phenotypic consequences of epistatic interactions could be easily misinterpreted if correlated mutations compensate each other or entail multiplicative fitness effects. It has been shown, for example, that variations in olfactory behavior are not only associated with polymorphisms of the major chemosensory gene families (Rollmann et al. 2010; Wang et al. 2010), but also with alleles that cause subtle variations in the olfactory neuron projections (Swarup et al. 2013; Arya et al. 2015). Further functional data is therefore indispensable to fully uncover the fitness landscape of complex traits, and the epistatic interactions between the upstream and downstream components of the chemosensory pathway (He et al. 2016). Meanwhile, we have integrated the information already available for *D. melanogaster*, which it is acknowledged to be partial, and thus potentially biased at some extent. It is well known, for example, that FlyBase is not completely updated, and lacks some CREs controlling the transcription of the chemosensory genes (Jafari et al. 2012; Barish and Volkan 2015). This could cast some fair doubts on whether the number of CREs is really lower at the chemosensory than at the rest of upstream regions (supplementary tables S4 and S5, Supplementary Material online). Conversely, as the chemosensory is a widely studied system, it is also possible that their CREs were over-represented in databases, making our comparison rather conservative. Nevertheless, given that the lower number of CREs in the chemosensory upstream regions is consistently supported by two independent sources of functional information (supplementary tables S4 and S5, Supplementary Material online), and that it correlates with selective parameters solely inferred from sequence data, we believe that the conclusion is sound, although claiming for a cautious interpretation.

It could be also thought that the surveyed expression data (representing a single experimental condition, young adult flies) may not be especially suitable for identifying transcriptional changes of chemosensory genes. Indeed, their

expression is often restricted to a few sensilla, leading to an overall intensity signal far below the detectable threshold (Zhang et al. 2014). Nevertheless, we conditioned our statistical test on SNPs that are significantly associated with an expression fold change (fig. 5); i.e., we tested if, among the eQTLs, there is an overrepresentation at extreme or intermediate allele frequencies. This obviously reduces the sample size, but prevents the bias caused by eQTLs that are undetectable at the chemosensory upstream regions.

We detected that many chemosensory-eQTLs regulate chemosensory genes transcribed in reproductive organs, especially OBPs. This is well in agreement with a previous study, which revealed that a large fraction of eQTLs are male-specific, and mainly control the transcription of genes in testes and accessory glands (Massouras et al. 2012). Actually, we found that the only two chemosensory-eQTLs segregating at high allele frequencies are located at the *Ir68a* and *Obp51a* upstream regions, two genes transcribed in male accessory glands. These results are consistent with the fast evolutionary rates reported for sex-related genes in *Drosophila* (Haerty et al. 2007; Larracuenté et al. 2008; Yeh et al. 2012), as well as with the transcriptional and functional diversification observed for the OBP family members (Chintapalli et al. 2007; Librado and Rozas 2013).

#### Evolutionary Rates across the Chemosensory Gene Families

The comparison of the evolutionary rates at different time-scales is a fundamental approach to gain insights into the origin and function of the major chemosensory multigene families (Missbach et al. 2014; Pelosi et al. 2014). Until now, however, such studies have been focused on the chemosensory gene turnover rates ( $\lambda$ ) and on the selective constraint at their protein-coding regions ( $\omega$ ). In agreement with previous studies (Sanchez-Gracia et al. 2009; Almeida et al. 2014), we found two distinctive modes of evolution. On the one hand, the OBP and especially the GR, OR and dIR upstream regions exhibit the highest  $p_{\alpha}$  values (table 2). By their own functional idiosyncrasy, the chemoreceptor upstream regions may provide the ideal substrate for rapid adaptations. For example, the misexpression of chemoreceptors into pre-existing sensory neurons may activate completely different downstream neural circuits, eliciting even the opposite (potentially adaptive) behavior given the same chemical stimuli (Cande et al. 2013). On the other hand, the aIR and CSP upstream regions mainly evolve under a strong negative selection (high  $-N_{es}$ ), with only a few mutations being advantageous (low  $p_{\alpha}$ ; table 2). Indeed, these gene families have an ancient origin (Vieira and Rozas 2011), and likely commit a basal function in sensing chemical cues (Croset et al. 2010), or even in other biological processes (Nomura et al. 1992). Overall, our comparison among chemosensory families recapitulates the genome-wide pattern, whereby biological processes

controlled by a large number of genes are more likely to undergo weak polygenic selection.

### Supplementary Material

Supplementary tables S1–S5 and figures S1–S5 are available at Genome Biology and Evolution online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgments

We thank Marta Morey and Alejandro Sánchez-Gracia for their helpful discussions and comments on the article, and Roman Arguello for sharing his article prior to publication. Work supported by grants from the Ministerio de Economía y Competitividad of Spain (BFU2010-15484 and CGL2013-45211), and from the Comissió Interdepartamental de Recerca i Innovació Tecnològica of Spain (2014SGR1055). J.R. was partially supported by ICREA Academia (Generalitat de Catalunya), whereas P.L. by a FPI predoctoral fellowship (Ministerio de Economía y Competitividad of Spain; BES-2008-002059).

### Literature Cited

- Almeida FC, Sánchez-Gracia A, Campos JL, Rozas J. 2014. Family size evolution in *Drosophila* chemosensory gene families: a comparative analysis with a critical appraisal of methods. *Genome Biol Evol.* 6:1669–1682.
- Andrews JC, et al. 2014. Octopamine Neuromodulation Regulates Gr32a-Linked Aggression and Courtship Pathways in *Drosophila* Males. *PLoS Genet.* 10:e1004356. doi: 10.1371/journal.pgen.1004356.
- Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 437:1149–1152.
- Arguello JR, et al. 2016. Extensive local adaptation within the chemosensory system following *Drosophila melanogaster*'s global expansion. *Nat Commun.* 7:ncomms11855.
- Arya GH, et al. 2015. The genetic basis for variation in olfactory behavior in *Drosophila melanogaster*. *Chem Senses* 40:233–243.
- Ayroles JF, et al. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet.* 41:299–307.
- Barish S, Volkan PC. 2015. Mechanisms of olfactory receptor neuron specification in *Drosophila*. *Wiley Interdiscip Rev Dev Biol.* 4:609–621.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 57:289–300.
- Cande J, Prud'homme B, Gompel N. 2013. Smells like evolution: the role of chemoreceptor evolution in behavioral change. *Curr Opin Neurobiol.* 23:152–158.
- Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134:1289–1303.
- Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet.* 39:715–720.
- Clark AG, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.
- Croset V, et al. 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* 6:e1001064.

- Depetris-Chauvin A, Galagovsky D, Grosjean Y. 2015. Chemicals and chemoreceptors: ecologically relevant signals driving behavior in *Drosophila*. *Chem Ecol*. 3:41.
- Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2012. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* 193:291–301.
- Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Res*. 36:W157–W162.
- Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol*. 26:2097–2108.
- Fan P, et al. 2013. Genetic and Neural Mechanisms that Inhibit *Drosophila* from Mating with Other Species. *Cell*. 154:89–102. doi: 10.1016/j.cell.2013.06.008.
- Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. *Genetics* 158:1227–1234.
- Ferretti L, Raineri E, Ramos-Onsins S. 2012. Neutrality tests for sequences with missing data. *Genetics* 191:1397–1401.
- Fiston-Lavier AS, Singh ND, Lipatov M, Petrov DA. 2010. *Drosophila melanogaster* recombination rate calculator. *Gene* 463:18–20.
- Foret S, Maleszka R. 2006. Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Res*. 16:1404–1413.
- Garud NR, Messer PW, Buzbas EO, Petrov DA. 2015. Recent selective sweeps in North American *Drosophila melanogaster* show signatures of soft sweeps. *PLOS Genet*. 11:e1005004.
- Gentleman RC, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 5:R80.
- Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol*. 4:658–667.
- Haddrill PR, Bachtrog D, Andolfatto P. 2008. Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol Biol Evol*. 25:1825–1834.
- Haerty W, et al. 2007. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* 177:1321–1335.
- He X, Zhou S, St. Armour GE, Mackay TFC, Anholt RRH. 2016. Epistatic partners of neurogenic genes modulate *Drosophila* olfactory behavior. *Genes Brain Behav*. 15:280–290.
- Hildebrand JG, Shepherd GM. 1997. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu Rev Neurosci*. 20:595–631.
- Huang DW, Sherman BT, Lempicki RA. 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4:44–57.
- Huang W, et al. 2012. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci*. 109:15553–15559.
- Hutter S, Vilella AJ, Rozas J. 2006. Genome-wide DNA polymorphism analyses using VariScan. *BMC Bioinformatics* 7:409.
- Jafari S, et al. 2012. Combinatorial activation and repression by seven transcription factors specify *Drosophila* odorant receptor expression. *PLoS Biol*. 10:e1001280.
- Jenkins DL, Ortori CA, Brookfield JF. 1995. A test for adaptive change in DNA sequences controlling transcription. *Proc Biol Sci*. 261:203–207.
- Jensen JD, Bachtrog D. 2011. Characterizing the influence of effective population size on the rate of adaptation: Gillespie's Darwin domain. *Genome Biol Evol*. 3:687–701.
- Joseph RM, Carlson JR. 2015. *Drosophila* chemoreceptors: a molecular interface between the chemical world and the brain. *Trends Genet*. 31:683–695.
- Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177:2251–2261.
- Keightley PD, Lercher MJ, Eyre-Walker A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol*. 3:e42.
- Keller A. 2007. *Drosophila melanogaster*'s history as a human commensal. *Curr Biol*. 17:R77–R81.
- Kharchenko PV, et al. 2011. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471:480–485.
- Kopp A, et al. 2008. Evolution of gene expression in the *Drosophila* olfactory system. *Mol Biol Evol*. 25:1081–1092.
- Krieger MJ, Ross KG. 2002. Identification of a major gene regulating complex social behavior. *Science* 295:328–332.
- Laisue PP, Vossball LB. 2008. The olfactory sensory map in *Drosophila*. In: Technau GM, editor. *Brain development in Drosophila melanogaster*. *Advances in Experimental Medicine and Biology*. New York: Springer. p. 102–114 [cited 2016 May 12]. Available from: [http://link.springer.com/chapter/10.1007/978-0-387-78261-4\\_7](http://link.springer.com/chapter/10.1007/978-0-387-78261-4_7).
- Larracuente AM, et al. 2008. Evolution of protein-coding genes in *Drosophila*. *Trends Genet*. 24:114–123.
- Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013. Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genet*. 9:e1003527.
- Lee Y, Kim SH, Montell C. 2010. Avoiding DEET through Insect Gustatory Receptors. *Neuron*. 67:555–561. doi: 10.1016/j.neuron.2010.07.006.
- Librado P, Rozas J. 2013. Uncovering the functional constraints underlying the genomic organization of the odorant-binding protein genes. *Genome Biol Evol*. 5:2096–2108.
- Linz J, et al. 2013. Host plant-driven sensory specialization in *Drosophila erecta*. *Proc R Soc Lond B Biol Sci*. 280:20130626.
- Mackay TF, et al. 2012. The *Drosophila melanogaster* genetic reference panel. *Nature* 482:173–178.
- Marygold SJ, et al. 2013. FlyBase: improvements to the bibliography. *Nucleic Acids Res*. 41:D751–DD757.
- Massouras A, et al. 2012. Genomic variation and its impact on gene expression in *Drosophila melanogaster*. *PLoS Genet*. 8:e1003055.
- Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. 2007. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol*. 5:e118.
- 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.
- Missbach C, et al. 2014. Evolution of insect olfactory receptors. *eLife* 3:e02115.
- Nielsen R, Bauer DuMont VL, Hubisz MJ, Aquadro CF. 2007. Maximum likelihood estimation of ancestral codon usage bias parameters in *Drosophila*. *Mol Biol Evol*. 24:228–235.
- Nomura A, Kawasaki K, Kubo T, Natori S. 1992. Purification and localization of p10, a novel protein that increases in nymphal regenerating legs of *Periplaneta americana* (American cockroach). *Int J Dev Biol*. 36:391–398.
- Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P. 2010. On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. *Mol Biol Evol*. 27:1226–1234.
- Pavlidis P, Metzler D, Stephan W. 2012. Selective sweeps in multilocus models of quantitative traits. *Genetics* 192:225–239.
- Pelosi P. 1996. Perireceptor events in olfaction. *J Neurobiol*. 30:3–19.
- Pelosi P, Iovinella I, Feliccioli A, Dani FR. 2014. Soluble proteins of chemical communication: an overview across arthropods. *Front Physiol*. 5:1–13.

- Poh YP, Ting CT, Fu HW, Langley CH, Begun DJ. 2012. Population genomic analysis of base composition evolution in *Drosophila melanogaster*. *Genome Biol Evol.* 4:1245–1255.
- Pritchard JK, Pickrell JK, Coop G. 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 20:R208–R215.
- Ramia M, Librado P, Casillas S, Rozas J, Barbadilla A. 2012. PopDrowser: the population *Drosophila* browser. *Bioinformatics* 28:595–596.
- R Core Team. 2015. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from: <http://www.R-project.org/>.
- Rollmann SM, et al. 2010. Odorant receptor polymorphisms and natural variation in olfactory behavior in *Drosophila melanogaster*. *Genetics* 186:687–697.
- Roy S, et al. 2010. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330:1787–1797.
- Sánchez-Gracia A, Rozas J. 2008. Divergent evolution and molecular adaptation in the *Drosophila* odorant-binding protein family: inferences from sequence variation at the OS-E and OS-F genes. *BMC Evol Biol.* 8:323.
- Sánchez-Gracia A, Vieira FG, Rozas J. 2009. Molecular evolution of the major chemosensory gene families in insects. *Heredity* 103:208–216.
- Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD. 2011. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* 189:1427–1437.
- Shiao M-S, et al. 2015. Expression divergence of chemosensory genes between *Drosophila sechellia* and its sibling species and its implications for host shift. *Genome Biol Evol.* 7:2843–2858.
- Smadja C, Butlin RK. 2009. On the scent of speciation: the chemosensory system and its role in premating isolation. *Heredity* 102:77–97.
- Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.* 23:23–35.
- Stephan W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Mol Ecol.* 25:79–88.
- Stephan W, Wiehe THE, Lenz MW. 1992. The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theor Popul Biol.* 41:237–254.
- Stortkuhl KF, Kettler R, Fischer S, Hovemann BT. 2005. An increased receptive field of olfactory receptor Or43a in the antennal lobe of *Drosophila* reduces benzaldehyde-driven avoidance behavior. *Chem Senses* 30:81–87.
- Swarup S, Huang W, Mackay TFC, Anholt RRH. 2013. Analysis of natural variation reveals neurogenetic networks for *Drosophila* olfactory behavior. *Proc Natl Acad Sci U S A.* 110:1017–1022.
- Torres-Oliva M, Almeida FC, Sánchez-Gracia A, Rozas J. 2016. Comparative genomics uncovers unique gene turnover and evolutionary rates in a gene family involved in the detection of insect cuticular pheromones. *Genome Biol Evol.* 8:1734–1747.
- Van Dongen S. 2008. Graph clustering via a discrete uncoupling process. *SIAM J Matrix Anal Appl.* 30:121–141.
- Vieira FG, Rozas J. 2011. 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.* 3:476–490.
- Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. 2005. VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* 21:2791–2793.
- Wang P, Lyman RF, Mackay TF, Anholt RR. 2010. Natural variation in odorant recognition among odorant-binding proteins in *Drosophila melanogaster*. *Genetics* 184:759–767.
- Whiteman NK, Pierce NE. 2008. Delicious poison: genetics of *Drosophila* host plant preference. *Trends Ecol Evol.* 23:473–478.
- Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet.* 13:59–69.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yeh S-D, et al. 2012. Functional evidence that a recently evolved *Drosophila* sperm-specific gene boosts sperm competition. *Proc Natl Acad Sci.* 109:2043–2048.
- Zhang Y, Zheng Y, Li D, Fan Y. 2014. Transcriptomics and identification of the chemoreceptor superfamily of the pupal parasitoid of the oriental fruit fly, *Spalangia endius* Walker (Hymenoptera: Pteromalidae). *Plos One* 9:e87800.

Associate editor: Josefa Gonzalez