

Molecular evolution and population genetics of two *Drosophila mettleri* cytochrome P450 genes involved in host plant utilization

JEREMY M. BONO,* LUCIANO M. MATZKIN,* SERGIO CASTREZANA† and THERESE A. MARKOW*

*Department of Ecology and Evolutionary Biology, University of Arizona, 1041 E. Lowell Street, Tucson, AZ 85721, USA,

†*Drosophila* Tucson Stock Center, University of Arizona, Tucson, AZ 85721, USA

Abstract

Understanding the genetic basis of adaptation is one of the primary goals of evolutionary biology. The evolution of xenobiotic resistance in insects has proven to be an especially suitable arena for studying the genetics of adaptation, and resistant phenotypes are known to result from both coding and regulatory changes. In this study, we examine the evolutionary history and population genetics of two *Drosophila mettleri* cytochrome P450 genes that are putatively involved in the detoxification of alkaloids present in two of its cactus hosts: saguaro (*Carnegiea gigantea*) and senita (*Lophocereus schottii*). Previous studies demonstrated that *Cyp28A1* was highly up-regulated following exposure to rotting senita tissue while *Cyp4D10* was highly up-regulated following exposure to rotting saguaro tissue. Here, we show that a subset of sites in *Cyp28A1* experienced adaptive evolution specifically in the *D. mettleri* lineage. Moreover, neutrality tests in several populations were also consistent with a history of selection on *Cyp28A1*. In contrast, we did not find evidence for positive selection on *Cyp4D10*, although this certainly does not preclude its involvement in host plant use. A surprising result that emerged from our population genetic analyses was the presence of significant genetic differentiation between flies collected from different host plant species (saguaro and senita) at Organ Pipe National Monument, Arizona, USA. This preliminary evidence suggests that *D. mettleri* may have evolved into distinctive host races that specialize on different hosts, a possibility that warrants further investigation.

Keywords: adaptation, cytochrome P450, host plant, positive selection, saguaro cactus, senita cactus

Received 31 December 2007; revision received 14 March 2008; accepted 21 April 2008

Introduction

Understanding the genetic basis of adaptation is one of the primary goals of evolutionary biology. Available evidence suggests that adaptive evolution is driven by changes in both the regulation and structure of genes, although the relative importance of these two mechanisms is an unresolved issue (Hoekstra & Coyne 2007; Wray 2007). One of the difficulties associated with identifying the genetic basis of an adaptation is that it requires not only ecological information on how a particular phenotype is adaptive, but also on the genetic mechanisms, either structural or regulatory, that underlie the phenotype. In many systems, this link between genetics and phenotype is elusive, usually because of

challenges associated with identifying specific genes responsible for a particular phenotype (Vasemagi & Primmer 2005).

The evolution of xenobiotic resistance in insects, which includes resistance to plant allelochemicals and insecticides, has proven to be an excellent system for investigating the genetic basis of adaptation. Insect genes involved in detoxification typically belong to three gene superfamilies: esterases, cytochrome P450 monooxygenases (P450s), and glutathione S-transferases (Despres *et al.* 2007; Li *et al.* 2007). The identification of candidate genes involved in resistance is further facilitated by the fact that detoxification genes are sometimes transcriptionally induced by xenobiotics or other chemicals, thus making the connection between an adaptive phenotype and the genetic mechanisms underlying it more apparent. Indeed, a number of genes involved in xenobiotic resistance have been identified, and resistant phenotypes have been attributed to several genetic

Correspondence: Jeremy M. Bono, Fax: 520 6263522; E-mail: jbono@email.arizona.edu

mechanisms including sequence amplifications, regulatory changes, and mutations in coding regions (Li *et al.* 2007).

Several species of *Drosophila* that are distributed throughout the deserts of southwestern North America and northern Mexico use rotting cacti as their main breeding substrate. The chemical composition of cactus rots includes a number of highly toxic compounds at concentrations that are intolerable to nonresident fly species (Kircher *et al.* 1967; Frank & Fogleman 1992; Fogleman 2000), implying that flies obligately associated with necrotic cactus have evolved mechanisms of resistance. The range of cactus hosts and their unique chemical compositions, coupled with diverse patterns of host specialization among cactophilic flies, makes the Sonoran Desert *Drosophila* system especially suitable for studying the genetic basis of adaptation (Fogleman *et al.* 1998; Fogleman & Danielson 2001). In fact, several recent studies on *Drosophila mojavensis* have implicated a number of detoxification genes in the ability of this species to tolerate the toxic environments of its hosts (Matzkin & Eanes 2003; Matzkin 2004, 2005, 2008; Matzkin *et al.* 2006).

Drosophila mettleri is a cactophilic fly that has an unusual preference for breeding in soil soaked with cactus rot exudates rather than in the rotting tissue itself. This behavioural preference means that, due to water loss through high evaporation rates in the desert, larvae of this species are exposed to concentration of toxic compounds that can be an order of magnitude higher than that experienced by other cactophilic species (Fogleman *et al.* 1982; Meyer & Fogleman 1987). Although *D. mettleri* is tolerant of the toxic compounds present in all the major cactus hosts (Fogleman *et al.* 1998), in nature the main hosts are saguaro (*Carnegiea gigantea*), cardón (*Pachycereus pringlei*), and to a lesser extent senita (*Lophocereus schottii*). Saguaro is present only on the mainland, while cardón and senita are found on the mainland and Baja peninsula. A population on Santa Catalina Island (off the coast of southern California), where none of the mainland columnar hosts occur, uses prickly pear (*Opuntia littoralis*) exclusively. The primary toxins in all the major hosts of *D. mettleri* are isoquinoline alkaloids, although the specific alkaloids and their concentrations vary for the different plant species (Kircher 1982). Most notably, the senita alkaloids lophocerine and pilocereine constitute up to 20% of dry tissue weight (compared to 1–2% for the alkaloids cargenine and gigantine in saguaro tissue), making senita especially toxic (Kircher 1969, 1982).

Several lines of evidence demonstrate the involvement of cytochrome P450s in the ability of *D. mettleri* to detoxify alkaloids present in saguaro and senita (Fogleman *et al.* 1998). First, experiments using the P450 inhibitor piperonyl butoxide (PBO) showed a complete loss of larval viability in substrates with alkaloids and PBO (Frank & Fogleman 1992). Subsequent *in vitro* experiments revealed significant induction of alkaloid detoxification metabolism upon exposure to senita, which was reduced in the presence of

PBO (Frank & Fogleman 1992; Danielson *et al.* 1994). Most convincingly, Danielson *et al.* (1997, 1998) demonstrated that a small number of *D. mettleri* P450s were up-regulated in response to saguaro and senita alkaloids. Of these five genes, three were only moderately induced, one of which was cross-induced by both saguaro and senita alkaloids. In contrast, one gene (*Cyp4D10*) was strongly induced only by saguaro alkaloids (Danielson *et al.* 1998) and another (*Cyp28A1*) was strongly induced by senita but not saguaro alkaloids (Danielson *et al.* 1997). Together, these studies suggest that alkaloid tolerance in *D. mettleri* may result from the collective actions of multiple P450s including a few 'generalist' genes that respond to a variety of compounds, and, perhaps more importantly, 'specialists' that respond only to a specific compound. Danielson *et al.* (1998) argued that these narrowly responsive genes represent the best candidates for direct involvement in the detoxification of cactus compounds.

These studies clearly demonstrate the important role of gene regulation in adaptation to cactus hosts. In this study, we ask whether changes in the coding sequence of the putative 'specialist' P450s (*Cyp4D10* and *Cyp28A1*) may also play a role in cactus host detoxification. Specifically, we examined whether these two genes show signatures of positive selection in their protein coding regions, as might be expected given their apparent specialized role in detoxification of alkaloids from specific host plants, and given that *D. mettleri*'s closest relatives, *D. micromettleri* and *D. eremophila*, use different host plant species.

In order to investigate potential population-specific patterns of selection, we also collected sequence data for both genes from five populations spanning the mainland, Baja peninsula, and Santa Catalina Island. These populations are not only geographically separated, but also differ in the species of host plants that are available, which could have important consequences for patterns of selection on these two putative 'specialist' P450 genes. Most notably, given that saguaro and senita are not available on Santa Catalina Island and that this population is genetically distinct from all other populations (Markow *et al.* 2002; Hurtado *et al.* 2004), it is possible that *Cyp4D10* and *Cyp28A1* have experienced a unique evolutionary history in this location that might be reflected in current patterns of variation. In addition, the absence of saguaro on the Baja peninsula (where cardón is the main host) could result in divergent selection pressures on *Cyp4D10* (the gene induced by saguaro), although this potential might be mitigated by the chemical similarity of saguaro and cardón (Frank & Fogleman 1992), and by substantial gene flow between some mainland and Baja populations (Hurtado *et al.* 2004). Finally, in one population (Organ Pipe National Monument, Arizona, USA), we collected flies directly from senita and saguaro necroses, which allowed us to test for genetic differentiation between individuals collected from different hosts in the same location.

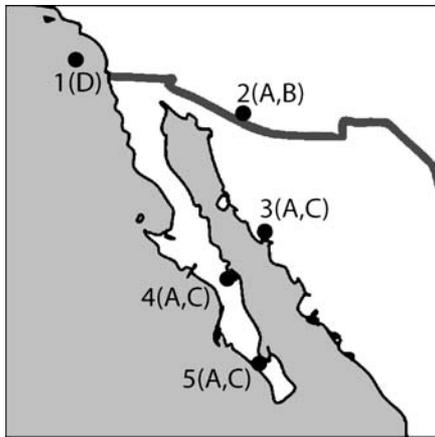


Fig. 1 Map of populations sampled and host plant species present at each location. (1) Santa Catalina Island, (2) Organ Pipe National Monument, (3) Guaymas, (4) Bahia Concepción, (5) Punta Conejo. (A) senita (*Lophocereus schottii*), (B) saguaro (*Carnegiea gigantea*), (C) cardón (*Pachycereus pringlei*), (D) prickly pear (*Opuntia littoralis*).

Materials and methods

Locations sampled

We used *Drosophila mettleri* samples collected from five localities: Guaymas (mainland Sonora, Mexico), Bahia Concepción (Baja peninsula, Mexico), Punta Conejo (Baja peninsula, Mexico), Santa Catalina Island (California, USA), and Organ Pipe National Monument (OPNM, Arizona, USA) (Fig. 1). Samples from the first four localities were a subset of those from an earlier *mtCOI* study (Hurtado *et al.* 2004), but the OPNM samples were not included in the previous study. We collected flies using traps with rotting cardón cactus (Guaymas, Bahia Concepción, Punta Conejo) or fermented bananas (Santa Catalina Island), or, in the case of OPNM, flies were collected directly from cactus necroses. At OPNM, several flies were collected from four different senita necroses located in the area known as Senita Basin, while other flies were collected from a saguaro necrosis approximately 3.5 km away.

DNA extractions, PCR, and sequencing

We used the DNeasy kit (QIAGEN), following the manufacturer's protocols, to extract total DNA from whole flies. Using sequences posted on GenBank, we designed primers that amplified multiple overlapping fragments to obtain nearly complete sequences for *Cyp4D10* and *Cyp28A1* using polymerase chain reaction (PCR). The products were then sequenced with an ABI 3700 analyser. Information on primers used for PCR and sequencing is given in Table S1, Supplementary material, and the numbers of flies sequenced for each population and each gene are reported in Table 1.

Table 1 Number of individuals sequenced from each locality for two genes: *Cyp4D10* and *Cyp28A1*

Locality	<i>Cyp4D10</i>	<i>Cyp28A1</i>
Guaymas	11	11
OPNM saguaro	10	9
OPNM senita	10	10
Bahia Concepción	12	11
Punta Conejo	13	15
Santa Catalina	12	12

We also obtained sequences for one individual from each of two outgroup species, *D. micromettleri* (Drosophila Tucson Stock Center ID 15081-1346.02) and *D. eremophila* (Drosophila Tucson Stock Center ID 15081-1292.00). *D. mettleri*, *D. micromettleri*, and *D. eremophila* form a monophyletic group within the repleta group, but the relationships among these species are somewhat uncertain. Data from Durando *et al.* (2000) suggest that *D. mettleri* and *D. micromettleri* are sister species, with *D. eremophila* as the outgroup. However, the sister relationship between *D. mettleri* and *D. micromettleri* is only moderately well-supported (bootstrap support = 75), and data from our study also produced ambiguous results (data not shown). Whatever the true relationships among these species, it is clear that speciation events were not particularly recent. Averaging the average number of nucleotide substitutions per site (Nei 1987) for the two genes used in our study revealed a considerable amount of divergence between all three species ($D_{xy} = 0.163$ for *D. mettleri* and *D. micromettleri*; $D_{xy} = 0.172$ for *D. mettleri* and *D. eremophila*; and $D_{xy} = 0.157$ for *D. micromettleri* and *D. eremophila*). In order to amplify *Cyp4D10* and *Cyp28A1* in the *D. micromettleri* and *D. eremophila*, we had to place primers within the coding region, thus resulting in a small amount of sequence loss. Again, we used primers that amplified multiple overlapping fragments, some of which were cloned (see Table S1) using the TOPO TA cloning kit (Invitrogen), in order to obtain clean sequence.

By querying *D. mettleri* *Cyp4D10* and *Cyp28A1* sequences against the *D. mojavensis* genome database, we determined the probable chromosomal locations of the two P450 genes. From this, we inferred that *Cyp4D10* is X-linked, which is consistent with the fact that none of the males in our data set were heterozygotes. In contrast, *Cyp28A1* is autosomal, being found on the third chromosome in *D. mojavensis* (Muller element B). We used SEQUENCHER version 4.1 (Gene Codes) to initially align and edit sequences. All *Cyp4D10* sequences were truncated to 2073 bp (entire sequence, 2142 bp), which included two introns (596 bp and 68 bp in length). All *Cyp28A1* sequences were truncated to 1774 bp (entire sequence, 1817 bp), which included five introns (60 bp, 59 bp, 58 bp, 58 bp, 57 bp in length). Truncated sequences for both genes corresponded to the length of

sequences obtained for the outgroups (no polymorphisms in either gene were observed in the regions cut from *D. mettleri*). Because we had heterozygotes in our data set and some of the subsequent analyses required that linkage phase be known (frequency spectrum tests), we used the ELB algorithm (Excoffier *et al.* 2003) implemented in ARLEQUIN version 3.01 (Excoffier *et al.* 2005) to reconstruct haplotypes. We randomly selected one reconstructed haplotype for each individual and used these in all subsequent analyses. While all tests comparing polymorphism and divergence would be robust to errors in haplotype reconstruction, tests based on the frequency spectrum would most likely be sensitive to errors, but it is unclear what sort of bias, if any, this might introduce into these analyses. A file with the reconstructed haplotypes is found in the Supplementary material. All sequences have been deposited in GenBank under Accession nos EU658939–EU659078.

Population structure

We used ARLEQUIN version 3.01 (Excoffier *et al.* 2005) to calculate pairwise F_{ST} for all pairs of populations separately for each locus. Significance was assessed using the permutation procedure available in ARLEQUIN (1000 permutations). We calculated F_{ST} in three ways: (i) with introns only, (ii) with synonymous sites only (in case of selective constraint on introns), and (iii) with nonsynonymous sites only. The rationale behind this was that discrepancies in inferred structure between noncoding/synonymous and nonsynonymous sites could help to determine whether population structure resulted from selection for different protein haplotypes in different populations, or whether population structure resulted from neutral changes. For example, strong structure at nonsynonymous, but not at noncoding or synonymous sites might indicate divergent selection in different populations, while structure at noncoding or synonymous, but not nonsynonymous sites might indicate selective constraint on protein structure.

Neutrality tests based on the frequency spectrum

We used population data to conduct three different tests of selective neutrality for each locus: Tajima's D (Tajima 1989), and Fu and Li's D and F (Fu & Li 1993). These tests analyse the allele frequency spectrum within populations in order to detect departures from expectations under selective neutrality. All tests were conducted with the computer program DNASP version 4.10 (Rozas *et al.* 2003). Given the significant structure we observed between populations in this study and in the Hurtado *et al.* (2004) study, we conducted separate analyses for each population. We used *D. micromettleri* as an outgroup for calculation of Fu and Li's D and F . Significance was assessed using 10 000 coalescent simulations using the Hudson's recombination

parameter (Hudson 1997) estimated in DNASP. Significant tests provide support for departures from a standard neutral model, but are unable to differentiate between the effects of selection and demography. Thus, while significantly negative values may be the result of a recent selective sweep, they may also reflect a recent population expansion following a bottleneck. Likewise, significantly positive values may indicate balancing selection, but could also result from a decrease in population size. The presence of population structure could skew the frequency spectrum in either direction depending on how the subpopulations are sampled. The effects of demographic processes such as bottlenecks are expected to be particularly evident in X-linked genes (*Cyp4D10* in this study), because of the lower effective populations size for such loci (Wall *et al.* 2002).

To further investigate what processes might underlie significant departures from neutral expectations detected by these tests, we performed the three tests separately on synonymous and nonsynonymous sites. The impetus for this approach is that events such as population expansion or a selective sweep should have similar effects on both synonymous and nonsynonymous variation, while a process such as weak purifying selection would be expected to result in a negatively skewed frequency spectrum for nonsynonymous but not synonymous sites (Hahn *et al.* 2002). We used a heterogeneity test (Hahn *et al.* 2002) to formally evaluate whether the frequency spectra of nonsynonymous and synonymous sites differed (using 5000 coalescent simulations).

Neutrality tests based on polymorphism and divergence

We also used our population data to conduct the Hudson–Kreitman–Aguade (HKA) test (Hudson *et al.* 1987), which compares the ratio of intraspecific polymorphism to interspecific divergence for at least two loci. Under neutrality, this ratio is expected to be the same for different genes, and thus, any departures from this pattern can provide evidence for selection. By comparing multiple loci, the HKA test is able to better differentiate between the effects of selection and demography because population-level processes should affect all genes equally, while selection is expected to be locus-specific. We used the HKA program (<http://lifesci.rutgers.edu/~heylib/HeylabSoftware.htm>) to conduct the test separately for *Cyp4D10* and *Cyp28A1*, in both cases using *mtCOI* as the second locus (assumed to be evolving neutrally; but see Ballard & Rand 2005), and *D. micromettleri* as the outgroup. Sequences for *mtCOI* were obtained from the Hurtado *et al.* (2004) data set (666 bp), with the exception of the OPNM and Punta Conejo populations, for which we sequenced 594 bp and 630 bp, respectively (GenBank Accession nos EU659079–659106) using primers reported in Folmer *et al.* (1994). We designated *Cyp4D10* as X-linked and *Cyp28A1* as autosomal in this analysis.

To further evaluate whether the two P450 genes showed evidence of selection, we also conducted the McDonald–Kreitman (MK) test (McDonald & Kreitman 1991), which uses population data from the focal species and at least one outgroup to compare the ratio of nonsynonymous to synonymous polymorphisms within a species to that same ratio for fixed differences between species. An excess of nonsynonymous fixed differences between species provides evidence for positive selection, while an excess of nonsynonymous polymorphism is consistent with balancing selection. This test is less sensitive to demographic assumptions than tests based on the frequency spectrum (Nielsen 2005), although the presence of slightly deleterious mutations in an expanding population can produce a pattern that is similar to that seen for positive selection (Eyre-Walker 2002). Because few stocks of *D. micrometleri* and *D. eremophila* were available, we were unable to gather polymorphism data for either outgroup species. We thus conducted a lineage-specific test using both outgroup sequences to identify changes that occurred specifically in the *D. mettleri* lineage. Given this strategy, uncertainties in the relationships among the focal species would only affect our analysis if *D. mettleri* is, in fact, the outgroup to *D. micrometleri* and *D. eremophila*. Considering this possibility, we also report the results of the standard MK tests using *D. mettleri* and each of the two other species as separate tests, which would not rely on the tree topology being correct.

Branch-site test of positive selection

We also tested for evidence of positive selection using a comparative approach based on comparing the rate of nonsynonymous substitutions to the rate of synonymous substitutions (dn/ds or ω). Specifically, we conducted the improved branch-site test for positive selection (Zhang *et al.* 2005) using the CODEML program available in the Phylogenetic Analysis by Maximum Likelihood (PAML) package, version 3.14 (Yang 1997). This analysis tests the a priori hypothesis that some sites in a designated lineage of a multispecies alignment have been under positive selection (i.e. $\omega > 1$). In our analysis, we tested the a priori

hypothesis that some sites in the *D. mettleri* lineage had experienced positive selection using branch-site model A as the alternative hypothesis and branch-site model A with $\omega_2 = 1$ as the null hypothesis.

We used the Clustal platform in the computer program BIOEDIT to generate multispecies alignments for each gene including sequences from as many *Drosophila* species with full genome sequence (*Drosophila* 12 Genome Consortium 2007) as we could confidently align. For *Cyp4D10*, this included sequences from *D. mettleri*, *D. micrometleri*, *D. eremophila*, *D. mojavensis*, *D. virilis* and *D. grimshawi* (sequences from other species were difficult to align and therefore excluded). For *D. mojavensis*, our BLAST produced hits to two genes, so both were included in the alignment (phylogenetic analysis indicated that these two genes were most closely related to each other and thus probably represent a duplication event specific to the *D. mojavensis* lineage). For *Cyp28A1*, our alignment included: *D. mettleri*, *D. micrometleri*, *D. eremophila*, *D. mojavensis*, *D. virilis*, *D. grimshawi* (two apparent paralogs), *D. willistoni*, *D. pseudoobscura*, *D. erecta*, *D. yakuba*, *D. sechellia*, *D. simulans*, and *D. melanogaster*. In order to account for the uncertainty in the phylogenetic relationships among *D. mettleri*, *D. micrometleri*, and *D. eremophila*, we performed the analysis with this group coded as a tritomy and also separately with each of the three alternative groupings. All analyses were conducted using the ‘clean data’ setting, which excludes from the analysis sites with gaps in the alignment.

Results

Across all populations, 33 haplotypes (including coding and noncoding regions) representing nine distinct protein haplotypes were found for *Cyp4D10*, while only 23 haplotypes (including coding and noncoding) representing nine distinct protein haplotypes were found for *Cyp28A1*. In nearly all populations, the number of segregating sites (S), nucleotide diversity (π) and θ (per site), were substantially higher for *Cyp4D10* than for *Cyp28A1*, indicating greater within population variation at this locus (Table 2). Separate estimates of S , π , θ for nonsynonymous, synonymous,

Population	Segregating sites (S)		Nucleotide diversity (π)		θ (per site)	
	<i>Cyp4D10</i>	<i>Cyp28A1</i>	<i>Cyp4D10</i>	<i>Cyp28A1</i>	<i>Cyp4D10</i>	<i>Cyp28A1</i>
Guaymas	47	22	0.0088	0.0026	0.0105	0.0042
OPNM all	26	8	0.0065	0.0018	0.0047	0.0013
OPNM saguaro	18	5	0.0025	0.0010	0.0041	0.0010
OPNM senita	0	4	0	0.0008	0	0.0008
Bahia Concepción	40	19	0.0084	0.0025	0.0085	0.0037
Punta Conejo	36	5	0.0052	0.0013	0.0076	0.0009
Catalina	0	1	0	0.0001	0	0.0002

Table 2 Number of segregating sites (S), Nucleotide diversity (π), and θ (per site) for *Cyp4D10* and *Cyp28A1*. Estimates of π and θ for *Cyp4D10* were obtained by multiplying by 4/3 to account for the fact that this gene is X-linked

Table 3 Pairwise F_{ST} for all pairs of populations using either noncoding or nonsynonymous sites from *Cyp4D10*. Significant differences ($P < 0.05$) are indicated in bold

	Guaymas	OPNM saguaro	OPNM senita	Bahia Concepción	Punta Conejo
Guaymas					
OPNM saguaro	0.32/0.37				
OPNM senita	0.29/0.89	0.67/0.98			
Bahia Concepción	-0.05/-0.03	0.36/0.54	0.33/0.92		
Punta Conejo	0.08/0.08	0.42/0.82	0.41/0.98	0.13/0.04	
Catalina	0.91/0.69	0.98/0.95	1.00/1.00	0.92/0.76	0.88/0.93

Table 4 Pairwise F_{ST} for all pairs of populations using either noncoding or nonsynonymous sites from *Cyp28A1*. Significant differences ($P < 0.05$) are indicated in bold

	Guaymas	OPNM saguaro	OPNM senita	Bahia Concepción	Punta Conejo
Guaymas					
OPNM saguaro	0.04/0.14				
OPNM senita	0.53/0.20	0.57/0.63			
Bahia Concepción	0.01/0.03	0.03/0.10	0.36/0.15		
Punta Conejo	0.05/-0.02	0.20/0.19	0.70/0.16	0.08/-0.01	
Catalina	0.60/0.23	0.71/0.66	0.88/0.00	0.51/0.14	0.80/0.19

and intronic sites are presented in Tables S2 and S3, Supplementary material.

Population structure

The most striking result from the population structure analysis was the significant genetic subdivision between OPNM flies collected from different hosts (senita and saguaro), which was evident in coding and noncoding portions of both loci (Tables 3 and 4). The flies collected from senita all had a divergent haplotype at *Cyp4D10* that was distinguished by the presence of three unique substitutions, including a 3-bp indel in the coding region and another nonsynonymous substitution. Moreover, we examined both reconstructed haplotypes for each senita fly at *Cyp28A1* and found that all shared at least one copy of a unique haplotype at this locus, although the protein haplotype was not unique. Three individuals were heterozygous with the second allele being found in the flies collected from saguaro and also in other populations. In contrast, flies collected from saguaro had only one unique synonymous substitution at *Cyp4D10*, and no unique substitutions at *Cyp28A1*. To further investigate the apparent host-associated genetic differentiation in this population, we also computed F_{ST} using *mtCOI*. The results from this locus provided further evidence for genetic subdivision between flies collected from the different hosts ($F_{ST} = 0.88$; $P < 0.05$). Indeed, all flies collected from senita shared the same haplotype, which was rare in the Hurtado *et al.* (2004) data set. In contrast, eight out of the nine individuals collected

from saguaro shared the main *D. mettleri* haplotype from the Hurtado *et al.* (2004) study, while one individual had the same haplotype as those collected from senita.

Pairwise F_{ST} estimates using noncoding regions and synonymous sites of both genes indicated significant genetic structure among nearly all populations, with the exception of the Bahia Concepción/Guaymas population pair and the Punta Conejo/Guaymas pair, which were not significantly differentiated at either locus, and the OPNM saguaro/Guaymas and OPNM saguaro/Bahia Concepción pairs, which were not differentiated at *Cyp28A1* (Tables 3 and 4; data for synonymous sites were not shown because results were similar to noncoding). For *Cyp4D10*, F_{ST} estimates for nonsynonymous sites revealed a similar pattern, with nearly all estimates being significant. In contrast, for *Cyp28A1*, five of the 11 population pairs that were significantly differentiated at noncoding sites were not differentiated at nonsynonymous sites (Table 4).

Neutrality tests based on the frequency spectrum

Significant departures from expectations under neutrality were detected for several populations for both *Cyp4D10* and *Cyp28A1* (Table 5). In nearly all cases, where significant deviations were found, the values were negative, indicating an excess of rare alleles in the population. Heterogeneity tests on differences between the frequency spectra of nonsynonymous and synonymous site classes revealed only a few cases where these differed (*Cyp4D10* Bahia Concepción Tajima's D and Fu and Li's F ; *Cyp4D10* Organ

Table 5 Results of Tajima's *D*, Fu and Li's *D* and *F*, and HKA tests on *Cyp4D10* and *Cyp28A1* for the five populations used in this study. HKA tests used *mtCOI* as the second locus

	Tajima's <i>D</i>		Fu and Li's <i>D</i>		Fu and Li's <i>F</i>		HKA (X^2)	
	<i>Cyp4D10</i>	<i>Cyp28A1</i>	<i>Cyp4D10</i>	<i>Cyp28A1</i>	<i>Cyp4D10</i>	<i>Cyp28A1</i>	<i>Cyp4D10</i>	<i>Cyp28A1</i>
Guaymas	-0.78**	-1.76**	-0.64	-2.89**	-0.82*	-3.08**	0.01	6.35**
OPNM (all)	1.51**	1.30	0.14	1.32	0.81	1.58*	1.47	0.39
OPNM saguaro	-1.83**	-0.14	-0.42	0.42	-0.97	0.31	0.72	—
OPNM senita	—	0.02	—	1.15	—	1.03	—	—
Bahia Concepción	-0.06	-1.40**	0.09	-2.11**	0.08	-2.38**	0.02	2.53*
Punta Conejo	-1.37*	1.47	-1.77**	1.18	-1.96**	1.44	0.36	6.12**
Catalina	—	-1.14	—	-1.42	—	-1.57	—	—

* $P < 0.10$; ** $P < 0.05$.

—, polymorphism data insufficient to conduct test.

Pipe all samples, Fu and Li's *F*; Table S4, Supplementary material). In the case of the Bahia Concepción population, the detected heterogeneity resulted from an excess of low frequency variants at nonsynonymous sites, which is consistent with weak purifying selection leading to mildly deleterious mutations segregating in the population. This was not, however, a general pattern for *Cyp4D10* across all populations, which makes it difficult to generalize about the possible effects of purifying selection on this gene overall. We do note, however, that because *Cyp4D10* is sex-linked, and thus the effective population size is smaller, a higher frequency of slightly deleterious mutations segregating in populations is expected relative to autosomal genes if weak purifying selection is occurring. The frequency spectrum observed for the OPNM population as a whole (including flies collected from both saguaro and senita) was unique, since it was the only population that contained positive values for all statistics at both loci, some of which were significant (Table 5). This likely resulted from the population subdivision between flies collected from the two host species. We therefore report values for this population as a whole and also for flies collected from each host species separately, although some tests were not possible due to a lack of polymorphism.

Neutrality tests based on polymorphism and divergence

None of the HKA tests with *Cyp4D10* detected any significant departures from the neutral expectation (Table 5; HKA table shown in Table S5, Supplementary material). In contrast, for *Cyp28A1*, significant departures from neutrality were detected in all populations for which the test could be conducted with the exception of the OPNM population as a whole (Table 5; HKA table, Table S5). In this analysis, we assumed that *mtCOI* was evolving neutrally; any violations of this assumption could have implications for the interpretation of our results. In fact, data from Hurtado *et al.*

(2004) indicated that *mtCOI* was not evolving neutrally in a few *D. mettleri* populations, as some values of Tajima's *D* were significantly negative. A negative Tajima's *D* could be the result of a number of underlying processes, each having different implications for the interpretation of our HKA results. First, a negative Tajima's *D* could result from a recent population expansion. Because demographic processes are expected to affect all loci in a similar manner, this is unlikely to have any significant consequences for the interpretation of our results. Alternatively, a negative Tajima's *D* could indicate a history of positive selection on *mtCOI*. If true, this would make our HKA test conservative. Most problematic would be the possibility that the negative values are the result of mildly deleterious alleles segregating in these populations, something that has been shown previously for mitochondrial genes (Ballard & Rand 2005). To further investigate this possibility, we reanalysed the Hurtado *et al.* (2004) *mtCOI* data set, this time partitioning nonsynonymous and synonymous variation and performing the heterogeneity test described previously (Hahn *et al.* 2002). Although this analysis was not possible for all populations due to a lack of polymorphism in one of the site classes, the results for three populations (Guaymas, Bahia Concepción, and OPNM as a whole) gave no indication of excess nonsynonymous polymorphism (data not shown). We also note that similar results were obtained with an HKA test comparing *mtCOI*, *Cyp4D10*, and *Cyp28A1* (data not shown). None of the lineage-specific or standard MK tests were significant for either gene (Table S6, Supplementary material; data for standard tests not shown).

Branch-site tests of positive selection

The branch-site test for *Cyp28A1* indicated that some sites in the *D. mettleri* branch have experienced positive selection (Table 6). The Bayes empirical Bayes (BEB) follow-up analysis identified two sites that had high probabilities of

Table 6 Log-likelihood values and log ratio tests (LRT) for the branch-site test of positive selection, with *Drosophila mettleri* designated as the foreground branch

Gene	Model	Log likelihood	Parameter estimates	LRT	Positively selected sites†
<i>Cyp4D10</i>	Branch-site model A, $\omega_2 = 1$ fixed	-5949.80	$p_0 = 0.80, p_1 = 0.16,$ $\omega_0 = 0.06, \omega_1 = \omega_2 = 1$		
	Branch-site model A, $\omega_2 > 1$	-5949.55	$p_0 = 0.82, p_1 = 0.16,$ $\omega_0 = 0.06, \omega_2 = 3.37$	0.49	None
<i>Cyp28A1</i>	Branch-site model A, $\omega_2 = 1$ fixed	-9523.28	$p_0 = 0.82, p_1 = 0.12,$ $\omega_0 = 0.08, \omega_1 = \omega_2 = 1$		
	Branch-site model A, $\omega_2 > 1$	-9521.28	$p_0 = 0.86, p_1 = 0.13,$ $\omega_0 = 0.08, \omega_2 = 6.26$	3.99*	141Q, 479G

* $P < 0.05$.

†from the Bayes empirical Bayes (BEB) analysis.

belonging to the positively selected site class (residue 141, $P = 0.99$; residue 479, $P = 0.99$; Table 6). These two amino acid substitutions were unique to *D. mettleri*, with the first involving a change from asparagine to glutamine, and the second from threonine to glycine. Furthermore, both sites were fixed in all *D. mettleri* populations. Although we report only the results of the analysis that coded the relationships among *D. mettleri*, *D. micromettleri*, and *D. eremophila* as a tritomy, the other analyses that considered all possible relationships among these species produced similar results with the same two sites being significant (data not shown). In contrast, the branch-site test for *Cyp4D10* was not significant for any of the possible groupings (Table 6).

Discussion

In this study, we demonstrate that the *Drosophila mettleri* gene *Cyp28A1*, which previously has been shown to be up-regulated in response to host plant (senita) alkaloids (Danielson *et al.* 1997), also has been under positive selection in the *D. mettleri* lineage. Our results are thus consistent with the growing body of evidence that both coding and regulatory changes play pivotal roles in adaptive evolution (Hoekstra & Coyne 2007; Wray 2007). Although our analysis identified only a small number of sites in this gene that have been under positive selection, the amino acid change at one of these (threonine to glycine) involves the substitution of a nonpolar for a polar amino acid, which could alter the three-dimensional structure of the protein. This, coupled with the fact that as little as a single amino acid change can drastically alter the substrate specificity of P450s (Von Wachenfeldt & Johnson 1995), suggests that these positively selected changes could have important consequences for protein functionality. Moreover, other nonsynonymous substitutions that have occurred in the *D. mettleri* lineage since the divergence from *D. micromettleri*/*D. eremophila* (18 total) also may be functionally important,

even if they were not specifically identified as being positively selected.

Several lines of evidence from our population genetic data also suggest a history of selection on *Cyp28A1*. First, we found significantly negative values for Tajima's D , and Fu and Li's D and F in two populations (Bahia Concepción and Guaymas). While an excess of low frequency variants would be expected for a relatively short period of time following a selective sweep, differentiating between this and possible alternative explanations such as population expansion or weak purifying selection, is challenging. Nevertheless, the fact that the excess polymorphism was not found exclusively at nonsynonymous sites is inconsistent with weak purifying selection being the primary explanation. Moreover, although it is difficult to disentangle the effects of selection and demography, examining multiple loci can aid in this process because the demographic history of a population should have a similar effect on all loci, while the effects of selection should be locus-specific. Given that some neutrality tests using *Cyp4D10*, and *mtCOI* (Hurtado *et al.* 2004), also indicated a significantly skewed frequency spectrum toward low frequency variants, it seems reasonable to conclude that demographic processes have at least partially shaped the pattern of allele frequency variation in *D. mettleri* populations, although this could also result from pervasive positive selection and/or hitchhiking throughout the genome (Hahn 2008). Despite this, HKA test results were still consistent with a scenario of selection on *Cyp28A1* (or a gene linked to it) in all but one population. In contrast, all MK tests, which would provide the most direct evidence for selection, were nonsignificant. One possible explanation for this discrepancy is that the MK test would have little power to detect positive selection when the number of fixations, both synonymous and nonsynonymous, is high (e.g. due to a long divergence time), especially if only a small number of nonsynonymous substitutions provided the selective advantage.

Taken together, the results of the battery of tests for selection we performed suggest a history of positive selection on *Cyp28A1*. The strongest support for this inference comes from tests that have more power to detect selection episodes that occurred over longer timescales (branch-site test and HKA test), while tests aimed at detecting more recent selection provided either modest support or no support at all for a scenario of selection on this gene (frequency spectrum statistics and MK tests). This, coupled with the fact that the two amino acid sites that were identified as being positively selected are fixed in all populations, leads us to conclude that the signatures of positive selection that we detected are probably the result of events that occurred in the distant past.

Subsequent to apparent historical positive selection, it does not appear that *Cyp28A1* has been under divergent selection pressures in different populations. Indeed, while analysis of noncoding regions (introns) and synonymous sites revealed strong population structure between most population pairs, this structure was reduced or disappeared altogether when only amino acid coding sites were considered. This result is not particularly surprising, given that senita (the host associated with *Cyp28A1*) is present on both the mainland and Baja peninsula. Interestingly, flies from Santa Catalina Island, where senita is not available, were still significantly genetically differentiated from almost all other populations, even when only amino acid changing sites were considered. This structure does not, however, result from nonsynonymous fixed differences between the Santa Catalina Island population and the others, but rather from the fact that several nonsynonymous sites that are polymorphic in the other populations are fixed in the Santa Catalina Island population. Given this, it does not appear that the observed pattern of structure reflects the action of divergent selection between these populations, or relaxed constraint on Santa Catalina Island.

In contrast to an apparent history of selection on *Cyp28A1*, we found no strong evidence for positive selection on *Cyp4D10* using either comparative or population genetic approaches. Although some tests based on the frequency spectrum produced patterns consistent with positive selection, subsequent analyses did not allow us to rule out demographic effects as the cause for violations of neutral expectations. This, of course, in no way precludes an important role for this gene in the detoxification of host plant compounds, especially considering the fact that the gene was up-regulated in response to saguaro alkaloids in an earlier experiment (Danielson *et al.* 1998). Furthermore, although *Cyp4D10* did not show a signature of adaptive amino acid evolution, eight amino acid substitutions have occurred and become fixed in the *D. mettleri* lineage, any of which could be important functionally.

A surprising result from this study was the significant level of genetic differentiation between OPNM flies collected from two different host plant species (saguaro and senita).

Host-switching and subsequent specialization has long been identified as one of the primary mechanisms for generating the high levels of diversity observed in plant-feeding arthropods (Bernays & Graham 1988; Mitter *et al.* 1988; Berlocher & Feder 2002; Funk *et al.* 2002). Until now, it has been assumed that *D. mettleri* is a generalist capable of using several different host species. Our results, however, suggest that flies breeding in saguaro and senita may actually represent genetically differentiated host races. Our inference is limited to some degree by small sample sizes, but even so, the fact that flies collected from senita, in addition to being genetically differentiated from flies collected from saguaro, also had a unique *Cyp4D10* protein haplotype that was quite divergent, suggests that our inference is probably not a sampling artefact.

The divergent *Cyp4D10* haplotype (induced by saguaro) is particularly interesting because it could be the result of relaxed selective constraint or selection for a novel function if, in fact, these flies have become specialized on senita. Without data from more locations, we cannot say whether the pattern of genetic differentiation we observed is a general one across the entire range of *D. mettleri*, or whether the results from OPNM are an anomaly. More extensive geographical sampling along with genetic, behavioural, and ecological studies on *D. mettleri* collected from different host species are necessary in order to fully evaluate the possible association between host plant use and genetic differentiation.

Overall patterns of genetic differentiation among populations at the two P450 loci largely corroborate the findings reported by Hurtado *et al.* (2004) for *mtCOI*. Most notably, in that study, and in an earlier allozyme study (Markow *et al.* 2002), the population on Santa Catalina Island was highly differentiated from all other populations. Our data for both coding and noncoding regions of *Cyp4D10* and *Cyp28A1* show this same pattern, which is not unexpected, given that this is an island population and suitable cactus hosts are not present on neighbouring stretches of the mainland. Nevertheless, mating experiments between *D. mettleri* from Santa Catalina Island and flies from various Baja and mainland populations, however, show little, if any, sexual isolation (Markow *et al.* 2002). This situation is paralleled in *D. mojavensis*, another cactophilic species, where a highly genetically differentiated host race on Santa Catalina Island is not sexually isolated from other populations of this species (Markow *et al.* 1983; Machado *et al.* 2007; Reed *et al.* 2007).

The results of our study, together with previous work on *D. mettleri* (Frank & Fogleman 1992; Danielson *et al.* 1994, 1995, 1997, 1998) and more recent work on *D. mojavensis* (Matzkin & Eanes 2003; Matzkin 2004, 2005, 2008; Matzkin *et al.* 2006), suggest a prominent role for both regulatory and coding sequence changes in the adaptation of Sonoran Desert *Drosophila* to their cactus hosts. Differences in host

specialization among these, and other, cactophilic flies indicate that the Sonoran Desert *Drosophila* system holds great promise as a system for investigating the genetic basis of adaptation.

Acknowledgements

We thank members of the Markow laboratory for useful discussions and four anonymous reviewers for constructive comments on the manuscript. Funding for this study came from an NIH Postdoctoral Excellence in Research and Teaching Fellowship from the Center for Insect Science awarded to J.M.B.

References

- Ballard JWO, Rand DM (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 621–642.
- Berlacher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: moving beyond controversy. *Annual Review of Entomology*, **47**, 773–815.
- Bernays EA, Graham M (1988) On the evolution of host specificity in phytophagous arthropods. *Ecology*, **69**, 886–892.
- Danielson PB, Foster JLM, McMahlil MM, Smith MK, Fogleman JC (1998) Induction by alkaloids and phenobarbital of Family 4 Cytochrome P450s in *Drosophila*: evidence for involvement in host plant utilization. *Molecular and General Genetics*, **259**, 54–59.
- Danielson PB, Frank MR, Fogleman JC (1994) Comparison of larval and adult P-450 activity levels for alkaloid metabolism in desert *Drosophila*. *Journal of Chemical Ecology*, **20**, 1893–1906.
- Danielson PB, Macintyre RJ, Fogleman JC (1995) Isolation of novel cytochrome-P450 genes from alkaloid-tolerant cactophilic *Drosophila* and their phylogenetic relatives. *Journal of Cellular Biochemistry*, **196**–196.
- Danielson PB, MacIntyre RJ, Fogleman JC (1997) Molecular cloning of a family of xenobiotic-inducible drosophilid cytochrome P450s: evidence for involvement in host-plant allelochemical resistance. *Proceedings of the National Academy of Sciences, USA*, **94**, 10797–10802.
- Despres L, David JP, Gallet C (2007) The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution*, **22**, 298–307.
- Drosophila* 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, **450**, 203–218.
- Durando CM, Baker RH, Etges WJ *et al.* (2000) Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. *Molecular Phylogenetics and Evolution*, **16**, 296–307.
- Excoffier L, Laval G, Balding D (2003) Gametic phase estimation over large genomic regions using an adaptive window approach. *Human Genomics*, **1**, 7–19.
- Excoffier L, Laval G, Schneider S (2005) ARLEQUIN version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Eyre-Walker A (2002) Changing effective population size and the McDonald–Kreitman test. *Genetics*, **162**, 2017–2024.
- Fogleman JC (2000) Response of *Drosophila melanogaster* to selection for P450-mediated resistance to isoquinoline alkaloids. *Chemico-Biological Interactions*, **125**, 93–105.
- Fogleman JC, Danielson PB (2001) Chemical interactions in the cactus-microorganism-*Drosophila* model system of the Sonoran Desert. *American Zoologist*, **41**, 877–889.
- Fogleman JC, Heed WB, Kircher HW (1982) *Drosophila mettleri* and senita cactus alkaloids: fitness measurements and their ecological significance. *Comparative Biochemistry and Physiology*, **71**, 413–417.
- Fogleman JC, Danielson PB, MacIntyre RJ (1998) The molecular basis of adaptation in *Drosophila*: the role of cytochrome P450s. *Evolutionary Biology*, **30**, 15–77.
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.
- Frank MR, Fogleman JC (1992) Involvement of cytochrome-P450 in host-plant utilization by Sonoran Desert *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **89**, 11998–12002.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.
- Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica*, **116**, 251–267.
- Hahn MW (2008) Toward a selection theory of molecular evolution. *Evolution*, **62**, 255–265.
- Hahn MW, Rausher MD, Cunningham CW (2002) Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics*, **161**, 11–20.
- Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**, 995–1016.
- Hudson RR (1997) Estimating the recombination parameter of a finite population model without selection. *Genetic Research*, **5**, 245–250.
- Hudson RR, Kreitman M, Montserrat A (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics*, **116**, 153–159.
- Hurtado LA, Erez T, Castrezana S, Markow TA (2004) Contrasting population genetic patterns and evolutionary histories among sympatric Sonoran Desert cactophilic *Drosophila*. *Molecular Ecology*, **13**, 1365–1375.
- Kircher HW (1969) The distribution of sterols, alkaloids and fatty acids in senita cactus, *Lophocereus schottii*, over its range in Sonora, Mexico. *Phytochemistry*, **8**, 1481–1488.
- Kircher HW (1982) Chemical composition of cacti and its relationship to Sonoran Desert *Drosophila*. In: *Ecological Genetics and Evolution: the Cactus-Yeast-Drosophila System* (eds Barker JSF, Starmer WT), pp. 143–158. Academic Press, Sydney.
- Kircher HW, Heed WB, Russell JS, Grove J (1967) Senita cactus alkaloids: their significance to Sonoran Desert *Drosophila* ecology. *Journal of Insect Physiology*, **13**, 1869–1874.
- Li XC, Schuler MA, Berenbaum MR (2007) Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annual Review of Entomology*, **52**, 231–253.
- Machado CA, Matzkin LM, Reed LK, Markow TA (2007) Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. *Molecular Ecology*, **16**, 3009–2024.
- Markow TA, Castrezana S, Pfeiler E (2002) Flies across the water: genetic differentiation and reproductive isolation in allopatric desert *Drosophila*. *Evolution*, **56**, 546–552.
- Markow TA, Fogleman JC, Heed WB (1983) Reproductive isolation in Sonoran Desert *Drosophila*. *Evolution*, **37**, 649–652.
- Matzkin LM (2004) Population genetics and geographic variation of alcohol dehydrogenase (*Adh*) paralogs and glucose-6-phosphate dehydrogenase (*G6pd*) in *Drosophila mojavensis*. *Molecular Biology and Evolution*, **21**, 276–285.
- Matzkin LM (2005) Activity variation in alcohol dehydrogenase paralogs is associated with adaptation to cactus host use in cactophilic *Drosophila*. *Molecular Ecology*, **14**, 2223–2231.

- Matzkin LM (2008) Molecular basis of host adaptation in cactophilic *Drosophila*: molecular evolution of a glutathione S-transferase (*GstD1*) gene in *Drosophila mojavensis*. *Genetics*, **178**, 1073–1083.
- Matzkin LM, Eanes WF (2003) Sequence variation of alcohol dehydrogenase (*Adh*) paralogs in cactophilic *Drosophila*. *Genetics*, **163**, 181–194.
- Matzkin LM, Watts TD, Bitler BG, Machado CA, Markow TA (2006) Functional genomics of cactus host shifts in *Drosophila mojavensis*. *Molecular Ecology*, **15**, 4635–4643.
- McDonald JM, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, **351**, 652–654.
- Meyer JM, Fogleman JC (1987) Significance of saguaro cactus alkaloids in ecology of *Drosophila mettleri*, a soil-breeding, cactophilic drosophilid. *Journal of Chemical Ecology*, **13**, 2069–2081.
- Mitter C, Farrel B, Wiegmann B (1988) The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist*, **138**, 107–128.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Reed LK, Nyboer M, Markow TA (2007) Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Molecular Ecology*, **16**, 1007–1022.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Vasemagi A, Primmer CR (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology*, **14**, 3623–3642.
- Von Wachenfeldt C, Johnson EF (1995) Structures of eukaryotic cytochrome P450 enzymes. In: *Cytochrome P450* (ed. Ortiz de Motellano PR), pp. 183–223. Plenum, New York.
- Wall JD, Andolfatto P, Przeworski M (2002) Testing models of selection and demography in *Drosophila simulans*. *Genetics*, **162**, 203–216.
- Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics*, **8**, 206–216.
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in Biosciences*, **13**, 555–556.
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**, 2472–2479.

Jeremy M. Bono is a PERT postdoctoral fellow whose research focuses on the study of adaptation and how the process leads to diversification within and between species. Luciano M. Matzkin is a postdoctoral researcher interested in the genetic basis of adaptation, focusing on how the ecology of organisms can shape transcriptional and molecular variation. Sergio Castrezana is the curator and insect specialist at the National *Drosophila* stock center. He is interested in arthropod behavioural adaptations to desert ecology. Therese A. Markow is a Regents Professor who has worked extensively on the ecology and evolution of cactophilic *Drosophila*.

Supplementary material

The following supplementary material is available for this article:

Table S1 Sequences for primer combinations used to amplify and sequence *Cyp4D10* and *Cyp28A1* from *Drosophila mettleri*, *D. micromettleri*, and *D. eremophila*

Table S2 Number of segregating sites, nucleotide diversity (π), θ (per site) for nonsynonymous, synonymous, and intronic sites of *Cyp4D10*

Table S3 Number of segregating sites, nucleotide diversity (π), θ (per site) for nonsynonymous, synonymous, and intronic sites of *Cyp28A1*

Table S4 Results of Tajima's *D*, and Fu and Li's *D* and *F*, using nonsynonymous and synonymous sites of *Cyp4D10* and *Cyp28A1*

Table S5 Polymorphism and divergence data used as input for HKA tests

Table S6 Results of lineage-specific McDonald–Kreitman tests on *Cyp4D10* and *Cyp28A1*

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03823.x> (This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.