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.