A Locus in *Drosophila sechellia* Affecting Tolerance of a Host Plant Toxin

Eric A. Hungate,* Eric J. Earley,† Ian A. Boussy,‡ David A. Turissini,* Chau-Ti Ting,§ Jennifer R. Moran,* Mao-Lien Wu,* Chung-I Wu,* and Corbin D. Jones†,[1]

*Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, †Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, ‡Department of Biology, Loyola University, Chicago, Illinois 60660, and § Department of Life Science, National Taiwan University, Taipei, 10617 Taiwan

ABSTRACT Many insects feed on only one or a few types of host. These host specialists often evolve a preference for chemical cues emanating from their host and develop mechanisms for circumventing their host's defenses. Adaptations like these are central to evolutionary biology, yet our understanding of their genetics remains incomplete. *Drosophila sechellia*, an emerging model for the genetics of host specialization, is an island endemic that has adapted to chemical toxins present in the fruit of its host plant, *Morinda citrifolia*. Its sibling species, *D. simulans*, and many other *Drosophila* species do not tolerate these toxins and avoid the fruit. Earlier work found a region with a strong effect on tolerance to the major toxin, octanoic acid, on chromosome arm *3*R. Using a novel assay, we narrowed this region to a small span near the centromere containing 18 genes, including three odorant binding proteins. It has been hypothesized that the evolution of host specialization is facilitated by genetic linkage between alleles contributing to host preference and alleles contributing to host usage, such as tolerance to secondary compounds. We tested this hypothesis by measuring the effect of this tolerance locus on host preference behavior. Our data were inconsistent with the linkage hypothesis, as flies bearing this tolerance region showed no increase in preference for media containing *M. citrifolia* toxins, which *D. sechellia* prefers. Thus, in contrast to some models for host preference, preference and tolerance are not tightly linked at this locus nor is increased tolerance *per se* sufficient to change preference. Our data are consistent with the previously proposed model that the evolution of *D. sechellia* as a *M. citrifolia* specialist occurred through a stepwise loss of aversion and gain of tolerance to *M. citrifolia*'s toxins.

HALF of all insects interact with plants (Grimaldi and Engel 2005). Most phytophageous insects, however, use only a few plant genera for food, mating, and oviposition (Bernays and Chapman 1994). Changes in host use can result in both new species and new adaptations (Ehrlich and Raven 1964; Janz 2011). For example, the evolution of a new host specialization may have contributed to the formation of new species in pea aphids (*Acyrthosiphon*) among others (Via 2001; Matsubayashi *et al.* 2010). Adapting to a new host can drive genetic and phenotypic change that is critical for isolating nascent species. In some cases, specialization has a price: Increased performance on the new host correlates with reduced performance on other hosts (Futuyma and Moreno 1988; Jaenike 1990; Fry *et al.* 1996; Scheirs *et al.* 2005; Via and Hawthorne 2005). This scenario poses a new challenge for the nascent specialist, as it must keep together alleles for finding and selecting the appropriate host ("preference") along with those for utilizing that host ("performance," *e.g.*, physiologically adapting to that host's secondary compounds or nutritional content; Jaenike 1990; Janz 2011). Theory suggests that a genetic correlation between the preference and performance alleles, such as caused by pleiotropy or genetic linkage, can overcome this problem and facilitate the switch to a new host (Lande 1979; Jaenike 1990; Fry *et al.* 1996; Janz 2011).

Until recently, evidence for this "genetic linkage" hypothesis in phytophagous insects has been mixed. Early genetic data in *Drosophila* by Jaenike (1986, 1987, 1989) suggested that oviposition preference and "settling" behavior are unlinked in *Drosophila tripunctata*, while Taylor and Condra (1983) found linkage between preference and performance

doi: 10.1534/genetics.113.154773
Manuscript received June 26, 2013; accepted for publication July 19, 2013
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.154773/-/DC1.
Data deposited in the Dryad Repository: http://doi.org/10.5061/dryad.rp6gt.
[1]Corresponding author: CB 3280, 3159 Genome Sciences Bldg., University of North Carolina, Chapel Hill, NC 27599. E-mail: cdjones@email.unc.edu

in *D. pseudoobscura*. No linkage was found in other herbivorous species, such as *Callosobruchus maculatus* (southern cowpea weevil; Wasserman and Futuyma 1981), *Colias philodice* (butterfly; Tabashnik 1986), *Papilionidae* (swallowtail butterflies; Thompson 1988; Thompson *et al.* 1990), *Chrysomelidae* (leaf-feeding beetles; Keese 1996), *Nilaparvata lugens* (brown planthopper; Sezer and Butlin 1998a,b), and *Oreina elongata* (leaf beetle; Ballabeni and Rahier 2000). However, more recent QTL mapping data for aphids (Hawthorne and Via 2001; Caillaud and Via 2012; Sauge *et al.* 2012), and other genetic association studies in *Euphydryas editha* (Edith's checkerspot butterfly; Ng 1988; Singer *et al.* 1988), *Liriomyza sativae* (leafminer fly; Via 1986), *Phyllotreta nemorum* (flea beetle; Nielsen 1996), and *Papilio glaucus* (eastern tiger swallowtail butterfly; Bossart 2003), suggest that some preference and performance alleles can be genetically linked.

A major concern with many of these genetic studies is their low resolution. QTL and marker-association studies produce candidate regions with large confidence intervals, increasing the chance that preference and performance alleles will overlap. Other studies simply infer genetic linkage due to the apparent heritability of host preference to well-performing offspring (*e.g.*, Singer *et al.* 1988). Most finer resolution genetic studies of adaptive host specialization have focused on host preference *or* avoidance. Few studies have focused on the genetics of tolerance of a specific compound in the host plant because few species with obvious host adaptations have the requisite genetic tool kit needed to study these traits (except studies involving domesticated plants and agricultural pests, in which selection pressures are often different than in natural populations).

In contrast, *D. sechellia*'s specialization on *Morinda citrifolia* has both an obvious adaptation and a genetic tool kit. *D. sechellia* is endemic to Seychelles (Tsacas and Bachli 1981), specializes on the fruit produced by *M. citrifolia* (Louis and David 1986), and is closely related to *D. simulans*, a well-studied human commensal and habitat generalist. *M. citrifolia* contains octanoic acid (OA) (Legal *et al.* 1994), which is a fatty acid that is toxic to *D. simulans* and other insects but tolerated by *D. sechellia* (Rkha *et al.* 1991). OA typically comprises 58% of the volatile chemicals in a ripe *M. citrifolia* (Farine *et al.* 1996; Pino *et al.* 2010), which makes it the main toxic component of the fruit. *D. sechellia* also prefers *M. citrifolia* over other fruit for consumption and oviposition (Rkha *et al.* 1991; Legal *et al.* 1992; Matsuo *et al.* 2007). As a result of this adaptation, *D. sechellia* has limited competition for access to *M. citrifolia* and may be protected from predation (Jones 2005).

Prior work coarsely mapped several tolerance factors and identified some loci underlying the preference behavior (Jones 1998, 2001, 2004; Colson 2004; Matsuo *et al.* 2007; Earley and Jones 2011). As with other studies of host preference, the earlier work either did not assay tolerance or lacked the resolution needed to confidently test the genetic linkage hypothesis (*e.g.*, Matsuo *et al.* 2007; Earley and Jones 2011). Moreover, because a specific gene involved in OA tolerance was not identified, the specific mechanism of OA tolerance in *D. sechellia* remains unknown.

In this study, we ultrafine map OA tolerance by genotyping independently derived recombinants using visible markers and a panel of molecular markers, along with a new phenotypic assay that provides reliable doses of OA vapor to flies without allowing them to directly contact the toxic chemical. We then measured the preference behavior of these recombinants in a test of the genetic linkage hypothesis. As preference and tolerance are not tightly linked, we reject the linkage hypothesis for this region. We hypothesize that the evolution of *D. sechellia* into an *M. citrifolia* specialist occurred through a stepwise, gradual gain of tolerance and loss of behavioral aversion to *M. citrifolia*'s toxins.

Materials and Methods

D. simulans/D. sechellia introgression lines

Jones (1998) identified a region harboring resistance alleles on chromosome arm *3*R between two visible markers. As this interval had the greatest effect on resistance, we dissected it further by generating a set of *D. sechellia/D. simulans* introgression lines. We used these 15 **o**riginal **i**ntrogression **l**ines (OILs) (Figure 1) to recombine elements of *D. sechellia* into the *D. simulans* background, using the *Dsim\jv st e osp p* mutant line (14021-0251.173, *Drosophila* Species Stock Center, University of California, San Diego) and *D. sechellia* S9 (M. Ashburner stock collection, Cambridge, UK). The presence or absence of a *D. sechellia* introgression was monitored with these recessive visible markers. Introgressed regions were present if the dominant wild-type phenotype was seen (from *D. sechellia*) rather than the recessive visible mutation (from *D. simulans*).

The OILs were made by crossing *D. sechellia* females with the males of the *D. simulans* marker line. F_1 hybrid females were then backcrossed to *D. simulans* males from the same line. The recessive mutations in the *D. simulans* background were then visible in some of the F_2 recombinants. Individual females with the desired visible marker combinations were backcrossed to mutant *D. simulans* males for >20 generations to reduce the size of the *D. sechellia* introgression on chromosome *3* and eliminate *D. sechellia* contamination from the rest of the genome. The 15 OILs represent every combination of the four visible markers, with the presence of the *D. sechellia* introgression in each OIL indicated by a black line in Figure 1A. Each OIL was derived from a pool of many F_2 females sharing the same marker phenotype, so each OIL "line" is actually a population. While male F_1 *D. simulans/D. sechellia* hybrids are sterile, 20th generation introgression males are fertile. To stably maintain the *D. sechellia* introgressions of each OIL, we backcrossed 20th generation OIL males to *D. simulans* females. The *D. sechellia* introgressions were maintained as heterozygotes.

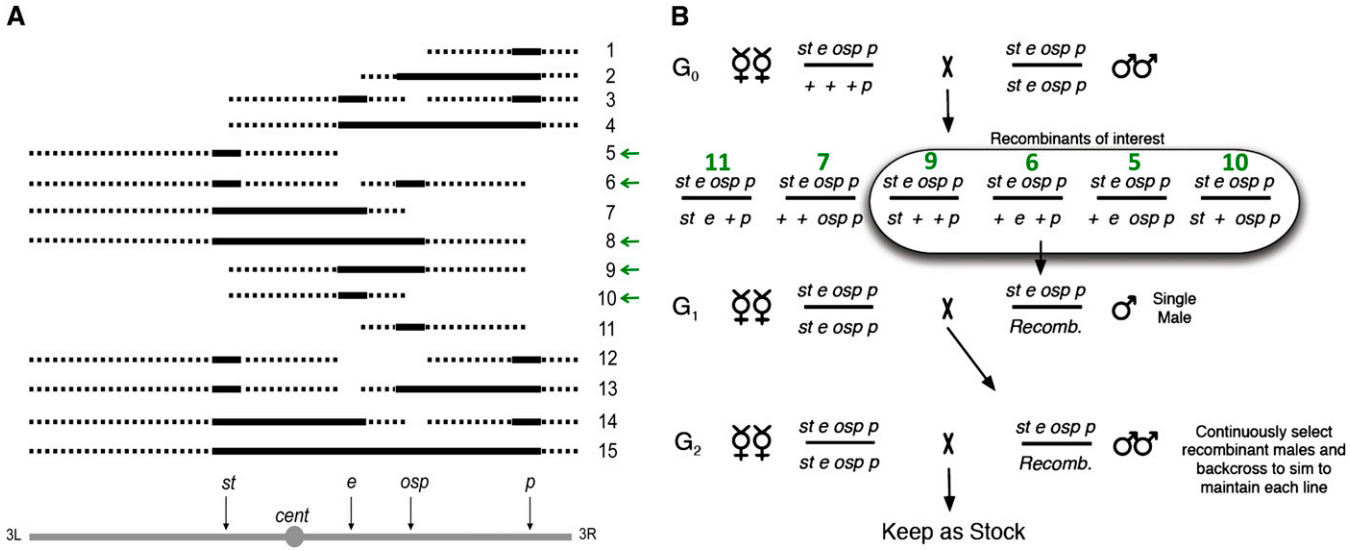

Figure 1 Description of OIL genotypes and the cross to construct additional recombinants. (A) Diagram of **o**riginal **i**ntrogression **l**ines (OILS). Solid black lines represent *D. sechellia* chromosome in *D. simulans* background. Dashed lines indicate the possible extent of the *D. sechellia* introgressions between visible markers. Four visible markers on chromosome *3* were used to categorize the OILs (*scarlet*, *st*; *ebony*, *e*; *outspread*, *osp*; *pink*, *p*). Introgression lines used for OA tolerance testing are denoted with green arrows (*i.e.*, have a breakpoint between either *st*/*e*, *e*/*osp*, or both). (B) Diagram of fly cross to generate new introgression lines for OA tolerance assay. G_0 females are from OIL 8 and males are pure *D. simulans*. The recombinants of interest (G_1 males) were backcrossed to *D. simulans* females. The six recombinant genotypes (second line of cross) correspond to OILs numbered in green. Visible mutant phenotype was continually selected in males to preserve the haplotype without recombination. *D. simulans* background indicated by recessive mutations (*st*, *e*, *osp*, and *p*). Presence of *D. sechellia* introgression denoted by symbol for wild type (+). *Recomb*, the recombinant chromosome inherited from the G_0 female.

Generation of two "gold standard" lines and large panel of recombinant lines

The OILs were tested for OA tolerance (data not shown) and it was determined that OIL 10 flies could be used to generate high and low gold standard lines. These lines, called "high 10" and "low 10," contain a *D. sechellia* introgression spanning the *e* locus, but exhibited different responses to OA exposure. We assayed many OIL 10 males and used three high and three low tolerance flies, respectively, to create high 10 and low 10. They were used to calibrate further tolerance assays.

Preliminary data indicated the OA tolerance locus was near *e*. We used OIL 8 to generate a huge population of individual recombinants in this region with unknown tolerance (Figure 1B). OIL 8 contains a large introgression spanning *st*, *e*, and *osp* loci. Three highly tolerant OIL 8 males were backcrossed to *D. simulans* females to create the line "high 8." To generate individual recombinant flies (Figure 2), we crossed high 8 females to *D. simulans* males and collected male offspring with introgressions at *st* and *osp* but not *e* (OIL 6), only at *e* (OIL 10), at *e* and *osp* (OIL 9), and only at *st* (OIL 5). These recombinants had a breakpoint between *e* and a neighboring marker (either *st* or *osp*). We gave each line an arbitrary number followed by a hyphen and the number of the OIL to which it was phenotypically identical (*e.g.*, 197-6 is an OIL 6 line). Initially, we generated 36 new recombinant lines to validate the OA tolerance assay (called the 36 "unknown" lines; see Supporting Information, File S1). Once it was clear that the assay was viable, we created another 700 recombinant lines and genotyped/phenotyped them as described below to finely map the OA tolerance locus.

Genotyping using CAPS

We genotyped recombinant lines using cleaved amplified polymorphic sequence (Konieczny and Ausubel 1993). Primers were designed to amplify both *D. simulans* and *D. sechellia* sequences that contained polymorphic restriction cut sites. PCR amplicons from *D. sechellia* sequence had an intact restriction site, whereas *D. simulans* amplicons did not. Recombinant flies with an introgression between flanking CAPS sites were propagated, and recombinant offspring underwent further CAPS genotyping with increasingly fine-mapped CAPS marker sites (Figure 2). Overall, 700 unique recombinant males were genotyped, using Acc65I, *Eco*RI, *Hin*dIII, HpyCH4IV, and *Spe*I.

Genomic DNA was extracted from males at each generation using a single fly purification method. Briefly, a single male fly was frozen at −80° and then homogenized with a pipette tip in a "squishing buffer" (10 mM Tris-HCL pH 8.2, 1 mM EDTA, 25 mM NaCl). To this mixture, 1 μl of 0.2 mg/ml of Proteinase K was added and incubated at 37° for 20 min and then inactivated at 95° for 2 min. The resulting DNA was PCR ready.

Genomic positions listed herein are the *D. melanogaster* positions from *D. melanogaster* Gene Models/Evidence Release 5 (FlyBase 1999) identified using a syntenic alignment with *D. simulans* and *D. sechellia*.

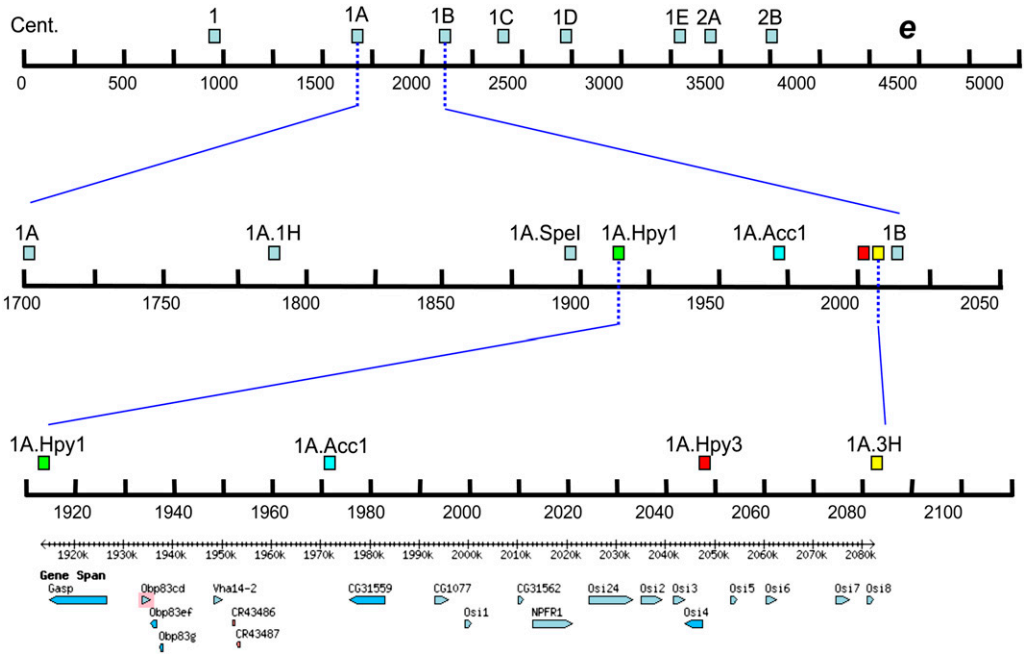

Figure 2 Map of CAPS markers used to narrow OA tolerance region on chromosome arm *3*R. The dashed and solid blue lines indicate zooming in (note the new scale on each level). The light blue boxes on the top level are markers using *Eco*RI spanning the region on *3*R between the centromere (Cent.) and *ebony* (*e*). Between the first and second levels, the focus becomes the region between markers 1A and 1B on the second level. The marker names ending in H are for *Hin*dIII; *Spe*I for *Spe*I; Hyp for HypCH4V; and Acc1 for Acc65I. The markers shaded with light blue are outside of the region of interest. The third level represents the ~170-kb region between markers 1A.Hpy1 and 1A.3H. These markers are color coded for easy identification between second and third levels. The bottom panel shows the names and positions of the 18 candidate genes (FlyBase). Scale for chromosome position: ×1000.

Octanoic acid tolerance assay

Tolerance to OA in *D. sechellia*, *D. simulans*, and hybrid recombinants was assayed using a vapor delivery system (Figure 3). A fish tank pump, regulated by a flow meter, pushed air through plastic tubing submerged in a tube of liquid OA at ~2.2 liter/min, followed by a second tube of OA, and finally into a third tube with flies (additional details are provided in File S1). To ensure full OA saturation, air was pumped for at least 1 hr before fly testing. The entire apparatus was in a fume hood with full light and ambient temperature (20–25°).

Flies were collected 4–7 days posteclosion with light CO_2 anesthesia no fewer than 4 days pretest. Between 10 and 60 flies were dumped in the test chamber and every 2 min the number of "knocked-down" (KD) flies were counted, up to a total of 30 min. Typically, OA exposure induces neurotoxin-like symptoms in flies: frantic whole-body movement, leg and wing twitching, and finally KD, where flies either invert their body or collapse while upright. When needed, we tapped the test chamber to distinguish tolerant flies at rest *vs.* KD flies.

Data analysis

Cumulative KD counts within a line were calculated as a proportion (no. flies KD/no. flies total) then logit transformed, and a linear model compared these values against log-transformed time of OA exposure (0–30 min at 2-min intervals). "Knock down 50%" (KD50) was calculated as the time at which 50% of the flies in a given assay were knocked down (R library MASS; R Development Core Team 2012). To determine the influence of genotype, mutant phenotype, fly test chamber density, air flow rate, and sex on KD50 values, we constructed a linear model and performed an ANOVA. Significant differences in KD50 between lines and sexes were calculated using Welch's *t*-test. All data have been deposited in Dryad (http://doi.org/10.5061/dryad.rp6gt).

Morinda tolerance assay

Five recombinant lines were tested for tolerance to ripe *Morinda* (high 8, high 10, low 10, 335-6, and 197-6) and pure *D. sechellia* and *D. simulans*. Naturally fallen fruit from *M. citrifolia* trees in a climate-controlled greenhouse were stored in plastic bags and used within 3 days. A pea-sized fruit pulp (no seeds) was spread across the top of a 60-mm Pyrex Petri dish. Individual flies were aspirated into this Petri dish and then placed in a growth chamber (25°, 60% humidity). After 30 min, flies were observed for KD every 10 min up to a maximum of 60 min (most flies survived to 30 min). "Survivorship" was measured as the first block of time during which knockdown was noticed. If no KD was observed at 60 min, that fly was scored as 70 and the test ended. Wilcoxon rank sum test compared differences between *D. simulans* and recombinant lines (Holm corrected *P*-value).

Behavior experiments

F_1 backcrossed flies aged 2–10 days posteclosion were subjected to a behavioral assay as in Dworkin and Jones (2009) and Earley and Jones (2011). Briefly, flies of mixed sex were introduced without anesthesia into the assay chamber (2L glass beaker, Fisher). Within the chamber were two glass

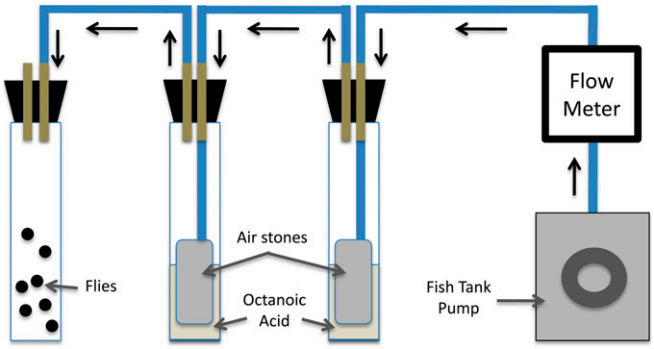

Figure 3 Diagram of OA tolerance test setup. OA vapor flows in the direction of the black arrows through plastic tubing (thick blue lines) and is bubbled through air stones twice to ensure saturation before reaching the flies. The OA/air stones and flies are housed in glass vials sealed with black stoppers. The stoppers have holes for glass tubing to pass through (brown lines), which are attached to the plastic tubing. The chamber containing flies is vented.

milk bottles, open on the top, containing either control or test food (22 ml diH20, 4.1 g instant fly media 4-24; Carolina Biological Supply), where test food was identical to control food save for acid inoculate (0.2% OA and 0.06% hexanoic acid, HA, v/v). The two food types were otherwise similar in color, texture, and water content. Cheese cloth was rubber banded over the entire chamber to prevent escape, and the chamber was placed in a growth chamber (25°, 60% humidity) overnight. The next day, flies located within each milk bottle were counted and sexed. A response index (RI) was calculated: RI = (no. flies in test bottle − no. in control bottle)/(no. in test bottle + no. in control bottle). This index was logit transformed to perform parametric tests. Data were analyzed using all-by-all *t*-tests with a false discovery rate (FDR) of 0.05 (R Development Core Team 2012), and a general linear model (GLM) (normal, model: RI = line + sex + line × sex + err; SAS Institute, Cary, NC). This assay captures fly settling behavior, which likely includes both positional preference and ovipositional preference—prior work suggests that they give qualitatively similar results (Earley and Jones 2011). Flies that chose a medium generally do not switch to the alternate medium (data not shown).

The recessive effect of introgressed loci was tested by selfing F_1 flies to create F_2 that segregated introgressed regions in both heterozygous (*D. simulans/D. sechellia*) and homozygous (*D. sechellia/D. sechellia*) states. Any F_2 flies expressing the recessive *D. simulans* markers (no introgression) were removed. F_2 flies were then pooled and tested in the same way as F_1.

Results

New high-throughput assay for volatile fatty acid tolerance

We developed and validated a new apparatus for measuring adult tolerance to OA and other volatile fatty acids. Adult foraging is important in host preference for some phytophagous insect species, such as the grass miner *Chromatomyia nigra* (Scheirs *et al.* 2005), the chrysomelid *Altica carduorum* (Scheirs *et al.* 2005), and *L. trifolii* (Scheirs *et al.* 2005). These species prefer to oviposit and feed on host plants best suited for adult performance. As Scheirs *et al.* (2005) point out, several studies that only consider larval performance also suggest that adult performance may have been affected by host quality (*e.g.*, Karowe 1990; Herr and Johnson 1992; Lu and Logan 1994). Because adult *D. sechellia* tend to feed on fresh *M. citrifolia* (Tsacas and Bachli 1981), which can be toxic even to *D. sechellia* larvae (Rkha *et al.* 1991), we believe that adult performance, rather than larval, is a key component of this adaptation.

Table 1 Welch's *t*-test *P*-values for comparisons of KD50 between lines and sexes

	Female			Male		
Line	*D. sim*	High 10	Low 10	*D. sim*	High 10	Low 10
Female						
D. sim	—	**<0.0001**	0.3324	**0.0068**	—	—
High 10	—	—	**<0.0001**	—	**0.0072**	—
Low 10	—	—	—	—	—	**0.0229**
Male						
D. sim	—	—	—	—	**<0.0001**	0.0635
High 10	—	—	—	—	—	**<0.0001**
Low 10	—	—	—	—	—	—

Significant *P*-value <0.05 are shown in boldface type; *D. sim*, *D. simulans*; KD50, time that 50% of flies are knocked down.

The high 10 and low 10 gold standard lines were used to initially validate our new "vapor" assay. We found a significant, and repeatable, difference in tolerance when high 10 males and females were compared to their respective sexes for low 10 and *D. simulans*, but not between the latter two lines for each sex (*P*-values in Table 1; KD50s plotted in Figure 4A; Figure S1). *D. simulans* was used as the low tolerance control, while *D. sechellia* was not knocked down after 3 consecutive hours of exposure. We used Welch's *t*-test because the variances for the high and low 10 male and the high and low 10 female comparisons were not equal (Bartlett test *P*-values = 0.0079 and 0.0034, respectively).

Our data show that knockdown is highly reproducible at a given flow rate, with higher flow rates resulting in more rapid knockdown (Figure S2). Fly density had no effect (see File S1). The assay could reliably and repeatedly classify 36 "unknown" lines for resistance (Figure 4, B and C; File S1). These lines were also used to test whether the visible genetic markers affected tolerance. While the visible marker *ebony* affected tolerance, this effect was generally weak and background specific (*e.g.*, the same marker did not always have the same effect across lines; File S1) and did not correlate with the number of markers in the genetic background.

High 10/low 10 gold standard lines show differential tolerance by sex

In all lines tested for male/female differences (high 10, low 10, and *D. simulans*; Table 1), the females were more

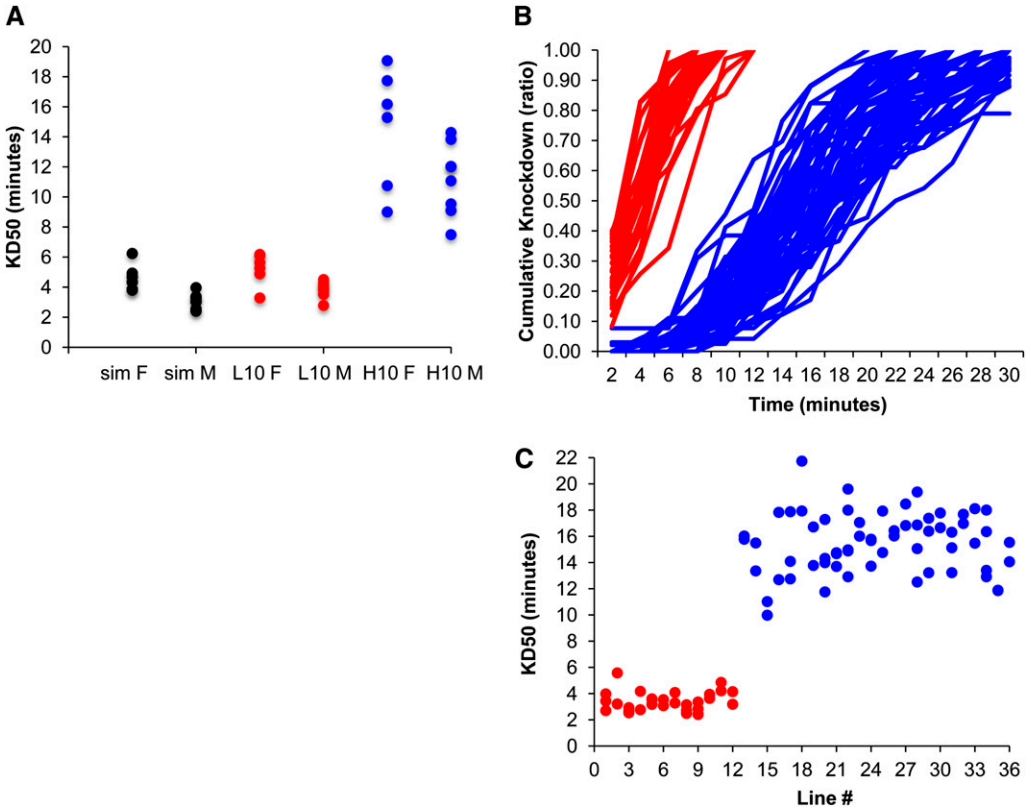

Figure 4 OA tolerance for gold standard and 36 unknown lines. (A) KD50 of all replicates for females (labeled "F") and males ("M"), separately, of *D. simulans* (*sim*, black), low 10 (L10, red), and high 10 (H10, blue). Plot of the (B) cumulative distribution of knockdown and (C) KD50 for all replicates of the 36 lines of unknown tolerance. The lines/dots clustered into two distinct groups. The blue lines/dots have been classified as high tolerance and the red as low tolerance. None of the lines/dots classified as high tolerance had replicates that clustered with the low tolerance lines and *vice versa*. Line numbers (1–36) in C correspond to the "Line #". column in Table S2 (mean KD50s).

tolerant to OA exposure than males. The percentage of cumulative knockdown over time of all individual replicates for high 10, low 10, and *D. simulans* are plotted by sex in Figure S3 (mean KD50s in Table S1).

The difference in tolerance by sex could be due to a different mechanism for tolerance in females than in males. Jones (1998) found the effects to be of different magnitudes between the sexes for every region he studied, along with an epistatic interaction between all three major chromosomes and one between markers *y* and *f* on the X chromosome in females, but not males. However, he did find effects for females in every region in which he found effects for males and explained the epistasis involving the X chromosome as possibly due to the X being hemizygous. Our results showed a significantly higher tolerance in females than males, but both sexes in highly tolerant lines exhibited significantly higher tolerances than their respective low tolerance counterparts. As the difference between the low and high tolerance lines was much larger than that between the sexes, it seems for this particular locus that the mechanism involved in OA tolerance is the same for both sexes. Instead it is likely that females were more tolerant due to their larger size, although we cannot specifically rule out that females have a differences in their tolerance mechanism compared to males.

Fine mapping of tolerance using recombinant line screening with CAPS markers

After validating the tolerance assay with the high/low 10 and 36 unknown lines, we focused on mapping tolerance factors in the region of interest. Using a marker panel consisting of 47 CAPS markers, we screened 700 independent *D. simulans/D. sechellia* recombinants. Twenty-four of these new recombinants helped reduce the target region by several hundred kilobases. Seven recombinants had a breakpoint that helped define the final 18-gene region (*i.e.*, had a boundary marker adjacent to it). We preserved the haplotypes of these seven "boundary lines" for further testing. The boundary lines clustered into clear low and high tolerance groups, similar to the 36 unknown lines, with a significant difference between the KD50s of the lines from each group (Welch's *P*-value = 4.385e-04; Figure S4). The final boundary markers of the region containing the tolerance locus are on *3*R at positions 1,913,252, defined by 335-6 (low tolerance), 697-6 (low), 505-10 (high), 525-10 (high), and 553-10 (high), and 2,082,441, defined by 197-6 (high) and 725-6 (high).

Formally, these effects could be a byproduct of the particular lines used in this introgression, the species chosen, or the hybrid background produced by this introgression. However, earlier work used a variety of different *D. simulans* and *D. sechellia* backgrounds and all were qualitatively similar (Amlou *et al.* 1998b; Jones 1998, 2001). Alternatively, *D. simulans* could be the outlier instead of *D. sechellia*. Analysis of *D. mauritiana*, a susceptible sister species of *D. sechellia* and *D. simulans*, shows that in this region *D. mauritiana* alleles tend to be more *D. simulans*-like than *D. sechellia*-like or none of the three species appears to be a strong outlier (Table S6 and Figure S5),

suggesting that *D. simulans* alleles in this region were likely not influencing tolerance any more than *D. mauritiana* alleles would have been.

Tolerance locus confirmed using fresh M. citrifolia fruit

Consistent with our vapor-based tolerance assay, the *M. citrifolia* tolerance assay showed clear differences between the *D. simulans* background line (*Dsim\jv st e osp p*) and several of the recombinant lines. *D. simulans* had a mean (±SE) KD of 31.67 ± 3.89 min in males and 32.50 ± 5 min in females, which is close to the minimum KD time allowed (*i.e.*, 30 min) in the experiment. A *D. sechellia* line (synA) remained upright for the entire 70 ± 0 min (maximum time allowed) in both males and females.

Lines that were tolerant to pure OA were generally also tolerant to *Morinda* fruit. In a mixed sex analysis comparing individual lines to their *D. simulans* background (Wilcoxon rank sum test), lines high 8, high 10, and 197-6 were all significantly more tolerant than *D. simulans*, whereas low 10 and 335-6 were not (Table 2). However, this difference appears to be driven by higher tolerance in females compared to males across all lines ($P < 0.001$, Wilcoxon; Figure 5, Table 2). This was particularly pronounced in the high tolerance lines. High 8 males were knocked down at 30 ± 0 min, while the females had a mean of 50 ± 4.71 min. The 197-6 males were down by 34 ± 1.89 min, while the females went down at 45.71 ± 3.42 min. High 10 males were down by 37.5 ± 2.7 min and females by 54.5 ± 5.1. This pattern was also seen in one of the low tolerance lines, low 10, where males averaged 30.53 ± 0.53 min and females averaged 36 ± 4.00 min. The 335-6 group was the exception, with the males going down at 30.83 ± 0.58 min and the females similarly at 30.69 ± 0.48 min. Of the highly tolerant recombinant lines, only 197-6 and high 10 had males stay upright beyond the initial 30-min check. Thus, the mixed sex results were mostly driven by the female tolerance.

Final tolerance region includes 18 genes, including Obp and Osiris families

A list of the 18 genes found in the ~170 kb tolerance region on *3*R is in Table S3. Two families of genes are represented: a cluster of three odorant-binding proteins (*Obp*) and nine *Osiris* genes. The three other named genes are *Gasp*, *Vha14-2*, and *NPFR1*. Of the 18 genes, only three remain unnamed.

The K_a/K_s ratios (Li 1993) from the *D. simulans* and *D. sechellia* lineages, as well as nonsynonymous sites and other sequence information, for the 18 genes are summarized in Table S4 and Table S5 (also see File S1). There is not a strong signature of positive selection at any of these loci. Only 7 of the 17 genes in *D. simulans* had nonsynonymous changes (41%), while 13 had such changes in *D. sechellia* (76%).

Earlier work contrasted levels of transcription of these genes across species, tissues, and treatments (Kopp *et al.* 2008; Dworkin and Jones 2009). None of the *Osiris* family genes showed differential expression. Across animal bodies, only CG31562 (a CHK kinase-like protein of unknown function) was significantly different between *D. simulans* and *D. sechellia* (*D. sechellia* expresses approximately one-third as much as *D. simulans*). In heads, *Obp83cd* was also significantly down in *D. sechellia* compared to *D. simulans* (approximately one-half), but not differentially expressed in antennae. *Obp83ef* was differentially expressed in antennae, but the difference between *D. simulans* and *D. sechellia* was substantially less than the difference between *D. simulans* and *D. melanogaster*, suggesting that expression of this gene is evolutionarily labile and the change is not associated with the shift to *M. citrifolia*. In an experiment looking at differential changes in gene expression associated with exposure to OA and HA, the gene CG1077 was weakly induced (~1.5 times).

Table 2 Wilcox rank-sum test (unpaired) *P*-values* comparing differences in *Morinda* tolerance between lines and *D. simulans* background

Line	Females	Males	Mixed
High10	**0.035**	0.076	**<<0.001**
High8	0.25	0.063	**0.03**
197-6	0.10	0.76	**0.0019**
Low10	0.86	0.32	0.96
335-6	0.27	0.48	0.16

* *P*-values are estimates because of rank ties between lines and *D. simulans* background. Bold type indicates significance at $P < 0.05$.

Tolerance alleles do not affect host preference behavior

D. sechellia exhibits strong preference for *M. citrifolia*'s fruit and its constituent fatty acids, OA and HA. These compounds, in contrast, are highly aversive to *D. simulans*. To test whether the tolerance conferred by the 18-gene region or the genes within this region affected behavior, we measured the preference of seven recombinant lines using our established preference assay (Earley and Jones 2011). Most lines did not differ from the *D. simulans* background control (Figure 6A; pairwise *t*-test and GLM, $P > 0.05$). The exceptions, high 8 (both males and females, $P = 0.017$) and low 10 (males only, $P = 0.011$), behaved in the opposite of expectation—lower tolerance resulted in less aversion. Similarly, the trend was for flies with higher tolerance to avoid the OA medium (Figure 6B), although this trend was not significant ($P = 0.3506$). To improve power, we pooled high and low tolerance lines and compared preference behavior between the two groups (replicates: 55 high, 26 low; 4993 flies). There was no effect for either sex and the trend was in the opposite of expectation (mean $RI_{high} = -1.49$; mean $RI_{low} = -1.21$; main effect $P = 0.148$, sex $P = 0.157$, interaction $P = 0.207$). Similarly, there was no difference between the pooled high or low lines and the *D. simulans* control (high, $P = 0.408$; low, $P = 0.121$). We confirmed for a subset of lines that this pattern was consistent for tolerance to *M. citrifolia* fruit (Figure 6C). These data suggest that the increased tolerance conferred by the introgressed

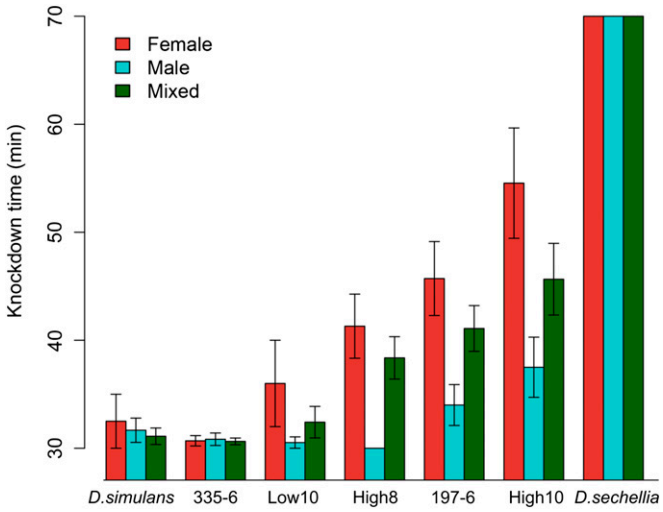

Figure 5 *M. citrifolia* tolerance assay. Introgression flies from high and low tolerant lines were assayed for tolerance of ripe *M. citrifolia* fruit. Individual flies were exposed, and knockdown was observed between 30 and 60 min every 10 min. Flies used: *D. simulans* (susceptible background of introgression lines), *D. sechellia* (SynA tolerant line), high 8, low 10, 335-6, 197-6, and high 10 (recombinant lines and high and low tolerant lines).

D. sechellia region is not sufficient to change behavior and excludes the possibility of an additive or dominant acting preference locus in this interval.

We assayed a subset of lines for a recessive preference factor linked to the tolerance factors. From the lines showing highest (high 10, 197-6) and lowest OA tolerance (low 10, 335-6), we sibmated F_1 flies to create F_2 progeny, segregating introgressions in homozygous (*D. sechellia/D. sechellia*) and heterozygous (*D. simulans/D. sechellia*) states, removing F_2's that had no introgression (*D. simulans/D. simulans*). If a recessive *D. sechellia* preference factor existed within an introgression line, then homozygous flies for that factor should have manifested higher OA preference and shifted F_2 behavior relative to a population of fully heterozygous introgressions. We pooled homozygote and heterozygote F_2's and found no significant difference in their behavior compared to fully heterozygous F_1 flies (Figure 7).

Discussion

Many insects feed on only one or a few types of host. Genetic linkage between alleles contributing to host preference and alleles contributing to host usage, such as tolerance of secondary compounds, has been suggested to facilitate the evolution of new host specializations. We used a forward genetic approach and a novel assay to see whether this type of genetic correlation contributed to the evolution of the host specialization in *D. sechellia*. We isolated a ~170-kb region on *3*R, harboring 18 genes that contains at least one locus affecting OA tolerance, a critical element of *D. sechellia*'s adaptation to the toxic fruit of its host plant, *M. citrifolia*. While the *D. sechellia* introgression conferred OA tolerance in a *D. simulans* background, it had little to no effect on host-seeking behavior.

M. citrifolia assay validates OA apparatus and methodology

To screen the thousands of flies needed for introgression mapping of tolerance loci, we developed a new assay for volatile fatty acid resistance that mimics the toxic effects of the fruit. Our analysis revealed that exposure to *M. citrifolia* fruit is quantitatively similar to the OA assay results, in terms of consistency in the lines that exhibited high and low tolerance behavior. However, the mixed sex results for lines exhibiting high tolerance to *M. citrifolia* were driven by tolerance of the females (*i.e.*, the males showed lower tolerance levels). Unfortunately, we do not know the concentration of OA in either the actual vapor of the OA assay or in the fruit itself (there is considerable variation among fruits and across ripening stages; Legal *et al.* 1992; Pino *et al.* 2010). Also, it is likely the fruit has a higher concentration of OA than the maximum our pump can produce. The consistently high tolerance of the females in both experimental setups suggests that the same tolerance mechanism is being assayed in both. Some of the variability among the sexes may reflect the larger size of the females. Experimentally, the high variability among lines in the *M. citrifolia* assay suggests that using the OA vapor methodology may be sensitive enough to detect moderate-to-weak effect loci undetectable using fruit.

Tolerance region harbors several candidate loci

Of the 18 genes in this region, two gene families represented two-thirds of the total and only three remain unnamed. A cluster of nine *Osiris* genes was present, along with three *Obps*. None of these 18 genes, however, showed a strong signature of positive selection that may be expected for a gene contributing to *D. sechellia* adaptation to its host. Likewise, gene expression data did not strongly implicate any one locus (but see *Obp83cd* below).

Osiris genes: According to Dorer *et al.* (2003), the Osiris gene family is clustered at the *Triplo-lethal* locus in *D. melanogaster*. All have endoplasmic reticulum signal peptides, may be integral to the plasma membrane, may have important housekeeping functions, and are highly dosage sensitive. In addition, their linkage and sequences are unusually highly conserved, as seen in *Anopheles gambiae* (Dorer *et al.* 2003). While none of these genes can be ruled out, it seems unlikely that their functionality can be appreciably altered without dire consequences to the individual.

Odorant-binding proteins: *Obp83cd*, *Obp83ef*, and *Obp83g* are members of a family of odorant binding proteins (OBPs) involved in olfactory perception, although their function is not fully understood (Vieira *et al.* 2007). OBPs are water soluble and exist in the aqueous lymph surrounding odorant receptors in the chemosensory sensilla of insects. They enhance the

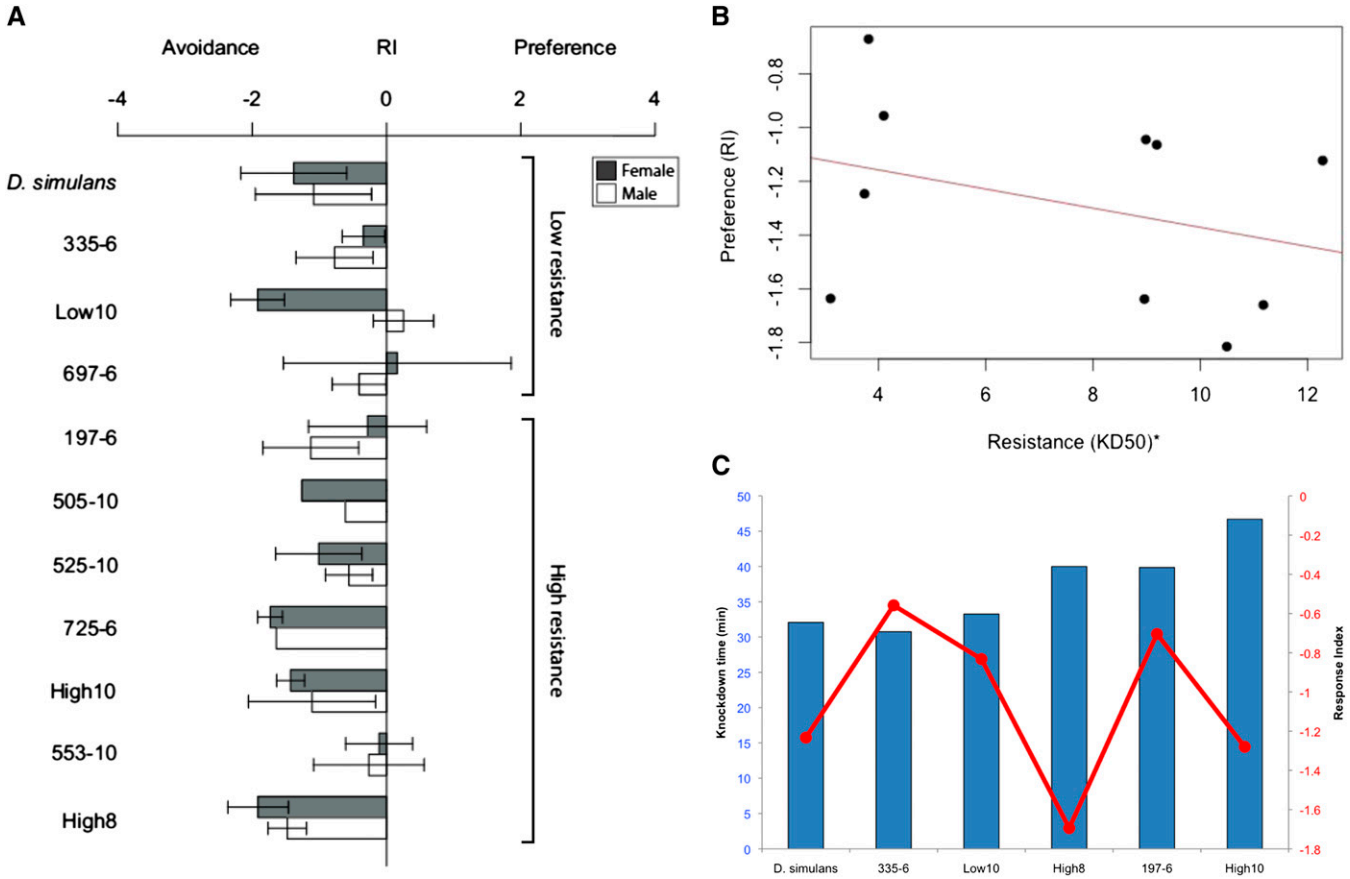

Figure 6 Tolerance alleles in the 18-gene region do not affect behavior in heterozygotes. (A) We measured the settling behavior of several introgression lines, as well as high/low 10 and *D. simulans* (also called "sim-marked"). The response index (RI) describes the degree of aversion (negative values) or preference (positive values). Most lines were not significantly different from controls, except for high 8 males and females, and low 10 males. Neither (B) OA tolerance nor (C) *M. citrifolia* fruit tolerance were positively correlated with OA preference.

solubility of hydrophobic odorants by binding to them and transporting them through the extracellular lymph to the dendritic membrane of neurons (Whiteman and Pierce 2008). *Obps* are typically found in olfactory tissues, but expression analysis shows that they are not limited to them (Pelosi *et al.* 2006). Besides transporting odorants, it has been suggested that *Obps* may also act as scavengers, removing toxic odorant molecules to prevent damage to cells (Steinbrecht 1998; Blomquist and Vogt 2003). The possible involvement of *Obps* in OA tolerance is consistent with the hypothesized link between *Obp* chemical detection and detoxification.

Obp83cd (Galindo and Smith 2001) and *Obp83ef* (Galindo and Smith 2001; Kopp *et al.* 2008) are significantly down-regulated in the head and up-regulated in the antennae, respectively. *Obp83cd* is expressed in the labellum (Galindo and Smith 2001), while *Obp83ef* is expressed in the antennae and other nonspecific tissues (Galindo and Smith 2001; Kopp *et al.* 2008). Intriguingly, *Obp83cd* shows a species-specific pattern of expression: It has reduced expression in *D. sechellia* relative to *D. simulans* (Dworkin and Jones 2009).

As with most genes, the regulatory regions of these *Obps* are not well characterized. The Regulatory Element Database for *Drosophila* v3.0 (Gallo *et al.* 2011) indicated that the regulatory regions for both *Obp83cd* and *Obp83ef* are 3 kb upstream of the start codons. We aligned these upstream regions, using *D. melanogaster* as the outgroup. For *Obp83cd*, there were 34 changes in *D. sechellia*, along with a 43-bp deletion. *Obp83ef* had 28 changes in *D. sechellia*. REDfly did not specify transcription factor binding sites, but any one of these upstream changes could alter the regulation of these genes in *D. sechellia*. No regulatory information was available for *Obp83g*.

Obp83cd and *Obp83g* have a *D. sechellia* K_a/K_s higher than the mean for the region, while *Obp83ef* has a K_a/K_s that is much lower than the mean. *D. simulans* alleles for *Obp83cd* and *Obp83g* have no nonsynonymous changes, while *D. sechellia* alleles have five and two, respectively. Of the genes that do not end prematurely or contain frameshifts, *Obp83cd* has the largest K_a (0.0076) of the *D. sechellia* alleles in this region. *Obp83ef* has one nonsynonymous change in both species, but a much higher K_a/K_s in *D. simulans* (0.6249 *vs.* 0.1073). Coding changes may not be involved in OA tolerance, but if they are, these genes are candidates with at least one amino acid change each.

Obps have previously been associated with OA avoidance behavior (Dworkin and Jones 2009), *M. citrifolia* preference

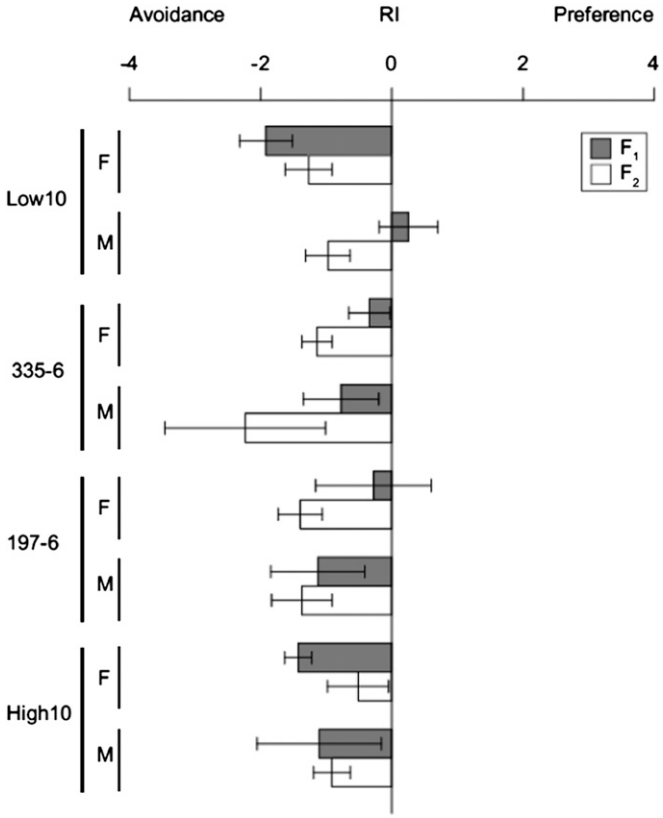

Figure 7 *D. sechellia* recessive alleles have no effect on behavior, as shown by the plot of the response index (RI). To test the effect of recessive alleles from *D. sechellia* on preference behavior, we selfed F_1 low (low 10 and 335-6) and high (high 10 and 197-6) tolerant introgression lines to make F_2 progeny. Flies possessing both homozygous and heterozygous *D. sechellia* introgressions were pooled. Their behavior was not different from purely heterozygous F_1. Females ("F") and males ("M") were assayed for all four lines.

(Matsuo *et al.* 2007), and host plant preference in *Drosophila* (Kopp *et al.* 2008), but not to tolerance of the normally toxic effects of OA experienced by most *Drosophila* species. While we are unable to eliminate the other 15 genes, the three *Obps* are our strongest candidates for future analysis.

No evidence for genetic linkage between preference and tolerance loci in this region

A positive genetic correlation between the preference and performance alleles due to pleiotropy or genetic linkage can facilitate the evolution of a new host specialization. This 18-gene region spans only ~170,000 bp and has an estimated recombination rate of 0.28 cM. These genes are therefore tightly linked. However, we find no evidence that harboring the *D. sechellia* tolerance alleles in this region causes the flies to behave in a more *D. sechellia*-like manner. Indeed the high 8 line, which spans a much larger interval (up to 23.3 cM on chromosome *3*, although the actual boundaries have not been mapped), does not appear to be significantly different from the controls and is marginally more aversive than some low tolerance lines. Previous work noted that *D. sechellia* preference for OA was recessive to *D. simulans* aversion (Higa and Fuyama 1993; Amlou *et al.* 1998a). We tested for recessive effect loci on high and low tolerance lines by selfing F_1 hybrids to segregate F_2 progeny with pooled homozygous and heterozygous *D. sechellia* introgressions (removing any F_2 with recessive homozygous *D. simulans* markers—hence, no introgression). If recessive OA preference loci exist within these introgressions, then F_2 flies should exhibit increased preference relative to F_1 flies. We did not see this pattern (Figure 7).

The limited genetic resolution of the earlier studies that suggested linkage between preference and tolerance potentially biased these studies toward overlap between QTL for preference and performance. Our data suggest that if these studies had higher resolution—and thus smaller confidence intervals—the apparent linkage between preference and performance may be reduced.

In the current study we have a different concern: We are only looking at a single locus and the critical early association between behavior and tolerance could have evolved elsewhere in the genome. Recent work, however, has identified strong preference factors on chromosome *2*, which has the weakest effect on tolerance (Jones 1998; Matsuo *et al.* 2007; Dworkin and Jones 2009). Similarly, the X chromosome, which has at least two tolerance factors, has no effect on preference behavior (Jones 1998, 2001, 2004; Earley and Jones 2011). Furthermore, the tolerance region on *3*R was previously shown to be one of the two largest contributors to resistance (the other locus is flanking; Jones 1998). Preferring the toxic host without tolerance alleles in this region of *3*R would be deleterious. Together these data suggest that linkage between preference and tolerance factors did not play a major role in the evolution of *D. sechellia*'s specialization.

There are several examples of genetically unlinked development of host preference and performance. Earlier work by Jaenike (1989) ruled out tight linkage between preference and performance in *D. tripunctata*. The same was found in other herbivorous species (see Introduction). Theoretically, the quickest way to achieve speciation through host specialization is if host preference and performance each have a simple genetic architecture and are tightly linked to each other (Fry 2003). Jaenike (1987) posited that it is unlikely for linkage disequilibrium to establish a genetic correlation between preference and performance in a system with more than a few interacting alleles. More likely, pleiotropy would explain such a linkage. (In the case of *D. sechellia*, we also eliminate this possibility.) Assortative mating on the new host plant may allow unlinked evolution of preference and performance during sympatric divergence (Diehl and Bush 1989). This scenario would likely require a few new alleles of large effect, which are seen in *D. sechellia*'s adaptation to *Morinda* fruit, but also may have resulted in a fitness gap between the old and new hosts. In this case, the deleterious effects of the new host on the performance of the unadapted fly may have been overcome by reduced competition on the new host (Wallace 1968; Berlocher and Feder

2002), since OA is toxic to *D. sechellia*'s sibling species. Alternatively, since five or more loci may be involved in OA tolerance (*i.e.*, a polygenic architecture), it is possible that part or most of the speciation process occurred gradually in allopatry (Templeton 1981).

Evolution of D. sechellia's host specialization likely occurred in a stepwise manner

D. sechellia's loss of OA avoidance and its development of preference for *M. citrifolia* seemed to derive in part from the elimination of related *Obp*s. Dworkin and Jones (2009) found that *Obp56e* had a premature stop codon in *D. sechellia* and that *D. melanogaster* showed reduced avoidance of *M. citrifolia* when *Obp56e* was knocked down. Matsuo *et al.* (2007) found a 4-bp insertion upstream of *Obp57e* in the *D. sechellia* allele affecting preference, which prevented expression when heterozygous within a *D. melanogaster Obp57e* deficiency line. If loss of functional *Obps* occurred first in the evolution of *D. sechellia*'s specialization on *M. citrifolia*, it would certainly have produced large selection pressure toward the development of tolerance on the flies that were coming into contact with the fruit.

Loss-of-function mutations are usually recessive, as they often result in the elimination or reduction of protein expression or of nonfunctional protein structures. However, Jones (1998) concluded that at least five loci involved in *D. sechellia* tolerance were dominant. Similarly, pesticide resistance factors are usually dominant or codominant (Ottea and Plapp 1984; Roush and McKenzie 1987; Houpt *et al.* 1988; ffrench-constant *et al.* 1993; Roush 1993). This asymmetry in dominance may reflect the different genetic mechanisms through which these two traits evolved. Loss of avoidance can be achieved by the loss or reduction of a sensory response, such as through pseudogenization of *Obps*. In fact, McBride (2007) discovered that *D. sechellia* harbors an unusually large number of pseudogenized chemosensory genes relative to *D. simulans* or *D. melanogaster*, even when accounting for *D. sechellia*'s small population size. Toxin resistance, in contrast, often requires *increased* expression of existing detoxification genes or *gain* of a new physiological mechanism. Both cases may result in additive-to-completely dominant phenotypes. This pattern may imply that during the early genetic steps of the evolution of a new specialization—or any adaptation—that "loss" of an ancestral trait may readily evolve from common recessive nulls segregating in the standing genetic variation. In contrast, dominant phenotypes associated with new traits and gain-of-function alleles may involve more new mutations and other relatively rare alleles, as these dominant alleles are expected to be at a lower frequency in the ancestral population (Orr and Betancourt 2001).

Acknowledgments

We thank the anonymous reviewers for their comments. We also thank Karna Mital for his contributions and Daniel Matute for his helpful advice. This work was supported by the National Science Foundation (NSF) DEB-0212686 (C.D.J.), NSF-Graduate Research Fellowship Program (E.J.E.), the National Institute of Diabetes and Digestive and Kidney Diseases grant P30DK056350 to the University of North Carolina Nutrition Obesity Research Center, and National Institutes of Health (NIH) R01 GM058686 and NIH ARRA R01 supplement GM058686-08S1.

Literature Cited

Amlou, M., B. Moreteau, and J. R. David, 1998a Genetic analysis of Drosophila sechellia specialization: oviposition behavior toward the major aliphatic acids of its host plant. Behav. Genet. 28: 455–464.

Amlou, M., B. Moreteau, and J. R. David, 1998b Larval tolerance in the Drosophila melanogaster species complex toward the two toxic acids of the D. sechellia host plant. Hereditas 129: 7–14.

Ballabeni, P., and M. Rahier, 2000 Performance of leaf beetle larvae on sympatric host and non-host plants. Entomol. Exp. Appl. 97: 175–181.

Berlocher, S. H., and J. L. Feder, 2002 Sympatric speciation in phytophagous insects: Moving beyond controversy? Annu. Rev. Entomol. 47: 773–815.

Bernays, E. A., and R. F. Chapman, 1994 *Host-Plant Selection by Phytophagous Insects*. Chapman and Hall, New York.

Blomquist, G. J., and R. G. Vogt, 2003 *Insect Pheromone Biochemistry and Molecular Biology: The Biosynthesis and Detection of Pheromones and Plant Volatiles*. Elsevier Academic Press, London.

Bossart, J., 2003 Covariance of preference and performance on normal and novel hosts in a locally monophagous and locally polyphagous butterfly population. Oecologia 135: 477–486.

Caillaud, M. C., and S. Via, 2012 Quantitative genetics of feeding behavior in two ecological races of the pea aphid, Acyrthosiphon pisum. Heredity 108: 211–218.

Colson, I., 2004 Drosophila simulans' response to laboratory selection for tolerance to a toxic food source used by its sister species D-sechellia. Evol. Ecol. 18: 15–28.

Diehl, S. R., and G. L. Bush, 1989 *The Role of Habitat Preference in Adaptation and Speciation*. Sinauer Associates, Sunderland, MA.

Dorer, D. R., J. A. Rudnick, E. N. Moriyama, and A. C. Christensen, 2003 A family of genes clustered at the Triplo-lethal locus of *Drosophila melanogaster* has an unusual evolutionary history and significant synteny with *Anopheles gambiae*. Genetics 165: 613–621.

Dworkin, I., and C. D. Jones, 2009 Genetic changes accompanying the evolution of host specialization in *Drosophila sechellia*. Genetics 181: 721–736.

Earley, E. J., and C. D. Jones, 2011 Next-Generation Mapping of Complex Traits with Phenotype-Based Selection and Introgression. Genetics 189: 1203.

Ehrlich, P. R., and P. H. Raven, 1964 Butterflies and plants: a study in coevolution. Evolution 18: 586–608.

Farine, J. P., L. Legal, B. Moreteau, and J. L. LeQuere, 1996 Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. Phytochemistry 41: 433–438.

ffrench-Constant, R. H., J. C. Steichen, T. A. Rocheleau, K. Aronstein, and R. T. Roush, 1993 A single-amino acid substitution in a gamma-aminobutyric-acid subtype-a receptor locus is associated with cyclodiene insecticide resistance in Drosophila populations. Proc. Natl. Acad. Sci. USA 90: 1957–1961.

FlyBase, 1999 The FlyBase database of the Drosophila Genome Projects and community literature. The FlyBase Consortium. Nucleic Acids Res. 27: 85–88.

Fry, J. D., 2003 Multilocus models of sympatric speciation: Bush *vs.* Rice *vs.* Felsenstein. Evolution 57: 1735–1746.

Fry, J. D., S. L. H. Einsohn, and T. F. C. M. Ackay, 1996 The contribution of new mutations to genotype-environment interaction for fitness in Drosophila melanogaster. Evolution 50: 2316–2327.

Futuyma, D. J., and G. Moreno, 1988 The evolution of ecological specialization. Annu. Rev. Ecol. Syst. 19: 207–233.

Galindo, K., and D. P. Smith, 2001 A large family of divergent Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. Genetics 159: 1059–1072.

Gallo, S. M., D. T. Gerrard, D. Miner, M. Simich, B. Des Soye *et al.*, 2011 REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic Acids Res. 39: D118–D123.

Grimaldi, D., and M. S. Engel, 2005 *Evolution of Insects*. Cambridge University Press, Cambridge, UK.

Hawthorne, D. J., and S. Via, 2001 Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412: 904–907.

Herr, J. C., and M. W. Johnson, 1992 Host plant preference of Liriomyza sativae (Diptera: Agromyzidae) populations infesting green onion in Hawaii. Environ. Entomol. 21: 1097–1102.

Higa, I., and Y. Fuyama, 1993 Genetics of food preference in Drosophila sechellia. I. Responses to food attractants. Genetica 88: 129–136.

Houpt, D. R., J. C. Pursey, and R. A. Morton, 1988 Genes-controlling malathion resistance in a laboratory-selected population of Drosophila melanogaster. Genome 30: 844–853.

Jaenike, J., 1986 Genetic complexity of host-selection behavior in Drosophila. Proc. Natl. Acad. Sci. USA 83: 2148–2151.

Jaenike, J., 1987 Genetics of oviposition-site preference in Drosophila-Tripunctata. Heredity 59: 363–369.

Jaenike, J., 1989 Genetic population structure of Drosophila tripunctata: patterns of variation and covariation of traits affecting resource use. Evolution 43: 1467–1482.

Jaenike, J., 1990 Host specialization in phytophagous insects. Annu. Rev. Ecol. Syst. 21: 243–273.

Janz, N., 2011 Ehrlich and Raven revisited: mechanisms underlying codiversification of plants and enemies. Annu. Rev. Ecol. Evol. Syst. 42(42): 71–89.

Jones, C. D., 1998 The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. Genetics 149: 1899–1908.

Jones, C. D., 2001 The genetic basis of larval resistance to a host plant toxin in Drosophila sechellia. Genet. Res. 78: 225–233.

Jones, C. D., 2004 Genetics of egg production in Drosophila sechellia. Heredity 92: 235–241.

Jones, C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica 123: 137–145.

Karowe, D. N., 1990 Predicting host range evolution: colonization of Coronilla varia by Colias philodice (Lepidoptera: Pieridae). Evolution 44: 1637–1647.

Keese, M. C., 1996 Feeding responses of hybrids and the inheritance of host-use traits in leaf feeding beetles (Coleoptera: Chrysomelidae). Heredity 76: 36–42.

Konieczny, A., and F. M. Ausubel, 1993 A procedure for mapping Arabidopsis mutations using codominant ecotype-specific pcr-based markers. Plant J. 4: 403–410.

Kopp, A., O. Barmina, A. M. Hamilton, L. Higgins, L. M. McIntyre *et al.*, 2008 Evolution of gene expression in the Drosophila olfactory system. Mol. Biol. Evol. 25: 1081–1092.

Lande, R., 1979 Quantitative genetic-analysis of multivariate evolution, applied to brain–body size allometry. Evolution 33: 402–416.

Legal, L., J. R. David, and J. M. Jallon, 1992 Toxicity and attraction effects produced by *Morinda citrifolia* fruits on the *Drosophila melanogaster* complex of species. Chemoecology 3: 125–129.

Legal, L., B. Chappe, and J. M. Jallon, 1994 Molecular-basis of Morinda-Citrifolia (L): toxicity on Drosophila. J. Chem. Ecol. 20: 1931–1943.

Li, W. H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36: 96–99.

Louis, J., and J. R. David, 1986 Ecological specialization in the Drosophila-Melanogaster species subgroup: a case-study of Drosophila-Sechellia. Acta Oecologica-Oecologia Generalis 7: 215–229.

Lu, W. H., and P. Logan, 1994 Genetic-variation in oviposition between and within populations of Leptinotarsa decemlineata (Coleoptera, Chrysomelidae). Ann. Entomol. Soc. Am. 87: 634–640.

Matsubayashi, K. W., I. Ohshima, and P. Nosil, 2010 Ecological speciation in phytophagous insects. Entomol. Exp. Appl. 134: 1–27.

Matsuo, T., S. Sugaya, J. Yasukawa, T. Aigaki, and Y. Fuyama, 2007 Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia. PLoS Biol. 5: 985–996.

McBride, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in Drosophila sechellia. Proc. Natl. Acad. Sci. USA 104: 4996–5001.

Ng, D., 1988 A novel level of interactions in plant–insect systems. Nature 334: 611–612.

Nielsen, J. K., 1996 Intraspecific variability in adult flea beetle behaviour and larval performance on an atypical host plant, pp. 160–162 in *Proceedings of the 9th International Symposium on Insect-Plant Relationships*. Springer-Verlag, New York.

Orr, H. A., and A. J. Betancourt, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics 157: 875–884.

Ottea, J. A., and F. W. Plapp, 1984 Glutathione S-transferase in the housefly: biochemical and genetic changes associated with induction and insecticide resistance. Pestic. Biochem. Physiol. 22: 203–208.

Pelosi, P., J. J. Zhou, L. P. Ban, and M. Calvello, 2006 Soluble proteins in insect chemical communication. Cell. Mol. Life Sci. 63: 1658–1676.

Pino, J. A., E. Marquez, C. E. Quijano, and D. Castro, 2010 Volatile compounds in noni (Morinda citrifolia L.) at two ripening stages. Ciencia e Tecnologia de Alimentos 30: 183–187.

R Development Core Team, 2012 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Rkha, S., P. Capy, and J. R. David, 1991 Host plant specialization in the Drosophila melanogaster species complex: a physiological, behavioral, and genetic analysis. Proc. Natl. Acad. Sci. USA 88: 1835–1839.

Roush, R. T., 1993 Occurrence, genetics and management of insecticide resistance. Parasitol. Today 9: 174–179.

Roush, R. T., and J. A. McKenzie, 1987 Ecological genetics of insecticide and acaricide resistance. Annu. Rev. Entomol. 32: 361–380.

Sauge, M. H., P. Lambert, and T. Pascal, 2012 Co-localisation of host plant resistance QTLs affecting the performance and feeding behaviour of the aphid Myzus persicae in the peach tree. Heredity 108: 292–301.

Scheirs, J., K. Jordaens, and L. De Bruyn, 2005 Have genetic trade-offs in host use been overlooked in arthropods? Evol. Ecol. 19: 551–561.

Sezer, M., and R. Butlin, 1998a The genetic basis of oviposition preference differences between sympatric host races of the brown planthopper (Nilaparvata lugens). Proc. R. Soc. Lond. B Biol. Sci. 265: 2399–2405.

Sezer, M., and R. K. Butlin, 1998b The genetic basis of host plant adaptation in the brown planthopper (*Nilaparvata lugens*). Heredity 80: 499–508.

Singer, M. C., D. Ng, and C. D. Thomas, 1988 Heritability of oviposition preference and its relationship to offspring performance within a single insect population. Evolution 42: 977–985.
Steinbrecht, R. A., 1998 Odorant-binding proteins: expression and function. Olfaction and Taste Xii 855: 323–332.
Tabashnik, B. E., 1986 *Evolution of Host Plant Utilization in Colias Butterflies. Evolutionary Genetics of Invertebrate Behavior*, pp. 173–184. Plenum, New York.
Taylor, C. E., and C. Condra, 1983 Resource partitioning among genotypes of Drosophila pseudoobscura. Evolution 37: 135–149.
Templeton, A. R., 1981 Mechanisms of speciation: a population genetic approach. Annu. Rev. Ecol. Syst. 12: 23–48.
Thompson, J. N., 1988 Evolutionary ecology of the relationship between oviposition preference and performance of offspring in phytophagous insects. Entomol. Exp. Appl. 47: 3–14.
Thompson, J. N., W. Wehling, and R. Podolsky, 1990 Evolutionary genetics of host use in swallowtail butterflies. Nature 334: 148–150.
Tsacas, L., and G. Bachli, 1981 *Drosophila sechellia*. n. sp., huitieme espece du sous-groupe melanogaster des iles Seychelles (Diptera, Drosophilidae). Rev. Fr. Entomol. 3: 146–150.
Via, S., 1986 Genetic covariance between oviposition preference and larval performance in an insect herbivore. Evolution 40: 778–785.
Via, S., 2001 Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol. Evol. 16: 381–390.
Via, S., and D. J. Hawthorne, 2005 Back to the future: genetic correlations, adaptation and speciation. Genetica 123: 147–156.
Vieira, F. G., A. Sanchez-Gracia, and J. Rozas, 2007 Comparative genomic analysis of the odorant-binding protein family in 12 Drosophila genomes: purifying selection and birth-and-death evolution. Genome Biol. 8: R141.
Wallace, B., 1968 Polymorphism, population size, and genetic load, pp. 87–108 in *Population Biology and Evolution*, edited by R. C. Lewontin. Syracuse University Press, Syracuse, NY.
Wasserman, S. S., and D. J. Futuyma, 1981 Evolution of host plant utilization in laboratory populations of the southern cowpea weevil, Callosobruchus maculatus Fabricius (Coleoptera: Bruchidae). Evolution 35: 605–617.
Whiteman, N. K., and N. E. Pierce, 2008 Delicious poison: genetics of Drosophila host plant preference. Trends Ecol. Evol. 23: 473–478.

Communicating editor: K. Nichols

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.154773/-/DC1

A Locus in *Drosophila sechellia* Affecting Tolerance of a Host Plant Toxin

Eric A. Hungate, Eric J. Earley, Ian A. Boussy, David A. Turissini, Chau-Ti Ting, Jennifer R. Moran, Mao-Lien Wu, Chung-I Wu, and Corbin D. Jones

DOI: 10.1534/genetics.113.154773

Data deposited in the Dryad Repository: http://doi.org/10.5061/dryad.rp6gt

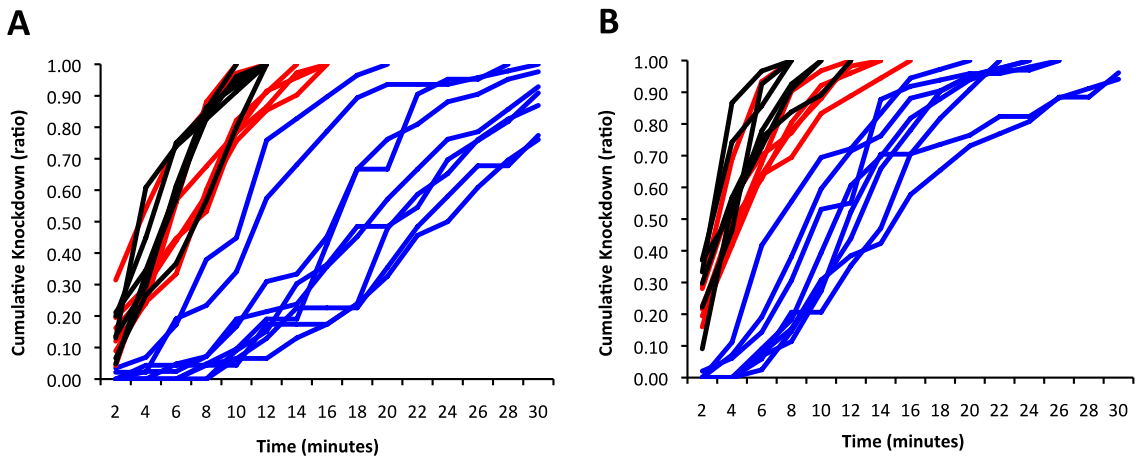

Figure S1 Plot of cumulative distribution of OA tolerance of all replicates for the comparison between *D. simulans* (black), Low 10 (red), and High 10 (blue), (A) females and (B) males.

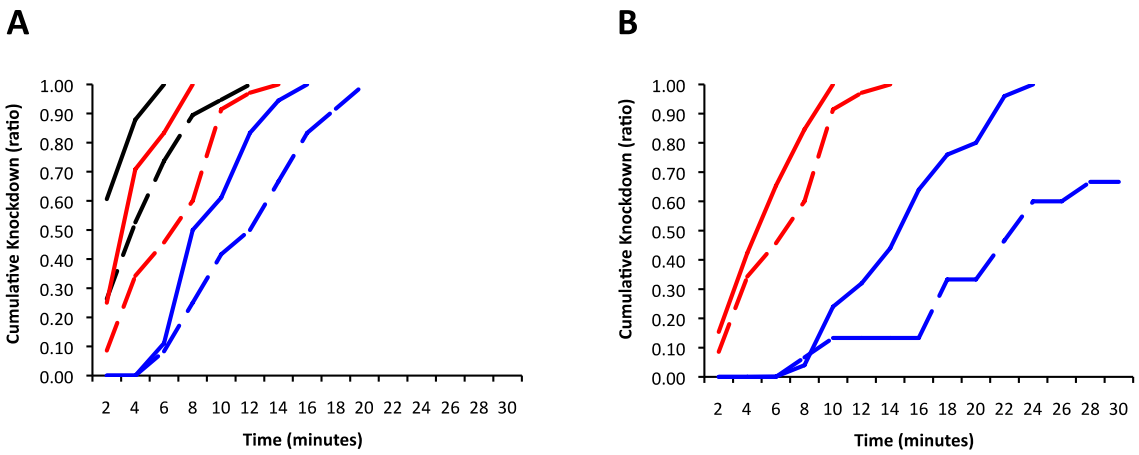

Figure S2 Comparison of impact on tolerance using a low flow rate (0.8 L/min) and a high flow rate (2.8 L/min) with three different genotypes (High 10, blue; Low 10, red; and *D. simulans*, black) and by sex. The dashed lines represent low flow and the solid lines are high flow. (A) Comparison of males between the three genotypes at the two flow rates and (B) comparison of High 10 and Low 10 females at the two flow rates.

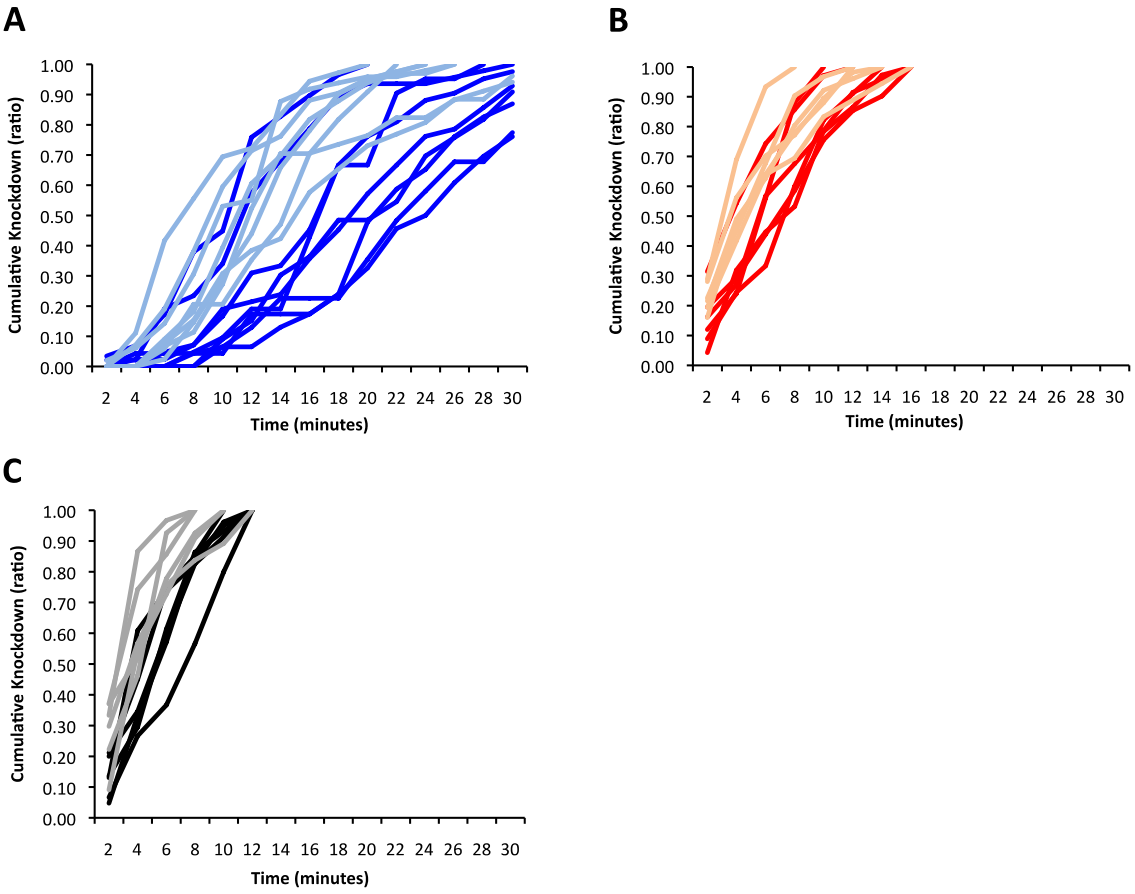

Figure S3 Plot of cumulative distribution of OA tolerance of all replicates for the comparison between (A) High 10 males (light blue) and females (dark blue), (B) Low 10 males (orange) and females (red), and (C) *D. simulans* males (gray) and females (black).

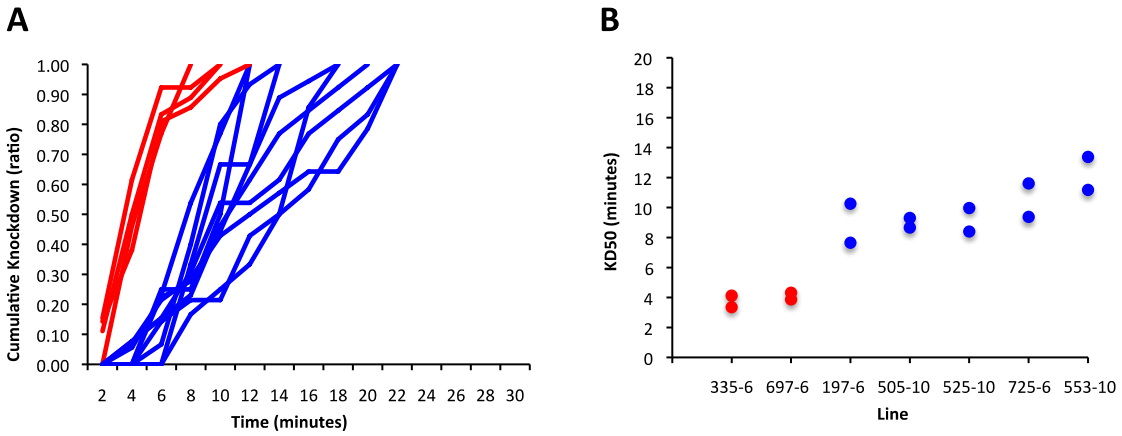

Figure S4 Cumulative distributions of percent knockdown over time for males of the seven boundary lines. Each of these lines has a boundary marker immediately flanking the 18-gene region (i.e. the OA tolerance locus). The introgressions for 335-6 and 697-6 stop short of including the OA tolerance locus, while the other five lines contain it. (A) Plot of the cumulative distributions of all replicates of the seven boundary lines (two replicates per line). (B) Plot of the KD50 (in minutes) for all replicates of the boundary lines. Blue represents line called as high tolerance, while red is for low tolerance. These lines were used for the *M. citrifolia* tests described in the Materials and Methods.

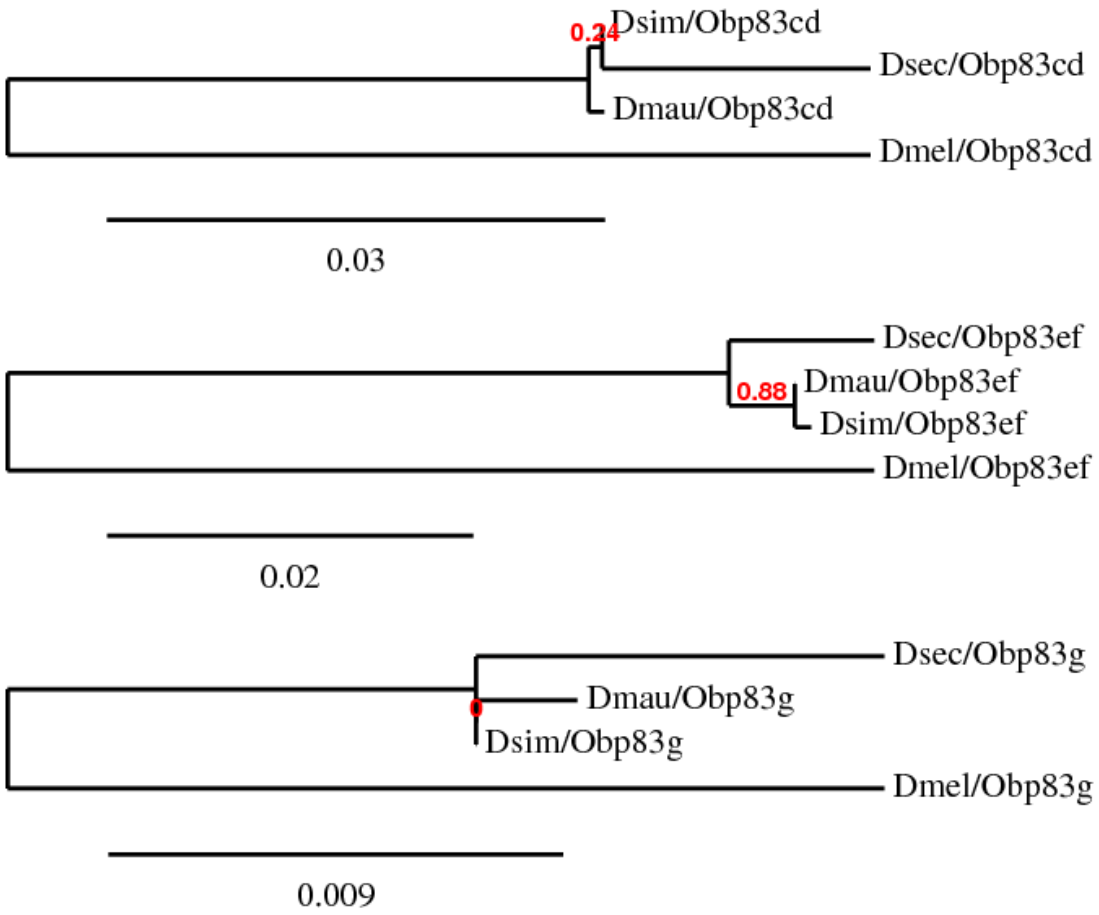

Figure S5 Maximum likelihood phylogenies of Odorant Binding Proteins within the tolerance region. Although two of three trees are unresolved, the *Obp83ef* trees shows that *D. sechellia* tends to have the unusual allele. Phylogenies were made using http://www.phylogeny.fr/. Bootstrap values are in red.

File S1

Materials and Methods

Reproducibility of knock-down: To demonstrate the reproducibility of the results from the OA vapor apparatus, we extensively tested the control lines called Low 10 and High 10, along with *D. simulans* and *D. sechellia*. We set the flow rate so most of the recombinant flies were knocked down within 30 minutes, but still allowed adequate resolution to see significant differences between tolerant and non-tolerant recombinant lines. Under these conditions, *D. simulans* flies could only tolerate the acid vapor for a few minutes, while *D. sechellia* flies from three replicate tests were not knocked-down after six consecutive hours of exposure. The "tolerant" recombinant flies had an intermediate tolerance, which was typically less than 30 minutes, but significantly more than that of *D. simulans*.

Effect of density on tolerance: We used linear regression analysis to determine if the density of flies in the test vial had any impact on OA tolerance. There was no significant effect detected for either gender in any of the three lines based on the slope of the linear regression not differing from zero when the number of flies in the vial was between 10 and 60 (High 10 male p-value = 0.2175, female p-value = 0.6740; Low 10 male p-value = 0.6756, female p-value = 0.2000; and *D. simulans* male p-value = 0.5881, female p-value = 0.8801).

Flow rate: Tolerance at two extreme flow rates (0.8 and 2.8 liters/minute) was assayed for High 10 and Low 10 males and females, with *D. simulans* males as the control. The cumulative distribution curves are plotted in Figure S2. Since only one replicate was used for each genotype at each flow rate, the KD50s are plotted without error bars. However, the KD50 for each line decreased when the flow rate increased. This effect was especially large in High 10, where the KD50 in females dropped from 23.18 to 14.00 minutes and from 10.96 to 8.60 minutes in males. The effect was fairly large in *D. simulans* males, as well, where the KD50 dropped from 5.46 to 1.82 minutes. The smallest effect was seen in Low 10, where the KD50 in females dropped from 6.19 to 4.21 minutes and from 3.49 to 2.96 minutes in males. Despite the small sample size, the consistency of the results across lines indicated that flow rate, which is a proxy for concentration, had a substantial impact on tolerance. The significance of this conclusion will be explained during the discussion of the *M. citrofolia* assay results.

36 unknown lines and the effect of markers on tolerance: The cumulative knockdown curves of each replicate for all 36 unknown recombinant lines are plotted in Figure 4B. The graph revealed obvious clustering of two distinct groups. In fact, none of the lines had replicates split between the high tolerance and low tolerance groups. In other words, a given line was distinctly highly tolerant or lowly tolerant across its replicates. 12 of the 36 lines exhibited low tolerance, while the other 24 were highly tolerant. The mean KD50s for each line are listed in Table S2

and are plotted in Figure 4C. Since most of the lines were assayed with only two replicates, we plotted every replicate for each line instead of using error bars. The highest low tolerance replicate had a KD50 of 5.58, while the lowest high tolerance replicate was 9.98. We did not run an ANOVA to statistically classify each line as low or high tolerance because of the small number of replicates per line, as well as the obvious clustering pattern in the data.

Since the unknown lines had the same phentoypes as OILs 5, 6, 8, 9, and 10 (Figure 1), we was able to group them by phenotype to determine if their visible mutations had any effect on tolerance. We separately analyzed the high and low tolerance clusters. It was necessary to confirm that the different phenotypic groups had equal variances, since the samples sizes varied largely. Each pair of phenotype groups was analyzed with the Bartlett Test, which showed no significant differences between the variances within any of the pairs (all p-values > 0.05; data not shown). The ANOVA for the low tolerance cluster indicated that phenotype (i.e. visible mutations used as markers) did have a significant impact on tolerance (p-value = 0.0016). A Tukey HSD Test showed that pairs exhibiting the significant differences were OIL phenotypes 6 and 10 (p-value = 0.0014) and 5 and 10 (p-value = 0.0386). The ANOVA for the high tolerance cluster indicated that phenotype did not play a significant role on tolerance (p-value = 0.0775). However, since the p-value was very close to 0.05, we completed the Tukey analysis anyway and found that OIL phenotype pairs 6/8 (p-value = 0.0754) and 6/10 (p-value = 0.0801) were nearly significant.

The data from the 36 unknown lines showed that lowly tolerant recombinant lines expressing *ebony* (i.e. lines not containing the *D. sechellia* introgression at *ebony*) were significantly less tolerant (phenotype from OILs 5 and 6) than those without *ebony* expression (OILs 9 and 10). The pattern held, but with non-significant p-values, for the highly tolerant lines. Coyne (1984) showed that the *Dsim\jv st e osp p* markers do not affect backcross hybrid size, so there should not have been any inherent difference in size between the OILs 5, 6, 9, and 10. However, *ebony* is known to reduce viability to about 80% of the wild type (Lindsley and Zimm, 1992). The reduced viability probably had an impact on OA tolerance since the flies were being exposed to a toxic chemical and *ebony* has been shown to be pleiotropically involved in neural function (Hovemann et al., 1998).

Ka/Ks: We calculated *Ka/Ks* for genes within the final candidate region (Bergen Center for Computational Science's *Ka/Ks* Calculation tool), where *Ka* is the number of non-synonymous changes in a codon divided by the number of non-synonymous sites, and *Ks* is the same as *Ka*, but for synonymous sites. We compared *D. simulans* and *D. sechellia* coding sequences with *D. melanogaster* as the outgroup.

The Ka/Ks ratios (Li 1993) from both the *D. simulans* and *D. sechellia* lineages, as well as non-synonymous sites and other sequence information, for the 18 genes are summarized in Tables S4 and S5. *Osi4* was not included in

the following results, due to having an undefined Ka/Ks for *D. simulans*. Using McBride's (2007) control group means for Ka, Ks, and Ka/Ks in both lineages, there were 4 genes in *D. simulans* with Ka values higher than the control value of 0.002, while there were 7 genes in *D. sechellia* higher than the control of 0.004. There was only one gene with a Ks above the control value of 0.023 in *D. simulans*, whereas there were two genes above the control of 0.030 in *D. sechellia*. 6 genes in *D. simulans* had a Ka/Ks higher than the control value of 0.117, while 10 genes (not including *Osi4*) in *D. sechellia* had a Ka/Ks > 0.145. Thus, 59% of the genes in the region in *D. sechellia* had an enrichment of non-synonymous changes relative to synonymous changes when compared to the control group, while only 35% of the genes in *D. simulans* showed the same. Moreover, only 7 of the 17 genes in *D. simulans* had any non-synonymous changes (41%), while 13 had such changes in *D. sechellia* (76%). In sum, there is not a strong signature of positive selection at any of these loci.

References

Coyne, J. A., 1984 Genetic basis of male sterility in hybrids between two closely related species of *Drosophila*. Proc. Natl. Acad. Sci. USA **51:** 4444–4447.

Hovemann B. T., R. P. Ryseck, U. Walldorf, K. F. Störtkuhl, I. D. Dietzel, et al., 1998 The *Drosophila ebony* gene is closely related to microbial peptide synthetases and shows specific cuticle and nervous system expression. Gene **221:** 1-9.

Lindsley, D. L., and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, Inc., San Diego.

McBride, C. S., 2007 Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. Proc. Natl. Acad. Sci. USA **104:** 4996–5001.

Thibault, S. T., M. A. Singer, W. Y. Miyazaki, B. Milash, N.A. Dompe, et al., 2004 A complementary transposon tool kit for *Drosophila melanogaster* using P and *piggyBac*. Nat. Genet. **36:** 283-287.

Table S1 KD50 values for *D. sim*, High 10, and Low 10 for both sexes

	Females				*Males*			
Line	*replicates*	*# flies*	*mean KD50**	*S.E.*	*replicates*	*# flies*	*mean KD50**	*S.E.*
D.sim	6	190	4.64	0.37	6	216	3.10	0.23
High 10	9	337	17.14	1.64	8	305	11.17	0.83
Low 10	6	230	5.23	0.44	6	270	3.81	0.25

D.sim, *D. simulans*
KD50, time that 50% of flies are knocked down
*KD50 units: minutes

Table S2 Mean KD50 for males for 36 recombinant lines of unknown tolerance

Line name	Line #	Mean KD50	S.D.	S.E.	# replicates
1-H6	1	3.37	0.64	0.37	3
10-L10	2	4.40	1.67	1.18	2
14-L5	3	2.71	0.20	0.12	3
3-L10	4	3.48	0.99	0.57	2
3-L5	5	3.45	0.23	0.13	3
4-L6	6	3.31	0.33	0.23	2
5-L5	7	3.69	0.57	0.40	2
5-L6	8	2.73	0.30	0.15	4
6-L6	9	2.76	0.44	0.22	4
7-L10	10	3.84	0.18	0.10	3
8-L10	11	4.54	0.45	0.32	2
9-L10	12	3.67	0.68	0.48	2
1-H10	13	15.90	0.17	0.12	2
1-H9	14	14.43	1.51	1.07	2
1-L6	15	15.27	3.99	2.82	2
1-L9	16	15.32	3.63	2.57	2
2-H10	17	14.91	2.66	1.53	3
2-H9	18	19.84	2.69	1.90	2
2-L6	19	15.25	2.08	1.47	2
2-L9	20	14.34	2.27	1.14	4
3-H9	21	14.38	0.59	0.34	3
4-H10	22	16.08	2.68	1.20	5
4-H9	23	16.54	0.74	0.52	2
4-L10	24	15.06	1.16	0.67	3
5-H9	25	16.35	2.24	1.58	2
5-L9	26	16.23	0.28	0.20	2
6-H10	27	17.65	1.16	0.82	2
6-L8	28	15.96	2.90	1.45	4
6-L9	29	15.67	2.17	1.25	3
7-H10	30	17.22	0.79	0.56	2
7-H9	31	14.18	2.18	1.26	3
7-L8	32	17.33	0.50	0.35	2
7-L9	33	16.80	1.86	1.32	2
8-H10	34	15.17	2.43	1.21	4
8-L9	35	11.88	0.02	0.01	2

9-L9	36	14.80	1.04	0.74	2

S.D., standard deviation
S.E., standard error
KD50, time that 50% of flies are knocked down
*KD50 units: minutes

Table S3 Gene ontology of genes in tolerance region

CG	Symbol	Gene Ontology Terms
CG10287	*Gasp*	Chitin binding and metabolic process; structural component of peritrophiv membrane
CG15582	*Obp83cd*	Odorant binding; sensory perception of chemical stimulus
CG31557	*Obp83ef*	Odorant binding; sensory perception of chemical stimulus
CG31558	*Obp83g*	Odorant binding; sensory perception of chemical stimulus
CG1076	*Vha14-2*	Hydrogen-exporting ATPase activity, phosphorylative mechanism
CG31559	CG31559	Electron carrier activity; protein disulfide oxidoreductase activity
CG1077	CG1077	Protease inhibitor
CG15585	*Osi1*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG31562	CG31562	Unknown function; protein features similar to DUF227, CHK kinase-like
CG1147	*NPFR1*	Neuropeptide F, Y, and tachykinin receptor activity; GPCR signaling pathway
CG15589	*Osi24*	Unknown function
CG1148	*Osi2*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG1150	*Osi3*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG10303	*Osi4*	Unknown function; integral to plasma membrane
CG15590	*Osi5*	Unknown function; integral to plasma membrane
CG1151	*Osi6*	Unknown function; protein features similar to DUF1676
CG1153	*Osi7*	Unknown function; protein features similar to DUF1676; integral to plasma membrane
CG15591	*Osi8*	Unknown function; integral to plasma membrane

Table S4 Number of non-synonymous and syntenic sites for genes in tolerance region

		Non-synonymous sites		Syntenic sites included in:		
CG	**Symbol**	***D. sim***	***D. sech***	**Ka/Ks**	**Alignment**	**Missing seq**
CG10287	*Gasp*	1	1	708	708	0
CG15582	*Obp83cd*	0	5	729	729	0
CG31557	*Obp83ef*	1	1	738	738	0
CG31558	*Obp83g*	0	2	441	441	0
CG1076	CG1076	0	1	111[a]	390	0
CG31559	CG31559	2	2[b]	1365	1365	0
CG1077	CG1077	11[c]	6	2193	2193	0
CG15585	*Osi1*	1	2[d]	513	513	414
CG31562	CG31562	2	5	765[e]	792	27
CG1147	*NPFR1*	0	1	1458	1458	0
CG15589	CG15589	2	5	1447[f]	1602	0
CG1148	*Osi2*	0	0	1173	1173	0
CG1150	*Osi3*	0	0	867	867	0
CG10303	*Osi4*	3	1	717[g]	1182	0
CG15590	*Osi5*	0	1	609	609	0
CG1151	*Osi6*	0	0	939	939	0
CG1153	*Osi7*	0	0	115	115	752
CG15591	*Osi8*	6	1[i]	309[h]	825	0

[a] Frameshift deletion at 112 in both *D. simulans* (10 bp) and *D. sechellia* (8 bp)
[b] In-frame deletion at 646 in *D. sechellia* (9 bp)
[c] In-frame deletion at 436 in *D. simulans* (6 bp)
[d] In-frame deletion at 154 in *D. sechellia* (3 bp)
[e] Pre-mature stop codon ending at 765 in *D. sechellia*
[f] Pre-mature stop codon ending at 1447 in *D. sechellia*
[g] Frameshift deletion at 718 in *D. simulans* (1 bp)
[h] Pre-mature stop codon ending at 309 in *D. simulans*
[i] One additional non-syn site in *D. sechellia* from *D. melanogaster* after pre-mature stop in *D. simulans*

Table S5 Ka/Ks calculation for genes in tolerance region for *D. simulans* and *D. sechellia*

		D. simulans			*D. sechellia*		
CG	**Symbol**	**Ka/Ks**	**Ka**	**Ks**	**Ka/Ks**	**Ka**	**Ks**
CG10287	*Gasp*	0.2652	0.0022	0.0083	0.0696	0.0022	0.0316
CG15582	*Obp83cd*	0	0	0.0039	0.2321	0.0076	0.0325
CG31557	*Obp83ef*	0.6249	0.0016	0.0026	0.1073	0.0022	0.0207
CG31558	*Obp83g*	0	0	0	0.4153	0.0054	0.0131
CG1076	CG1076	0	0	0	0.6269	0.0164	0.0262
CG31559	CG31559	0.1304	0.0020	0.0156	0.1880	0.0023	0.0124
CG1077	CG1077	0.6445	0.0067	0.0104	0.3237	0.0042	0.0130
CG15585	*Osi1*	0	0	0	0.1791	0.0049	0.0271
CG31562	CG31562	0.7708	0.0045	0.0058	0.6480	0.0075	0.0116
CG1147	*NPFR1*	0	0	0.0114	0.0495	0.0008	0.0171
CG15589	CG15589	0.5857	0.0019	0.0033	0.3146	0.0036	0.0115
CG1148	*Osi2*	0	0	0.0024	0	0	0.0099
CG1150	*Osi3*	0	0	0.0100	0	0	0.0131
CG10303	*Osi4*	2.7557	0.0028	0.0000	1.1981	0.0046	0.0038
CG15590	*Osi5*	0	0	0.0045	0.1989	0.0026	0.0132
CG1151	*Osi6*	0	0	0.0080	0	0	0.0142
CG1153	*Osi7*	0	0	0	0	0	0
CG15591	*Osi8*	0.3306	0.0274	0.0830	0.2063	0.0050	0.0241
	Mean	0.1972	0.0027	0.0100	0.2094	0.0038	0.0171

Table S6 Catalog of *D. simulans* and *D. mauritiana* allelic differences at the tolerance locus [a]

***D.simulans* locus**	***D. melanogaster* homolog**	**Location in Gene (bp)**	**Location on Chromosome (bp)**	***D.simulans* allele**	***D. mauritiana* allele**
Dsim\GD19843	mel:NPFR1	447	3R:2030917	sim:AGGG	mau:AGG
Dsim\GD19843	mel:NPFR1	550	3R:2031020	sim:AAGCTCAGC	mau:AAGC
Dsim\GD19843	mel:NPFR1	1175	3R:2031645	sim:GGTG	mau:GG
Dsim\GD19843	mel:NPFR1	1405	3R:2031875	sim:GCCCGATACCCGATACCCGAT	mau:GCCCGATACCCGAT
Dsim\GD19843	mel:NPFR1	1759	3R:2032229	sim:CT	mau:C
Dsim\GD19843	mel:NPFR1	2184	3R:2032654	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	2309	3R:2032779	sim:CA	mau:CAA
Dsim\GD19843	mel:NPFR1	2329	3R:2032799	sim:C	mau:A
Dsim\GD19843	mel:NPFR1	2519	3R:2032989	sim:A	mau:G
Dsim\GD19843	mel:NPFR1	2601	3R:2033071	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	2850	3R:2033320	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	3018	3R:2033488	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	3522	3R:2033992	sim:CTTTTTT	mau:CTTTTT
Dsim\GD19843	mel:NPFR1	3998	3R:2034468	sim:GAAAA	mau:GAAAAA
Dsim\GD19843	mel:NPFR1	4077	3R:2034547	sim:TA	mau:T
Dsim\GD19843	mel:NPFR1	4112	3R:2034582	sim:A	mau:AATAGGATTCCAAG
Dsim\GD19843	mel:NPFR1	4292	3R:2034762	sim:A	mau:G
Dsim\GD19843	mel:NPFR1	4334	3R:2034804	sim:G	mau:A
Dsim\GD19843	mel:NPFR1	4463	3R:2034933	sim:C	mau:T,G
Dsim\GD19843	mel:NPFR1	4475	3R:2034945	sim:T	mau:A
Dsim\GD19843	mel:NPFR1	4562	3R:2035032	sim:T	mau:C
Dsim\GD19843	mel:NPFR1	4883	3R:2035353	sim:T	mau:C
Dsim\GD19849	mel:Osi8	89	3R:2094767	sim:ACCCC	mau:ACCC
Dsim\GD19849	mel:Osi8	275	3R:2094953	sim:GT	mau:G
Dsim\GD19849	mel:Osi8	513	3R:2095191	sim:A	mau:AGG
Dsim\GD19849	mel:Osi8	743	3R:2095421	sim:CTTTTTTTTT	mau:CTTTTTTTT

Dsim\GD19849	mel:Osi8	868	3R:2095546	sim:T	mau:C
Dsim\Obp83ef	mel:Obp83ef	141	3R:1963207	sim:T	mau:A
Dsim\Obp83ef	mel:Obp83ef	760	3R:1963826	sim:TTATATTTTTTATATA	mau:TTATA
Dsim\Obp83ef	mel:Obp83ef	821	3R:1963887	sim:TGG	mau:TGGGG
Dsim\Obp83ef	mel:Obp83ef	910	3R:1963976	sim:G	mau:C
Dsim\GD19841	mel:Osi1	401	3R:2022030	sim:C	mau:T
Dsim\GD19841	mel:Osi1	521	3R:2022150	sim:G	mau:A
Dsim\GD19841	mel:Osi1	581	3R:2022210	sim:A	mau:C
Dsim\GD19841	mel:Osi1	691	3R:2022320	sim:T	mau:C
Dsim\GD19846	mel:Osi3	106	3R:2057905	sim:A	mau:G
Dsim\GD19846	mel:Osi3	190	3R:2057989	sim:C	mau:T
Dsim\GD19846	mel:Osi3	205	3R:2058004	sim:CTTTTTTTTT	mau:CTTTTTTTT,CTTTTTTT
Dsim\GD19846	mel:Osi3	322	3R:2058121	sim:A	mau:T
Dsim\GD19846	mel:Osi3	405	3R:2058204	sim:GTATTTAT	mau:GTATTTATTTAT
Dsim\GD19846	mel:Osi3	1121	3R:2058920	sim:CAAAA	mau:CAAAAA
Dsim\GD19846	mel:Osi3	1244	3R:2059043	sim:G	mau:A
Dsim\GD19846	mel:Osi3	1380	3R:2059179	sim:A	mau:G
Dsim\GD19846	mel:Osi3	1439	3R:2059238	sim:C	mau:T
Dsim\GD19846	mel:Osi3	1490	3R:2059289	sim:A	mau:G
Dsim\GD19846	mel:Osi3	1589	3R:2059388	sim:C	mau:T
Dsim\GD19848	mel:Osi6	177	3R:2077595	sim:T	mau:C
Dsim\GD19848	mel:Osi6	441	3R:2077859	sim:C	mau:T
Dsim\GD19848	mel:Osi6	561	3R:2077979	sim:G	mau:A
Dsim\GD19848	mel:Osi6	812	3R:2078230	sim:T	mau:C
Dsim\GD19848	mel:Osi6	903	3R:2078321	sim:C	mau:T,G
Dsim\GD19848	mel:Osi6	1353	3R:2078771	sim:C	mau:T
Dsim\GD19557	mel:CG31559	172	3R:2006506	sim:C	mau:G
Dsim\GD19557	mel:CG31559	628	3R:2006962	sim:G	mau:A
Dsim\GD19	mel:CG31559	1007	3R:2007341	sim:G	mau:A

557					
Dsim\GD19557	mel:CG31559	2077	3R:2008411	sim:TC	mau:TCATTGC
Dsim\GD19557	mel:CG31559	2183	3R:2008517	sim:A	mau:T
Dsim\GD19557	mel:CG31559	2299	3R:2008633	sim:CAAAAAA	mau:CAAAAA
Dsim\GD19557	mel:CG31559	2759	3R:2009093	sim:GAGCAACAGCAGCAGCAACAGCAGCAGCA	mau:GAGCAACAGCAGCAGCA
Dsim\GD19557	mel:CG31559	2952	3R:2009286	sim:T	mau:C
Dsim\GD19557	mel:CG31559	4240	3R:2010574	sim:GC	mau:G
Dsim\GD19557	mel:CG31559	4440	3R:2010774	sim:AT	mau:A
Dsim\GD19557	mel:CG31559	4995	3R:2011329	sim:A	mau:C
Dsim\GD19557	mel:CG31559	5053	3R:2011387	sim:ATAT	mau:ATATTAT
Dsim\GD19557	mel:CG31559	5344	3R:2011678	sim:T	mau:A
Dsim\GD19557	mel:CG31559	5486	3R:2011820	sim:AATTTATAT	mau:AAT
Dsim\GD19557	mel:CG31559	5754	3R:2012088	sim:A	mau:G,C
Dsim\GD19557	mel:CG31559	5812	3R:2012146	sim:T	mau:C
Dsim\GD19557	mel:CG31559	6063	3R:2012397	sim:C	mau:G
Dsim\GD19557	mel:CG31559	6112	3R:2012446	sim:T	mau:C
Dsim\GD19557	mel:CG31559	6214	3R:2012548	sim:CGGG	mau:CGG
Dsim\GD19557	mel:CG31559	6276	3R:2012610	sim:C	mau:T
Dsim\GD19557	mel:CG31559	6474	3R:2012808	sim:T	mau:G
Dsim\GD19561	mel:Gasp	252	3R:1946656	sim:T	mau:A
Dsim\GD19561	mel:Gasp	440	3R:1946844	sim:C	mau:T
Dsim\GD19561	mel:Gasp	706	3R:1947110	sim:G	mau:A
Dsim\GD19561	mel:Gasp	1086	3R:1947490	sim:C	mau:G
Dsim\GD19561	mel:Gasp	1149	3R:1947553	sim:TA	mau:TAA,TGGTGCATCCCTAA
Dsim\GD19561	mel:Gasp	1316	3R:1947720	sim:A	mau:G
Dsim\GD19561	mel:Gasp	1493	3R:1947897	sim:T	mau:C
Dsim\GD19561	mel:Gasp	1519	3R:1947923	sim:AGGTGTAGGGAATGGG	mau:AGG
Dsim\GD19561	mel:Gasp	1943	3R:1948347	sim:T	mau:A
Dsim\GD19	mel:Gasp	2223	3R:1948627	sim:ATGTGTGTGTG	mau:ATGTGTGTGTG

561				TGTGTG	TGTGTGTG
Dsim\GD19561	mel:Gasp	2766	3R:1949170	sim:A	mau:G
Dsim\GD19561	mel:Gasp	3083	3R:1949487	sim:A	mau:T
Dsim\GD19561	mel:Gasp	3307	3R:1949711	sim:A	mau:G
Dsim\GD19561	mel:Gasp	3494	3R:1949898	sim:CAA	mau:CAAA
Dsim\GD19561	mel:Gasp	3595	3R:1949999	sim:GCC	mau:GC
Dsim\GD19561	mel:Gasp	3998	3R:1950402	sim:A	mau:G
Dsim\GD19561	mel:Gasp	4514	3R:1950918	sim:GC	mau:GCC
Dsim\GD19561	mel:Gasp	4667	3R:1951071	sim:AGCTCCTGCTCCTGCTCCTGCT	mau:AGCTCCTGCTCCTGCT
Dsim\GD19561	mel:Gasp	4986	3R:1951390	sim:T	mau:A
Dsim\GD19561	mel:Gasp	5312	3R:1951716	sim:C	mau:A
Dsim\GD19561	mel:Gasp	5885	3R:1952289	sim:T	mau:C,G
Dsim\GD19561	mel:Gasp	5905	3R:1952309	sim:TAAAAAAAA	mau:TAAAAAAAAA
Dsim\GD19561	mel:Gasp	6071	3R:1952475	sim:GCCCCCCCC	mau:GCCCCCCC
Dsim\GD19561	mel:Gasp	6652	3R:1953056	sim:C	mau:T
Dsim\GD19561	mel:Gasp	6715	3R:1953119	sim:T	mau:C
Dsim\GD19561	mel:Gasp	6793	3R:1953197	sim:G	mau:T
Dsim\GD19561	mel:Gasp	7329	3R:1953733	sim:TTTTCGTTTCGTTTC	mau:TTTTCGTTTCGTTTCGTTTC
Dsim\Obp83cd	mel:Obp83cd	65	3R:1961592	sim:C	mau:A
Dsim\Obp83cd	mel:Obp83cd	389	3R:1961916	sim:T	mau:C
Dsim\Obp83cd	mel:Obp83cd	792	3R:1962319	sim:A	mau:T
Dsim\Obp83cd	mel:Obp83cd	867	3R:1962394	sim:AA	mau:AAAATGA
Dsim\GD19844	mel:Osi24	210	3R:2046272	sim:G	mau:A
Dsim\GD19844	mel:Osi24	264	3R:2046326	sim:T	mau:C
Dsim\GD19844	mel:Osi24	849	3R:2046911	sim:AATGTTTAT	mau:AAT
Dsim\GD19844	mel:Osi24	2001	3R:2048063	sim:G	mau:C
Dsim\GD19844	mel:Osi24	2317	3R:2048379	sim:G	mau:A
Dsim\GD19844	mel:Osi24	2399	3R:2048461	sim:C	mau:T
Dsim\GD19838	mel:-	432	3R:1989928	sim:CTTTTTTTTTT	mau:CTTTTTTTT,CTTTTTTTTTTT

Dsim\GD19838	mel:-	641	3R:1990137	sim:T	mau:TAAC
Dsim\GD19838	mel:-	2101	3R:1991597	sim:A	mau:G
Dsim\GD19838	mel:-	2127	3R:1991623	sim:TCCCC	mau:TCCC
Dsim\GD19838	mel:-	2340	3R:1991836	sim:G	mau:T
Dsim\GD19838	mel:-	2393	3R:1991889	sim:C	mau:A
Dsim\GD19838	mel:-	2767	3R:1992263	sim:A	mau:T
Dsim\GD19838	mel:-	2774	3R:1992270	sim:A	mau:G
Dsim\GD19838	mel:-	2839	3R:1992335	sim:C	mau:A
Dsim\GD19554	mel:Osi4	77	3R:2060376	sim:C	mau:T
Dsim\GD19554	mel:Osi4	316	3R:2060615	sim:TTCTTCACTTATAAAATATC	mau:T
Dsim\GD19554	mel:Osi4	645	3R:2060944	sim:A	mau:G
Dsim\GD19554	mel:Osi4	765	3R:2061064	sim:CTT	mau:CT
Dsim\GD19554	mel:Osi4	954	3R:2061253	sim:TCCCCCC	mau:TCCCCC
Dsim\GD19554	mel:Osi4	1072	3R:2061371	sim:A	mau:C
Dsim\GD19554	mel:Osi4	1173	3R:2061472	sim:A	mau:G
Dsim\GD19554	mel:Osi4	1297	3R:2061596	sim:C	mau:T
Dsim\GD19554	mel:Osi4	1399	3R:2061698	sim:C	mau:G
Dsim\GD19554	mel:Osi4	1585	3R:2061884	sim:A	mau:AAAC
Dsim\GD19554	mel:Osi4	1687	3R:2061986	sim:G	mau:GT
Dsim\GD19554	mel:Osi4	2069	3R:2062368	sim:GC	mau:GCGCTGCTGCC
Dsim\GD19554	mel:Osi4	2368	3R:2062667	sim:C	mau:T
Dsim\GD19554	mel:Osi4	2840	3R:2063139	sim:A	mau:T
Dsim\GD19554	mel:Osi4	2855	3R:2063154	sim:A	mau:G
Dsim\GD19554	mel:Osi4	2895	3R:2063194	sim:C	mau:T
Dsim\GD19842	mel:CG33301,CG16898	333	3R:2025889	sim:C	mau:T
Dsim\GD19842	mel:CG33301,CG16898	861	3R:2026417	sim:G	mau:A
Dsim\GD19842	mel:CG33301,CG16898	876	3R:2026432	sim:C	mau:T
Dsim\GD19842	mel:CG33301,CG16898	906	3R:2026462	sim:GGAATTGGAG	mau:GG
Dsim\GD19	mel:CG33301,CG	1180	3R:2026736	sim:G	mau:A

842	16898				
Dsim\GD19842	mel:CG33301,CG16898	1232	3R:2026788	sim:C	mau:G
Dsim\GD19558	mel:-	56	3R:1969285	sim:TGTCCT	mau:TGTCCTTGCCGTCCT
Dsim\GD19558	mel:-	96	3R:1969325	sim:T	mau:A
Dsim\GD19558	mel:-	107	3R:1969336	sim:A	mau:G
Dsim\GD19558	mel:-	420	3R:1969649	sim:TAAAA	mau:TAAA
Dsim\GD19558	mel:-	436	3R:1969665	sim:ACTGCTAGT	mau:ACTGCTAGTTCTGCTAGT
Dsim\GD19555	mel:-	74	3R:2042841	sim:GTGT	mau:GTGTTTGT
Dsim\GD19556	mel:-	64	3R:2020986	sim:C	mau:T
Dsim\GD19847	mel:Osi5	40	3R:2069929	sim:A	mau:G
Dsim\GD19847	mel:Osi5	387	3R:2070276	sim:G	mau:A
Dsim\GD19847	mel:Osi5	609	3R:2070498	sim:C	mau:T
Dsim\GD19847	mel:Osi5	661	3R:2070550	sim:TGACATCGACATC	mau:TGACATC
Dsim\GD19847	mel:Osi5	979	3R:2070868	sim:C	mau:T
Dsim\Obp83g	mel:Obp83g	103	3R:1965138	sim:G	mau:T
Dsim\Obp83g	mel:Obp83g	410	3R:1965445	sim:A	mau:G
Dsim\GD19837	mel:Vha14-2	402	3R:1976141	sim:G	mau:A
Dsim\GD19840	mel:CG1077	207	3R:2015194	sim:C	mau:T
Dsim\GD19840	mel:CG1077	246	3R:2015233	sim:T	mau:G
Dsim\GD19840	mel:CG1077	508	3R:2015495	sim:C	mau:G
Dsim\GD19840	mel:CG1077	565	3R:2015552	sim:T	mau:C
Dsim\GD19840	mel:CG1077	602	3R:2015589	sim:C	mau:A
Dsim\GD19840	mel:CG1077	610	3R:2015597	sim:G	mau:A
Dsim\GD19840	mel:CG1077	1098	3R:2016085	sim:T	mau:G
Dsim\GD19840	mel:CG1077	1239	3R:2016226	sim:C	mau:T
Dsim\GD19840	mel:CG1077	1347	3R:2016334	sim:T	mau:C
Dsim\GD19840	mel:CG1077	1706	3R:2016693	sim:A	mau:T
Dsim\GD19840	mel:CG1077	2368	3R:2017355	sim:A	mau:G
Dsim\GD19840	mel:CG1077	2495	3R:2017482	sim:T	mau:C

Dsim\GD19840	mel:CG1077	2506	3R:2017493	sim:G	mau:A
Dsim\GD19840	mel:CG1077	2716	3R:2017703	sim:G	mau:C
Dsim\GD19840	mel:CG1077	2832	3R:2017819	sim:A	mau:G
Dsim\GD19839	mel:-	74	3R:2009286	sim:T	mau:C
Dsim\GD19845	mel:Osi2	267	3R:2053474	sim:C	mau:T
Dsim\GD19845	mel:Osi2	483	3R:2053690	sim:C	mau:G
Dsim\GD19845	mel:Osi2	1091	3R:2054298	sim:G	mau:C
Dsim\GD19845	mel:Osi2	1581	3R:2054788	sim:C	mau:A

[a] Because a published reference of *D. mauritiana* is not available, we produced these variant calls using public and other (unpublished) short read sequencing.