GENETICS, 2022, 220(3), iyab238
https://doi.org/10.1093/genetics/iyab238
Advance Access Publication Date: 6 January 2022
Investigation

Powerful, efficient QTL mapping in *Drosophila melanogaster* using bulked phenotyping and pooled sequencing

Stuart J. Macdonald [1,2] Kristen M. Cloud-Richardson,[1] Dylan J. Sims-West,[1] and Anthony D. Long[3,*]

[1]Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66045, USA
[2]Center for Computational Biology, University of Kansas, Lawrence, KS 66047, USA,
[3]Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, CA 92697, USA

*Corresponding author: Department of Ecology and Evolutionary Biology, University of California at Irvine, Irvine, CA 92697, USA. Email: tdlong@uci.edu

Abstract

Despite the value of recombinant inbred lines for the dissection of complex traits, large panels can be difficult to maintain, distribute, and phenotype. An attractive alternative to recombinant inbred lines for many traits leverages selecting phenotypically extreme individuals from a segregating population, and subjecting pools of selected and control individuals to sequencing. Under a bulked or extreme segregant analysis paradigm, genomic regions contributing to trait variation are revealed as frequency differences between pools. Here, we describe such an extreme quantitative trait locus, or extreme quantitative trait loci, mapping strategy that builds on an existing multiparental population, the *Drosophila* Synthetic Population Resource, and involves phenotyping and genotyping a population derived by mixing hundreds of *Drosophila* Synthetic Population Resource recombinant inbred lines. Simulations demonstrate that challenging, yet experimentally tractable extreme quantitative trait loci designs (≥4 replicates, ≥5,000 individuals/replicate, and selecting the 5–10% most extreme animals) yield at least the same power as traditional recombinant inbred line-based quantitative trait loci mapping and can localize variants with sub-centimorgan resolution. We empirically demonstrate the effectiveness of the approach using a 4-fold replicated extreme quantitative trait loci experiment that identifies 7 quantitative trait loci for caffeine resistance. Two mapped extreme quantitative trait loci factors replicate loci previously identified in recombinant inbred lines, 6/7 are associated with excellent candidate genes, and RNAi knock-downs support the involvement of 4 genes in the genetic control of trait variation. For many traits of interest to drosophilists, a bulked phenotyping/genotyping extreme quantitative trait loci design has considerable advantages.

Keywords: QTL mapping; multiparental populations; complex traits; DSPR; caffeine resistance

Introduction

There has been tremendous progress in understanding the genetic underpinnings of complex, polygenic trait variation in the last 20–30 years. This has been particularly true for complex human disease (Visscher et al. 2017), where the pioneering work of the Wellcome Trust Case-Control Consortium (Wellcome Trust Case Control Consortium 2007) led to an ever expanding catalog of thousands of replicable genome-wide association study (GWAS) hits for a wide spectrum of human disorders and disease-relevant traits (Buniello et al. 2019). Equally, the efficiency and power of studies dissecting complex trait variation in animal and plant systems have been enhanced by inexpensive, sequencing-based genotyping solutions (e.g. Davies et al. 2016), and an array of readily available genetic mapping populations (e.g. Kover et al. 2009; Aylor et al. 2011; Svenson et al. 2012; King et al. 2012b; Baud et al. 2013; Huang et al. 2014; Noble et al. 2021).

Whereas GWAS is the dominant mapping paradigm in humans, the experimental flexibility of animal and plant systems—particularly, the traditional "model systems" (e.g. *Drosophila melanogaster*, *Caenorhabditis elegans*, mice)—has facilitated the development of an array of mapping approaches with varied strengths and weaknesses. GWAS-based methods, such as the *Drosophila* Genetic Reference Panel (DGRP; Huang et al. 2014), a set of 200 inbred lines derived from wild-caught flies, can have high power to identify intermediate-frequency loci of large effect (e.g. Magwire et al. 2012), and fine mapping resolution since they leverage the extensive ancestral recombination experienced by the sampled population. But as with all GWAS approaches such methods will struggle when functional alleles are rare (Pritchard 2001; Spencer et al. 2009; Thornton et al. 2013), or variant effects are low (Long et al. 2014; Mackay and Huang 2018). Other strategies in experimental plants and animals use multiparental populations (MPPs), collections of recombinant individuals, or inbred lines derived from intercrossing a modest number of founding genotypes for several generations. One example of a *D. melanogaster* MPP is the *Drosophila* Synthetic Population Resource (DSPR; King et al. 2012a, 2012b), which consists of 2 sets of RILs, each derived from an independent set of 8 founding

Received: November 11, 2021. **Accepted:** December 19, 2021

inbred strains. MPP-based methods sample less population diversity than GWAS designs, yet can have high quantitative trait loci (QTL) mapping power (King et al. 2012a; Gatti et al. 2014; Keele et al. 2019). This is true even when causative variants are rare in the population, since those captured in the founders are present at higher frequency in the mapping panel. In addition, the design and analysis of MPPs explicitly enables the detection of loci segregating for multiple functional alleles (Long et al. 2014); when directly assessed, such allelic heterogeneity appears to be common for complex traits (King et al. 2014; Hormozdiari et al. 2017).

Despite the range of tools available, a complete empirical picture of the genetic basis of complex traits is lacking. One practical barrier to the use of available mapping panels in the *D. melanogaster* system is that power is a strong function of sample size, so several hundred strains must be maintained, and multiple individuals must be phenotyped from each strain. Methods that mitigate some of the burden of phenotyping can be enabling.

One attractive alternative to depending on inbred lines is to employ a mapping approach that leverages a bulked phenotyping strategy (Fig. 1). Such methods are clearly inspired by the bulked segregant mapping approaches widely used in plant and insect genetics (Michelmore et al. 1991; Martin et al. 2012; Zou et al. 2016) and are akin to the extreme QTL (X-QTL) approach used by yeast geneticists (Ehrenreich et al. 2010). These approaches entail creating a large "base" population of recombinant individuals and subjecting a sample to a bulked phenotyping regime that selects the top (say) 5–10% of individuals. The selection step enriches for haplotypes that harbor phenotype-increasing alleles closely linked to a causative gene. Haplotype frequencies are then estimated in pooled DNA samples obtained from the base and selected populations, and at sliding windows throughout the genome, a statistical test identifies regions showing significant divergence between the populations. In contrast to mapping using an individual- or RIL-based strategy, where single individuals or sets of genetically identical individuals are phenotyped, an X-QTL approach is attractive for several reasons. First, an investigator need only handle/maintain/track a single population as opposed to several hundred separate RILs. Second, deriving a

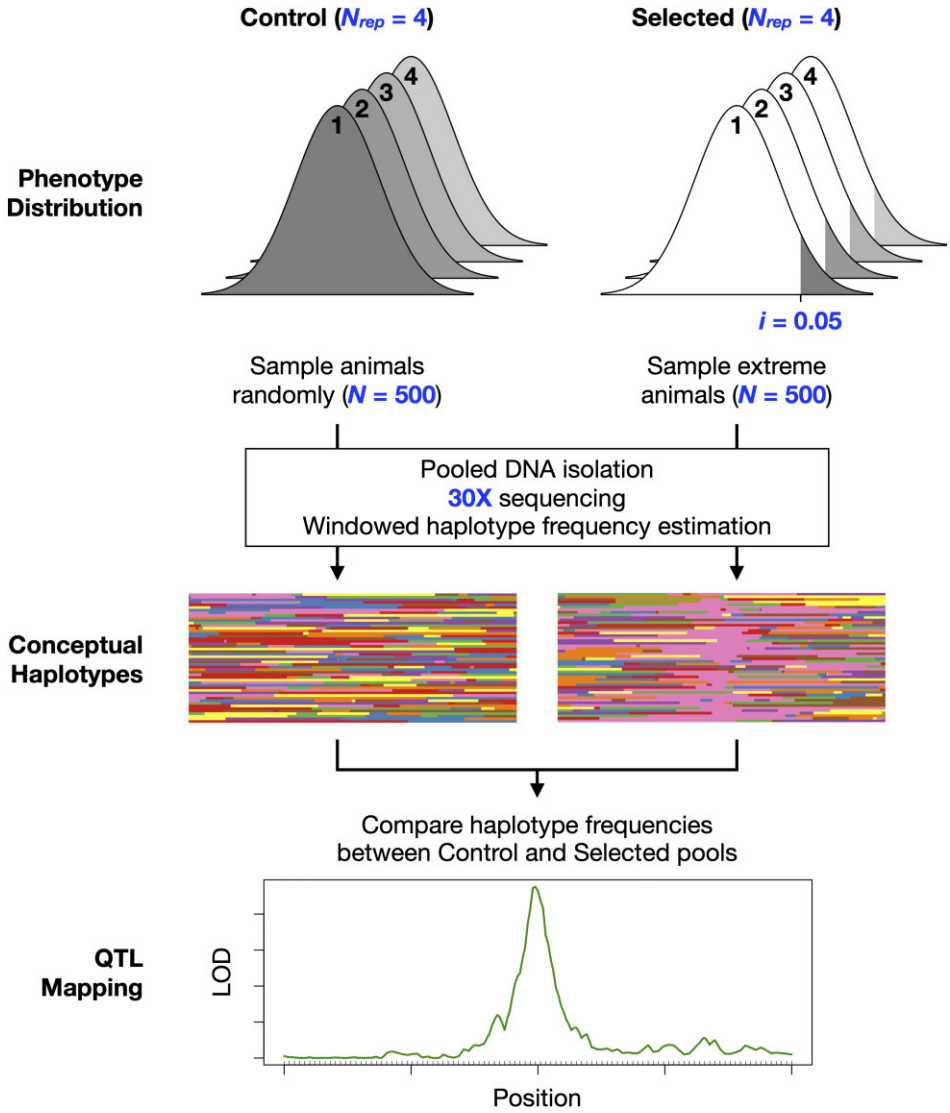

Fig. 1. Overview of the DSPR X-QTL approach. A segregating base population is created by mixing a series of DSPR RILs. For each of several replicates (N_{rep}), a set of N_{DNA} individuals are randomly sampled from the base population (control pool), along with a set of N_{DNA} phenotypically extreme individuals (selected pool) in—for instance—the top 5% of the phenotype distribution (i). Following DNA extraction, each pool is sequenced to ~30× coverage, and haplotypes are estimated in windows through the genome. Haplotype frequencies are then compared between paired control and selected pools, and QTL are evident as significant differences in frequency.

selected population has the potential to be straightforward, especially for traits that "phenotype themselves," such as survival or preference traits. Third, phenotyped individuals are outbred diploids raised in a common garden environment, so any effects of inbreeding depression are alleviated, and batch/vial effects are mitigated. Fourth, the sequencing effort required to accurately estimate haplotype frequencies from pooled samples initiated via a small number of founders, perhaps counter-intuitively, is modest and explored in this work.

Here, we consider an X-QTL experiment utilizing the set of DSPR RILs (King et al. 2012b). Our design consists of intercrossing several hundred RILs (N_{RIL}) to create a base population, expanding this population for several generations to both grow the total number of flies available for testing, and further recombine the RILs utilized (Fig. 1). We then impose a selective regime on the base population, select for the *i*% most extreme individuals, obtain pooled DNA preps (of size N_{DNA} individuals) from both the base and selected samples, sequence, and repeat the entire experiment N_{rep} times. Based on the observed frequencies of SNPs throughout the genome, and the known sequence of the 8 highly characterized inbred DSPR founders (King et al. 2012b; Chakraborty et al. 2019), a vector of 8 haplotype frequencies can be estimated at locations throughout the genome for each pooled sample. Finally, at those same locations we carry out a statistical test for differentiation in those haplotype frequencies between the selected and control populations. Loci associated with large values of the test statistic indicate the presence of a QTL. The X-QTL strategy we propose here is analogous to earlier work (Ehrenreich et al. 2010; Huang et al. 2012; Burga et al. 2019) but differs in having multiple founder alleles segregating in the base population, and in establishing the base from RILs derived from an advanced generation intercross to obtain higher mapping resolution. Below we explore—in theory and in practice—the utility of this approach relative to typical RIL-based mapping as a function of achievable experimental designs.

Materials and methods

The *Drosophila* Synthetic Population Resource

The DSPR consists of a series of RILs derived from a pair of advanced generation synthetic, recombinant populations (populations A and B, or pA and pB), each of which was founded by a set of 8 highly inbred founder strains (King et al. 2012b; Chakraborty et al. 2019). The synthetic populations were each maintained as duplicate subpopulations (pA1, pA2, pB1, pB2) at large census size for 50 generations of free recombination, after which RILs were derived with 25 generations of strict brother-sister mating. The collection of RILs is highly inbred. Only an estimated 1–2% of the genome, primarily in difficult to homogenize and characterize centromeric regions, is segregating in any given RIL (King et al. 2012b).

Power and false positive rates

To assess the power of the X-QTL approach, in comparison with RIL-based mapping using the DSPR, we simulated the distribution of $-\log_{10}(P)$ scores for a complex trait with 50% heritability, 2% heritability due to a focal gene, and variation at this gene due to 8 equally frequent founder haplotypes with additive Gaussian genotypic effects. These assumptions are reasonable since QTL exhibiting >2% heritability are commonly observed in RIL-based QTL mapping studies employing the DSPR (see for instance Najarro et al. 2015; Everman et al. 2019). Beavis effects are subtle for large experiments (Beavis et al. 1991; Beavis 1994; King and Long 2017), and QTL mapped in the DSPR often appear to exhibit an allelic series (King et al. 2014). Changing the trait heritability while holding the heritability of the focal gene constant should not impact the power of the X-QTL approach. But for RIL-based mapping, replicate phenotypic measures for each strain will increase power, since the ratio of environmental to genetic variation is decreased. To estimate power at a marker located in a single focal QTL, we need only simulate the QTL itself and can ignore any flanking information. To efficiently calculate the distribution of $-\log_{10}(P)$ values for a range of experimental designs, for these power simulations we only track and test genotypes at the causative gene and assume all 8 haplotypes are equally frequent.

To simulate an RIL-based experiment, for each of 1,000 replicates we simulate a 500 RIL mapping experiment, obtain 10 phenotype measures per RIL, and derive genome-wide *P*-values via one-way ANOVAs with founder allele as a factor and mean phenotype as a response variable. To simulate the X-QTL experiment we create 4, 8, or 12 large replicate (N_{rep}) diploid base populations by mixing the above 500 RILs, instantaneously expand the population, and then choose the 5% or 10% most phenotypically extreme individuals from each replicate population to create selected pools. Instantaneous expansion implies no recombination occurs during the expansion of the base population (recombination is not relevant as we only consider tests at the causative site), nor does sampling variation change the frequency of QTL alleles. These assumptions are relaxed below where we explore QTL localization for a smaller number of replicates and a more modest set of parameter combinations. For each replicate we randomly choose 150, 300, or 600 diploid individuals from the selected pool, and an equivalent number from the expanded base population to create a control pool [control vs selected = treatment (*trt*)]. The parameter combinations explored were chosen to encompass a series of experimental designs that could realistically be carried out. The 8 *known* founder haplotype (*h*) arcsine square-root transformed frequencies (*asf*) at each position are tested for differentiation between control and selected pools using an ANOVA (i.e. *asf*~*h***trt*, with *h*:*trt* tested using *replicate*:*h*:*trt* as an error term). Minus log base 10 *P*-values, $-\log_{10}(P)$, associated with the X-QTL- and RIL-based statistical tests are compared.

We repeated the entire set of power simulations with a QTL explaining zero percentage of phenotypic variation to derive a distribution of test statistics under the null hypothesis of no QTL being located at the marker. For each parameter set, we calculate the per chromosome false positive rate (FPR) as the proportion of scans out of 250 pure replicates in which a single marker has a $-\log_{10}(P)$ score of greater than 4. Similarly, we calculate the genome-wide false positive rate as $\text{gFPR} = 1-(1-\text{FPR})^5$, since our test chromosome 3L is roughly 1/5th of the *D. melanogaster* genome. At a $-\log_{10}(P)$ threshold of 4, the gFPR is 7.1% (i.e. 7.1% of all genome scans will result in a single interval with a $-\log_{10}(P) > 4$). But the false positive rate is a subtle function of the manipulated parameters, with the number of flies contributing to a DNA pool having a measurable impact on the gFPR. Conditioning on effective pools of 300 or 600 flies, mimicking census sizes of perhaps 500 or 1,000 (discussed below), results in an gFPR of 3.9%. Thus, a $-\log_{10}(P)$ score of 4 seems to be an appropriate threshold over a much of our recommended parameter space, but it is important to note that smaller pools may result in a slightly liberal false positive rate.

Localization

The power simulations above that only consider a causative marker locus are efficient and can explore a large parameter space but lack the necessary realism to explore localization. We thus carried out additional simulations with fewer replicates and a reduced parameter space where we tracked entire chromosome 3L haplotypes. The chromosome 3L haplotypes used to initiate the simulated base populations were sampled from the actual chromosome-wide DSPR RIL genotypes (King et al. 2012b), and thus, realistically model founder haplotype frequencies and patterns of linkage in experimental DSPR-derived mixed populations. We sampled (N_{RIL}) 200, 400, 600, or 800 DSPR pA RILs (http://www.flyrils.org), extracting haplotype probabilities every 200 kb along chromosome 3L. At each location haplotype, calls are represented by 8 additive dose probabilities (i.e. the vector sums to 1), but here we more simply define the founder haplotype at each position in each RIL as the largest of the 8 dosage calls. Since doses are typically near 0 or 1 (King et al. 2012b), switching from "soft" to "hard" calls does not discard a great deal of information while greatly speeding up simulations. Sampling only every 200 kb (as opposed to the native 10-kb marker spacing in the DSPR) makes the simulation machinery 20× more efficient, and since adjacent test statistics are often correlated at sub-cM sampling densities, there is no great need to sample at a higher marker density for much of the parameter space we explore (see Supplementary Fig. 2a). We simulated a Gaussian QTL in the middle of the chromosome contributing 2% to heritable variation. In addition, we placed additive Gaussian polygenic "background" loci of equal variance at each of the 200 marker locations such that the total background variance was 8%, and environmental variation was set equal to 90% (V_e in this manner includes genetic variation due to the remainder of the genome as well as variance due purely to the environment). Unlike for power calculations (above), after expansion of the base population to 20 thousand individuals, we allow 4, 8, or 16 generations of random mating, with exactly 1 recombination event per female chromosome per generation to mimic recombination events occurring during population expansion.

For each starting number of RILs (N_{RIL}), we simulated the proportion of individuals selected—the intensity—as 5%, 10%, and 20% (i) and randomly sampled 2 sets of 150, 300, or 600 individuals (N_{DNA}) from a set of 20,000 simulated individuals (i.e. a mock treatment vs control) to "make DNA." Errors in the haplotype estimator and/or differences in the number of DNA molecules per individual that make their way into the Illumina sequencing library are likely to reduce the number of individuals being tested. Thus, we consider the simulated values of N_{DNA} to reflect the idea of *an effective sample size* that is lower than the actual number of individuals that would be used in an experiment. We guess, without evidence, that N_{DNA} values of 150/300/600 perhaps represent census sizes of 250/500/1,000. The concept of an effective sample size smaller than the census size will reduce both the power and localization ability of the X-QTL approach and is thus a conservative assumption. We repeated the simulation 4, 8, or 12 times for each parameter combination mimicking an experiment replicated to different levels, with each experiment producing the same $-\log_{10}(P)$ as the power calculations described above. We estimated localization ability in kilobases as the number of markers within a 3 $-\log_{10}(P)$ drop of the most-significant marker (MSM) in a scan, conditional on the MSM being greater than 4, multiplied by 200 (since markers are spaced every 200 kb).

False positives due to polygenic background

In the simulation described above, for some "high power" parameter combinations, we further examined the distribution of $-\log_{10}(P)$ scores at markers greater than 7 Mb from the causative site. Elevated $-\log_{10}(P)$ scores at such loosely linked sites likely reflect signal due to linkage disequilibrium between founder haplotypes at the locus and chromosome-wide polygenic scores. In the results below, we show that this is a weak force, provided there are a few generations of recombination during establishment of the base population.

Higher-resolution localization

For some of the above parameter combinations, the localization ability was approaching the marker density of one every 200 kb. As a result we repeated the localization experiment for a subset of the parameter combinations, but changed the marker density to one every 10 kb, and only examined the middle third of chromosome 3L. We further only considered 4 generations of random mating following derivation of the base population. These changes held the considerable computation effort roughly constant allowing us to more precisely examine localization ability. For each set of experimental design parameters (i.e. N_{RIL}, N_{DNA}, i, N_{rep}) we estimated the localization ability in kilobases as 10 times the number of markers (since markers are spaced every 10 kb) within a 2 or 3 $-\log_{10}(P)$ drop of the MSM in a scan, conditional on the MSM being greater than 4. This served as a benchmark for the idea that a 2 or 3 $-\log_{10}(P)$ drop from the MSM includes the QTL 95% of the time (analogous to the 2-LOD drop widely used in interval QTL mapping). Next, we carried out ANOVAs on the average interval size as a function of the parameters we varied (N_{RIL}, i, N_{DNA}, N_{rep}), and included the average $-\log_{10}(P)$ score per parameter combination as a covariate to identify factors that impact localization ability.

Development of a mixed DSPR population

To create an experimental mixed population of RILs, we manually collected 10 eggs from each of 663 pA RILs (365 from subpopulation pA1, 298 from subpopulation pA2), placing ∼400 eggs into each of a series of *Drosophila* stock bottles (6 oz; ThermoFisher, AS355) containing ∼50ml of cornmeal-yeast-molasses medium. All bottles were placed inside a 1 cubic foot (12 in. × 12 in. × 12 in.) population cage, and approximately every 2 weeks old bottles were removed from the cage and replaced with 9–12 fresh bottles. The first cage generation had a maximum of 6,630 animals, and anecdotally the density of flies in the cage did not markedly vary, and stayed very high, through the experiments described below.

Rearing and collecting experimental animals

Four replicates of the experiment were carried out, following 1 (replicate A), 3 (rep. B), 4 (rep. C) and 5 (rep. D) generations of population cage maintenance. In each case we followed the same basic procedure outlined below (also see Supplementary Table 1).

We placed 5–7 apple juice agar dishes (Supplementary Text 1) supplemented with a small quantity of live yeast paste (Supplementary Text 1) into the population cage for several hours. Afterward, any yeast paste was removed from the surface of the agar, the dishes were filled with 1× phosphate-buffered saline (PBS), and a paintbrush was used to dislodge the eggs from the surface. Subsequently, the egg suspension from all plates was poured into a 50-ml tube and the eggs allowed to fall to the bottom. The majority of the PBS was poured off and replaced 2–3 times until the PBS was relatively clear and free of

debris. Subsequently, a standard 100-µl pipettor was set to 12 µl, and a wide-bore tip used to move aliquots of the egg suspension to rearing vials (narrow, polystyrene; ThermoFisher, AS515) containing ~10 ml of cornmeal-yeast-molasses media (Supplementary Text 1). (Pipetting very aggressively ensures that eggs are moved along with PBS.) For each replicate, we collected eggs over 2 consecutive days to generate a sufficient number of vials/experimental females. While we do not know the exact number of eggs per vial, the level of larval activity across vials was visually consistent, and in a sample of 10 rearing vials for which all emerging females were counted, the number of adult females averaged ~70 per vial.

Following development, and 2 days after the first evidence of adults, all flies from each rearing vial were tipped to a mixed-sex housing vial. Given the density of flies in these vials they were placed on their side. Two days later, experimental female flies were manually collected over CO_2 anesthesia and sorted into single-sex housing vials in groups of 19–30.

Caffeine resistance assay

The assay used largely replicated that employed previously to identify QTL in the set of inbred DSPR RILs (Najarro et al. 2015). For each of the 4 replicates (A, B, C, and D), we carried out the following procedure.

The day prior to the initiation of the experiment we generated assay media (Supplementary Text 1), cooled 1 l to ~55°C, and thoroughly mixed in 10 g of caffeine (SigmaAldrich, C0750) achieving a final caffeine concentration of 1%. Media was then poured into a series of 100-mm Petri dishes and allowed to cool. Subsequently, bundles of polycarbonate *Drosophila* activity monitor tubes (5-mm diameter × 65-mm length; TriKinetics, PPT5x65) were filled to ~10-mm height with solidified media by pressing them into the media surface, wiped clean, and then sealed by plunging the tubes individually into molten paraffin wax.

On the day of the experiment, 3–5-day-old females ($N = 2{,}337$–$2{,}572$ per replicate) were moved from single-sex housing vials and loaded into monitor tubes and tubes were capped with small foam plugs cut from Droso-Plugs (Genesee Scientific, 59-200). Rather placing the monitor tubes into an automated monitoring system (Najarro et al. 2015), tubes were placed on their side along the base of a series of cardboard/plastic narrow fly vial trays (Genesee Scientific, 32-124 or 59-163B), with ~160 tubes per tray. Starting the day following setup, the tubes were surveyed for dead animals twice each day and dead animals were counted and removed.

For the first replicate (replicate A), 43% of the animals were loaded into monitor tubes via manual, oral aspirators, while 57% were loaded via an automated instrument (FlySorter). By segregating tubes into different trays based on loading method, we noticed that automatically loaded animals had lower lifespan on average (Supplementary Fig. 6). Since we selected the most resistant animals on a tray-by-tray basis (see below), the impact of the loading method may be limited (indeed, dropping replicate A from the final X-QTL analysis has little effect over dropping replicate B—Supplementary Fig. 4). However, for all 3 later replicates (replicates B, C, and D), we solely loaded flies by manual aspiration.

Collection of pools of control and caffeine-resistant animals

For each replicate, we collected 250 "control" females from the set of single-sex housing vials by randomly selecting 25 vials, aspirating 10 females per vial into a 15-ml tube on ice, and then freezing the tube at −20°C to await DNA isolation.

Ideally the pool of "selected" females for each replicate would be the 250 females surviving the longest under caffeine conditions. This was not possible for 2 reasons. First, we intermittently monitored the number of dead animals and had to judge the appropriate time to collect the final set of females to avoid collecting many more, or many fewer, than the 250 desired. Second, the exposure cannot start for all experimental females at the same time; with 2–3 investigators it took 6–9 h to load all females into tubes. Since our target was the 250 most resistant animals (i.e. 10% of the starting sample), and since tubes were sequentially arrayed out into trays at ~160/tray, we targeted the collection of ~16 females per tray. The mortality trajectory of each tray was fairly consistent (Supplementary Fig. 6), aside from trays of flies that were automatically loaded (see above). When we reached the target number of live flies per tray (subject to the intermittent monitoring constraint noted), those monitor tubes were frozen at −20°C, and subsequently, all selected flies were pooled into a single 15-ml tube and refrozen at −20°C. We collected 228–254 females per selected pool (9.0–10.3% of the starting set of flies).

DNA isolation

DNA was isolated from each of the 8 samples of flies using an adaptation of the Gentra Puregene Cell Kit protocol (Qiagen, 158767). Briefly, each pool of flies was homogenized via glass beads in 1× PBS and subjected to several strokes of a dounce homogenizer. Then, a small amount of the homogenate was taken forward through cell lysis, protein precipitation, an RNase step, and DNA precipitation and resuspension (see Supplementary Text 2 for more details). DNA integrity was confirmed by running a small amount of each sample on a 1.5% agarose gel, and DNA was quantified using a fluorometer (Qubit dsDNA BR Assay Kit, ThermoFisher, Q32853).

Library preparation and sequencing

Each DNA sample was diluted, and a 50-µl aliquot containing between 832 and 841 ng was provide to the KU Genome Sequencing Core for library construction using the NEBNext Ultra II DNA Library Prep Kit (NEB, E7645L), incorporating unique dual-indexing (NEB, E6440S). Final library fragment sizes ranged from 303 to 327 bp (Agilent TapeStation 2200), and the genomic DNA-derived library insert sizes are expected to be around 200 bp. Libraries were pooled at equal concentrations and sequenced with PE101 reads on a single S4 lane of an Illumina NovaSeq 6000 (UC Irvine Genomics High-Throughput Facility). We further estimated average insert sizes from the alignment files (using samtools stats) at 197–221 bp over libraries. Given the average insert sizes relative to read lengths, the F and R reads overlapped for a large fraction of read pairs. This is not a problem for bcftools mpileup, but in future work, we plan to increase the library fragment size to greater than 500 bp, as fewer reads from a larger insert library should provide the same information.

Haplotype estimation from experimental sequencing data

Haplotypes are estimated using slight modifications to published software (Linder et al. 2020). In short, we use `bwa-mem` (Li 2013) to generate a BAM file for each sequenced pooled sample, and for each of the highly inbred founder lines from which the pA DSPR population was derived (King et al. 2012b). We then use `bcftools mpileup` and `bcftools call` (Li 2011; bcftools version 1.9) to generate a file that reports REF and ALT counts at all SNPs. Finally,

we employ `bcftools query`, along with a custom perl script, to output the frequency of the REF allele in each sample for the set of SNPs that are not polymorphic within any given founder strain.

For each pooled sample, for a window size of 200 kb and a step size of 10 kb, we use the `limSolve` R library (Soetaert et al. 2009) to estimate the optimal vector of 8 founder haplotype frequencies ($\boldsymbol{f}$) that minimizes the sum of the squares between observed allele frequencies and predicted allele frequencies. In other words, for each window $\boldsymbol{Y} = \boldsymbol{Xf}$, where $\boldsymbol{Y}$ is the vector of allele frequencies in the pool, and $\boldsymbol{X}$ is an F column matrix (largely 0 vs 1) of genotypes associated with the F founders. The `limSolve` library allows us to constrain solutions to those whose frequencies sum to one, and where each frequency is >0.0003 (this small mass for every founder avoids convergence issues). For any given window, the number of SNP positions is large and dropping subsets of SNPs does not impact the estimator very strongly. The script to do this is described more fully in prior work (Linder et al. 2020) and is on GitHub.

Establishing the level of error in haplotype estimation

We created a pseudo-pool with an average sequencing coverage of 240× by combining an equal number of reads from the replicate A and B caffeine experiment control pools (N_{DNA} = 250 in each case). Using this high coverage data, we estimated allele frequencies at a genome-wide set of SNPs each private to a single founder. We then down-sampled the number of reads in the pool to more modest coverage (70×, 35×, 17×) and ran our haplotype frequency estimator. The frequency of a private SNP at 240× coverage and the frequency of the founder harboring that private SNP at the window nearest the SNP location are estimators of the frequency of that founder in the sample. We calculated the mean square difference between the 2 estimators as a function of coverage as: 0.000708 (full 240× coverage), 0.000711 (70×), 0.000717 (35×), and 0.000725 (17×). The mean squared difference between frequency estimates (Var_{diff}) should equal the variance in SNP estimating frequency (Var_{SNP}) plus the variance contributed by the haplotype estimator (Var_{HAP}). Since Var_{diff} varies little as a function of coverage; this suggests that error in the haplotype frequency estimator is quite small.

We estimate the expectation of Var_{SNP} as pq/N over all SNPs at 240× to be 0.000322. But if we alternatively obtain the expectation of Var_{SNP} as the variance among SNP frequencies within the same founder allele and within 10 kb of one another (such SNPs should have almost identical frequency), we obtain 0.000611, a variance roughly twice as large. We conclude that the error in SNP frequency estimates obtained from a pooled DNA sample has an over-dispersed variance relative to binomial expectations, as is commonly claimed in the literature (King et al. 2012b; Wei et al. 2017; Zhang and Emerson 2019). If we assume 0.000611 more accurately estimates Var_{SNP}, this implies the variance in our haplotype estimator is quite small, with an average absolute error of ∼0.01 [i.e. sqrt(Var_{diff}-Var_{SNP})] irrespective of coverage.

Functional testing of potential candidates

We used the Gal4-UAS-RNAi system to functionally test a series of genes implicated by mapped X-QTL. We used the ubiquitous, *Actin 5C* (*Act5C*) promoter-driven Gal4 strain (Bloomington Drosophila Stock Center number 25374), and a strain expressing Gal4 in the adult anterior midgut (1099 from Nicholas Buchon, flygut.epfl.ch; Buchon et al. 2013). All UAS-RNAi strains were from the VDRC (Vienna Drosophila Resource Center) "GD" collection, which each harbor a *P*-element-derived UAS transgene (Dietzl et al. 2007). Transgenes targeted genes *Crys* (stock ID 37736), *Cyp12d1* (50507), *Cyp6d5* (12138), *Ugt36A1* (9489), *E23* (2620), *osy* (38661), *Tlk* (46424), and *Vha100-5* (6121). We compared each to a background, transgene-free control strain (stock ID 60000). The UAS-RNAi and control strains have the same genetic background, aside from the transgene. The 2 Gal4 driver strains have distinct backgrounds, which could affect phenotype, but we do not attempt to compare phenotypes between ubiquitous and gut-specific knockdown of any gene.

We tested the same set of experimental Gal4-UAS genotypes across 2 batches. In the first batch, we crossed ∼7 male Gal4 animals to 10 female UAS (or control) animals, setting up 2 replicate vials per Gal4/UAS combination (18 genotypes total). Two days following the first emergence of adults, all flies from each vial were tipped to a mixed-sex housing vial. After a further 2 days, experimental females were sorted over CO_2 into single-sex vials housing 20 experimental females. The following day 32 experimental, 3–5-day-old females (16 per replicate cross vial) were loaded without anesthesia into 1% caffeine monitor tubes (see above), and we used the *Drosophila* activity monitoring system to track movement of each individual, yielding an accurate lifespan for each (see Najarro et al. 2015).

In the second batch, we crossed 10 female Gal4 animals to ∼7 male UAS (or control) animals—i.e. the reciprocal cross direction from the first batch. We set up 3 replicate vials per Gal4/UAS combination (16 genotypes total; 2 crosses involving the *Act5C*-Gal4 strain were not reattempted since in the first batch no experimental Gal4-UAS-RNAi animals were obtained, Supplementary Table 2). As described for the first batch, we collected 3–5-day-old experimental females and loaded 48 animals per genotype (16 per cross vial) into monitors. Caffeine resistance phenotypes for each Gal4-UAS-RNAi genotype and each batch were compared to the relevant control genotype via Dunnett's tests.

Environmental conditions used to rear and assay X-QTL and RNAi flies

The source X-QTL population cage, the vials used to rear and house experimental X-QTL and RNAi animals, and the monitor tubes used to assay caffeine resistance were all resident in the same laboratory incubator and maintained at 25°C, 50% relative humidity, on a 12-h light:12-h dark cycle. Media used in population cage bottles and in both rearing and housing vials was a cornmeal-yeast-molasses mix (Supplementary Text 1), while media used for the resistance assay was a cornmeal-yeast-dextose mix supplemented with caffeine (Supplementary Text 1).

Results

Powerful DSPR-based X-QTL mapping in *Drosophila*

We simulated QTL mapping power under a typical RIL-based QTL mapping experiment as well as several different X-QTL designs. Figure 2 depicts the distribution of $-\log_{10}(P)$ scores at a QTL contributing 2% to trait variation. The distribution of the X-QTL $-\log_{10}(P)$ scores is presented as a function of the number of individuals sampled to create the selected and control DNA pools (150, 300, 600), the proportion of individuals selected during phenotyping (10% and 5%), and the number of replicate base populations (4, 8, 12). We see the expected null distribution of X-QTL test statistics when comparing 2 equally sized draws from the control/base population (Supplementary Fig. 1), and provided

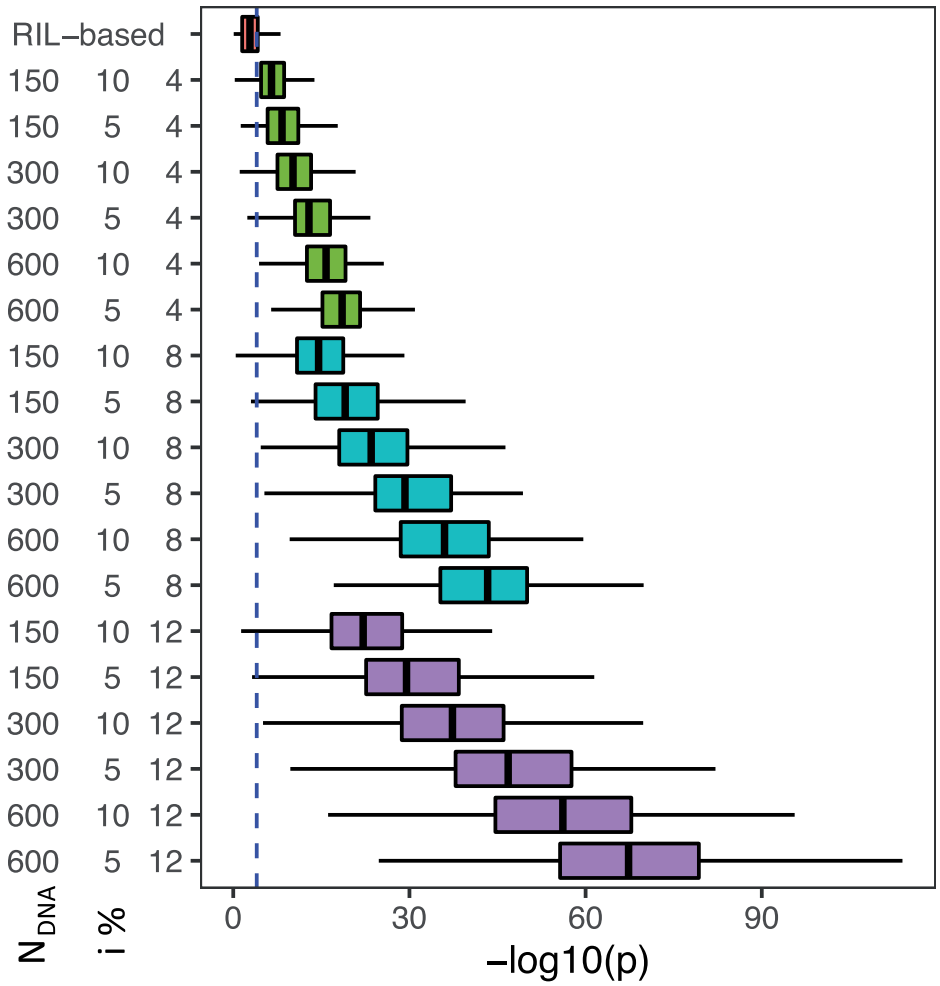

Fig. 2. Distribution of $-\log_{10}(P)$ test statistics for RIL-based mapping compared to X-QTL. Analysis assumes a QTL contributing 2% to heritability. X-QTL $-\log_{10}(P)$ distributions are shown as a function of the effective number of individuals contributing to the DNA pools (150, 300, 600), proportion of individuals selected (10%, 5%), and the number of experimental replicates (4, 8, 12). *P*-values are routinely much higher under X-QTL mapping compared to mapping directly with RILs. Larger pools, more extreme selection, and a greater number of replicates further increase X-QTL power. The vertical dotted blue line represents a 4 $-\log_{10}(P)$ threshold for significance.

DNA is prepared from a sufficiently large pool of animals (≥300), a $-\log_{10}(P)$ threshold of 4 holds the genome-wide false positive rate at ≤5%. Using a threshold of 4, the proportion of $-\log_{10}(P)$ scores greater than 4 in Fig. 2 is the power to map a QTL. Power with the X-QTL framework is routinely higher than a typical RIL-based experiment employing 500 RILs and 10 replicate phenotype measures per RIL. Stronger selection, more replicates, and a greater number of individuals contributing to the DNA pools strongly impact X-QTL power (Fig. 2) and should be maximized in any experiment.

The simulated pool sizes of 150, 300, or 600 individuals sampled to make genomic DNA should be thought of as the *effective number of individuals* in the pool. Factors such as uneven representation of individuals in the pool (e.g. variation in body size, differential lysis or degradation during the bulk DNA prep), and errors in haplotype frequency estimation will result in the effective number of individuals in the pool being lower than the census number. By only choosing flies of a single sex (to somewhat control for body size variation), using fresh tissue from living animals, and generating 20–40× sequencing coverage per pool (see experimental results below), we believe that the difference between the effective and census population sizes can be minimized. The relationship between the census and effective numbers of sampled flies is difficult to estimate, and likely somewhat a function of the exact details of how flies are collected and processed. We guess that effective sizes of 150, 300, and 600 may approximate census sizes of 250, 500, and 1,000 individuals, respectively. Overall, our simulations suggest that an X-QTL experiment employing DSPR RILs can be quite powerful if certain experimental parameters are carefully optimized.

X-QTL mapping resolution can be high

Although the power of an X-QTL experiment can be high, if the approach suffers from poor QTL localization ability it may not be suitable for the dissection of complex traits. We used the actual genotypes of DSPR RILs (King et al. 2012b) to simulate a base population used for X-QTL mapping. Figure 3 depicts the average $-\log_{10}(P)$ profiles for a simulated QTL on chromosome 3L as a function of simulated parameters for a base population instantaneously expanded to a large size, randomly mated for 4 generations, and then phenotyped/genotyped in an 8-fold replicated experiment. (Additional generations of random mating were also simulated and are discussed below.) Localization is not greatly impacted by the number of individuals used in the DNA prep (N_{DNA}), the number of RILs used to generate the base population (N_{RIL}), or the proportion

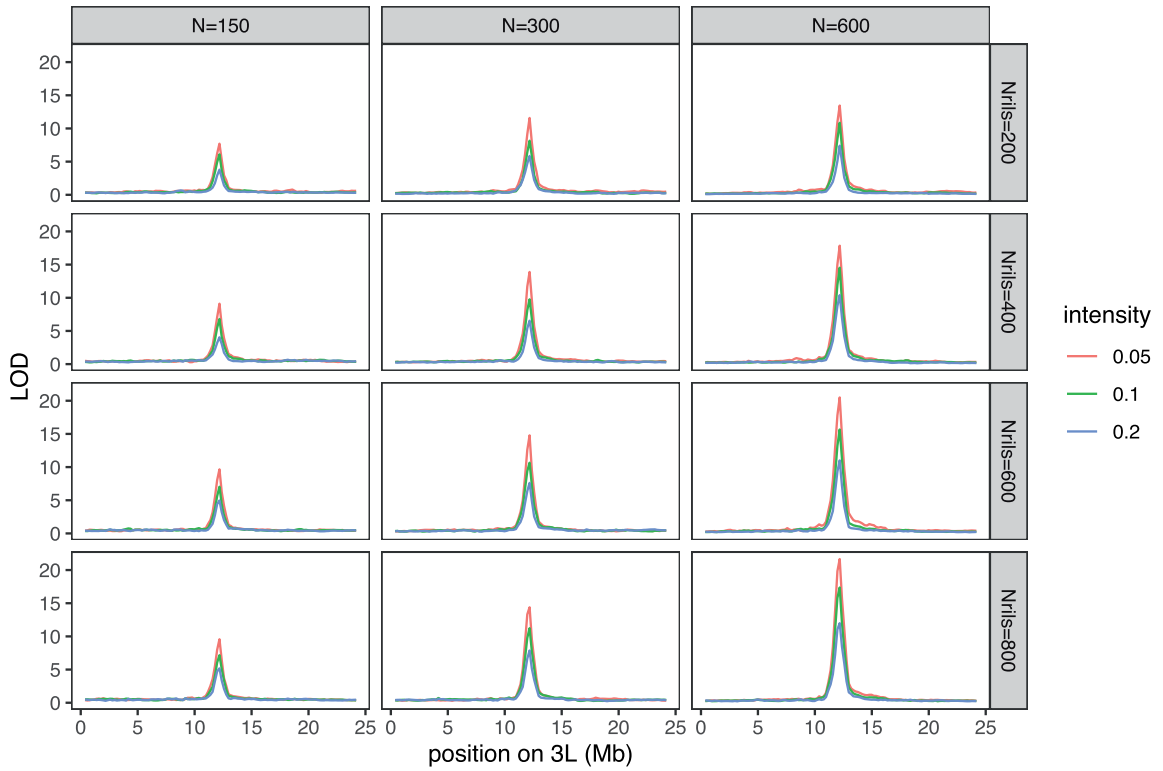

Fig. 3. Average QTL localization for an X-QTL experiment. Simulated base populations were derived from DSPR population A RILs and are created via diallel crosses of the RILs, instantaneous expansion, and then 4 generations of random mating. We only consider 200 positions spanning chromosome 3L, simulating a QTL contributing 2% heritability in the middle of the chromosome, and an 8-fold replicated experiment. Localization is presented as a function of the effective number of individuals contributing to the DNA pools (N = 150, 300, 600), the number of RILs used to derive the base population (N_{RIL} = 200, 400, 600, 800), and the proportion selected (i = 5%, 10%, 20%).

of individuals selected (i). As was clear above, power is considerably higher and the $-\log_{10}(P)$ score at the QTL peak is greater, for larger values of N_{DNA} and i, so these parameters should be maximized regardless. Supplementary Fig. 2a depicts the $-\log_{10}(P)$ profiles in a manner similar to Fig. 3, except shows only a single realization of our simulation. This illustrates the considerable variation in the $-\log_{10}(P)$ profile of any particular simulation realization, or indeed of any real QTL mapping experiment.

We carried out additional localization simulations of a smaller region, at a higher marker density, considering only DNA preps of 300 or 600 flies, and 4 generations of random mating following base population creation (focusing on a narrow set of parameters reduced computational effort considerably). Supplementary Fig. 2b plots the average 3 $-\log_{10}(P)$ support window size as a function of average QTL peak $-\log_{10}(P)$ score, with each point the average over 250 pure replicate simulations for any particular parameter combination. This relationship shows that localization ability is a strong function of peak $-\log_{10}(P)$ score, with highly significant peaks being better localized, much like QTL mapping in general. An ANOVA (see the GitHub site) examining the effect of different experimental parameters (N_{RIL}, N_{DNA}, i, and N_{rep}) on localization after including average $-\log_{10}(P)$ score as a covariate, shows that only increasing the number of experimental replicates further impacts localization ability. That is, although the manipulated parameters all impact localization ability, their impact on localization largely parallels their impact on power. It is noteworthy that localization ability eventually reaches a plateau (at ~100 kb) that is likely related to the average unrecombined block size defining the DSPR RILs used to create the base population. Figure 3 and Supplementary Fig. 2b demonstrate that in situations where the X-QTL design is powerful, there is considerable localization ability, in many cases approaching sub-cM QTL intervals. For more modestly sized experiments, localization ability is on the order of 1 cM, comparable to RIL-based mapping with >500 RILs and multiple individuals measured per RIL (King et al. 2012a).

Our statistical test for differences in haplotype frequencies between selected and control pools results in a $-\log_{10}(P)$ statistic that summarizes the evidence for a change in allele frequency. Provided the degrees of freedom distinguishing the null and alternative hypotheses are small, $-\log_{10}(P)$ statistics are closely related to traditional LOD scores. Thus, we used the 10-kb resolution simulations above to count the proportion of simulation replicates in which the interval harboring the simulated QTL was within a 2 or 3 $-\log_{10}(P)$ drop of the most significant marker interval (MSM; conditional on the MSM being >4) as a function of N_{RIL}, i, N_{DNA}, and N_{rep}. An ANOVA showed that only N_{RIL} significantly impacts the probability of a 2 or 3 $-\log_{10}(P)$ drop including the true location of the QTL, so we pooled the simulation results over the other 3 factors to more precisely estimate properties of 2 and 3 $-\log_{10}(P)$ drops. We observed that a 2 $-\log_{10}(P)$ drop includes the true location of the QTL 82% of the time when N_{RIL} = 200, increasing to 89% of the time when N_{RIL} = 800. We conclude that a "2 $-\log_{10}(P)$ drop" is slightly liberal in terms of identifying an interval harboring the true location of the QTL ~95% of the time (as is traditionally believed). However, provided the $N_{RIL} \geq 400$, a 3 $-\log_{10}(P)$ drop includes the true location of the QTL 94–96% of the time. It is noteworthy that when the number of RILs used to construct the base population is only 200, a 3 $-\log_{10}(P)$ only includes the true location of the QTL 90% of the

time, suggesting that the more RILs used to found the base population the stronger localization ability is under our X-QTL strategy.

The impact of the number of founding RILs

It is not surprising that increasing the number of replicates of an experiment yields an increased $-\log_{10}(P)$ score at the QTL. It is also expected that increasing the number of flies contributing to the DNA pool similarly increases the average QTL $-\log_{10}(P)$ score. As the number of flies contributing to a DNA pool increases, haplotype frequency estimates in the pool approach those of the population from which the pool is sampled, thus increasing the power of the hypothesis test. It is also intuitive that increasing the selection intensity while holding N_{DNA} constant will result in greater power. However, we further observe that the number of RILs used to create the base population has a large impact on power, particularly between 200 and 600 founding RILs (Fig. 3). It is unclear why this is the case. Irrespective of the number of RILs employed to found the base population the simulated QTL is scaled to always account for 2% of total phenotypic variation (while in truth sampling would result in the QTL explaining more or less of total variation), so one's intuition is that power would not be a function of N_{RIL}. This being said, creating the base population using a greater number of founder RILs will result in founder alleles being better represented in terms of their frequencies in the base population, and allows for better estimates of the frequency changes associated with each allele.

In our initial power simulations, the false positive rate and significance threshold are obtained from a comparison of 2 equally sized draws without phenotypic selection from the base population, with other parameters varied. To rule out the possibility that certain parameter combinations give only the appearance of increased power, we further examined the distribution of $-\log_{10}(P)$ scores at markers loosely linked to the causative site (i.e. markers >7 Mb from the causative site; see Supplementary Fig. 2c). Given the mapping resolution of our experiment, the distribution of $-\log_{10}(P)$ scores at positions distant from the QTL location should approach that of 2 equally sized draws from the base population. However, $-\log_{10}(P)$ scores could be elevated at loosely linked markers if a limited sample of founder RILs led to linkage disequilibrium between causative loci and the polygenic background.

In Supplementary Fig. 2c, we observe slightly elevated $-\log_{10}(P)$ scores for certain high power parameter combinations (i.e. those employing strong selection, large pool sizes, and many replicates) when the number of generations of random intercrossing following base population establishment is ≤4, consistent with some X-QTL signal coming from polygenic background variation. Despite our simulations being able to detect this potential source of false positive mapping signal at loosely linked sites, this factor appears to have only a subtle impact on the genome-wide false positive rate, and the force is extremely weak for many realistic parameter combinations. It is also apparent that 4–8 generations of free recombination after establishment of the base population is very effective at removing this source of variation. More than 8 generations of free recombination result in little additional gain, and we speculate that this is because recombination events in an expanding base population remain at low frequency and do not have nearly the impact on localization as events present in the RILs.

Theoretical arguments aside, in practice a few generations of population expansion and recombination are likely unavoidable given the size of the base population necessary to select harshly (5% or 10% of individuals) while maintaining a large number of individuals in the selected pool (ideally up to 600). A corollary is that maintaining very large populations for extended periods is challenging and can allow for significant haplotype frequency shifts (King et al. 2012), so there will likely be a tradeoff between ensuring several of generations of intercrossing to limit long-range LD and avoiding major changes to haplotype frequencies via selection and/or drift.

Accurate haplotype frequency estimation with modest genome-wide sequencing coverage

Haplotype frequency estimates from pooled resequencing data with known founders can be quite accurate, depending on the window size, the number of founder haplotypes, the sequence divergence among haplotypes, and the level of recombination the population has experienced (Kessner et al. 2013; Tilk et al. 2019; Linder et al. 2020). We developed a method to estimate haplotype frequencies from pooled samples composed of known founders in the context of 4- and 18-way multiparental yeast populations (Cubillos et al. 2013; Burke et al. 2014b; Linder et al. 2020). In contrast to the yeast system, where the DNA pool consists of billions of cells/individuals, sampling in *Drosophila* is finite, and technical factors (body size differences, variation in the level of lysis in a bulk DNA isolation, etc.) will further lower the effective population size in a pool of individuals. On the other hand, the SNP density relative to the size of nonrecombined haplotype blocks can be higher in flies than in yeast. Furthermore, different yeast strains can share long, highly similar genomic segments (Peter et al. 2018) making unique haplotype assignment difficult. We sought to quantify the accuracy of haplotype frequency estimation in flies as a function of the particular founders and RILs actually employed.

Although we cannot measure the effective pool size in any *Drosophila* X-QTL experiment, we can measure the average squared difference between 2 estimates of each founder's haplotype frequency: the frequency of SNPs private to a single founder in a high coverage (240×) dataset, and the haplotype frequency estimate for the window nearest this SNP from down-sampled data. We detail this experiment in the *Materials and Methods* to maintain readability. The principal result is that the average squared difference over millions of private SNPs is not very sensitive to the level of downsampling, implying that estimates of haplotype frequency are extremely accurate at ∼35× coverage. Furthermore, based on the variance in SNP frequency estimates due to sampling, we estimate that the average absolute error in haplotype frequency estimates is ∼1%. This error is approximately constant for coverages from 17× to 240×.

It is important to point out a perhaps counter-intuitive corollary of the accuracy of haplotype estimates from pooled sequencing data. Under binomial sampling one would need to sequence to ∼1,000× coverage to directly estimate SNP frequencies with the same degree of accuracy we achieve for haplotypes with only 35× coverage. It is further important to note that the errors on haplotype frequency estimates are small enough (even at 35× coverage) that the finite number of flies contributing to a DNA pool can contribute to noise and lack of power. That is, with 500 individuals equally contributing DNA to a pool, the sampling error on the estimate of true population frequencies is the same order of magnitude as the error in haplotype estimation, supporting our claim that it is important to carry out X-QTL

experiments using large pools of 500–1,000 flies. Counterintuitively, estimating haplotype frequencies from pools requires only modest coverage to be effective.

Experimental test of X-QTL mapping identifies multiple QTL

We carried out an X-QTL experiment using a population constructed by mixing 663 DSPR pA RILs, allowing the population to undergo free recombination for 1–5 generations, and selecting for caffeine-resistant females in a 4-fold replicated experiment. Each replicate tested 2,337–2,572 animals, selecting the 9.0–10.3% most caffeine-resistant females in each case (228–254 animals), along with a sample of 250 control animals from the base population. Full details of each replicate are presented in Supplementary Table 1. We extracted DNA in bulk from each of the 8 resulting samples, constructed libraries, and generated ∼140× of alignable sequence per pool. We estimated frequencies of the 8 haplotypes in each pool over 200-kb windows with a 10-kb step size and, for each window, carried out a linear model testing for a change in the arcsin square root allele frequency between the selected and control pools (see *Materials and Methods* for full details).

Unlike the case of RIL- or individual-based QTL mapping, it is unclear how to create a permutation-based null distribution for X-QTL mapping (Doerge and Churchill 1996), as a result we employ the $-\log_{10}(p)$ threshold of 4 derived from our simulations. The $-\log_{10}(p)$ threshold of 4 derived from our simulations may be slightly anti-conservative in this caffeine resistance experiment, as the number of flies contributing to the DNA pool is <500. To further investigate an appropriate threshold, we simulated a dataset that mimicked the precise caffeine experiment carried out (effective number of individuals in the DNA pool = 150, number of replicates = 4, selection intensity = 10%, no QTL) and created QQ plots (Supplementary Fig. 3). Pools of ∼250 female flies (effective $N_{DNA} = 150$) can be associated with a slightly elevated false positive rate, and this is observed in the QQ plot with the simulated data showing some inflation. Nonetheless, for this parameter combination, a $-\log_{10}(P)$ threshold of 4 is associated with a genome-wide false positive rate of 0.22%. Supplementary Fig. 3 also plots the distribution of observed $-\log_{10}(P)$ scores, and $-\log_{10}(P)$ scores after the removal of all test statistics within 2 Mb of each of our 7 significant QTL peaks (Table 1—discussed below). The $-\log_{10}(P)$ scores associated with actual scans—with or without peak regions removed—show much greater inflation. The simplest explanations for inflation after mapped QTL regions are removed are that the signal of linkage extends over regions larger than 2 Mb, or that there exists a class of subtle effect QTL that individually do not reach significance, yet are contributing to signal throughout the genome.

Figure 4 plots the genome-wide $-\log_{10}(P)$ values comparing haplotype frequencies in caffeine-selected and control pools based on the full-coverage (∼142×) pooled sequencing dataset, as well as for datasets downsampled to ∼36× or ∼14×. Two patterns are immediately apparent: First, we see several significant, well-localized peaks (Table 1), consistent with our power and localization simulations (Figs. 2 and 3). It appears that we observe significant QTL despite our experiment being nonoptimal in terms of the number of replicates, the selection intensity, and the number of flies in each DNA pool. These differences were due to our analytical and experimental approaches moving forward contemporaneously, and because we elected to employ a fairly challenging caffeine resistance phenotyping regime for X-QTL mapping to more closely mimic that used originally in RIL-based mapping experiment (Najarro et al. 2015). Our assay involved loading individual flies into single *Drosophila* activity monitor tubes (as opposed to bulk testing multiple individuals in standard fly vials) was only done to copy our previous work. The benefit of our choice of character is that we can directly compare mapped QTL identified via a pooled population approach and an RIL-based approach.

Second, the locations and significance levels associated with each peak are very similar even for highly downsampled read data (compare the 3 panels of Fig. 4). This is consistent with our earlier observation that the error in haplotype frequency estimates are largely independent of coverage over the explored range. That the downsampled datasets yield similar results to the full dataset suggests that future experiments can be more efficiently and inexpensively carried out at 20–40× sequencing per pool. This may be counterintuitive to some but is consistent with other published claims that when founders are known, haplotype frequency estimation from pooled sequencing data can be quite accurate even with low coverage read data (Kessner et al. 2013; Tilk et al. 2019; Linder et al. 2020).

Figure 5 presents $-\log_{10}(P)$ scores and allele frequency changes at QTL, and Table 1 provides the properties of these mapped QTL, for the full-coverage (∼142×) dataset. Supplementary Fig. 5 presents mapped QTL and frequency changes using the ∼35× dataset. It is apparent that the properties of mapped QTL are largely the same, except for X-QTL:E, which shifts roughly 1 Mb to the left when using the downsampled data; the X-QTL:E region has 2 "peaks" that only just reach significance depending on the pooled sequencing dataset employed and has a fairly wide 3 $-\log_{10}(P)$ support interval. The *overall* genome-wide similarity of the $-\log_{10}(P)$ profiles, especially for the more significant regions,

Table 1. Properties of mapped QTL.

QTL	CHR	POS[a]	−log10(P)	3 −log10(P) CI		Size (kb)	Cand. genes	# genes[b]
				Left	Right			
A	X	3,716,075	5.36	2,916,075	4,736,075	1,820	Tlk	131
B	2L	3,370,610	4.72	3,030,610	4,290,610	1,260	E23, Ugt36A1	137
C	2L	11,730,610	4.59	10,710,610	12,170,610	1,460	Vha100-5, Crys	187
D	2R	10,848,099	13.1	10,628,099	11,168,099	540	Cyp12d1	51
E	3L	13,108,478	4.99	12,328,478	16,368,478	4,040	—	463
F	3R	14,067,353	8.81	13,447,353	14,347,353	900	Cyp6d5	105
G	3R	26,287,353	6.77	25,707,353	27,007,353	1,300	osy	144

[a] Positions (bp) given with respect to release 6 of the *D. melanogaster* reference genome.
[b] Number of protein-coding genes.
CHR, Chromosome; CI, Confidence Interval; POS, Position.

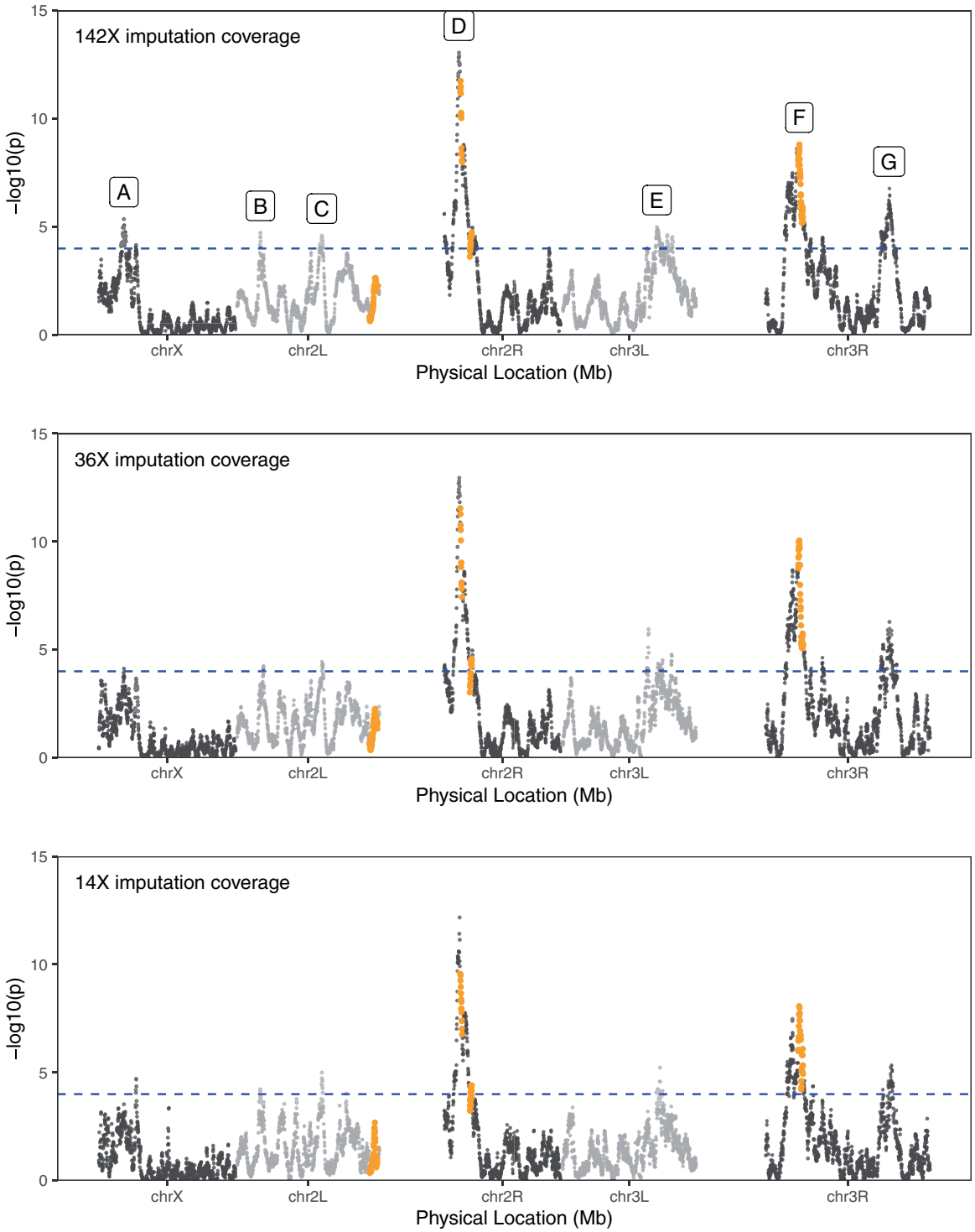

Fig. 4. X-QTL scan for caffeine resistance loci. Each panel shows $-\log_{10}(P)$ test statistics derived from comparisons between caffeine-selected and control pools using the full-coverage pooled sequencing data (~142× per sample), and data where read coverage was downsampled to ~36× or ~14×. Points on different chromosome arms are shaded differently, the x-axis scale is proportional to physical location using the *D. melanogaster* reference release 6 genome coordinates, and heterochromatic regions (Supplementary Table S3) are not plotted (leading to gaps at centromeres). The statistical threshold for QTL detection (horizontal blue, dashed line) is set at $-\log_{10}(P) = 4$. Each QTL "peak" is given a letter code, corresponding to codes in Table 1 and Fig. 5. Orange points correspond to the intervals of caffeine resistance QTL mapped previously by Najarro et al. (2015) using the DSPR pA RILs directly.

supports our claim that little (but some) additional information is gained by sequencing libraries to higher coverage. If this experiment was repeated, the impact of additional experimental replicates, stronger selection, or more flies in the selected pools would likely dwarf that of additional sequence coverage (see above simulation results).

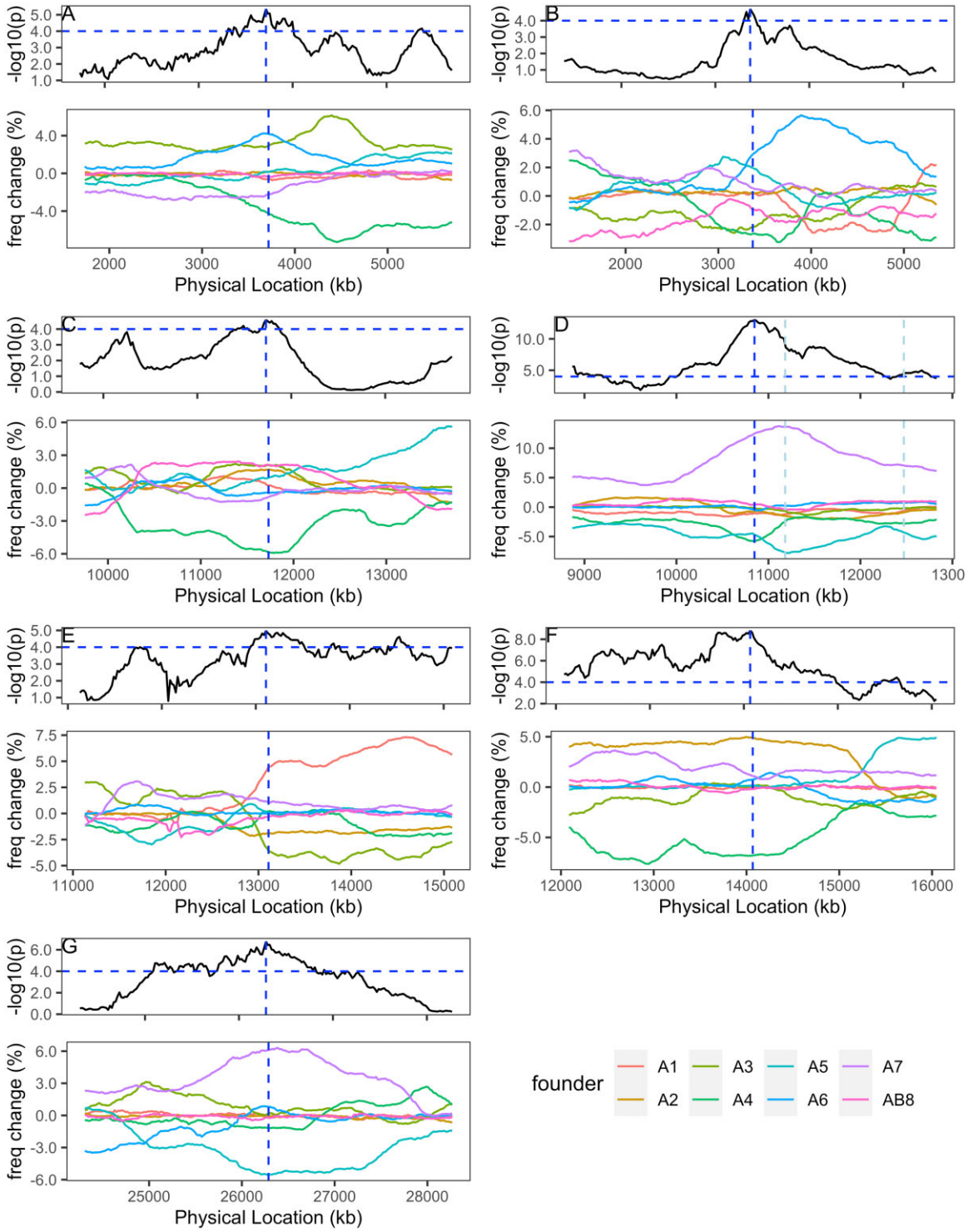

Fig. 5. Founder haplotype frequencies at caffeine X-QTL. A 2-Mb window of the genome, centered on each of the mapped X-QTL loci a–f) (see Fig. 4 and Table 1), is presented for the full-coverage sequencing data. The top panel for each X-QTL presents the $-\log_{10}(P)$ test statistics across the window, with the vertical dashed blue line at the QTL peak, and the horizontal dashed blue line at $-\log_{10}(P) = 4$. The bottom panel for each X-QTL presents the change in frequency of each haplotype, polarized such that positive values correspond to an increase in the frequency of that haplotype in the selected pool. In panel X-QTL:D, the vertical light blue lines represent the locations of the Q2 (left) and Q3 (right) loci mapped by Najarro et al. (2015) illustrating the physical proximity of these loci, and the similarity of the haplotype frequencies, which together suggest that Q3 of the prior work may be spurious.

X-QTL replicates a subset of mapped caffeine resistance loci

Najarro et al. (2015) resolved several caffeine resistance loci via directly screening DSPR RILs, 4 of which were mapped in the pA population: Q1 mapped to chromosome 2L adjacent to the centromere ($-\log_{10}(P) = 9.9$, % heritability explained = 5.5%), Q2 mapped to the centromeric end of 2R ($-\log_{10}(P) = 27.2$, heritability = 14.4%), Q3 mapped ~2 cM to the right of Q2 on 2R ($-\log_{10}(P) = 11.3$, heritability = 6.3%), and Q9 mapped to 3R ($-\log_{10}(P) = 10.2$, heritability = 5.7%). The intervals implicated by these 4 mapped QTL in the original

study are reflected in our X-QTL map with orange symbols, and it appears that we replicate 2 of them (Fig. 4, top panel, X-QTL:D and X-QTL:F).

The largest-effect QTL mapped in RILs, Q2, was replicated by X-QTL:D (Fig. 4 and Table 1). Najarro et al. (2015) found that RILs carrying the A7 haplotype at this locus had the highest average caffeine resistance. Similarly, we estimate that, at this position in our X-QTL study, haplotype A7 exhibits the greatest frequency increase in the selected pools over the control pools (Fig. 5D). Notably, founder A7 is the only pA founder strain that possesses 2 copies of the *Cyp12d1* cytochrome P450 detoxification gene (as does the *D. melanogaster* genome reference strain, where the copies are termed *Cyp12d1-d* and *Cyp12d1-p*). *Cyp12d1* expression is known to be induced in response to caffeine (Willoughby et al. 2006; Coelho et al. 2015; Najarro et al. 2015), and experiments using ubiquitous Gal4-UAS-RNAi knockdown (under the control of an actin promoter), and adult-specific knockdown (using an RU486-inducible "GeneSwitch" actin Gal4 driver) indicated that *Cyp12d1* knockdown reduces caffeine resistance (Najarro et al. 2015). In addition, Najarro et al. (2015) found a significant association between *Cyp12d1* copy number variation (CNV) and caffeine resistance in both the pA and pB RIL panels. By controlling for *Cyp12d1* CNV status during RIL-based QTL mapping (via the addition of a covariate into the mapping model), the Q2 QTL was eliminated, suggesting that the CNV—or something in linkage disequilibrium with it—is causative.

We also replicated RIL-based QTL Q9 with X-QTL:F (Fig. 4 and Table 1). In our previous RIL-based study, RILs carrying founder haplotype A2 at this location showed higher average caffeine resistance, while RILs carrying A4 had among the lowest resistance. We recapitulate these findings, showing that A2 exhibits the greatest frequency increase in selected pools, while A4 shows a large frequency decrease in the selected pools (Fig. 5f). Both the Q9 and X-QTL:F intervals contain *Cyp6d5*, a gene that is transcriptionally induced in response to caffeine exposure (Najarro et al. 2015).

Q1 from our RIL-based study does not replicate here. We previously estimated its effect to be modest, so easy replication may not be anticipated (Zhou et al. 2020). In addition, the QTL is extremely close to the chromosome 2 centromere, a location where effective mapping is challenging (Noor et al. 2001). As a result, the original Q1 may not represent a true causative caffeine resistance locus. RIL-based Q3 also fails to replicate, and again it has a modest effect. In addition, Najarro et al. (2015) mapped Q3 to a position just 2.5 cM away from the large-effect Q2 QTL, so Q3 may not represent an independent locus. Indeed, the haplotype frequencies at both these positions are similar in our X-QTL populations (Fig. 5d compare frequencies at positions 11,182 and 12,472 kb, which are the locations of Q2 and Q3, respectively).

Functional testing plausible candidate caffeine resistance genes

Six of the 7 X-QTL we identify are mapped to 540–1,820-kb intervals, each encompassing 51–187 protein-coding genes (Table 1). In addition to the 2 candidates discussed above—*Cyp12d1* (X-QTL:D) and *Cyp6d5* (X-QTL:F)—a number of plausible, novel candidate genes emerge from mapped X-QTL. X-QTL:A includes *Tlk* (*Tousled-like kinase*), a gene involved in cell cycle progression and the DNA damage response, that also contains a caffeine-sensitive phosphorylation site (Groth et al. 2003). X-QTL:B includes *E23* (*Early gene at 23*), which encodes an ATP-binding cassette (ABC) transporter subunit (Hock et al. 2000)—ABC transporters have important roles in xenobiotic metabolism (Xu et al. 2005), and *Ugt36A1* (*UDP-glycosyltransferase family 36 member A1*), which is a UDP glycosyltransferase detoxification enzyme gene. X-QTL:C encompasses *Crys* (*Crystallin*), whose gene product is involved in the formation of the peritrophic matrix (Kuraishi et al. 2011), a structure that forms a protective layer lining the midgut epithelium (Terra 2001), and *Vha100-5* (*Vacuolar H^+ ATPase 100kD subunit 5*), which encodes a subunit of an ATP-dependent proton pump and is expressed in the fly midgut (Overend et al. 2016). X-QTL:G encompasses the gene *osy* (*oskyddad*), which encodes an ABC transporter subunit, is required to provide a barrier to xenobiotics in *D. melanogaster* larvae (Wang et al. 2020), and shows reduced gene expression in adult males exposed to caffeine (Coelho et al. 2015).

To evaluate the effects of each of these 8 genes, we employed both ubiquitous RNAi knockdown using Gal4 under the control of an actin promoter and expressed Gal4 specifically in the anterior portion of the adult midgut. Using both drivers, and both reciprocal cross directions (see *Materials and Methods*), both *Cyp12d1* and *Cyp6d5* show a significant reduction in lifespan on caffeine-containing media compared to the control, no-knockdown genotype (Dunnett's tests, $0.05 < P < 10^{-12}$, Fig. 6), reconfirming the strong relevance of this pair of detoxification genes in the response to caffeine challenge. The *Tlk* gene (X-QTL:A) only shows a significant reduction in resistance when knocked down in the gut in 1 of the 2 reciprocal crosses (cross direction-specific Dunnett's tests, $P < 10^{-4}$ and $P = 0.08$), while the ubiquitous knockdown was lethal (Supplementary Table 2). Of the 2 genes tested within the X-QTL:B interval (*E23*, *Ugt36A1*) the *E23* ABC transporter subunit gene appears to be the best candidate based on the RNAi; knockdown of this gene leads to significantly reduced lifespan in all 4 experiments (Dunnett's tests, $0.01 < P < 10^{-8}$, Fig. 6), while knockdown of *Ugt36A1* shows no effect on phenotype in the gut, and only a minor reduction in resistance in 1 of the 2 ubiquitous knockdown reciprocal crosses. Of the 2 genes tested beneath X-QTL:C, *Crys* shows no evidence for an effect (Fig. 6), while *Vha100-5* gut knockdown reduces resistance (cross direction-specific Dunnett's tests, $P < 10^{-4}$ and $P < 0.01$), and resistance is reduced in 1 of the 2 reciprocal crosses of the ubiquitous knockdown tests (Dunnett's tests, $P = 0.8$ and $P < 0.001$). Finally, *osy*, the sole gene tested within the X-QTL:G interval shows no effect with gut knockdown (Fig. 6), and ubiquitous knockdown led to lethality (Supplementary Table 2). Ultimately, the novel caffeine resistance candidates *E23* and *Vha100-5* appear particularly worthy of additional functional study.

Discussion

Here, we explore an alternative to RIL-based QTL mapping in *Drosophila* that is analogous to the X-QTL approach of yeast (Ehrenreich et al. 2010), and more recently the nematode model system *C. elegans* (Burga et al. 2019). Conceptually derived from the bulked segregant analysis mapping strategy (Michelmore et al. 1991), X-QTL mapping as employed in yeast utilizes extremely large pools of haploid segregants from a single-generation cross between 2 strains. Unlike this 2-parental strain, single-generation cross, the base population we employ is created by combining several hundred advanced intercross DSPR RILs (King et al. 2012a, 2012b), and up to 8 haplotypes segregate at any given genomic location in the DSPR X-QTL mapping population.

X-QTL mapping draws replicate samples of phenotypically extreme and control individuals from a large base population and identifies regions of the genome exhibiting consistent differences

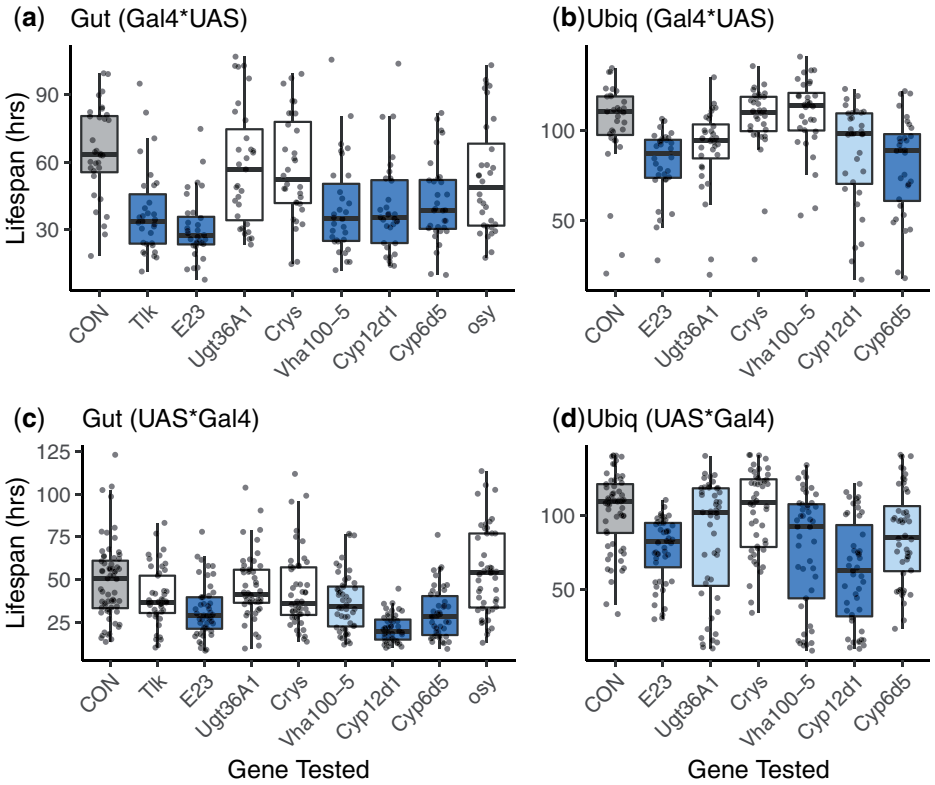

Fig. 6. Effects of candidate gene RNAi knockdown on caffeine resistance. Shows the lifespan (in hours) of replicate females from a series of Gal4-UAS-RNAi genotypes (and controls, CON) on 1% caffeine media. The genes impacted are listed by their symbols along the x-axes of the plots. Each dot represents the phenotype of a single individual, and for each genotype, the individual-level scores are overlaid with a boxplot (following the R ggplot2 "geom_boxplot" defaults). Each panel represents a different Gal4 driver and a different cross direction: a and c) anterior midgut driver and b and d) ubiquitous driver. a and b) Test individuals are the result of crossing male Gal4 to female UAS and c and d) test individuals are the result of crossing male UAS to female Gal4. Within each combination of driver and cross direction, all Gal4-UAS-RNAi genotypes were compared to the control genotype using a Dunnett's test. Box colors reflect these tests: gray = CON, white = not significantly different from control ($P > 0.05$), light blue = significantly different from control at $P < 0.05$, dark blue = significantly different from control at $P < 0.005$.

in founder haplotype frequencies between extreme and control pools following short-read sequencing (Fig. 1). For a QTL contributing 2% to variation in a complex trait, we explore the power and localization ability of this X-QTL mapping approach as a function of the number of experimental replicates, the number of RILs used to found the base population, the number of individuals contributing to a DNA pool, the proportion of individuals selected, and the number of generations of expansion/recombination prior to phenotyping the intercrossed base population. Our simulations suggest that for difficult, but experimentally achievable designs, the X-QTL approach can routinely have considerably higher power than traditional RIL-based mapping (Fig. 2), and at least as high QTL localization ability (Fig. 3). This being said, power and localization ability is impacted by the precise experimental design employed and the highest power is only achieved when the number of RILs contributing to a base is greater than 500, selection intensity is greater than 10% (ideally one should aim for ~5%), the number of flies contributing to the DNA pool is >500 (implying that the total number of individuals assayed is ≥5,000), and 4 (and ideally 8 or more) experimental replicates are used.

Several other mixed population mapping strategies have previously been executed in *Drosophila*. Huang et al. (2012) derived a population—termed "Flyland"—from 40 inbred strains of the DGRP (Mackay et al. 2012; Huang et al. 2014). This population was allowed to undergo recombination for 70 generations, 2,000 individuals were phenotyped for each of 3 complex traits, and allele frequencies were contrasted between sequenced pools of 300 phenotypically extreme individuals. A similar design has been employed in studies of a number of other phenotypes (e.g. Morozova et al. 2015; Zhou et al. 2017; Fochler et al. 2017), each resulting in tens to hundreds of variants exhibiting allele frequency differences between control and selected pools. No analytical details of the power or resolution of this approach are available, and in contrast to our simulated and experimental data, where increasing numbers of replicates greatly improve power, these studies are generally unreplicated. Flyland studies also typically phenotype fewer individuals than our simulations suggest is optimal for the highest power. And since the strategy relies on directly contrasting SNP allele frequency between pools, rather than using the haplotype approach we take (which improves allele frequency estimates, Tilk et al. 2019), unless sequencing coverage is high, the power of the Flyland approach may be low.

Another group has proposed the use of "hybrid swarm" populations. Here, many inbred strains—34 in the case of (Erickson et al. 2020)—are mixed for 4–5 generations, then recombinant animals are phenotyped and individually subjected to ultra-low pass sequencing, facilitating GWAS-type mapping analyses (Erickson et al. 2020; Weller et al. 2021). This approach has the advantage of not requiring RILs and needing only a small number of generations of population maintenance. But, while the total sequencing effort is reasonable, rather than working with DNA pools, an investigator would need to efficiently and

inexpensively generate thousands of individually indexed sequencing libraries.

Several studies have used various mixed population designs to genetically dissect body pigmentation, a classic *Drosophila* trait impacted by a small set of well-documented candidate genes (reviewed in Massey and Wittkopp 2016). Bastide et al. (2013) phenotyped 8,000 outbred animals derived from thousands of unknown founders and, after sequencing pools of extremely light or dark animals, successfully resolved large-effect loci to classic pigmentation loci (*bric-a-brac*, *ebony*, *tan*). Dembeck et al. (2015) created a population from 3 light and 3 dark pigmented DGRP lines, maintained this for a few generations and found allele frequency differences between sequencing pools generated from light and dark flies that localized to classic pigmentation gene locations, in addition to other regions of the genome. Bastide et al. (2016) used a series of 2-way advanced intercross, bulked segregant analysis mapping populations to yield a similar result; some peaks resolve to intervals harboring known pigmentation genes, while others might implicate novel candidates. Collectively, these studies suggest that there are intermediate frequency causative SNPs of large effect impacting *Drosophila* pigmentation, and as a result, frequency differences at single SNPs in mapping populations can be dramatic, and the effects at QTL sometimes be quite large. The applicability of these approaches to more polygenic traits is unknown.

Perhaps the most similar experiment to the one we describe here is Burke et al. (2014a), who sequenced pools of the 2% most extreme long-lived individuals in the synthetic populations from which the DSPR RILs were derived (King et al. 2012b). However, this aging study only obtained ∼100 flies for each sequencing pool, and while 4 populations were used, none were replicated. The simulations we present here suggest that despite some success, this longevity experiment was likely only modestly powered.

Key features of the X-QTL approach we recommend are that a large, highly recombinant population is created and that several pairs of matched extreme and control DNA pools are each generated from large numbers of animals and sequenced. A great deal of the power of the design comes from the populations being derived from a modest number of known founders, which allows tests for differentiation between control and extreme pools to be carried out on imputed founder haplotype frequencies, as opposed to using directly ascertained SNP frequencies.

A perhaps nonintuitive, but well-established result (Long et al. 2011; Kessner et al. 2013; Burke et al. 2014b; Tilk et al. 2019; Linder et al. 2020) is that by virtue of the base population ultimately consisting of genetic material from only 8 highly characterized "founders," sliding window haplotype frequency estimates are extremely accurate, even with as little as 20–40× of short-read sequencing. We show that at 35× sequencing coverage, for much of the genome, founder haplotype frequencies are estimated with absolute errors of roughly 0.01. Accurate haplotype frequency estimates strongly impact the power of the X-QTL approach in flies. If founder haplotypes are ignored or unavailable, and SNP frequencies are employed directly, one would require sequencing coverage of ∼1,000× to obtain similarly accurate allele frequency estimates. The modest sequencing coverage necessary under our design means that the sequencing budget for an 8-fold replicated experiment can be relatively low. A primary practical limitation of our approach is fly handling; the pools of flies used to make DNA must consist of 100s of individuals, and given the necessary harsh selection regime, thousands of individuals must initially be targeted (i.e. N_{DNA}/i).

Our 4-fold replicated proof of principle experiment used pools of ∼250 female flies either randomly chosen from a base population or in the top ∼10% of individuals surviving following exposure to high levels of caffeine. We chose this demonstration trait as it has been the subject of past research using an RIL-based mapping approach (Najarro et al. 2015), and we wished to determine if QTL identified by the 2 methods mapped to the same locations; but the assay method involves considerable fly handling and is not ideally suited to X-QTL mapping. As a result, our empirical test of the approach was carried out with parameters near the lower end of those our simulations suggest yield the highest power. Nonetheless, we still mapped 7 X-QTL at a genome-wide false positive rate of <5% (Table 1). Two of these X-QTL overlap with 2 of the 4 QTL mapped by Najarro et al. (2015) in the population A panel of the DSPR RILs, the progenitors of the X-QTL population employed in the present study. Not only is there overlap of the QTL intervals for this pair of QTL, but the founder effects are also consistent over studies; overlapping X-QTL:D and RIL Q2 is driven by founder A7 (Fig. 5), and overlapping X-QTL:F and RIL Q9 shows that founders A2 and A4 are associated with high and low resistance, respectively (Fig. 5). Thus, the apparent replication of these QTL is supported by their positions, but also the effects of the haplotypes on phenotype. In addition, the X-QTL results suggest that one or both of the previously mapped but unreplicated QTL may be spurious; one is very close to the centromere and the other is very close to a larger-effect, replicated locus. Finally, with X-QTL, we were able to map 5 new caffeine resistance QTL associated with 6 novel and preexisting candidate genes (Table 1).

Using both general and gut-specific Gal4 drivers, we attempted to functionally validate implicated genes; both candidates underlying the pair of replicated QTL found here and in Najarro et al. (2015)—*Cyp12d1* and *Cyp6d5*—yielded reduced caffeine resistance on knockdown. Furthermore, 2 new candidate genes—*E23* and *Vha100-5*—received functional support. Our RNAi results support the contention that these genes harbor segregating variation that impacts caffeine resistance, but it is obvious that RNAi cannot specifically demonstrate this. We contend that our RNAi-based replication of caffeine-associated effects at genes *Cyp12d1* and *Cyp6d5* is encouraging, since these genes have been previously implicated in caffeine resistance by others via orthogonal methods; both are transcriptionally induced in response to caffeine (Willoughby et al. 2006) and have been demonstrated to impact caffeine metabolism (Willoughby et al. 2006; Coelho et al. 2015). In addition, since we employ gut-restricted knockdown, our results may be less impacted by nonspecific RNAi-induced fitness defects than experiments using ubiquitous drivers, or those employing germline mutations. Validation of candidate genes via knockdown can only be considered suggestive, as genes chosen based on function in a genomic interval harboring dozens of genes are prone to phenocopies at an unknown rate. Nonetheless, we suggest that the genes we implicate here are excellent candidates for studies that can specifically test for the presence of caffeine-relevant allelic variation.

Comparing 2 caffeine resistance experiments employing near identical phenotyping regimes, it appears the results from the X-QTL study provided at least similar, if not superior genetic insight into the phenotype than the RIL-based study. And had we not wished to compare designs, enforcing use of the same phenotyping assay, a less cumbersome bulked phenotyping approach would likely have been possible for X-QTL mapping (e.g. exposing groups of flies to caffeine-supplemented media in vials vs testing flies singly in narrow tubes), yielding larger numbers

of phenotyped animals, and—as our simulations suggest—higher power.

Several caveats of X-QTL mapping in flies are apparent, many resulting from the pooled nature of the phenotyping and genotyping that is employed. First, since X-QTL are detected as haplotype frequency differences between pooled samples each consisting of hundreds of outbred animals, one does not obtain estimates of dominance or epistasis associated with mapped X-QTL. Dominance and epistasis can be estimated using RILs (or individuals), although we note that dominance cannot be estimated when RILs are directly phenotyped, the most common mapping strategy using RILs. Furthermore, in MPPs segregating for 8 founder alleles, allele-specific dominance and epistatic terms can be difficult to accurately estimate. Working directly with 8-way RILs, there are 64 two gene epistatic terms and, if the RILs are intercrossed to obtain heterozygous genotypes, there are potentially 28 dominance terms, and a very large number of epistatic terms. Investigators typically phenotype on the order of ∼500 DSPR RILs, so estimating dominance and epistasis in the MPP framework can quickly result in over-fitted models. The result is that in practice, dominance and epistasis are rarely studied using the DSPR or MPPs more generally (although see King et al. 2012b; Cogni et al. 2016). In addition, while any characterization of the genetic basis of complex trait variation will be incomplete without a description of dominance and epistasis (Ehrenreich 2017), the contribution of nonadditive effects to the variance of complex traits may not be high in most cases (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014; Bloom et al. 2015; Albert et al. 2018; Nag et al. 2020; Hivert et al. 2021). Finally, when epistatic interactions have been identified in powerful yeast QTL studies, they generally involve at least 1 locus that also has a main effect (Bloom et al. 2015; Albert et al. 2018), implying that an X-QTL strategy can identify major players involved in epistatic interactions, and follow-up studies could identify epistasis.

Second, in a study using inbred lines, one might collect multiple phenotypes on the same set of lines, allowing examination of genetic correlations among traits (e.g. Dickson et al. 2016; Everman et al. 2019). Furthermore, with inbred lines it is straightforward to include known covariates associated with each strain, for instance, infection status with the common *Wolbachia* microbe, or inversion genotype, during mapping (both these covariates are accounted for as standard by the DGRP analytical machinery; Huang et al. 2014). Equally, one can statistically account for larger-effect QTL, enabling more powerful mapping of smaller-effect loci (e.g. Cogni et al. 2016). Such analyses are not possible with X-QTL mapping given the pooled nature of the phenotyping and genotyping.

Third, screens of the mouse Collaborative Cross set of 8-way RILs have identified single RILs having particularly extreme or interesting phenotypes; for example, RILs having high susceptibility to epileptic seizures (Gu et al. 2020), or high susceptibility to *Salmonella* infection (Zhang et al. 2018). Such strains have potential utility as novel models of disease. But again, given the bulked phenotyping and genotyping approach that X-QTL relies on, such interesting phenotypes would not be captured in stable genotypes via X-QTL mapping.

Finally, not all target traits will be as amenable to bulk phenotyping as the resistance trait we empirically explore here. When every individual must be individually handled/scored (e.g. counting bristle number—Macdonald and Long 2007), an RIL-by-RIL approach may be more profitable, especially given some of the other caveats of X-QTL mapping (above). However, X-QTL mapping is attractive and can have high detection and localization ability for traits where some form of "self-selection" is possible (e.g. measures of toxin/stress resistance, and behaviors such as negative geotaxis, or oviposition preference). The list of traits that has been dissected using the DGRP (Table 3 of Mackay and Huang 2018) suggests that roughly half of the studied traits could have been phenotyped in bulk, some perhaps quite easily. That is, many of the trait drosophilists currently study are potentially suitable for dissection via an X-QTL approach.

Despite some negatives, several additional efficiencies are obtained with X-QTL mapping. By virtue of working with a single large population, as opposed to several thousand vials, the accurate tracking of individuals/vials/blocks/strains necessitated in an RIL-based study is largely avoided. Furthermore, since phenotypes are all obtained from outbred genotypes, the impact of rare recessive genotypes and inbreeding depression are muted. In this regard, the contribution of inbreeding depression to RIL-based mapping results has rarely been quantified. Furthermore, as all flies are reared in a common garden, X-QTL mapping can also be advantageous for traits for which block and/or vial effects are unavoidable. Lastly, our simulations suggest that the highest powered X-QTL designs can enjoy 3 $-\log_{10}(P)$ support intervals, roughly 95% confidence intervals on true QTL location, that approach 100 kb. This is approaching single gene resolution in *Drosophila*. Narrow 3 $-\log_{10}(P)$ support intervals can greatly accelerate the process of sifting through candidate genes, especially as the field moves toward stronger validation approaches that are more time consuming to carry out.

It is important to note that the resolution of the experiments we simulate and experimentally employ assuming the base population is initiated with highly recombinant genotypes. DSPR RILs were derived from a population allowed to intercross for 50 generations prior to RIL initiation, and less recombined material would be associated with much lower mapping resolution. That said, the availability of the DSPR need not be a constraint on the design we outline. Indeed, early MPP-based mapping in *D. melanogaster* did not employ RILs and instead directly interrogated a segregating population (Macdonald and Long 2007), much like the mouse Diversity Outbred population (Svenson et al. 2012). It would be possible to develop equivalent mixed base populations for X-QTL mapping by intercrossing the original 8 DSPR founder strains (or any set of inbred strains), followed by maintenance of the resulting population for 50 generations (∼2 years) to build up recombination events. Clearly the availability of RILs, and the ability to skip this population creation step, is a time- and labor-saving feature of the approach we outline.

X-QTL mapping can provide powerful, high-resolution mapping of QTL. In concert with emerging CRISPR-Cas9 homologous recombination-based allele replacements (Gratz et al. 2013, 2014; Ren et al. 2013; Port et al. 2014; Lamb et al. 2017), prime-editing strategies (Anzalone et al. 2019; Bosch et al. 2021), and recombinase-mediated cassette exchange following Cas9 (Bateman et al. 2006; Voutev and Mann 2018), X-QTL mapping may allow the field to move from mapped, main effect QTL to a more precise functional characterization of candidate genes than is possible via RNAi knockdowns. As these replacement/editing technologies become more mature, especially in model systems, QTL mapping may become more focused on the identification of high-confidence candidate genes underlying large, main-effect loci, with the accurate estimation of the effects associated with mapped factors being left to specific follow-up experiments. For many characters of interest to *Drosophila* geneticists, the X-QTL approach we describe may provide a blueprint for quickly and cost-effectively identifying candidate genes underlying additive QTL.

Data availability

Simulation, analysis code, and code to reproduce many of the figures are available from GitHub (https://github.com/tdlong/fly_XQTL). Short-read sequencing data for all 8 control and caffeine-selected pools can be obtained from the NCBI SRA under bioproject accession PRJNA714149. RNAi data and R analysis scripts (rnai_supp_data_analysis.zip) and the haplotype calls for each population (allhaps.200 kb.txt.gz) are available at figshare (data uploaded to figshare: https://doi.org/10.25386/genetics.16368381). The haplotype calls allow the mapping results to be reproduced and the approach explored without the computation effort or expertise associated with aligning raw reads to the genome, calling SNPs, and calling haplotypes.

Funding

This work was supported by NIH R01 OD010974 (to SJM and ADL), NIH R01 ES029922 (to SJM), and NIH R01 GM115562 (to ADL). We also thank the University of Kansas Genome Sequencing Core facility (funded by NIH P20 GM103638) for library construction, and the Kansas INBRE program (funded via NIH P20 GM103418) for computational support.

Conflicts of interest statement

The authors declare no conflict of interest.

Literature cited

Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. Genetics of trans-regulatory variation in gene expression. eLife. 2018;7:e35471.

Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149–157.

Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS, Ferris MT, Frelinger JA, Heise M, Frieman MB, et al. Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Res. 2011;21(8):1213–1222.

Bastide H, Betancourt A, Nolte V, Tobler R, Stöbe P, Futschik A, Schlötterer C. A genome-wide, fine-scale map of natural pigmentation variation in Drosophila melanogaster. PLoS Genet. 2013;9(6):e1003534.

Bastide H, Lange JD, Lack JB, Yassin A, Pool JE. A variable genetic architecture of melanic evolution in Drosophila melanogaster. Genetics. 2016;204(3):1307–1319.

Bateman JR, Lee AM, Wu C-T. Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. Genetics. 2006;173(2):769–777.

Baud A, Hermsen R, Guryev V, Stridh P, Graham D, McBride MW, Foroud T, Calderari S, Diez M, Ockinger J, et al.; Rat Genome Sequencing and Mapping Consortium. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. Nat Genet. 2013;45(7):767–775.

Beavis WD, Grant D, Albertsen M, Fincher R. Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. Theor Appl Genet. 1991;83(2):141–145.

Beavis WD. The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference. 1994;250:266.

Bloom JS, Kotenko I, Sadhu MJ, Treusch S, Albert FW, Kruglyak L. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nat Commun. 2015;6:8712.

Bosch JA, Birchak G, Perrimon N. Precise genome engineering in Drosophila using prime editing. Proc Natl Acad Sci U S A. 2021;118:e2021996118.

Buchon N, Osman D, David FPA, Fang HY, Boquete J-P, Deplancke B, Lemaitre B. Morphological and molecular characterization of adult midgut compartmentalization in Drosophila. Cell Rep. 2013;3(5):1725–1738.

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005–D1012.

Burga A, Ben-David E, Lemus Vergara T, Boocock J, Kruglyak L. Fast genetic mapping of complex traits in C. elegans using millions of individuals in bulk. Nat Commun. 2019;10:2680.

Burke MK, King EG, Shahrestani P, Rose MR, Long AD. Genome-wide association study of extreme longevity in Drosophila melanogaster. Genome Biol Evol. 2014a;6(1):1–11.

Burke MK, Liti G, Long AD. Standing genetic variation drives repeatable experimental evolution in outcrossing populations of Saccharomyces cerevisiae. Mol Biol Evol. 2014b;31(12):3228–3239.

Chakraborty M, Emerson JJ, Macdonald SJ, Long AD. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. Nat Commun. 2019;10(1):4872.

Coelho A, Fraichard S, Le Goff G, Faure P, Artur Y, Ferveur J-F, Heydel J-M. Cytochrome P450-dependent metabolism of caffeine in Drosophila melanogaster. PLoS One. 2015;10(2):e0117328.

Cogni R, Cao C, Day JP, Bridson C, Jiggins FM. The genetic architecture of resistance to virus infection in Drosophila. Mol Ecol. 2016;25(20):5228–5241.

Cubillos FA, Parts L, Salinas F, Bergström A, Scovacricchi E, Zia A, Illingworth CJR, Mustonen V, Ibstedt S, Warringer J, et al. High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. Genetics. 2013;195(3):1141–1155.

Davies RW, Flint J, Myers S, Mott R. Rapid genotype imputation from sequence without reference panels. Nat Genet. 2016;48(8):965–969.

Dembeck LM, Huang W, Carbone MA, Mackay TFC. Genetic basis of natural variation in body pigmentation in Drosophila melanogaster. Fly (Austin). 2015;9(2):75–81.

Dickson PE, Miller MM, Calton MA, Bubier JA, Cook MN, Goldowitz D, Chesler EJ, Mittleman G. Systems genetics of intravenous cocaine self-administration in the BXD recombinant inbred mouse panel. Psychopharmacology (Berl). 2016;233(4):701–714.

Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature. 2007;448(7150):151–156.

Doerge RW, Churchill GA. Permutation tests for multiple loci affecting a quantitative character. Genetics. 1996;142(1):285–294.

Ehrenreich IM.Epistasis: Searching for Interacting Genetic Variants Using Crosses. Genetics. 2017;206(2):531–535.

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. 2010;464(7291):1039–1042.

Erickson PA, Weller CA, Song DY, Bangerter AS, Schmidt P, Bergland AO. Unique genetic signatures of local adaptation over space and time for diapause, an ecologically relevant complex trait, in Drosophila melanogaster. PLoS Genet. 2020;16(11):e1009110.

Everman ER, McNeil CL, Hackett JL, Bain CL, Macdonald SJ. Dissection of complex, fitness-related traits in multiple drosophila mapping populations offers insight into the genetic control of stress resistance. Genetics. 2019;211(4):1449–1467.

Fochler S, Morozova TV, Davis MR, Gearhart AW, Huang W, Mackay TFC, Anholt RRH. Genetics of alcohol consumption in Drosophila melanogaster. Genes Brain Behav. 2017;16(7):675–685.

Gatti DM, Svenson KL, Shabalin A, Wu L-Y, Valdar W, Simecek P, Goodwin N, Cheng R, Pomp D, Palmer A, et al. Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda). 2014;4:1623–1633.

Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wildonger J, O'Connor-Giles KM. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics. 2013;194(4):1029–1035.

Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. Genetics. 2014;196(4):961–971.

Groth A, Lukas J, Nigg EA, Silljé HHW, Wernstedt C, Bartek J, Hansen K. Human Tousled like kinases are targeted by an ATM- and Chk1-dependent DNA damage checkpoint. EMBO J. 2003;22(7):1676–1687.

Gu B, Shorter JR, Williams LH, Bell TA, Hock P, Dalton KA, Pan Y, Miller DR, Shaw GD, Philpot BD, et al. Collaborative Cross mice reveal extreme epilepsy phenotypes and genetic loci for seizure susceptibility. Epilepsia. 2020;61(9):2010–2021.

Hivert V, Sidorenko J, Rohart F, Goddard ME, Yang J, Wray NR, Yengo L, Visscher PM. Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. Am J Hum Genet. 2021;108(5):786–798.

Hock T, Cottrill T, Keegan J, Garza D. The E23 early gene of Drosophila encodes an ecdysone-inducible ATP-binding cassette transporter capable of repressing ecdysone-mediated gene activation. Proc Natl Acad Sci U S A. 2000;97(17):9519–9524.

Hormozdiari F, Zhu A, Kichaev G, Ju CJ-T, Segrè AV, Joo JWJ, Won H, Sankararaman S, Pasaniuc B, Shifman S, et al. Widespread allelic heterogeneity in complex traits. Am J Hum Genet. 2017;100(5):789–802.

Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, et al. Epistasis dominates the genetic architecture of Drosophila quantitative traits. Proc Natl Acad Sci U S A. 2012;109(39):15553–15559.

Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 2014;24(7):1193–1208.

Keele GR, Crouse WL, Kelada SNP, Valdar W. Determinants of QTL mapping power in the realized collaborative cross. G3 (Bethesda). 2019;9(5):1707–1727.

Kessner D, Turner TL, Novembre J. Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. Mol Biol Evol. 2013;30(5):1145–1158.

King EG, Macdonald SJ, Long AD. Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. Genetics. 2012a;191(3):935–949.

King EG, Merkes CM, McNeil CL, Hoofer SR, Sen S, Broman KW, Long AD, Macdonald SJ. Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 2012b;22(8):1558–1566.

King EG, Sanderson BJ, McNeil CL, Long AD, Macdonald SJ. Genetic dissection of the Drosophila melanogaster female head transcriptome reveals widespread allelic heterogeneity. PLoS Genet. 2014;10(5):e1004322.

King EG, Long AD. The Beavis effect in next-generation mapping panels in Drosophila melanogaster. G3 (Bethesda). 2017;7:1643–1652.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. 2009;5(7):e1000551.

Kuraishi T, Binggeli O, Opota O, Buchon N, Lemaitre B. Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2011;108(38):15966–15971.

Lamb AM, Walker EA, Wittkopp PJ. Tools and strategies for scarless allele replacement in Drosophila using CRISPR/Cas9. Fly (Austin). 2017;11(1):53–64.

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987–2993.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv q-bio.GN, 2013.

Linder RA, Majumder A, Chakraborty M, Long A. Two synthetic 18-way outcrossed populations of diploid budding yeast with utility for complex trait dissection. Genetics. 2020;215(2):323–342. https://doi.org/10.1534/genetics.120.303202

Long Q, Jeffares DC, Zhang Q, Ye K, Nizhynska V, Ning Z, Tyler-Smith C, Nordborg M. PoolHap: inferring haplotype frequencies from pooled samples by next generation sequencing. PLoS One. 2011;6(1):e15292.

Long AD, Macdonald SJ, King EG. Dissecting complex traits using the Drosophila Synthetic Population Resource. Trends Genet. 2014;30(11):488–495.

Macdonald SJ, Long AD. Joint estimates of quantitative trait locus effect and frequency using synthetic recombinant populations of Drosophila melanogaster. Genetics. 2007;176(2):1261–1281.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482(7384):173–178.

Mackay TFC, Huang W. Charting the genotype–phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. Wiley Interdiscip. Rev. Dev. Biol. 2018;7:e289.

Magwire MM, Fabian DK, Schweyen H, Cao C, Longdon B, Bayer F, Jiggins FM. Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in Drosophila melanogaster. PLoS Genet. 2012;8(11):e1003057.

Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G, Jiggins CD, Kronforst MR, Long AD, McMillan WO, et al. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc Natl Acad Sci U S A. 2012;109(31):12632–12637.

Massey JH, Wittkopp PJ. The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species. Curr Top Dev Biol. 2016;119:27–61.

Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A. 1991;88(21):9828–9832.

Morozova TV, Huang W, Pray VA, Whitham T, Anholt RRH, Mackay TFC. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult drosophila. BMC Genomics. 2015;16:865.

Nag A, McCarthy MI, Mahajan A. Large-scale analyses provide no evidence for gene-gene interactions influencing type 2 diabetes risk. Diabetes. 2020;69(11):2518–2522.

Najarro MA, Hackett JL, Smith BR, Highfill CA, King EG, Long AD, Macdonald SJ. Identifying loci contributing to natural variation in xenobiotic resistance in Drosophila. PLoS Genet. 2015;11(11):e1005663.

Noble LM, Rockman MV, Teotónio H. Gene-level quantitative trait mapping in Caenorhabditis elegans. G3 (Bethesda). 2021;11: jkaa061.

Noor MA, Cunningham AL, Larkin JC. Consequences of recombination rate variation on quantitative trait locus mapping studies. Simulations based on the Drosophila melanogaster genome. Genetics. 2001;159(2):581–588.

Overend G, Luo Y, Henderson L, Douglas AE, Davies SA, Dow JAT. Molecular mechanism and functional significance of acid generation in the Drosophila midgut. Sci Rep. 2016;6:27242.

Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, Sigwalt A, Barre B, Freel K, Llored A, et al. Genome evolution across 1,011 Saccharomyces cerevisiae isolates. Nature. 2018; 556(7701):339–344.

Port F, Chen H-M, Lee T, Bullock SL. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. Proc Natl Acad Sci U S A. 2014;111(29):E2967–76.

Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet. 2001;69(1):124–137.

Ren X, Sun J, Housden BE, Hu Y, Roesel C, Lin S, Liu L-P, Yang Z, Mao D, Sun L, et al. Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. Proc Natl Acad Sci U S A. 2013;110(47):19012–19017.

Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–427.

Soetaert K, Van den Meersche K, van Oevelen D. Package limSolve, Solving Linear Inverse Models in R; 2009.

Spencer CCA, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 2009;5(5):e1000477.

Svenson KL, Gatti DM, Valdar W, Welsh CE, Cheng R, Chesler EJ, Palmer AA, McMillan L, Churchill GA. High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics. 2012;190(2):437–447.

Terra WR. The origin and functions of the insect peritrophic membrane and peritrophic gel. Arch Insect Biochem Physiol. 2001;47(2):47–61.

Thornton KR, Foran AJ, Long AD. Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PLoS Genet. 2013;9(2):e1003258.

Tilk S, Bergland A, Goodman A, Schmidt P, Petrov D, Greenblum S. Accurate allele frequencies from ultra-low coverage pool-seq samples in evolve-and-resequence experiments. G3 (Bethesda). 2019;9: 4159–4168.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22.

Voutev R, Mann RS. Robust ΦC31-mediated genome engineering in Drosophila melanogaster using minimal attP/attB phage sites. G3 (Bethesda). 2018;8(5):1399–1402.

Wang Y, Norum M, Oehl K, Yang Y, Zuber R, Yang J, Farine J-P, Gehring N, Flötenmeyer M, Ferveur J-F, et al. Dysfunction of Oskyddad causes Harlequin-type ichthyosis-like defects in Drosophila melanogaster. PLoS Genet. 2020;16(1):e1008363.

Wei KH-C, Reddy HM, Rathnam C, Lee J, Lin D, Ji S, Mason JM, Clark AG, Barbash DA. A pooled sequencing approach identifies a candidate meiotic driver in Drosophila. Genetics. 2017;206(1): 451–465.

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447:661–678.

Weller CA, Tilk S, Rajpurohit S, Bergland AO. Accurate, ultra-low coverage genome reconstruction and association studies in Hybrid Swarm mapping populations. G3 (Bethesda). 2021;11:jkab062.

Willoughby L, Chung H, Lumb C, Robin C, Batterham P, Daborn PJ. A comparison of Drosophila melanogaster detoxification gene induction responses for six insecticides, caffeine and phenobarbital. Insect Biochem Mol Biol. 2006;36(12):934–942.

Xu C, Li CY-T, Kong A-NT. Induction of phase I, II and III drug metabolism/transport by xenobiotics. Arch Pharm Res. 2005;28(3): 249–268.

Zhang J, Malo D, Mott R, Panthier J-J, Montagutelli X, Jaubert J. Identification of new loci involved in the host susceptibility to Salmonella Typhimurium in collaborative cross mice. BMC Genomics. 2018;19(1):303.

Zhang X, Emerson JJ. Inferring the genetic architecture of expression variation from replicated high throughput allele-specific expression experiments. bioRxiv 699074, 2019.

Zhou S, Luoma SE, St Armour GE, Thakkar E, Mackay TFC, Anholt RRH. A Drosophila model for toxicogenomics: genetic variation in susceptibility to heavy metal exposure. PLoS Genet. 2017;13(7): e1006907.

Zhou X, St Pierre CL, Gonzales NM, Zou J, Cheng R, Chitre AS, Sokoloff G, Palmer AA. Genome-wide association study in two cohorts from a multi-generational mouse advanced intercross line highlights the difficulty of replication due to study-specific heterogeneity. G3 (Bethesda). 2020;10(3):951–965.

Zou C, Wang P, Xu Y. Bulked sample analysis in genetics, genomics and crop improvement. Plant Biotechnol J. 2016;14(10):1941–1955.

Communicating editor: P. J. Wittkopp