

Fine-Mapping Nicotine Resistance Loci in *Drosophila* Using a Multiparent Advanced Generation Inter-Cross Population

Tara N. Marriage,* Elizabeth G. King,[†] Anthony D. Long,[†] and Stuart J. Macdonald*¹

*Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66045, and [†]Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697

ABSTRACT Animals in nature are frequently challenged by toxic compounds, from those that occur naturally in plants as a defense against herbivory, to pesticides used to protect crops. On exposure to such xenobiotic substances, animals mount a transcriptional response, generating detoxification enzymes and transporters that metabolize and remove the toxin. Genetic variation in this response can lead to variation in the susceptibility of different genotypes to the toxic effects of a given xenobiotic. Here we use *Drosophila melanogaster* to dissect the genetic basis of larval resistance to nicotine, a common plant defense chemical and widely used addictive drug in humans. We identified quantitative trait loci (QTL) for the trait using the DSPR (*Drosophila* Synthetic Population Resource), a panel of multiparental advanced intercross lines. Mapped QTL collectively explain 68.4% of the broad-sense heritability for nicotine resistance. The two largest-effect loci—contributing 50.3 and 8.5% to the genetic variation—map to short regions encompassing members of classic detoxification gene families. The largest QTL resides over a cluster of ten UDP-glucuronosyltransferase (UGT) genes, while the next largest QTL harbors a pair of cytochrome P450 genes. Using RNAseq we measured gene expression in a pair of DSPR founders predicted to harbor different alleles at both QTL and showed that *Ugt86Dd*, *Cyp28d1*, and *Cyp28d2* had significantly higher expression in the founder carrying the allele conferring greater resistance. These genes are very strong candidates to harbor causative, regulatory polymorphisms that explain a large fraction of the genetic variation in larval nicotine resistance in the DSPR.

A routine part of life for all organisms is avoiding, and if necessary metabolizing, toxic substances encountered in the environment. A common challenge for animals are those toxins produced by potential prey and plant hosts as chemical defenses against predation and herbivory (Glendinning 2002, 2007). Understanding how organisms overcome these defenses can give us insight into the evolution of host specialization, which can often involve an organism overcoming the defenses of a particular host, avoiding competition by making use of a resource toxic to other species (for example, Hungate *et al.* 2013). Many animals, especially insects, are also commonly exposed to chemical pesticides used to protect crop plants. As a consequence of this strong evolutionary pressure,

there are a number of examples of insecticide resistance arising in natural populations (Crow 1957). Understanding the biology and molecular genetics underlying resistance to insecticides (Perry *et al.* 2011; Ffrench-Constant 2013) is valuable in the design of pest management strategies. In addition, humans are frequently exposed to an array of potentially harmful compounds, notably pharmaceuticals. Given the desire to achieve maximal drug efficacy while minimizing dosage and avoiding adverse drug responses, elaborating the mechanisms of drug metabolism, and the genetic factors that affect it, is critically important for human health.

Animals use a cascade of enzymes to metabolize xenobiotic compounds into less harmful substances (Xu *et al.* 2005; Li *et al.* 2007), and organisms possess many hundreds of genes whose products are involved in detoxification reactions. The best known class of phase I detoxification enzymes are the cytochrome P450 genes (P450s) that carry out oxidation, and other reactions on a broad range of compounds, typically decreasing their toxicity. The products of P450 reactions become the substrates for phase II enzymes, such as glutathione-S-transferases and UGTs. These enzymes add

Copyright © 2014 by the Genetics Society of America
doi: 10.1534/genetics.114.162107

Manuscript received February 15, 2014; accepted for publication May 25, 2014
Raw sequencing reads have been deposited in the Sequence Read Archive (SRA)
under project accession number SRP036167.

Supporting information is available online at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162107/-/DC1>.

¹Corresponding author: Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Ave., Lawrence, KS 66045. E-mail: sjmac@ku.edu

large, charged groups to substrate molecules, and the resulting molecules are more hydrophilic, and thus more readily excreted. Finally, in phase III a range of membrane transporters, including ATP-binding cassette (ABC) transporters remove the conjugated products of phase II metabolism from the cell. Despite our general understanding of the series of molecular events involved in detoxification, hundreds of detoxification genes have been identified in sequenced genomes (for instance, Strode *et al.* 2008; Chung *et al.* 2009; You *et al.* 2013), and for most xenobiotics the precise series of enzymes involved in their metabolism *in vivo* are unknown.

Here we sought to explore the genetic factors responsible for metabolic resistance to nicotine in the fruitfly. We are interested in nicotine for three broad reasons: First, nicotine is generated by a number of plant species, for example tobacco (*Nicotiana* sp.), as a defense against herbivory (Steppuhn *et al.* 2004). Nicotine presents a potent toxin to most herbivores, and only a few insect species are known to feed on nicotine-producing plants. Notably, the facultative tobacco specialist *Manduca sexta* (the tobacco hornworm) detoxifies ingested nicotine by inducing P450 enzymes (Snyder and Glendinning 1996). Second, nicotine has itself been used as an insecticide, and various pesticides that are chemically similar to nicotine—neonicotinoid pesticides—are in wide use (Goulson 2013). Third, nicotine is an extensively used addictive compound in humans, and nicotine dependence leads to a considerable number of tobacco-related, preventable deaths (Mokdad *et al.* 2004; Jha *et al.* 2013).

The elite model organism *Drosophila* has proven a valuable system to understand the molecular genetics of drug responses (Kaun *et al.* 2012) and pesticide resistance (Perry *et al.* 2011). Indeed, one of the best described cases of the genetic control of xenobiotic resistance is the role of *Cyp6g1* in resistance to the pesticide dichlorodiphenyltrichloroethane (DDT) in *D. melanogaster*. Overexpression of this gene, as a result of a series of naturally occurring gene duplication events, and the recruitment of enhancer sequences carried by various transposable elements (TEs) inserted upstream of the gene, has a major effect on DDT resistance (Daborn *et al.* 2002; Chung *et al.* 2007; Schmidt *et al.* 2010).

Early work identified *D. melanogaster* strains showing reduced mortality in the presence of nicotine (Hall *et al.* 1978), and screens of wild-derived lines revealed considerable genetic variation for nicotine resistance (Carrillo and Gibson 2002; Passador-Gurgel *et al.* 2007). Furthermore, expression of a number of genes, including the P450 *Cyp6g1*, have been implicated in mediating increased survival time on nicotine (Passador-Gurgel *et al.* 2007; Li *et al.* 2012). However, the complete catalog of *Drosophila* genes has yet to be interrogated for any role in resistance to nicotine.

A powerful method of identifying naturally segregating genetic variants contributing to complex trait variation is to use a large set of recombinant genotypes derived from an advanced generation intercross among several founding genomes (Churchill *et al.* 2004; Macdonald and Long 2007; Kover *et al.* 2009; Svenson *et al.* 2012; Threadgill and

Churchill 2012; Baud *et al.* 2013). We have constructed the DSPR to allow such genetic dissection in *D. melanogaster* (King *et al.* 2012a,b). The DSPR is a stable genetic reference panel consisting of >1700 recombinant inbred lines (RILs) from two different eight-way synthetic recombinant populations. These populations were maintained for 50 generations before initiating inbred lines, and the multiple rounds of recombination to which the populations were subjected allows high QTL mapping resolution (Darvasi and Soller 1995; Cheng *et al.* 2010). In addition, the increased number of founders relative to traditional two-way QTL mapping gives better representation of the allelic diversity in the species and the potential to characterize the phenotypic effects of multiple alleles at a locus (for example, Kislukhin *et al.* 2013).

Using a larval nicotine resistance assay, we screened >800 DSPR RILs, and succeeded in mapping QTL that collectively explained almost 70% of the broad-sense heritability for nicotine resistance, including one QTL that alone contributed ~50% of the genetic variation. Following genomewide expression profiling via RNAseq, we identified genes within mapped QTL intervals that showed differential gene expression between founders predicted to harbor low- and high-resistance alleles and/or between nicotine-free and nicotine-supplemented media. The large-effect QTL maps to a cluster of 10 UGT genes, with *Ugt86Dd* the most likely expression candidate, showing significantly higher constitutive expression in the resistant founder, and expression induction following exposure to nicotine. The second largest effect QTL maps to a pair of P450 genes (*Cyp28d1* and *Cyp28d2*) both of which show higher constitutive expression in the resistant founder. The small set of genes we identify in this initial screen represent a valuable starting point to dissect the process of nicotine metabolic detoxification in *Drosophila* and will ultimately facilitate the resolution of the precise causative sequence variants.

Materials and Methods

Mapping population

We employed RILs from the DSPR to map QTL for nicotine resistance. Full details of the resource and its development can be found in King *et al.* (2012a), and details of the analytical methods, including simulations demonstrating the power and mapping resolution of the DSPR, are presented in King *et al.* (2012b).

Briefly, the DSPR consists of >1700 RILs derived from a pair of eight-way synthetic populations (pA and pB), each generated by intercrossing a different set of 7 highly inbred founder lines (A1–A7 or B1–B7), along with one founder common to both populations (AB8). Each synthetic population was maintained as two independent replicate subpopulations (pA1, pA2, pB1, and pB2) at large population size for 50 generations, undergoing multiple rounds of recombination. Subsequently, >800 RILs were generated from each

population via 25 generations of full sibling mating. This design results in a set of RILs whose genomes are a fine-scale mosaic of segments from 8 founder lines. All 15 founder lines have been completely resequenced to $\sim 50\times$, and the RILs were genotyped via restriction site-associated DNA (RAD) tags, which along with a hidden Markov model (HMM) allowed the underlying founder haplotype structure of each RIL to be determined. HMM-derived RIL genotypes are freely available at www.FlyRILs.org and on Dryad (doi.org/10.5061/dryad.r5v40).

Larval nicotine resistance assay

Flies from each RIL were allowed to lay eggs for 48 hr on standard cornmeal–yeast–molasses food, supplemented with 0.5% activated charcoal to aid visualization of larvae, and active yeast paste to encourage females to lay. We constructed custom egg-laying chambers by cutting standard narrow polypropylene vials ~ 1 inch from the base, filling the base with food, and taping the two sections of the vial back together. This allowed flies to lay in a familiar environment and facilitated manual larval collection from the surface of the media in the base of the chamber under a stereoscope. Following egg laying, 60 first instar larvae were collected from each line.

Our assay counts the number of first instar larvae that ultimately emerge as adults in the presence of nicotine. To increase throughput we elected to raise test individuals in 24-well plates (EK-2053, E&K Scientific). Each well contained 3 ml of standard fly media, with 12 wells on each plate holding control food, and 12 holding food supplemented with 0.18 $\mu\text{l/ml}$ nicotine (N3876, Sigma). This nicotine concentration was selected after a series of *ad hoc* test experiments seeking to identify a value that would discriminate among RILs. Media for all experimental replicates were prepared fresh 24 hr before larval collection, and nicotine was added to the molten fly food at $\sim 50^\circ$ to minimize volatilization. For each line, we placed 30 first instar larvae in a control food well, and another 30 in a nicotine food well, assaying 12 lines in a single plate. Plates were sealed with porous, breathable tape (3M Transpore white surgical tape), and maintained at 25° and 50% relative humidity on a 12 hr light/12 hr dark cycle for 14 days. At this point, plates were frozen at -20° , and we counted the number of adults that had emerged in each well. We did not sex the emerging adults.

We carried out the assay described above on 810 pA lines, testing each line a maximum of once per block, and in three to nine replicate blocks (mean = 4.9). Lines were arbitrarily assigned to replicate blocks, to assay plates within blocks, and to positions in plates, through the use of anonymous barcodes. Given the very large effect associated with the QTL on 3R, we additionally screened 464 pB lines using a much lower level of replication (one to four replicates, mean = 1.6) in an attempt to validate this QTL. Raw phenotype data are presented in [Supporting Information, Table S1](#).

Heritability

We estimated the broad-sense heritability of our nicotine resistance measure (the fraction of 30 first instar larvae emerging as adults on nicotine food) by calculating the genetic and phenotypic variance components from a linear mixed model using the `lme` and `VarCorr` functions in the `nlme` package in R (Pinheiro *et al.* 2011). The model used was

$$y_{i,j} = \mu + g_i + \varepsilon_{i,j},$$

where $y_{i,j}$ is the j th observation of the i th RIL, μ is the grand mean, g_i is the RIL-specific random effect, and $\varepsilon_{i,j}$ is the associated error. We estimated the heritability of RIL means as the estimated genetic variance component over the total variance of RIL means. We obtained a heritability estimate solely for the pA population due to the paucity of replicate assays carried out for the pB RILs.

QTL mapping

The analytical framework for QTL mapping in the DSPR is explained in King *et al.* (2012a,b). Briefly, for each region in each RIL our HMM assigns a probability that the genotype is one of the 36 possible founder genotype combinations (eight homozygous possibilities, 28 heterozygous possibilities). Since the vast majority of the RIL genomes are homozygous we converted these probabilities to eight additive probabilities by assuming any heterozygous states are intermediate between the respective pair of homozygous states. To perform QTL mapping, we regress the mean line phenotype on the eight additive probabilities, including subpopulation as a covariate, analyzing the pA and pB panels separately. Genomewide significance thresholds were determined via 1000 permutations of the data (Churchill and Doerge 1994). Confidence intervals on the locations of QTL were taken as 2-LOD support intervals, which simulations suggest provides 93–95% confidence intervals given our experimental design, sample size, and mapped QTL effect sizes (King *et al.* 2012a). All data and R scripts required to regenerate our QTL mapping results are available on Dryad (doi.org/10.5061/dryad.r5v40).

QTL phasing

After estimating the effect of each founder genotype at each QTL for each population, we categorized the founders into QTL allelic classes. We first assigned a hard genotype to RILs that had a $>95\%$ probability of harboring a particular founder at a QTL, ignoring those RILs where the founder of origin was uncertain. We then ranked founder genotype means and fit a series of models testing all possible two-way partitions of the founders, choosing the partition with the highest F -statistic. Using an F -test, this optimal two-class partition model was compared to the model with all founders belonging to a single group. Assuming the two-group model provides a better fit than the single group model ($P < 0.0001$), this partition was fixed and all possible

three-class models (that include the fixed optimal two-class partition) were tested. This process continued until adding an additional partition did not result in a significantly improved fit. We fit the general model

$$y_{i,j} = \mu + A_i + \varepsilon_{i,j},$$

where $y_{i,j}$ is the j th observation of the i th allelic group (the number of allelic groups depends on the number of partitions being tested), μ is the grand mean, A_i is the allelic group effect, and $\varepsilon_{i,j}$ is the associated error.

RNAseq

Our mapping identified a pair of large-effect QTL in pA, and we sought to determine whether any genes implicated in these intervals have expression differences between (a) lines that appear to harbor “high” and “low” alleles, and (b) control-food and nicotine-food treatments. We collected first instar larvae from founder lines A3 (low nicotine resistance) and A4 (high nicotine resistance) and placed them onto control media or nicotine-containing media (0.18 μ l/ml) for 4 hr. One hundred larvae were then harvested for each strain/treatment combination and snap frozen in liquid nitrogen. RNA from each of the four samples was isolated using TRIzol reagent (15596-018, Life Technologies) and cleaned through RNeasy Mini columns (74104, Qiagen) following the manufacturer’s protocols. RNAseq libraries were constructed using the Illumina TruSeq kit and sequenced over four lanes of an Illumina HiSeq2500 instrument to generate single-end 100-bp reads (Genome Sequencing Facility, University of Kansas Medical Center).

We used sickle (version 1.200, github.com/najoshi/sickle) to trim raw sequencing reads, and TopHat (version 2.0.9, tophat.cbcb.umd.edu; Trapnell *et al.* 2009; Kim *et al.* 2013) to assemble reads from each sample to the *D. melanogaster* reference genome (National Center for Biotechnology Information build 5.3, tophat.cbcb.umd.edu/igenomes.shtml). Following quality trimming we had 180–183 million reads per sample, and 79.5–84.0% of these aligned to the reference genome. Subsequently we used Cuffdiff (version 2.1.1, cufflinks.cbcb.umd.edu; Trapnell *et al.* 2010, 2013) to identify differentially expressed genes (see [File S1](#) for details of the code and settings). For each comparison between lines and/or treatments we considered only those tests that were successfully executed (“status” column in “gene_exp.diff” Cuffdiff output file has “OK” flag), and unless otherwise stated only considered genes to be significantly differentially expressed if they survived a Benjamini–Hochberg correction for multiple testing ($q < 0.05$).

Testing effects of variants under QTL

We have previously identified SNPs segregating among the founders (King *et al.* 2012a), and annotated these variants with SnpEff (Cingolani *et al.* 2012). Using the estimated mosaic haplotype structure of each RIL, we inferred for every variant present in the founders the probability that each

RIL harbors the minor allele. For each variant within mapped QTL intervals we then fit a model associating phenotype with genotype, while simultaneously correcting for subpopulation, to test whether individual variants are capable of explaining the QTL peaks (see Kislukhin *et al.* 2013).

Results

Phenotypic variation among RILs

For a large number of DSPR RILs we measured the fraction of first instar larvae that ultimately emerged as adults in two different conditions; control medium, and medium supplemented with nicotine. Our initial goal in examining the fraction of adults to emerge under nicotine-free conditions was to control for any natural variation in larva-to-adult viability among RILs. Indeed, we succeeded in mapping QTL for viability on control media ([Figure S1](#), [Table S2](#)). However, across pA RILs we observed a very strong correlation between mean viability on nicotine-supplemented food and the ratio of this measure to the mean viability on control food (Pearson’s $r = 0.83$, $P < 0.001$). In addition, following QTL analysis in pA, genome-wide LOD scores for both measures are nearly identical, and the same QTL are mapped (data not shown). Hence, we use the simpler measure—larva-to-adult viability on nicotine food, averaged over replicate trials—as our measure of nicotine resistance for the remainder of this study.

Nicotine resistance shows marked variation across RILs ([Figure 1](#)) with the mean phenotype showing a minimum value of zero (no larvae emerge as adults) and a maximum value of 0.96 (the vast majority of larvae emerge as adults) across pA RILs. There is a significant difference between the two pA subpopulations in mean RIL nicotine resistance (Welch’s $t = 9.3$, $P < 0.001$) and we accounted for this in our subsequent QTL mapping by including subpopulation as a covariate in the analysis.

We estimated the broad-sense heritability of our measure of nicotine resistance (the fraction of 30 first instar larvae that emerge as adults on nicotine food) in the pA population as 0.70, with similar values obtained from separate analyses of the subpopulations (data not shown). We employed RIL means for QTL mapping, reducing the environmental variance, and the broad-sense heritability of this mean measure of resistance is 0.91 in pA.

Characterizing nicotine resistance QTL

Our genome scan in the pA population revealed four autosomal QTL contributing to nicotine resistance ([Figure 2](#), [Table 1](#), [Table S3](#)), including one major-effect QTL on 3R that alone explains around half of the broad-sense heritability for the trait. We also identified this latter QTL, Q4, in the pB population despite scoring fewer RILs using fewer assay replicates (see *Materials and Methods*), likely due to the magnitude of the effect associated with this locus. The other three QTL identified in pA were not reconfirmed in pB,

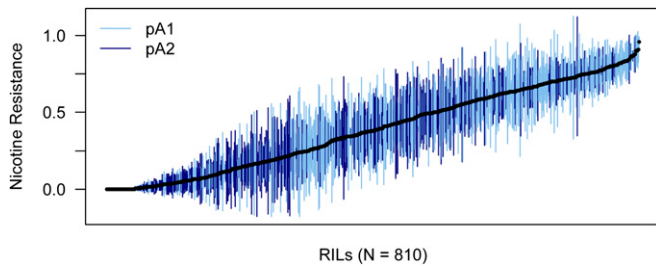


Figure 1 Means (filled circles) and 1 SDs (vertical lines) for nicotine resistance across 810 pA RILs (463 pA1 RILs and 347 pA2 RILs). Mean nicotine resistance is the fraction of first instar larvae that emerge as adults on nicotine-supplemented media, averaged over replicate tests.

although we do not know whether this is due to the absence of the causative variants in the pB population or whether the reduced mapping power in pB precluded their detection. In the pA population, our mapped QTL explain 68.4% of the broad-sense heritability for larval nicotine resistance.

Three of the four QTL (Q1, Q2, and Q4) were mapped to relatively short chromosomal regions of 150–260 kb harboring 15–40 protein-coding genes, raising the possibility of identifying a very small number of highly likely candidate causative genes. In contrast, Q3 maps to a broad 1.4-Mb interval since it lacks a well-defined LOD peak, and the LOD score at the peak is only slightly above threshold (Figure 2). In addition, unlike the other three QTL, nicotine resistance QTL Q3 overlaps a QTL contributing to viability on control food (Figure S1, Table S2). This implies Q3 most likely represents a locus contributing to general viability rather than strictly to viability in the presence of nicotine.

For each of the four mapped loci, we estimated the phenotypic effects associated with the eight founder genotypes (Figure 3) and attempted to “phase” the QTL by placing founders into QTL allelic groups (see *Materials and Methods*). The founder effects plot for Q1 suggests this QTL is biallelic, with founders A2 and A4 harboring the high allele, and the remaining six founders harboring the low allele. Our phasing supports this, highlighting two founder groups, A2/A4 and A1/A3/A5/A6 (the phasing of A7 and AB8 is ambiguous since the probability that these founders are members of the most likely group is <0.95). For Q2, our phasing also suggests two allelic groups are present, A4/A5 and A2/A3/A6/A7/AB8 (founder A1 is ambiguous), with a similar result for Q3 (A4/A5 and A2/A6/A7/AB8, with founders A1 and A3 being ambiguous). Q4 does not show a simple biallelic pattern of founder effects in pA, and our phasing shows strong support for three classes, A2, A3, and A4/A7, whereas in pB the founders fall into two groups, B2/B3/B4 and B6/B7/AB8. Since strain AB8, which was used to found both panels of RILs, is only present at an appreciable frequency at the Q4 locus in population pB, clarifying the relationship between the groups we identify in pA and in pB is difficult. It is plausible that the effect at Q4 is due to an allelic series, as we (Kislukhin *et al.* 2013) and others (Baud *et al.* 2013) have observed previously for QTL mapped in

panels founded by multiple parental genotypes. However, given that Q4 maps to a region containing a family of detoxification genes (see below), it is perhaps more likely that multiple genes are collectively responsible for generating the signal at Q4.

Identifying candidate causative genes

As shown in Table 1, the number of protein-coding genes implicated in each mapped QTL interval is relatively small. In an effort to further refine the list of candidates, we used an approach inspired by Wayne and McIntyre (2002) and conducted an RNAseq experiment using tissue from whole first instar larvae from founders A3 and A4 and two experimental treatments: control media and media containing nicotine. These two strains were selected because our phasing above suggested A3 harbors a low, relatively susceptible allele, and A4 harbors a high, relatively resistant allele, for three QTL: Q1, Q2, and Q4. Notably, while founders A3 and A4 have similar larva-to-adult viabilities under control conditions, they show a strong difference in nicotine resistance in our assay (Welch's $t = 5.5$, $P < 10^{-4}$; Table S4); Following 11 replicate trials, A3 had a mean resistance phenotype of 0.09 (1 SD = 0.177) and A4 a mean phenotype of 0.54 (1 SD = 0.199).

Following RNA isolation and cleanup, Illumina library generation and sequencing, read trimming, and alignment of the reads to the reference genome, we identified differentially expressed genes across four contrasts: (a) A3 control vs. A4 control, (b) A3 nicotine vs. A4 nicotine, (c) A3 control vs. A3 nicotine, and (d) A4 control vs. A4 nicotine. The first two contrasts allow us to detect genes that are expressed at different levels in the two lines, while the last two contrasts allow us to detect genes that change in expression in response to nicotine. Summing over all four QTL regions, a total of 30 genes showed at least a nominal ($P < 0.05$) change in expression in at least one contrast (Q1 = 3, Q2 = 4, Q3 = 17, and Q4 = 6), and a total of 9 genes exhibited differential expression that survived a per-contrast genome-wide false discovery rate (FDR) threshold of 5% in at least one contrast (Q1 = 2, Q2 = 1, Q3 = 4, and Q4 = 2). Details of this latter set of genes are presented in Table 2, and full details of all differentially expressed genes ($P < 0.05$) identified over all contrasts are presented in Table S5.

Q1 contains 34 protein-coding genes (Table 1) of which two, *Cyp28d1* and *Cyp28d2*, are members of the cytochrome P450 family of detoxification enzymes. Genes of this class are strong *a priori* candidates to contribute to variation in resistance to xenobiotics, and both of these genes show significantly higher constitutive gene expression in A4 than in A3 (Table 2, Figure 4), lines predicted to harbor high and low alleles at Q1, respectively (Figure 3). Furthermore, in founder A3, the *Cyp28d1* gene appears to be induced in response to nicotine (Table 2, Figure 4). These data are consistent with one or both of these genes being the causative factor(s) underlying Q1, with higher gene expression corresponding to a higher level of resistance to nicotine.

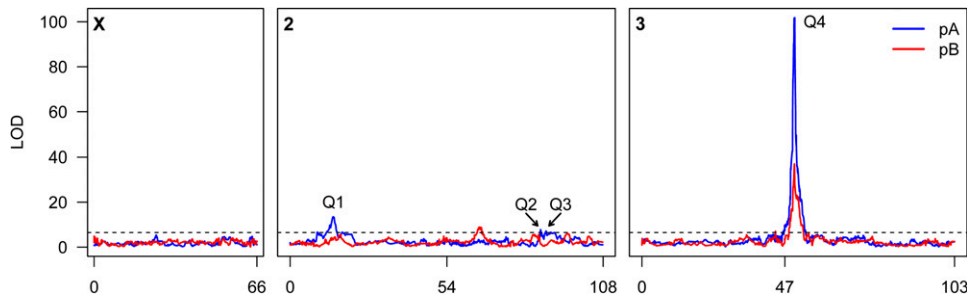


Figure 2 Genome scan for nicotine resistance QTL. Solid curves are the scans for the pA (blue) and pB (red) data, and the horizontal dotted line represents the genome-wide 5% permutation threshold (LOD = 6.5 for both populations). Genetic distances along the chromosomes are indicated along the x-axis. The centromeres are at positions 54 and 47 on chromosomes 2 and 3, respectively. The positions of the four QTL we describe in the text are indicated

on the plot for ease of reference. (Note that while we do identify an above-threshold peak in pB on chromosome 2 at genetic position 65 (LOD = 8.9), we are reluctant to consider this a confirmed QTL given the limited number of lines and replicate assays employed in the B population. See *Discussion*.)

Q2 encompasses 40 genes, including one cytochrome P450 family member, *Cyp12b2*. This gene showed no significant expression variation, even at a nominal level ($P < 0.05$), in our RNAseq study, although it does harbor two nonsynonymous sites that segregate among the pA founders (King *et al.* 2012a). Thus, while *Cyp12b2* appears not to contribute to nicotine resistance via a change in expression level, we cannot rule out a causative role for structural variation at this locus. The best expression candidate for Q2 is *CG15080*, a gene that shows higher constitutive expression on control food in line A4 than in line A3, and induction of gene expression in response to nicotine in A3 (Table 2). There is very limited information available on *CG15080* in FlyBase (Marygold *et al.* 2013), with just one report noting a cuticle defect following RNAi knockdown (Adler *et al.* 2013). Two other genes within the Q2 interval showed expression variation due to nicotine treatment in line A3 at the 5% level; *CG15093* and *l(2)03709* both decreased in expression in response to nicotine ($P = 0.025$ and $P = 0.048$, respectively). All other contrasts for these three genes were nonsignificant.

Q3 spans the largest number of protein-coding genes of any QTL mapped in this study (143, Table 1) and also encompasses large arrays of 5S rRNA and tRNA genes. None of the genes represent classic families of detoxification genes (e.g., P450s, glutathione S-transferases, and so on). This is perhaps to be expected since Q3 likely represents a general viability locus, rather than a locus specifically involved in nicotine resistance. In the RNAseq experiment, four genes under Q3 showed strong differential expression in at least one of the four founder/treatment contrasts (Table 2). However, in contrast to the other QTL, it is not clear from our data that lines A3 and A4 possess different alleles at Q3, so it is possible that these differentially expressed genes within the Q3 interval are not causally related to differential viability.

The largest effect QTL we identified, Q4 maps to just 32 genes in pA and 15 in pB. Ten of these are UGT genes. This gene family, of which a total of 33 are present in the *D. melanogaster* genome (Luque and O'Reilly 2002), are involved in glucuronidation, the addition of glucuronic acid to a substrate to make it more soluble in water, and thus easier to excrete. Under Q4, the only protein-coding genes that showed expression variation among samples at $P < 0.05$ were UGT genes. Four UGT genes showed no significant

expression changes (*Ugt86Da*, *Ugt86Dg*, *Ugt86Di*, and *Ugt86Dj*), although in the case of *Ugt86Di*, this could simply be a result of low expression, since FPKM (fragments per kilobase of transcript per million mapped reads) values in all samples are < 1 . Four additional UGT genes showed a slight change in expression in at least one contrast: *Ugt86De* exhibited lower expression in A4 compared to A3 in both control food and nicotine food ($P < 0.01$ in both cases), *Ugt35a* had higher expression in A4 in both treatment conditions ($P < 0.05$ in both), *Ugt86Dh* had higher expression in A4 than in A3 on control food only ($P < 0.05$), and *Ugt35b* had higher expression in A4 on nicotine food only ($P < 0.05$). Again, low read counts may have reduced our power to detect differential gene expression for *Ugt35b*.

Two UGT genes under Q4 showed differential expression in at least one contrast that survived correction for multiple testing (Table 2). *Ugt86Dc* had lower expression in A4 than in A3 in both control and nicotine food ($P < 0.001$ and $P < 0.0001$, respectively). Since A4 is predicted to carry the more resistant allele, this is not the pattern expected under the assumption that higher expression equates to higher amounts of gene product, leading to greater resistance. *Ugt86Dd* does show this pattern (Table 2, Figure 4) with the gene showing higher constitutive expression in A4 ($P < 0.0001$) in both treatment conditions. In addition, nicotine appeared to slightly induce *Ugt86Dd* expression in both lines ($P < 0.05$ in both cases; Table 2, Figure 4). Thus, *Ugt86Dd* represents the best expression candidate underlying Q4. Nevertheless, the range of expression variation for the other UGT genes at this locus, the lack of a simple biallelic pattern of founder effects in pA (Figure 3), and the presence of strong associations at nonsynonymous SNPs in a number of UGT genes (below) suggest that more than one UGT gene may be responsible for the effect observed at Q4.

Screening for candidate causative variants

In the DSPR, and other Multiparent Advanced Generation Inter-Cross (MAGIC), QTL are identified by associating variation among segregating founder haplotypes with variation in line phenotype. When a QTL signal is due to a single variant, in addition to the QTL being explained by haplotype variation, it will also be explained by the genotype at the causative variant. To attempt to discover such variants, we

Table 1 Details of mapped QTL

Name	LOD score	Chr	Peak cM (2-LOD C.I.) ^a	Peak Mb (2-LOD C.I.) ^a	No. of genes ^b	Percent of H^2 ^c
Q1 (pA)	13.5	2L	14.8 (14.3–15.3)	5.10 (4.99–5.25)	34	8.5
Q2 (pA)	7.8	2R	86.3 (86.1–86.7)	14.57 (14.51–14.73)	40	5.0
Q3 (pA)	7.2	2R	88.2 (87.5–91.7)	15.26 (15.03–16.42)	143	4.6
Q4 (pA)	101.8	3R	50.3 (50.1–50.4)	7.04 (6.83–7.06)	32	50.3
Q4 (pB)	37.0	3R	50.2 (50.1–50.3)	6.94 (6.85–7.00)	15	45.2

^a 2-LOD C.I. indicates the 2-LOD support interval (~95% confidence interval) of the QTL.

^b The number of protein-coding genes in the 2-LOD support interval.

^c The percentage of broad-sense heritability (H^2) explained by the QTL.

carried out association tests at SNPs under all four mapped QTL, using SNP genotypes imputed using the mosaic haplotype structure of each RIL. Results of these local, QTL-centric association scans are presented in Figure S2.

In no case is there a single SNP (or physically close group of SNPs) that shows a marked stronger association with phenotype than other SNPs in the region, and no associations clearly implicate particular genes within QTL intervals. Given the linkage disequilibrium (LD) structure of the DSPR, it is likely that large numbers of SNPs tag the exact same haplotypes, or combinations of haplotypes, making delineation of the causative SNPs via association in the DSPR very difficult. This phenomenon is easily seen in Figure S2, where large collections of SNPs spanning a QTL region lead to nearly identical P -values.

One notable feature of the localized association tests is that SNPs predicted to have severe impacts on protein function (loss of start codons, loss and gain of stop codons, changes in splice sites) do not show stronger associations with phenotype than other SNPs (Figure S2). Twenty sites under the four QTL are assigned such “high impact” effects by SnpEff (Cingolani *et al.* 2012), and none have a P -value in the top 9% of associations per QTL interval, suggesting such events contribute little to the overall genetic variation for nicotine resistance in the DSPR.

A number of segregating nonsynonymous SNPs (nsSNPs) are present beneath mapped QTL (Q1 = 330, Q2 = 238, Q3 = 523, and Q4 = 149). Although no nsSNP shows the strongest association signal in the QTL-centric association tests (Figure S2), an nsSNP has a P -value in the top 0.2% of associations for QTL Q1, Q2, and Q3, and in the top 3% of associations for Q4. There are a number of nsSNPs within the P450 genes implicated by our QTL (*Cyp28d1* and *Cyp28d2* under Q1, and *Cyp12b2* under Q2), but none are strongly associated with phenotype. The most significant nsSNP in these genes has a P -value five orders of magnitude above the minimum value in each interval. In contrast, a number of highly significant nsSNPs are present in the UGT gene cluster in Q4, including several in the top 5% of associations in the region. Thus, while we can exclude nsSNPs in P450 genes as major contributors to nicotine resistance variation in the DSPR, structural variation in UGT genes could play an important role.

Given the association between TE insertion, heightened expression of *Cyp6g1*, and resistance to DDT in *Drosophila*

(Daborn *et al.* 2002; Chung *et al.* 2007; Schmidt *et al.* 2010), and previous observations of the contribution of TE insertions to complex trait variation (Mackay and Langley 1990; Shrimpton *et al.* 1990; Long *et al.* 2000; Gruber *et al.* 2007), we sought to identify TEs segregating in the DSPR in the P450 and UGT candidate gene regions implicated by our mapping and RNAseq. The only TE in the *Cyp28d1/Cyp28d2* region is present in the *Cyp28d1* gene solely in founder A1 (Cridland *et al.* 2013). This TE is unlikely to contribute to Q1 because RILs carrying the A1 haplotype at this position do not have a unique phenotype distinguishable from all other RILs (Figure 3). In addition, our RNAseq study that revealed significant differential expression of *Cyp28d1* (Table 2) did not employ founder A1.

Seven of the 10 UGT genes under Q4 have a TE in or within 1 kb of the gene (Figure S3). Five of these seven TEs are carried by pA founders represented at the chromosomal position of Q4 (A2, A3, A4, or A7; Figure 3), and one is carried by a pB founder present at Q4 (B6; Figure 3). These TEs could conceivably contribute to Q4, but since no single TE is present at appreciable frequencies in both populations at this position, the same TE cannot be causative in both pA and pB.

Discussion

A large number of genes have the potential to be involved in variable responses to xenobiotic compounds: the target receptors of the compounds (Ffrench-Constant *et al.* 1991, 1994), the enzymes involved in the three phases of drug metabolism, such as P450s, and those *trans*-acting factors that initiate the detoxification response by inducing P450 expression (Misra *et al.* 2011). Here we made use of a large mapping population—the DSPR—derived from a collection of naturally derived alleles, to carry out an unbiased genetic screen for loci contributing to variation in xenobiotic toxicity in *Drosophila*.

Larval nicotine resistance is highly heritable in the DSPR, and in the pA panel of ~800 RILs we successfully mapped four QTL that collectively explain 68.4% of the broad-sense heritability for the trait (Table 1). Our estimates of the QTL effects are possibly biased upwards due to the Beavis effect (Beavis 1994), although the large sample sizes we employed (~800 RILs) implies any overestimation is minimal (Xu 2003). Regardless, QTL Q4 contributes a very large fraction

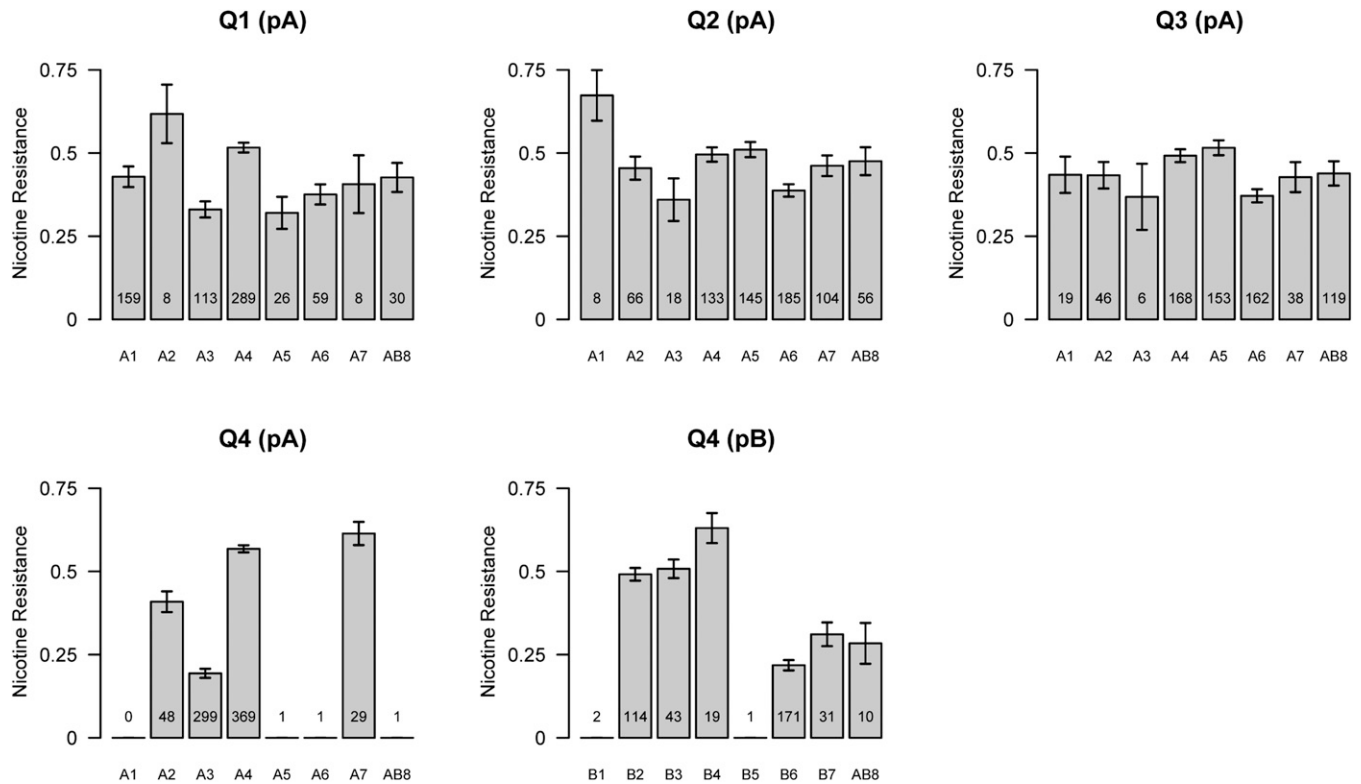


Figure 3 Founder haplotype means and 1 SDs at all mapped QTL. The number of RILs for which we confidently assign a founder genotype (probability > 0.95) is listed at the bottom of each bar, and only founder means associated with at least five observations are presented. We note that at the genomic region harboring Q4, several founder alleles have been almost completely purged from our panels of RILs.

to the genetic variation, perhaps explaining as much as 50.3% of the heritability for nicotine resistance. Our followup RNA-seq experiment allowed us to identify several genes under mapped QTL that showed differential gene expression between founder lines predicted to harbor high and low alleles, and/or exhibited changes in gene expression on exposure to nicotine (Table 2). The pair of cytochrome P450 genes under Q2 (*Cyp28d1* and *Cyp28d2*) and a UDP-glycosyltransferase under Q4 (*Ugt86Dd*) represent likely candidates for a causative role in nicotine resistance in *Drosophila*.

Our identification of detoxification pathway genes beneath the two largest-effect QTL implies that the primary response of *Drosophila* larvae to nicotine exposure, at least in our assay, is to attempt to metabolize the compound into nontoxic, excretable molecules. The RNAseq data further imply that regulatory, rather than structural variation at these detoxification loci is responsible for a large fraction of the variation in the resistance phenotype. This interpretation is supported by experiments showing that P450 expression can be induced in the presence of xenobiotics in flies (Willoughby *et al.* 2006; Chung *et al.* 2011) and other organisms (Snyder and Glendinning 1996), and by studies demonstrating that lines resistant to insecticides can show higher levels of expression of one or more P450 genes than susceptible lines (for example, Daborn *et al.* 2002). Studies demonstrating that *cis*-regulatory variation at P450 loci associates with both increased expression levels and heightened insecticide resistance are also consistent

with our observations (Dombrowski *et al.* 1998; Daborn *et al.* 2002; Chung *et al.* 2007; Schmidt *et al.* 2010). Nonetheless, a number of nsSNPs in UGT genes exhibit strong associations with phenotype in our QTL-centric association scans. Given the haplotype structure of the DSPR, such association tests have difficulty discriminating causative sites from those non-functional polymorphisms that are in LD with true causative variants. Thus, we cannot exclude a role for nsSNPs in UGT genes in nicotine resistance. Future work will be required to delineate the contributions of regulatory variation, particularly at *Ugt86Dd*, and structural variation at UGT genes to the major effect on phenotype observed at QTL Q4.

Comparison with previous studies on nicotine in flies

Previously, Passador-Gurgel *et al.* (2006) used microarray-based expression profiling of 4385 genes in adult female heads from a large sample of wild-derived inbred lines, testing each line on control food, and after a short, 8-hr exposure to nicotine. Genes putatively involved in nicotine resistance were defined as those whose transcript abundance was correlated with survival time on nicotine across the lines. None of the genes implicated in this study are within our mapped QTL intervals. However, in some cases we observe significant expression variation at genes implicated by Passador-Gurgel *et al.* (2006). For instance, the alkaline phosphatase-encoding gene *CG8147* whose expression correlated with survival in Passador-Gurgel *et al.* (2006) showed strongly increased

Table 2 Differential expression of candidate genes under QTL

QTL	Gene	FPKM ^a				Fold change (<i>P</i> -value) ^b			
		A3 (C)	A4 (C)	A3 (N)	A4 (N)	A3 (C) vs. A4 (C)	A3 (N) vs. A4 (N)	A3 (C) vs. A3 (N)	A4 (C) vs. A4 (N)
Q1	<i>Cyp28d1</i>	3.5	186.2	55.6	172.5	5.75 (<0.0001*)	1.63 (0.005)	4.01 (<0.0001*)	-0.11 (0.843)
Q1	<i>Cyp28d2</i>	6.0	54.7	2.7	50.6	3.18 (<0.0001*)	4.22 (<0.0001*)	-1.15 (0.102)	-0.11 (0.829)
Q2	<i>CG15080</i>	3.7	14.0	16.0	24.8	1.91 (0.002)	0.63 (0.215)	2.11 (<0.001*)	0.83 (0.102)
Q3	<i>CG18607</i>	21.1	1.1	6.4	0.2	-4.26 (0.001*)	-4.70 (0.002)	-1.71 (0.019)	-2.15 (0.072)
Q3	<i>Cpr56F</i>	12.9	38.5	62.5	82.5	1.58 (0.007)	0.40 (0.441)	2.28 (<0.001*)	1.10 (0.036)
Q3	<i>mus209</i>	111.3	101.6	29.8	51.0	-0.13 (0.793)	0.77 (0.146)	-1.90 (<0.001*)	-0.99 (0.054)
Q3	<i>Obp56d</i>	53.6	26.4	292.0	188.4	-1.02 (0.092)	-0.63 (0.215)	2.45 (<0.001*)	2.84 (<0.0001*)
Q4	<i>Ugt86Dc</i>	4.5	0.4	6.3	0.3	-3.65 (0.001)	-4.37 (<0.001*)	0.50 (0.455)	-0.22 (0.827)
Q4	<i>Ugt86Dd</i>	4.0	51.9	10.6	128.9	3.70 (<0.0001*)	3.60 (<0.0001*)	1.41 (0.039)	1.31 (0.012)

^a Fragments per kilobase of transcript per million mapped reads (values from Cuffdiff output files) for each of the four line/treatment combinations. C, control food; N, nicotine food.

^b The log₂ fold change (second sample divided by first sample) in gene expression for each of the four contrasts, and the associated *P*-value, not corrected for multiple tests. An asterisk indicates the test survives a per-contrast FDR of 5% (*i.e.*, $q < 0.05$).

expression following nicotine exposure in line A4 (Table S5). The general lack of correspondence between the two studies could be due to a range of issues. First, Passador-Gurgel *et al.* (2006) survey less than half of the genes in the *Drosophila* genome, and specifically their array did not include *Cyp28d2*, and included only 3 of the 10 UGT genes under Q4. Second, even with 800 RILs phenotyped, the power to map small-effect QTL (*i.e.*, those contributing $\leq 2.5\%$ to the phenotypic variance) in the DSPR is low (King *et al.* 2012b), and it is possible any contributions of loci identified by Passador-Gurgel *et al.* (2006) to nicotine resistance in the DSPR are simply too small to yield above-threshold QTL. Furthermore, the differences in expression observed by Passador-Gurgel *et al.* (2006) could be at genes downstream of those harboring segregating variation generating those differences, and would not be mapped in our study. Third, the studies used different life stages/tissues (first instar larvae *vs.* adult heads), and it is possible adults and larvae respond to nicotine challenge via different pathways. Finally, the panel of lines used to found the DSPR is of worldwide distribution, while Passador-Gurgel *et al.* (2006) sampled lines from North Carolina and California. Any population specificity in the response to nicotine could render an effect difficult to detect in another sample. In this context, it is notable that the two population samples used by Passador-Gurgel *et al.* (2006) showed different distributions of survival time on nicotine, and there was no overlap in the transcripts correlating with survival time in each population.

A previous study in *D. melanogaster* showed that over-expression of the *Cyp6g1* gene (via insertion of an *Accord* TE upstream of the gene) leads to nicotine resistance (Li *et al.* 2012). We find no evidence of a QTL surrounding this gene in the pA population, despite the *Accord* insertion being present in founder A6. However, this founder haplotype is only represented in pA by five RILs, likely making any true effect difficult to detect. In pB, we do observe a small, above-threshold peak (Figure 2; 2R, 64.5–66.2 cM, 7.50–8.44 Mb, 8.9 LOD) that encompasses 160 protein-coding genes including *Cyp6g1*. We chose not to focus on this QTL given the small number of pB RILs assayed, the much lower level

of assay replication employed in the pB population compared to the pA population, and the absence of this QTL in pA. Nevertheless, it is possible the *Cyp6g1* gene has an effect on nicotine resistance in the pB panel of DSPR RILs, and that additional RIL screening would clearly map a QTL implicating this gene. One factor to consider is that while this small pB-specific peak appears to be driven by founder haplotype B5 (Figure S4), in that carriers of this haplotype have higher resistance than other RILs, only founder B4 possesses the known *Accord* enhancer of *Cyp6g1* gene expression and the proposed mediator of nicotine resistance at this locus (Li *et al.* 2012).

Resolution of expression candidate genes

Since regulatory variation is thought to be responsible for a large fraction of complex trait variation (Gibson and Weir 2005; Gilad *et al.* 2008; Cookson *et al.* 2009), we chose to employ RNAseq on founders to confirm that genes within QTL peaks showed differential expression on exposure to nicotine and/or between lines harboring different resistance QTL alleles. While our results point to strong *a priori* candidates within QTL Q1 and Q4, clearly such an experiment is unable to detect genes harboring structural changes that affect phenotype without affecting transcript levels. Although it appears that severe changes to protein-coding genes, such as premature stop codon polymorphisms, do not contribute to our identified QTL (Figure S2), there are many nonsynonymous sites within QTL that could, individually, collectively, or in addition to the regulatory changes we observe, contribute to nicotine resistance in the DSPR. Genes harboring functional, structural variants could be identified by screening all genes within QTL statistically via quantitative complementation tests (Long *et al.* 1996; Pasyukova *et al.* 2000), assuming appropriate deficiencies and/or single gene mutations are available. It is also possible to functionally test all genes in an interval via RNAi to yield the causative gene (for instance, Bergland *et al.* 2012), although off-target effects could be of particular concern when testing closely related members of gene families such as the tandemly duplicated P450 genes we implicate here.

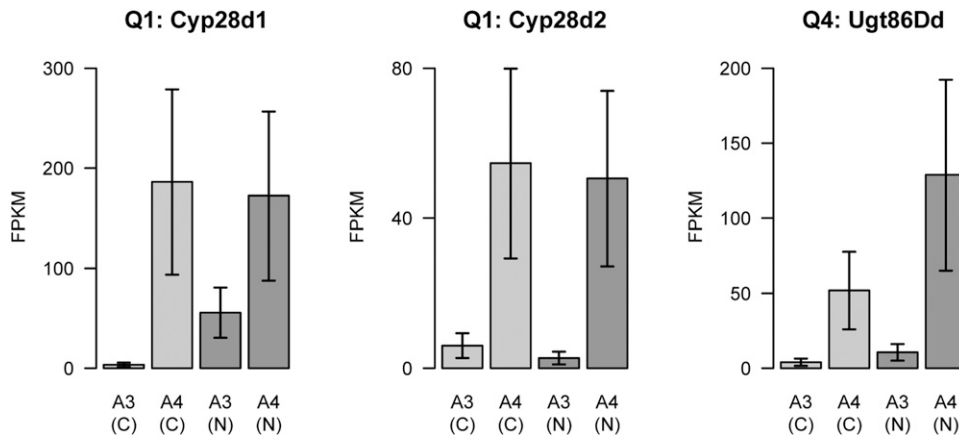


Figure 4 Expression of three candidate genes in first instar larvae from lines A3 and A4 on both control (C) and nicotine-supplemented (N) food. Bars show the expression level—represented by the FPKM score—for the gene, and the vertical lines are the 95% confidence interval on these values (all taken directly from Cuffdiff output files). Statistical tests comparing expression measures across lines and treatments are presented in Table 2.

Our RNAseq study should be taken as preliminary evidence for the role of expression variation at *Cyp28d1/Cyp28d2* and *Ugt86Dd* in influencing nicotine resistance. We selected lines A3 and A4 specifically because our mapping data suggested they harbor alternate alleles at QTL Q1 and Q4. However, the lines obviously harbor different alleles at most other loci, and these allelic differences will frequently result in variation in transcript abundance (Stamatoyannopoulos 2004). Thus, while unlikely, it is not impossible that the between-line differential expression of detoxification genes under QTL that we see is not due to *cis*-regulatory variants in promoter or enhancer regions of these genes, but is instead due to variants in *trans*-regulatory proteins. In contrast, the induction of *Cyp28d1* expression, and to a lesser extent that of *Ugt86Dd*, we see in response to nicotine is perhaps more compelling evidence for these genes' role in nicotine resistance. However, a number of other genes outside of QTL intervals show differential expression in response to nicotine (Table S5), and some of these are also known to be involved in detoxification. The presence of strong expression candidates within QTL intervals greatly increases our confidence that such genes harbor causative polymorphisms, but merely seeing a change in expression in response to a toxin does not guarantee these genes causally mediate differential resistance to the drug.

The lack of tissue specificity in our RNAseq dataset, since we extracted RNA from groups of whole first instar larvae, does not allow us to determine the tissue, or tissues where the increased gene expression we see is relevant to detoxification. However, it is likely that the key tissues involved in xenobiotic metabolism are the midgut, the fat body, and the malpighian tubules (Perry *et al.* 2011). Using *in situ* hybridization for the majority of the P450 genes in *Drosophila*, Chung *et al.* (2009) found that *Cyp28d1* was expressed in the malpighian tubules of late third instar larvae, while *Cyp28d2* was expressed in third instar larval gonads. The malpighian tubules are involved in excretion and osmoregulation, so this expression data suggest that *Cyp28d1* are more likely to have a causative role in nicotine resistance than *Cyp28d2*. Nevertheless, since *Cyp28d1* appears to generally have higher expression than *Cyp28d2*

across many tissues (Yang *et al.* 2007), it is conceivable that *in situ* hybridization experiments are unable to reliably distinguish low levels of *Cyp28d2* gene product from background. Future work, for example, by directing RNAi against our candidate genes to the malpighian tubules, is required to determine the tissue specificity of gene action.

Potential for finding the causative loci

The DSPR enables causative loci to be mapped to relatively short genetic intervals encompassing small numbers of genes (Table 1), but the haplotype structure of the population will nevertheless typically preclude easy identification of the precise causative site, particularly if such sites are regulatory in origin, as is suggested by our RNAseq results. Nevertheless, our localized, QTL-specific association tests can be ranked by the strength of the association, and those sites with associations considerably below the highest score in the region could be reasonably excluded from consideration, leaving relatively short lists of candidate associated sites. Considering only those associations with *P*-values within five orders of magnitude of the smallest *P*-value in each QTL region, we implicate 570/4243 (Q1), 1409/5457 (Q2), 4805/25464 (Q3), 251/3265 (Q4 mapped in pA), and 49/3265 (Q4 mapped in pB) SNPs as plausible candidate nicotine resistance loci. Furthermore, since a total of only 110 and 270 SNPs are present in the regions surrounding *Cyp28d1/Cyp28d1* and *Ugt86Dd*, respectively, resolving the causative, perhaps regulatory, nucleotide polymorphisms in these genes is less daunting.

If we assume the causative changes are indeed regulatory in nature, one possible way to narrow the list of likely candidate genes and sites further would be to make use of DNase I sequencing (Boyle *et al.* 2008) to measure chromatin accessibility—DNase I-hypersensitive sites (DHSs)—in first instar larvae from a number of inbred strains and map the location of putative regulatory elements (Thomas *et al.* 2011). Since expression QTL are often associated with DHSs (Degner *et al.* 2012; Gaffney *et al.* 2012), any polymorphism segregating within a DHS represents an excellent candidate to regulate nearby genes and contribute to phenotypic variation.

Resolving the actual causative loci underlying our QTL might also be achieved by carrying out a genome-wide association study (GWAS) using a panel of naturally derived inbred genotypes (Mackay *et al.* 2012), by comparing allele frequency variation between pools of susceptible and resistant genotypes (Lai *et al.* 2007; Huang *et al.* 2012, Bastide *et al.* 2013) or by measuring allele frequency divergence in divergently selected populations (Burke *et al.* 2010; Turner *et al.* 2011, 2013). Often these studies are complicated by the very conservative statistical thresholds that must be employed to avoid false positives. For instance, with close to 2.5 million SNPs segregating within the DGPR (Mackay *et al.* 2012), statistical thresholds approaching $P < 10^{-8}$ are required to correct for multiple tests in a genome-wide scan. However, these thresholds can be relaxed considerably, based on the positions of mapped QTL or expression candidates or by giving additional weight to those SNPs we identify here as having strong associations in the DSPR. For example, adequate control of the type I error rate in a series of association tests localized to the *Cyp28d1/Cyp28d1* region could be achieved with $P < 10^{-4}$. Assuming the genetic architecture of trait variation is preserved between mapping panels, these additional methodologies could succeed in mapping the precise causative variants.

In summary, using the DSPR we have dissected the genetic basis of nicotine toxicity in flies, identifying a small number of relatively large-effect factors that appear to explain the bulk of the genetic variance. This observation is in stark contrast to the general picture from human (Park *et al.* 2010; Yang *et al.* 2010) and *Drosophila* GWAS (Mackay *et al.* 2012) where large numbers of very small-effect associations contribute to trait variation. This could simply be due to our choice of trait, since some human GWASs have identified large-effect causative variants, for instance those contributing to variation in warfarin dose (Cooper *et al.* 2008; Takeuchi *et al.* 2009). Given the resolution possible with the DSPR (King *et al.* 2012a,b) we were able to resolve causative loci to intervals encompassing fairly small numbers of genes, and RNAseq allowed us to implicate three strong candidate detoxification genes—*Cyp28d1*, *Cyp28d1*, and *Ugt86Dd*—for future confirmation. The haplotype structure of the DSPR precludes identification of the exact causative nucleotide variants as is sometimes possible with a population-based GWAS approach. Nevertheless, we can employ local, QTL-centric association tests using imputed RIL genotypes to rank variants within QTL intervals and target subsequent genetic and functional validation to the most likely candidate loci.

Acknowledgments

We thank Carla Burford, Denny Swartzlander, and Mike Najarro for technical support during nicotine resistance data collection. This work was supported by National Institutes of Health grant R01 RR024862/OD010974 to S.J.M. and A.D.L.

Literature Cited

- Adler, P. N., L. F. Sobala, D. Thom, and R. Nagaraj, 2013 *dusky-like* is required to maintain the integrity and planar cell polarity of hairs during the development of the *Drosophila* wing. *Dev. Biol.* 379: 76–91.
- Bastide, H., A. Betancourt, V. Nolte, R. Tobler, P. Stobe *et al.*, 2013 A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. *PLoS Genet.* 9: e1003534.
- Baud, A., R. Hermsen, V. Guryev, P. Stridh, D. Graham *et al.*, 2013 Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat. Genet.* 45: 767–775.
- Beavis, W. D., 1994 The power and deceit of QTL experiments: lessons from comparative QTL studies. *Proceedings of the 49th Annual Corn and Sorghum Industry Research Conference*. American Seed Trade Association, Washington, DC, pp. 250–266.
- Bergland, A. O., H. S. Chae, Y. J. Kim, and M. Tatar, 2012 Fine-scale mapping of natural variation in fly fecundity identifies neuronal domain of expression and function of an aquaporin. *PLoS Genet.* 8: e1002631.
- Boyle, A. P., S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies *et al.*, 2008 High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132: 311–322.
- Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467: 587–590.
- Carrillo, R., and G. Gibson, 2002 Unusual genetic architecture of natural variation affecting drug resistance in *Drosophila melanogaster*. *Genet. Res.* 80: 205–213.
- Cheng, R., J. E. Lim, K. E. Samocha, G. Sokoloff, M. Abney *et al.*, 2010 Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics* 185: 1033–1044.
- Chung, H., M. R. Bogwitz, C. McCart, and A. Andrianopoulos, R. H. French-Constant *et al.*, 2007 *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* 175: 1071–1077.
- Chung, H., T. Sztal, S. Pasricha, M. Sridhar, P. Batterham *et al.*, 2009 Characterization of *Drosophila melanogaster* cytochrome P450 genes. *Proc. Natl. Acad. Sci. USA* 106: 5731–5736.
- Chung, H., A. Boey, C. Lumb, L. Willoughby, P. Batterham *et al.*, 2011 Induction of a detoxification gene in *Drosophila melanogaster* requires an interaction between tissue specific enhancers and a novel *cis*-regulatory element. *Insect Biochem. Mol. Biol.* 41: 863–871.
- Churchill, G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963–971.
- Churchill, G. A., D. C. Airey, H. Allayee, J. M. Angel, A. D. Attie *et al.*, 2004 The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat. Genet.* 36: 1133–1137.
- Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w¹¹¹⁸; iso-2; iso-3*. *Fly (Austin)* 6: 80–92.
- Cookson, W., L. Liang, G. R. Abecasis, M. Moffatt, and M. Lathrop, 2009 Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* 10: 184–194.
- Cooper, G. M., J. A. Johnson, T. Y. Langae, H. Feng, I. B. Stanaway *et al.*, 2008 A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* 112: 1022–1027.
- Cridland, J. M., S. J. Macdonald, A. D. Long, and K. R. Thornton, 2013 Abundance and distribution of transposable elements in

- two *Drosophila* QTL mapping resources. *Mol. Biol. Evol.* 30: 2311–2327.
- Crow, J. F., 1957 Genetics of insect resistance to chemicals. *Annu. Rev. Entomol.* 2: 227–246.
- Daborn, P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil *et al.*, 2002 A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* 297: 2253–2256.
- Darvasi, A., and M. Soller, 1995 Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141: 1199–1207.
- Degner, J. F., A. A. Pai, R. Pique-Regi, J. B. Veyrieras, D. J. Gaffney *et al.*, 2012 DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482: 390–394.
- Dombrowski, S. M., R. Krishnan, M. Witte, S. Maitra, C. Diesing *et al.*, 1998 Constitutive and barbital-induced expression of the *Cyp6a2* allele of a high producer strain of *CYP6A2* in the genetic background of a low producer strain. *Gene* 221: 69–77.
- Ffrench-Constant, R. H., 1994 The molecular and population genetics of cyclodiene insecticide resistance. *Insect Biochem. Mol. Biol.* 24: 335–345.
- Ffrench-Constant, R. H., 2013 The molecular genetics of insecticide resistance. *Genetics* 194: 807–815.
- Ffrench-Constant, R. H., D. P. Mortlock, C. D. Shaffer, R. J. MacIntyre, and R. T. Roush, 1991 Molecular cloning and transformation of cyclodiene resistance in *Drosophila*: an invertebrate γ -aminobutyric acid subtype A receptor locus. *Proc. Natl. Acad. Sci. USA* 88: 7209–7213.
- Gaffney, D. J., J. B. Veyrieras, J. F. Degner, R. Pique-Regi, A. A. Pai *et al.*, 2012 Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* 13: R7.
- Gibson, G., and B. Weir, 2005 The quantitative genetics of transcription. *Trends Genet.* 21: 616–623.
- Gilad, Y., S. A. Rifkin, and J. K. Pritchard, 2008 Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* 24: 408–415.
- Glendinning, J. I., 2002 How do herbivorous insects cope with noxious secondary plant compounds in their diet? *Entomol. Exp. Appl.* 104: 15–25.
- Glendinning, J. I., 2007 How do predators cope with chemically defended foods? *Biol. Bull.* 213: 252–266.
- Goulson, D., 2013 An overview of the environmental risks posed by neonicotinoid insecticides. *J. Appl. Ecol.* 50: 977–987.
- Gruber, J. D., A. Genissel, S. J. Macdonald, and A. D. Long, 2007 How repeatable are associations between polymorphisms in *achaete-scute* and bristle number variation in *Drosophila*? *Genetics* 175: 1987–1997.
- Hall, L. M., R. W. Von Borstel, B. C. Osmond, S. D. Hoeltzli, and T. H. Hudson, 1978 Genetic variants in an acetylcholine receptor from *Drosophila melanogaster*. *FEBS Lett.* 95: 243–246.
- Huang, W., S. Richards, M. A. Carbone, D. Zhu, R. R. Anholt *et al.*, 2012 Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc. Natl. Acad. Sci. USA* 109: 15553–15559.
- Hungate, E. A., E. J. Earley, I. A. Boussy, D. A. Turissini, C. T. Ting *et al.*, 2013 A locus in *Drosophila sechellia* affecting tolerance of a host plant toxin. *Genetics* 195: 1063–1075.
- Jha, P., C. Ramasundarahettige, V. Landsman, B. Rostron, M. Thun *et al.*, 2013 21st-century hazards of smoking and benefits of cessation in the United States. *N. Engl. J. Med.* 368: 341–350.
- Kaun, K. R., A. V. Devineni, and U. Heberlein, 2012 *Drosophila melanogaster* as a model to study drug addiction. *Hum. Genet.* 131: 959–975.
- Kim, D., G. Perrea, C. Trapnell, H. Pimentel, R. Kelley *et al.*, 2013 TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14: R36.
- King, E. G., C. M. Merkes, C. L. McNeil, S. R. Hooper, S. Sen *et al.*, 2012a Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* 22: 1558–1566.
- King, E. G., S. J. Macdonald, and A. D. Long, 2012b Properties and power of the *Drosophila* Synthetic Population Resource for the routine dissection of complex traits. *Genetics* 191: 935–949.
- Kislukhin, G., E. G. King, K. N. Walters, S. J. Macdonald, and A. D. Long, 2013 The genetic architecture of methotrexate toxicity is similar in *Drosophila melanogaster* and humans. *G3 (Bethesda)* 3: 1301–1310.
- Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich *et al.*, 2009 A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet.* 5: e1000551.
- Lai, C. Q., J. Leips, W. Zou, J. F. Roberts, K. R. Wollenberg *et al.*, 2007 Speed-mapping quantitative trait loci using microarrays. *Nat. Methods* 4: 839–841.
- Li, X., M. A. Schuler, and M. R. Berenbaum, 2007 Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. *Annu. Rev. Entomol.* 52: 231–253.
- Li, X., S. Bai, and B. N. Cass, 2012 *Accord* insertion in the 5' flanking region of *CYP6G1* confers nicotine resistance in *Drosophila melanogaster*. *Gene* 502: 1–8.
- Long, A. D., S. L. Mullaney, T. F. Mackay, and C. H. Langley, 1996 Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. *Genetics* 144: 1497–1510.
- Long, A. D., R. F. Lyman, A. H. Morgan, C. H. Langley, and T. F. Mackay, 2000 Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. *Genetics* 154: 1255–1269.
- Luque, T., and D. R. O'Reilly, 2002 Functional and phylogenetic analyses of a putative *Drosophila melanogaster* UDP-glycosyltransferase gene. *Insect Biochem. Mol. Biol.* 32: 1597–1604.
- Macdonald, S. J., and A. D. Long, 2007 Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of *Drosophila melanogaster*. *Genetics* 176: 1261–1281.
- Mackay, T. F., and C. H. Langley, 1990 Molecular and phenotypic variation in the *achaete-scute* region of *Drosophila melanogaster*. *Nature* 348: 64–66.
- Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482: 173–178.
- Marygold, S. J., P. C. Leyland, R. L. Seal, J. L. Goodman, J. Thurmond *et al.*, 2013 FlyBase: improvements to the bibliography. *Nucleic Acids Res.* 41: D751–D757.
- Misra, J. R., M. A. Horner, G. Lam, and C. S. Thummel, 2011 Transcriptional regulation of xenobiotic detoxification in *Drosophila*. *Genes Dev.* 25: 1796–1806.
- Mokdad, A. H., J. S. Marks, D. F. Stroup, and J. L. Gerberding, 2004 Actual causes of death in the United States, 2000. *JAMA* 291: 1238–1245.
- Park, J. H., S. Wacholder, M. H. Gail, U. Peters, K. B. Jacobs *et al.*, 2010 Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* 42: 570–575.
- Passador-Gurgel, G., W. P. Hsieh, P. Hunt, N. Deighton, and G. Gibson, 2007 Quantitative trait transcripts for nicotine resistance in *Drosophila melanogaster*. *Nat. Genet.* 39: 264–268.
- Pasyukova, E. G., C. Vieira, and T. F. Mackay, 2000 Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. *Genetics* 156: 1129–1146.
- Perry, T., P. Batterham, and P. J. Daborn, 2011 The biology of insecticidal activity and resistance. *Insect Biochem. Mol. Biol.* 41: 411–422.

- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar; and R Development Core Team, 2011 nlme: linear and nonlinear mixed effects models. R package version 3.1–101.
- Schmidt, J. M., R. T. Good, B. Appleton, J. Sherrard, G. C. Raymant *et al.*, 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. *PLoS Genet.* 6: e1000998.
- Shrimpton, A. E., T. F. Mackay, and A. J. Brown, 1990 Transposable element-induced response to artificial selection in *Drosophila melanogaster*: molecular analysis of selection lines. *Genetics* 125: 803–811.
- Snyder, M. J., and J. I. Glendinning, 1996 Causal connection between detoxification enzyme activity and consumption of a toxic plant compound. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 179: 255–261.
- Stamatoyannopoulos, J. A., 2004 The genomics of gene expression. *Genomics* 84: 449–457.
- Steppuhn, A., K. Gase, B. Krock, R. Halitschke, and I. T. Baldwin, 2004 Nicotine's defensive function in nature. *PLoS Biol.* 2: E217.
- Strode, C., C. S. Wondji, J. P. David, N. J. Hawkes, N. Lumjuan *et al.*, 2008 Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. *Insect Biochem. Mol. Biol.* 38: 113–123.
- Svenson, K. L., D. M. Gatti, W. Valdar, C. E. Welsh, R. Cheng *et al.*, 2012 High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics* 190: 437–447.
- Takeuchi, F., R. McGinnis, S. Bourgeois, C. Barnes, N. Eriksson *et al.*, 2009 A genome-wide association study confirms *VKORC1*, *CYP2C9*, and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genet.* 5: e1000433.
- Thomas, S., X. Y. Li, P. J. Sabo, R. Sandstrom, R. E. Thurman *et al.*, 2011 Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biol.* 12: R43.
- Threadgill, D. W., and G. A. Churchill, 2012 Ten years of the collaborative cross. *G3 (Bethesda)* 2: 153–156.
- Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25: 1105–1111.
- Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan *et al.*, 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28: 511–515.
- Trapnell, C., D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn *et al.*, 2013 Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31: 46–53.
- Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet.* 7: e1001336.
- Turner, T. L., P. M. Miller, and V. A. Cochrane, 2013 Combining genome-wide methods to investigate the genetic complexity of courtship song variation in *Drosophila melanogaster*. *Mol. Biol. Evol.* 30: 2113–2120.
- Wayne, M. L., and L. M. McIntyre, 2002 Combining mapping and arraying: an approach to candidate gene identification. *Proc. Natl. Acad. Sci. USA* 99: 14903–14906.
- Willoughby, L., H. Chung, C. Lumb, C. Robin, P. Batterham *et al.*, 2006 A comparison of *Drosophila melanogaster* detoxification gene induction responses for six insecticides, caffeine and phenobarbital. *Insect Biochem. Mol. Biol.* 36: 934–942.
- Xu, S., 2003 Theoretical basis of the Beavis effect. *Genetics* 165: 2259–2268.
- Xu, C., C. Y. Li, and A. N. Kong, 2005 Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Arch. Pharm. Res.* 28: 249–268.
- Yang, J., C. McCart, D. J. Woods, S. Terhzaz, K. G. Greenwood *et al.*, 2007 A *Drosophila* systems approach to xenobiotic metabolism. *Physiol. Genomics* 30: 223–231.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42: 565–569.
- You, M., Z. Yue, W. He, X. Yang, G. Yang *et al.*, 2013 A heterozygous moth genome provides insights into herbivory and detoxification. *Nat. Genet.* 45: 220–225.

Communicating editor: B. E. Huang

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162107/-/DC1>

Fine-Mapping Nicotine Resistance Loci in *Drosophila* Using a Multiparent Advanced Generation Inter-Cross Population

Tara N. Marriage, Elizabeth G. King, Anthony D. Long, and Stuart J. Macdonald

File S1

Code used for RNAseq analysis.

Sickle (version 1.200)

Run on each of the four FASTQ files: A3-control (A3C), A3-nicotine (A3N), A4-control (A4C), and A4-nicotine (A4N)

```
sickle se -f A3C.fastq.gz -t sanger -o A3C.sickle.fastq -q 30 -l 30 -n  
gzip A3C.sickle.fastq
```

TopHat (version 2.0.9)

Run on each of the four trimmed FASTQ files

```
tophat2 -p 12 -G genes.gtf -o ./Assembly_A3C/ --no-novel-juncs --library-type  
fr-unstranded genome A3C.sickle.fastq.gz
```

Cufflinks (version 2.1.1)

Run on all four TopHat assemblies simultaneously

```
cuffdiff -N -o ./CuffDiff_Output/ -b genome.fa -p 12 -u genes.gtf  
-L A3C,A3N,A4C,A4N ./Assembly_A3C/accepted_hits.bam ./Assembly_A3N/accepted_hits.bam  
./Assembly_A4C/accepted_hits.bam ./Assembly_A4N/accepted_hits.bam
```

Table S1 Raw first-instar larval phenotypes measured on DSPR RILs.

Available for download as a .csv file at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162107/-/DC1>

The "Population" column indicates which population (A or B) the line is from. The "PopulationReplicate" column indicates which subpopulation (A1, A2, B1, or B2) the line is from. The "RIL" column is the line number. The "NicotineViability" column is the fraction of 30 first-instar larvae that emerge as adults on nicotine-supplemented media. The "ControlViability" column is the fraction of 30 first-instar larvae that emerge as adults on control, nicotine-free media. There are multiple rows for any given RIL, each row containing the data from a single replicate.

Table S2 Details of QTL mapped for larva-to-adult viability on control food

Name	LOD score	Chr	Peak cM (2-LOD CI) ^a	Peak Mb (2-LOD CI) ^a	Number of genes ^b
i (pA)	8.1	2R	82.8 (82.6–83.4)	13.44 (13.38–13.60)	50
ii (pA) ^c	8.9	2R	88.2 (88.0–89.9)	15.26 (15.19–15.86)	75
iii (pA)	9.2	3R	47.2 (47.2–47.3)	1.30 (1.21–1.47)	50
iv (pB) ^d	8.7	X	32.6 (31.8–33.2)	10.73 (10.52–10.89)	37
v (pB) ^d	7.5	X	39.4 (37.9–40.7)	12.44 (12.10–12.73)	41
vi (pB) ^d	8.4	2L-2R	53.8 (53.7–58.5)	2L:19.76 (2L:19.59–2R:4.44)	733

^a 2-LOD CI indicates the 2-LOD support interval of the QTL.

^b The number of protein-coding genes in the 2-LOD support interval.

^c The location of this control viability QTL overlaps that of Q3 (Table 1), a QTL contributing to viability on nicotine food.

^d Given the relatively small number of RILs phenotyped for population pB, coupled with the low level of within-line replication we recommend caution in interpreting QTL mapped in pB.

Table S3 LOD scores for genomewide QTL scans.

Available for download as a .csv file at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162107/-/DC1>

The "Chromosome" column indicates the chromosome arm of the position under test (X, 2L, 2R, 3L, and 3R). The "PhysicalPosition" and "GeneticPosition" columns indicate the physical and genetic positions of the site under test, respectively. The four "LODscore" columns present the LOD scores at each position for the two traits (nicotine resistance and larva-to-adult viability on control food) in each of the two mapping populations (pA and pB). The four "VarianceExplained" columns indicate the fraction of the phenotypic variance in trait/population that is explained by among-line genetic variation at the site under test.

Table S4 Viability measures for founder lines A3 and A4 used for RNAseq

Founder	Replicate ^a	Nicotine Viability ^b	Control Viability ^c
A3	1	0.233	0.367
A3	2	0	0.633
A3	3	0	0.6
A3	4	0	0.733
A3	5	0	0.867
A3	6	0	0.6
A3	7	0	0.833
A3	8	0	0.767
A3	9	0	0.867
A3	10	0.267	0.667
A3	11	0.533	0.733
		mean = 0.09	mean = 0.70
		SD = 0.177	SD = 0.147
A4	1	0.233	0.133
A4	2	0.3	0.9
A4	3	0.767	0.833
A4	4	0.833	0.8
A4	5	0.6	0.767
A4	6	0.5	0.7
A4	7	0.467	0.6
A4	8	0.533	0.5
A4	9	0.567	0.767
A4	10	0.767	0.733
A4	11	0.333	0.8
		mean = 0.54	mean = 0.68
		SD = 0.199	SD = 0.214

^a Each replicate assay involved 30 first-instar larvae.

^b The fraction of first-instar larvae emerging as adults on nicotine-supplemented media.

^c The fraction of first-instar larvae emerging as adults on control, nicotine-free media.

Table S5 All nominally significant ($p < 0.05$) differentially-expressed genes identified by RNAseq.

Available for download as a .csv file at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.162107/-/DC1>

The data is taken directly from the "gene_exp.diff" Cuffdiff output file, and simply trimmed to remove genes with per-contrast p -values above 5%. The "Gene" column is the symbol for the gene under test. The "Chromosome", "GeneStart", and "GeneStop" columns give the position of the gene in the *D. melanogaster* reference genome (NCBI build 5.3). The "Sample1" and "Sample2" columns give the names of the two samples being compared, with four contrasts possible: A3-control versus A4-control, A3-nicotine versus A4-nicotine, A3-control versus A3-nicotine, and A4-control versus A4-nicotine. The columns "Sample1.FPKM" and "Sample2.FPKM" give the FPKM (Fragments Per Kilobase of exon model per Million mapped fragments) values for each sample. The "Log2.FoldChange" column gives the \log_2 fold change in expression (Sample2 divided by Sample1). The "TestStatistic" column gives the test statistic used by Cuffdiff to compute the significance of the observed change in FPKM between samples, the "Pvalue" column provides the uncorrected p -value of the test statistic, and the "Qvalue" column provides the Benjamini-Hochberg FDR-adjusted p -value of the test statistic.

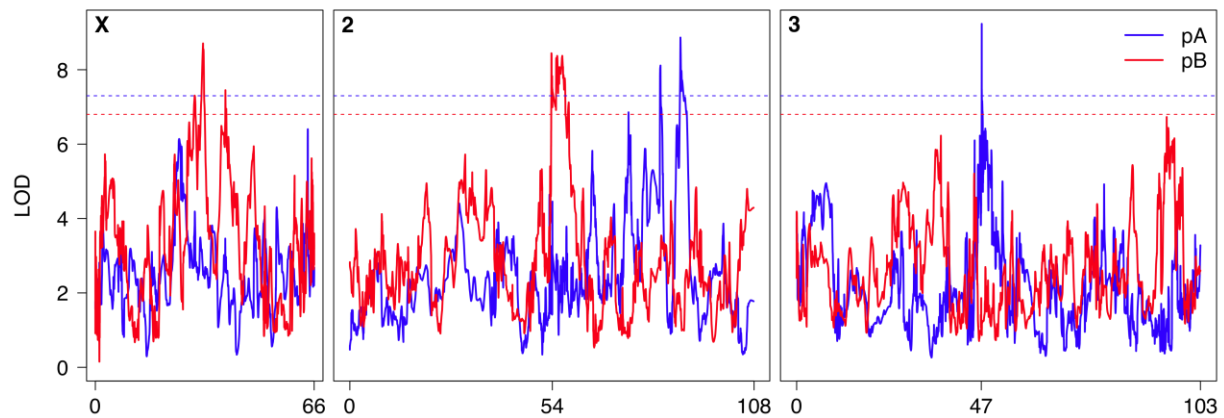


Figure S1 Genome scan for QTL contributing to larva-to-adult viability on control media. Solid curves are the scans for the pA (blue) and pB (red) data, and the horizontal dotted lines represent the genomewide 5% permutation threshold (LOD = 7.3 for pA and 6.8 for pB). Genetic distances along the chromosomes are indicated along the x-axis. The centromeres are at positions 54 and 47 on chromosomes 2 and 3, respectively. QTL identified in this scan are described in Table S2.

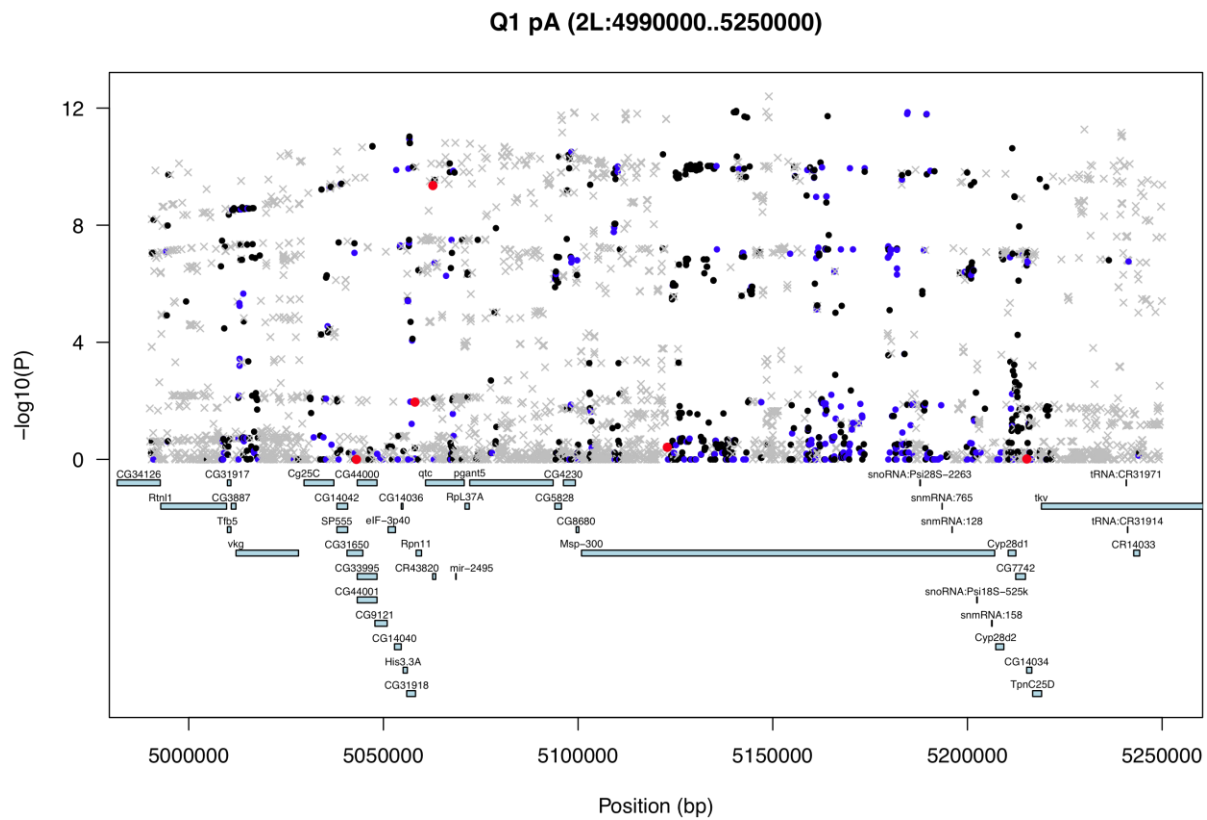


Figure S2 Association tests at SNPs residing beneath all four mapped QTL for nicotine resistance (Q1, Q2, Q3, and Q4 mapped in pA, and Q4 only mapped in pB). SNP genotypes in RILs were inferred from the mosaic founder haplotype structure of each RIL. The $-\log_{10}(p)$ value is plotted for each segregating site, with the plot symbol reflecting the impact of the variant: Red filled circles (loss of a start codon, loss or gain of a stop codon, changes in splice sites), blue filled circles (nonsynonymous changes), black filled circles (synonymous changes), and gray crosses (all other variants). The position of each gene within each QTL is shown as a light blue box, with the width of each box showing the distance from the start to the end of the gene model. Note that due to the number of genes under Q3, no gene names are provided for this plot. Note also the absence of tested variants at ~15,600-kb in the Q3 plot; This is due to a large array of 5S rRNA genes that likely precluded high-quality SNP identification in this region.

Q2 pA (2R:14510000..14730000)

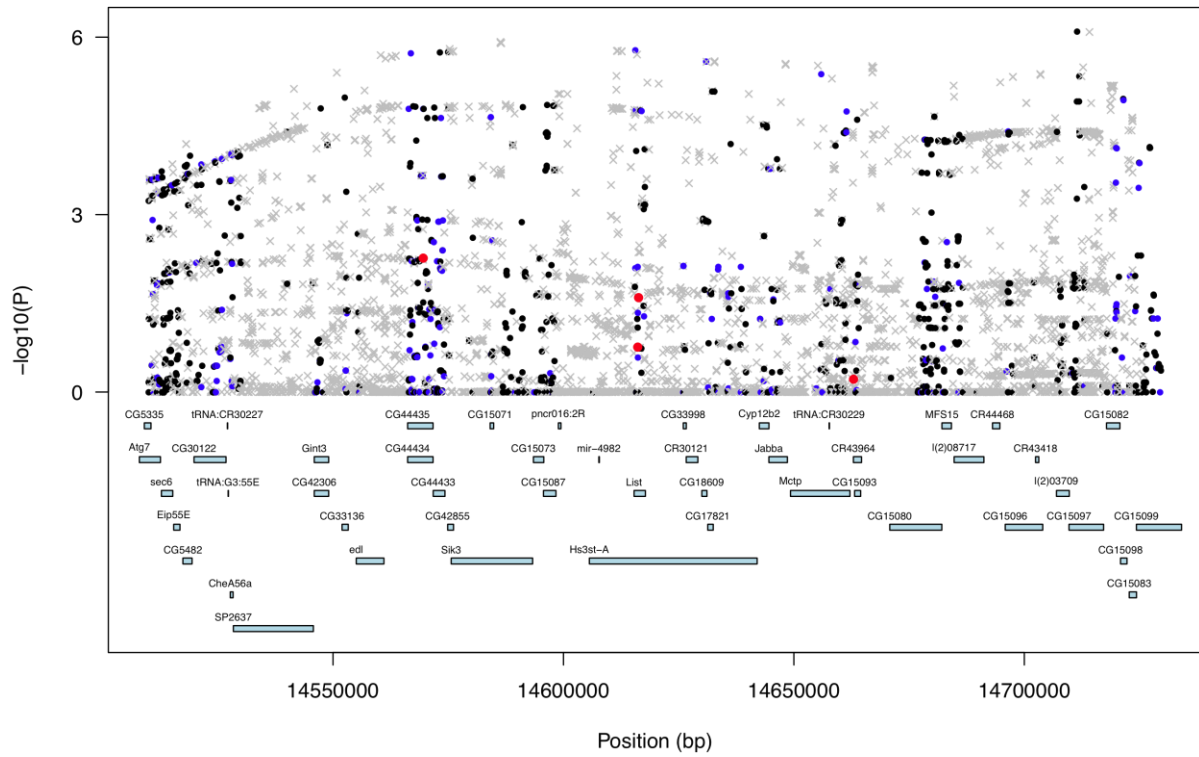


Figure S2 Continued.

Q3 pA (2R:15030000..16420000)

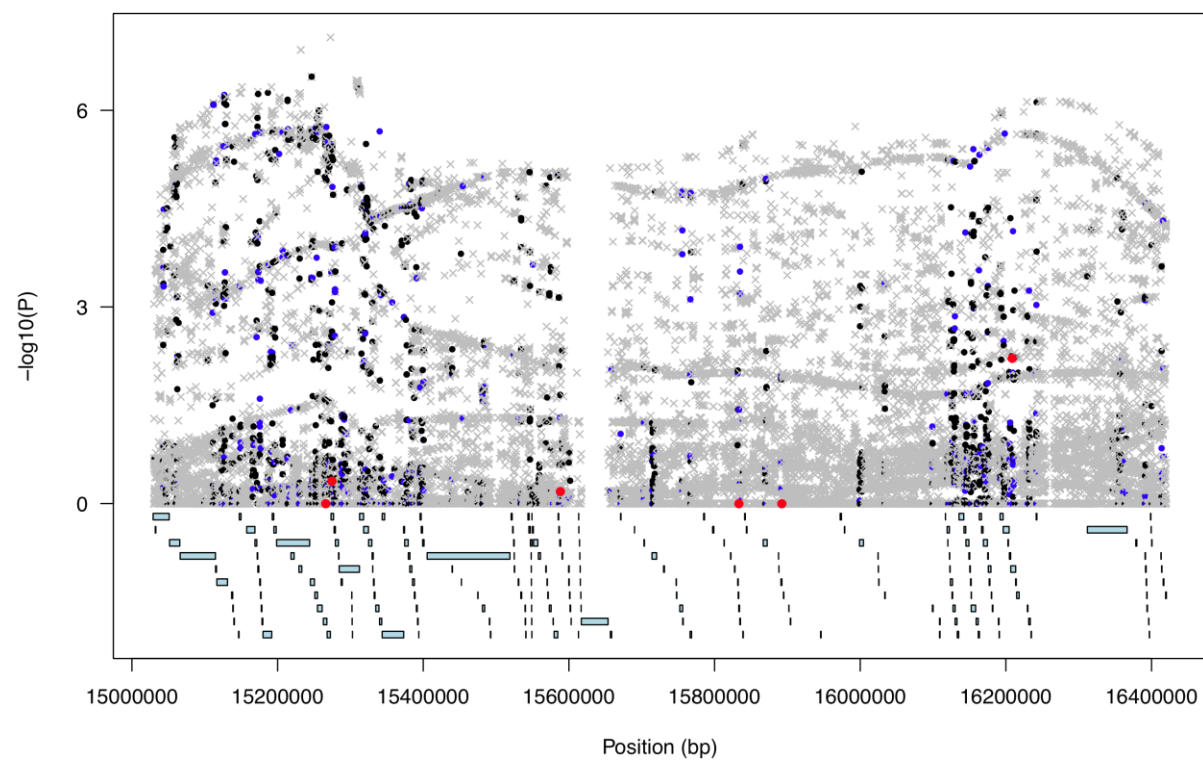


Figure S2 Continued.

Q4 pA (3R:6830000..7060000)

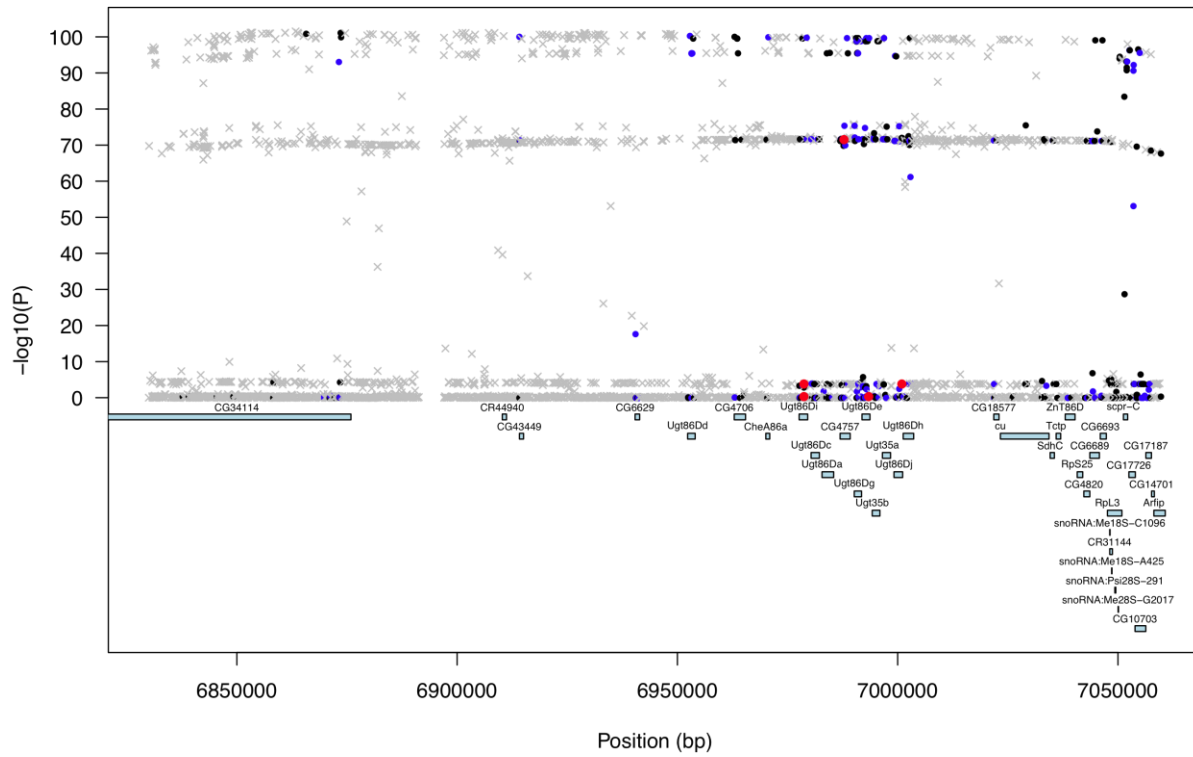


Figure S2 Continued.

Q4 pB (3R:6830000..7060000)

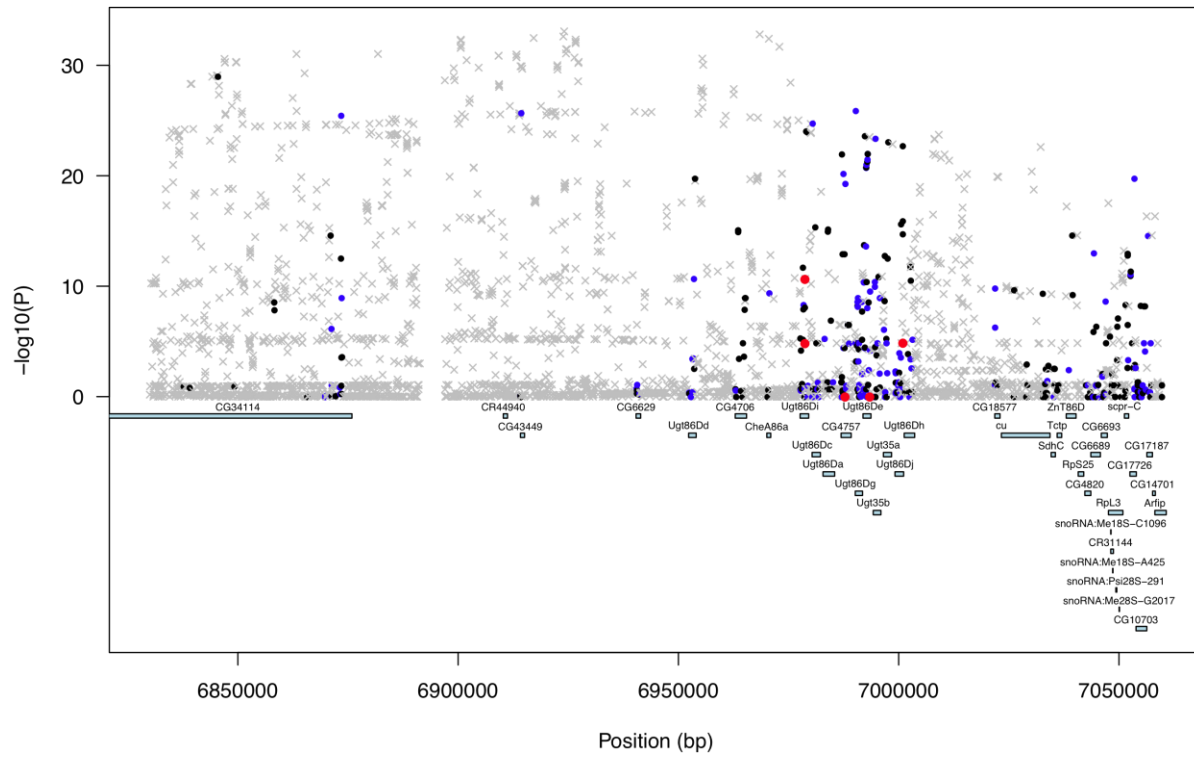


Figure S2 Continued.

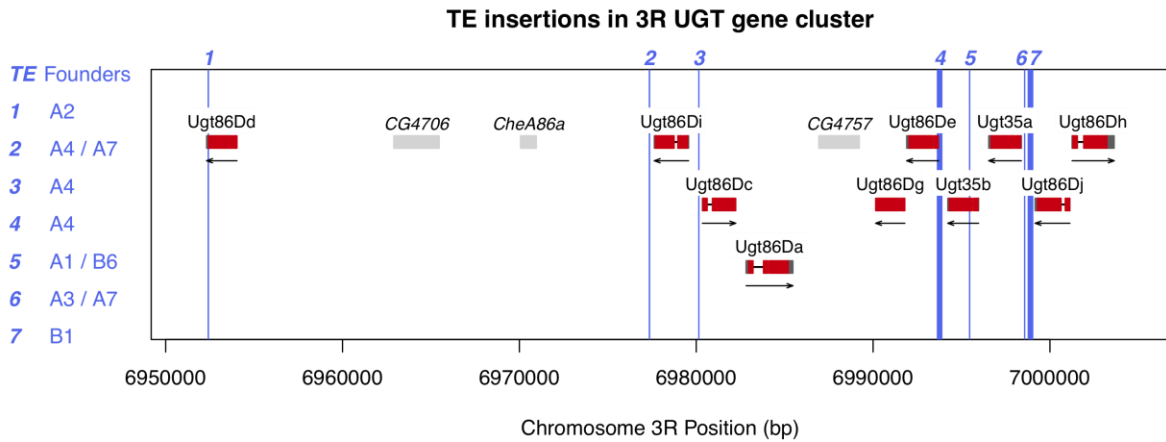


Figure S3 Positions of transposable elements (TEs) in or near the UGT genes implicated in Q4. The gene models for each of the 10 UGT genes are presented as red boxes (protein-coding regions) and dark gray boxes (untranslated regions), with arrows beneath the gene models showing the direction of transcription. The positions of other genes in the region of Q4 harboring UGT genes are presented as light gray boxes. The vertical blue bars are the positions of seven TEs that are either within, or within 1-kb of the start or end of, UGT genes (data taken from Supplementary Table 4 from Cridland *et al.* 2013) with the width of the bars representing ambiguity in the insertion position of the element. To the left of the plot are the founders that the elements are inserted into.

Peak in pB at 2R:7,500,000–8,440,000

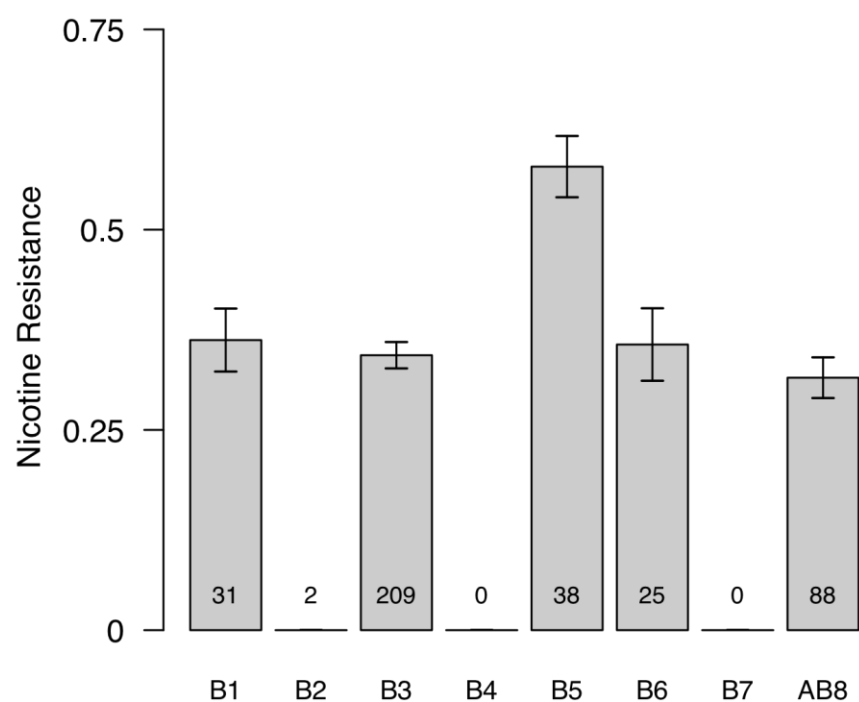


Figure S4 Founder haplotype means and 1-SDs at the pB-specific peak at position 7,500,000-8,440,000 on chromosome 2R. This interval harbors P450 gene *Cyp6g1*. The number of RILs for which we confidently assign a founder genotype (probability > 0.95) is listed at the bottom of each bar, and only founder means associated with at least 5 observations are presented.