The Illusion of Polygenicity in Poolseq studies: Insufficient Power Can Mask Simple Genetic Architectures

Anthony D. Long[1], Katherine M. Hanson[2], Stuart J Macdonald[2]

[1] Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697, USA

[2] Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045, USA

Corresponding Author:

Anthony D. Long
Ecology and Evolutionary Biology
321 Steinhaus Hall
University of California
Irvine, CA 92697

Email: tdlong@uci.edu

Abstract:

In Drosophila melanogaster, genome-wide association studies (GWAS) and quantitative trait locus (QTL) mapping have often yielded seemingly contradictory results, with GWAS typically suggesting highly polygenic architectures while QTL mapping identifies fewer loci of larger effect. We investigated this discrepancy using a published mapping study that contrasted pools of control flies with pools of flies representing the most zinc chloride-resistant larvae. The dataset provided a unique opportunity to compare three distinct analytical approaches: an X-QTL mapping strategy based on inferring known founder haplotype frequencies, a SNP-by-SNP GWAS using directly ascertained SNP counts, and a SNP-by-SNP GWAS using imputed SNP counts derived from estimated haplotype frequencies. Our analysis revealed that X-QTL and imputed SNP approaches uncovered a relatively simple genetic architecture dominated by several major-effect loci. In contrast, the traditional SNP-by-SNP GWAS suggested a highly polygenic trait with numerous small-effect loci dispersed throughout the genome. We demonstrate that this apparent discrepancy primarily stems from the limited statistical power of the SNP-by-SNP GWAS approach when applied to pooled sequencing data. Despite ~700X sequence coverage and greater than 3000 flies per treatment, the inferred polygenic architecture under the SNP-by-SNP approach appears to simply be an artifact of insufficient power to detect major effect loci. Through simulations, we illustrate that reliably detecting subtle allele frequency differences between pools requires substantially larger sample sizes and higher sequencing coverage than typically employed in current studies. Our findings highlight the importance of considering statistical power and methodological approaches when interpreting genetic architecture studies.

Introduction:

Genome Wide Association Studies (GWAS) have identified reproducible associations between genotype and phenotype for hundreds of human traits (Manolio *et al.* 2009; Buniello *et al.* 2019; Uffelmann *et al.* 2021). The majority of GWAS hits appear to be intermediate frequency variants that typically explain less, and oftentimes much less, than 1% of the variation in a complex trait (Boyle *et al.* 2017; Abdellaoui *et al.* 2023). In contrast, both bi-parental QTL (Quantitative Trait Locus) mapping studies, and more modern studies using multiparental panels (MPPs) have tended to map larger effect QTL, often explaining upwards of 5% of trait variation (Valdar *et al.* 2006; Gatti *et al.* 2014; Long *et al.* 2014; de Koning and McIntyre 2017; Scott *et al.* 2020). Some have argued that the apparent discrepancy between the genetic architectures uncovered by these different experimental approaches is due to mapped QTL representing dozens of smaller effect loci that just happen to cluster together in the genome (Nguyen Ba *et al.* 2022). However, it is challenging to reconcile GWAS and QTL mapping results since the best GWAS datasets exist for disease states in human populations, powerful MPP-based mapping datasets are routinely executed in genetic model systems or plant species important in agronomic contexts (Scott *et al.* 2020), and the highest resolution QTL mapping experiments are dominated by single generation, bi-parental QTL mapping of xenobiotic resistance and growth rate in yeast (Ehrenreich *et al.* 2012; Albert *et al.* 2014; Bloom *et al.* 2015). It remains difficult to compare the results of genetic dissection across studies, traits, systems, and kingdoms of life.

In *Drosophila melanogaster*, a handful of traits have been dissected using both an MPP-panel based mapping approach using the *Drosophila* Synthetic Population Resource (DSPR; (King *et al.* 2012)) and a GWAS-based analysis using collection of inbred lines called the *Drosophila* Genetic Reference panel (DGRP; (Mackay *et al.* 2012; Huang *et al.* 2014)). We are aware of 7 such traits: resistance to caffeine (Najarro *et al.* 2015), copper toxicity (Everman *et al.* 2021, 2023), malathion resistance (Battlay *et al.* 2018; Macdonald and Long 2022), boric acid resistance (Najarro *et al.* 2017), starvation stress (Everman *et al.* 2019), and susceptibility to sigma virus and Drosophila C virus (Magwire *et al.* 2012; Cogni *et al.* 2016). Of these traits, variation in Drosophila C virus resistance appears to be primarily due to a single, large-effect intermediate frequency locus that is detected via both approaches. For sigma virus, a major effect polymorphism identified in a DGRP GWAS was not replicated in the DSPR. However, the DSPR panel is fixed for the resistant allele at the likely causative gene (Cogni *et al.* 2016). For malathion resistance, a DGRP GWAS identified a pair of strong associations, however only one of these was replicated using a DSPR-based QTL approach. Interestingly, the likely causative gene at the un-replicated locus does not appear to segregate for any known causative resistance mutations in the DSPR, as malathion resistance likely evolved subsequent to the DSPR founders being sampled from the wild (Macdonald and Long 2022). For the remaining four traits - resistance to boric acid, caffeine, copper and starvation - DSPR mapping-based approaches have successfully identified loci of large effect, while in contrast DGRP association-based approaches have concluded that trait variation is highly polygenic and due to dozens to hundreds of loci of subtle effect spread throughout the genome. Notably these association hits are subtle enough that they rarely reach genome-wide “Bonferroni” statistical significance, and studies instead rely on a suggestive threshold, often a -log10(p-value) > 5, despite over 2 million tests being carried out. The highly polygenic signal in these four studies is consistent with DGRP-based GWAS outcomes more generally (Mackay and Huang 2018), as well as those in other Drosophila GWAS-type studies employing individual-based genotyping (Pallares *et al.* 2023). Overall, in Drosophila the genetic basis of trait variation dissected using GWAS-type approaches appears highly polygenic, while in contrast traits dissected using QTL-mapping type approaches appear more oligogenic, and this remains true even when holding the trait constant.

Historically, genetic dissection of complex traits in Drosophila has employed a line-by-line approach where a collection of genotyped ILs or RILs (Inbred Lines or Recombinant Inbred Lines) are phenotyped for a trait, and then the trait is mapped by regressing phenotype onto genotypes throughout the genome (Lander and Botstein 1989; Mackay *et al.* 2012; King *et al.* 2012). More recently, researchers are employing an approach more akin to human GWAS case-control studies: DNA is obtained from a pool of several hundred individuals from either a control or selection treatment, and short-read sequencing identifies frequency differences between the two pools. These experiments can take the form of selecting extreme versus control individuals from a large within-generation population (Bastide *et al.* 2013, 2016; Burke *et al.* 2014; Huang *et al.* 2020) or comparing control and experimentally-evolved populations derived from the same base population (Hardy *et al.* 2018; Barghi *et al.* 2019). These experiments can be performed on a population recently derived from several hundred, often unknown founders, contrasting control and selected pools via SNP-by-SNP tests, analogous to

the IL-based DGRP GWAS approach (Huang *et al.* 2014). Alternatively, the population can be derived from a more modest number of founders, founder alleles can be tracked, and differences in those founder haplotype frequencies contrasted between populations, analogous to the RIL-based DSPR QTL-mapping approach (Macdonald *et al.* 2022). Here we show that when the SNP-by-SNP GWAS-type approach is applied to a synthetic population derived from a small number of founders it uncovers a highly polygenic architecture, despite a QTL-mapping approach applied to the same dataset uncovering a much simpler architecture. We further show that this apparent difference in architectures is almost completely associated with the poor power of the GWAS approach given limited sequence coverage in pools of individuals.

Results/Discussion:

In prior work we employed an approach we term X-QTL, or “extreme” QTL mapping (Macdonald *et al.* 2022; Macdonald and Long 2022), to map QTL contributing to zinc resistance in a Drosophila synthetic population (Hanson *et al.* 2025). The X-QTL approach started with a large base population created by mixing several hundred DSPR RILs derived from 8 founder strains (King *et al.* 2012), exposed thousands of embryos from that synthetic, outbred recombinant population to media containing high levels of zinc chloride, selected the ~7% of adults surviving this treatment (or matched no-selection controls), and repeated this selection 12 times. DNA was isolated from pooled samples of individuals from each treatment and replicate, the resulting pools sequenced to modest levels of coverage, known founder haplotype frequencies imputed throughout the genome, and regions showing replicable shifts in haplotype frequency between treatments identified as contributing to trait variation. The idea of imputing haplotype frequencies per treatment/replicate at regular intervals along the genome and carrying out statistical testing on these haplotype frequency estimates (versus directly testing SNP frequencies) is an important aspect of the X-QTL approach. The design of the zinc chloride experiment resulted in the number of individuals sampled being quite large. Figure 1 shows that an average of 285 individuals were sampled per replicate over 12 replicates of the experiment, and each pool was sequenced to an average depth of 59X. That is, a total of 3420 individuals were sampled and sequenced to a total coverage of 700X per treatment. Figure 2A reproduces, using a modified Cochran–Mantel–Haenszel (CMH) test stratified by replicate, the X-QTL scan of Hanson et al (2025) based on imputed haplotype frequencies. It is apparent that resistance is associated with regions of major effect localized at centimorgan scales. The CMH test on pseudo-counts we employ results in inflated -log10(p-values) and as a result we do not define a significance threshold in the panels A&C, although the testing approach employed in Hanson (2025) shows these QTL to be highly significant. The dataset is consistent with other studies in which MPP-based mapping identifies regions of large effect (Macdonald *et al.* 2022; Macdonald and Long 2022), and one may be led to conclude that zinc resistance just happens to be a trait where several genes of large effect explain the variation.

Our haplotype inference X-QTL approach is not the norm in studies of trait variation based on pooled sequencing. A more common approach is to align reads to the genome, identify SNPs, count reference/alternative alleles, and contrast counts between selected/control samples in a “case-control” design. The total allele counts in the implied 2X2 table used for testing is the total

155 sequence coverage over treatments, as opposed to twice the total number of individuals, as
156 would be the case in a human case-control GWAS experiment with individual genotypes. The
157 total sequence coverage over SNPs in the zinc X-QTL dataset is 740X and 984X for the zinc-
158 treated versus control DNA pools respectively, so the described X-QTL experiment is a
159 surprisingly large case/control GWAS by Drosophila standards. At these coverages, a
160 case/control design may have greater power than a ~200 inbred line-based GWAS for a
161 moderate heritability trait, a design typical of historical DGRP-based studies (Mackay and
162 Huang 2018). And the design is perhaps comparable in power to the largest Drosophila
163 individual-based GWAS to date, that compared ~800 long-lived flies to ~1000 controls for two
164 starting populations and dietary treatments (Pallares *et al.* 2023), as both genotyping errors
165 (Gordon *et al.* 2002; Morris and Kaplan 2004; Moskvina *et al.* 2006), and uncorrected population
166 stratification (Lander and Schork 1994; Devlin and Roeder 1999; Clayton *et al.* 2005;
167 Georgiopoulos and Evangelou 2016) in that study would have an effect on power equivalent to
168 decreasing sample size.
169
170 The size of the zinc dataset motivated us to carry out a zinc-treatment versus control CMH test,
171 stratified by replicate, using the directly ascertained SNP counts of the dataset (Figure 2B).
172 Despite 642 SNPs exceeding a Bonferroni threshold of 7.44, we observe little apparent
173 localization of causative gene regions. We suspect many highly significant hits are SNPs that
174 have subtle alignment errors in the bam files. It is apparent from this figure that if we had only
175 carried out a GWAS experiment using directly ascertained SNP counts, we may have concluded
176 that this trait has a highly polygenic genetic basis. But this does not seem to be the case, and
177 even if zinc resistance *is* highly polygenic, our haplotype inference-based analysis (Figure 2A)
178 suggests those polygenes are not widely dispersed throughout the genome. Instead, the
179 simplest explanation for our observation is that the errors on directly ascertained SNP frequency
180 estimates (i.e., $\sqrt{\frac{pq}{Cov}}$) are large relative to the subtle allele frequency differences between pools.
181 Looking only at Chromosome 3R and excluding SNPs with -log10(p-values) greater than 15
182 (Supplementary Figure 1B) there is some evidence for larger p-values corresponding to the
183 locations of peaks resolved via the haplotype-based tests, but the enrichment is quite subtle
184 reflecting a problem with power despite ~700X sequencing coverage per treatment.
185
186 We have shown elsewhere that in modest founder number synthetic populations the error on
187 haplotype frequency estimates is proportional to ~75-100% of the binomial sampling errors on
188 twice the number of individuals per treatment (Macdonald *et al.* 2022; Macdonald and Long
189 2022). This claim about the accuracy of haplotype estimates is consistent with other studies
190 examining the problem of estimating known founder haplotype frequencies from pooled DNA
191 samples (Long *et al.* 2011; Kessner *et al.* 2013; Tilk *et al.* 2019; Linder *et al.* 2020). In the zinc
192 chloride experiment the average number of autosomal alleles per treatment (2*3420=6840) is
193 roughly ten-fold greater than the average sequence coverage per treatment (700X). In this
194 situation the binomial sampling variation associated with short read sequence coverage is much
195 greater than the binomial sampling associated with the 2N alleles sampled per pool. Thus, a
196 major difference between the SNP-based GWAS and MPP-based X-QTL mapping is simply
197 power. A problem with studies that contrast SNP frequencies between two treatments at total

198 sequence coverages less than 1000X is that the power to detect subtle shifts in allele
199 frequencies is quite limited given any sort of correction for multiple testing (see below). The
200 same problem applies if the number of fully ascertained individuals in a case control is less than
201 500.
202
203 To illustrate that a lack of power drives the inability of a SNP-based GWAS to uncover the
204 architecture of zinc-resistance, and not the decision to focus on the SNPs themselves, we
205 imputed SNP frequencies throughout the genome based on the known sequences of the
206 founders of our synthetic population and estimated haplotype frequencies. We then convert
207 imputed SNP frequencies to pseudo-counts based on the number of individuals in each pool
208 and carry out the CMH GWAS using imputed versus directly ascertained SNP counts. (Figure
209 2C). Although counter intuitive, *these imputed SNP frequencies are estimated much more*
210 *accurately than directly ascertained SNP frequencies*. It is apparent that imputed SNPs uncover
211 a similar architecture to the haplotype-based analysis (compare panels A&C of Figure 2). The
212 imputed-SNP-based scans seem to offer the same resolution as X-QTL based mapping, with
213 the resolution in the experiment largely determined by the size of unrecombined founder blocks
214 in the DSPR RILs used to found the synthetic population (King *et al.* 2012) and sample size. It is
215 noteworthy that this resolution is lower than a GWAS of a wild population, or wild-derived inbred
216 lines, where linkage disequilibrium is expected to extend over shorter distances (Macdonald *et*
217 *al.* 2005; Huang *et al.* 2014). Supplementary Figure 1C shows an interesting property of the
218 GWAS carried out using imputed SNPs. In significant regions different SNPs seem to fall into
219 significant "bins", with some SNPs being highly significant whilst others are not significant at all.
220 This is expected given the limited number of haplotypes in DSPR experiments, only the SNPs
221 that "tag" haplotypes showing large frequency shifts will be highly significant.
222
223 The lack of power of directly ascertained SNP case-control studies is consistent with simple
224 simulations. We simulated a pooled sequencing case/control experiment consisting of two pools
225 of N individuals, with each pool sequenced to C coverage. Our simulations assume a single
226 control or experimental pool of size N, although in practice several smaller replicates may be
227 carried out to achieve this total N. We carried out these simulations for a true difference in allele
228 frequency between the two pools of 2.5% or 5% (or a 0% control), which would represent fairly
229 large differences in allele frequency if one assumed variation is due to polygenes of somewhat
230 subtle effect. We further assume a rather liberal, but in Drosophila GWAS a widely employed
231 threshold for significance of -log10(p-value)>5 (Mackay and Huang 2018). We plot percent
232 power as a function of coverage and number of individuals in the pool for the two different effect
233 sizes (Figure 3A). It is apparent that in a SNP-based approach, power is very low if the true
234 difference between allele frequencies is 2.5%. Power only reaches ~20% for a 5% difference in
235 frequency between pools once sequence coverage reaches ~1000X per pool. A curious
236 observation from these simulations is that for a constant coverage and effect, power can be
237 inversely proportional to the number of individuals sampled (*e.g.* for Figure 3A, power is higher
238 for N=500 than for N=5000). This observation almost certainly arises from confounding the
239 problem of detecting differences in frequencies between pools versus differences in allele
240 frequency between treatments. Indeed, as the sequence coverage approaches the number of
241 individuals in the pool the false positive rate can be high (Figure 3B), since the chi-square test

242 becomes very good at detecting differences in frequency between the two pools, as opposed to
243 the populations from which those pools are drawn. It is apparent from Figure 3C that it is
244 generally a good practice for the sequence coverage to be much less than the number of
245 individuals in the pool to avoid this source of false positives.
246
247 A criticism that can be leveled at the illustrative example presented in this work is that the
248 statistical tests for directly ascertained SNPs use total counts derived from sequence coverage
249 (see Methods), whereas the two imputation-based tests use total counts associated with twice
250 the number of individuals in the pool from which DNA is made. Counts associated with the 2N
251 individuals in the DNA pools results in higher power, since 2N is ~9X total sequence coverage.
252 The total sequence coverage is the true count associated with directly ascertained SNPs,
253 whereas the 2N associated with imputation-based tests incorrectly assumes imputation is
254 equivalent to perfectly genotyping the N individuals. However, the decision to use 2N to obtain
255 counts from imputed haplotype frequencies impacts the -log10(p-value) as a multiplicative
256 constant throughout the genome. This "inflation" impacts the top and bottom Manhattan plots of
257 Figure 2 as an incorrect Y-axis scale, but it does not contribute to the noise of the middle plot,
258 so the patterns hold. A second criticism is that an MPP is not appropriate for a GWAS-type
259 analysis; a synthetic population derived from an 8-way cross is not the type of population
260 typically employed in a GWAS study. This is reflected in Figure 2 where for the two imputation-
261 based scans the "hits" extend over centimorgan sized intervals, a much larger region than the
262 localization signal typical of a GWAS. But this difference doesn't detract from the value of our
263 illustration. We observe that directly ascertained SNP frequencies at coverages approaching
264 700X per sample are insufficient to detect subtle frequency shifts, a result independent of the
265 genetic details of the population employed. If anything, the problem is more dire in a population
266 harboring lower levels of linkage disequilibrium as a function of distance, as in that case one is
267 trying to detect the signal associated with only a handful of SNPs in LD with a causative factor
268 (and a smaller p-value is required to control for false positives). When using directly ascertained
269 SNP frequencies, and experiments with fewer than several thousand individuals sequenced to
270 cumulative coverage of less than 1000X per treatment, a simple lack of power to detect
271 causative factors of fairly-large effect can lead to the inference of a highly polygenic
272 architecture.
273
274 Human "GWAS" geneticists have strongly advocated for a stringent significance testing
275 threshold of at least a -log10(p-value)>6 paired with greater than three thousand fully genotyped
276 case and control individuals in order to detect subtle effect risk alleles, with the logic nicely laid
277 out in Wellcome Trust Case Control Consortium (2007). Although human GWAS power
278 calculations are almost always in terms of relative risk, analytical calculations for a chi-square
279 test at a power of 80% and a false positive rate of $p<10^{-6}$ (to control for multiple tests) suggests
280 an experimental design with approximately two thousand fully genotyped individuals per pool
281 are required to detect a true 5% frequency shift, or five thousand individuals per pool to detect a
282 3% shift. In the context of testing for allele frequency differences between two pools of two
283 thousand flies sequenced to less than 4000X per pool (*i.e.*, less than full genotyping) one would
284 expect there to be a reduction in power. At a sequencing coverage much less than twice the
285 number of individuals sampled, the error in allele frequency estimation associated with

286 sequencing coverage will largely determine power. Thus, at sequencing coverages much less
287 than 2N some “trick” is required such that the error on an allele frequency estimate for each pool
288 is the same magnitude as the error associated with complete genotyping. In the special case of
289 synthetic populations derived from known founders, imputing founder haplotype frequencies is
290 one such trick, as SNP frequencies obtained from imputed founder haplotypes can have errors
291 approaching 75%-90% of those obtained from complete genotypes for a pool of size 2N
292 individuals.
293
294 **Methods:**
295
296 The zinc chloride dataset is described in Hanson et al. (Hanson *et al.* 2025). Briefly, the Zinc X-
297 QTL study utilized a cage population derived by intermating 663 Drosophila Synthetic
298 Population Resource (DSPR) "A" founder population recombinant inbred lines. After 31-36
299 generations of synthetic population maintenance, 12 replicate zinc selection assays were
300 performed. To perform the assays, 0–24-hour old embryos were collected from the cage
301 population and transferred to control (water) or treatment (25-mM zinc chloride) bottles.
302 Emerged adult females were collected, and pooled DNA samples were prepared for both control
303 and zinc-selected groups, DNA libraries were prepared using the Illumina DNA Prep protocol
304 and sequenced with PE150 reads on an Illumina NovaSeq 6000. Reads were mapped to the *D.*
305 *melanogaster* Release 6 reference genome, REF/ALT counts were obtained for every SNP, and
306 founder haplotype frequencies estimated in overlapping 1.5-cM windows across the genome for
307 each sample.
308
309 Here, we employ the raw SNP tables and haplotype frequency estimates from Hanson et al.
310 (2025). But unlike the earlier work all the statistical tests in the present study employ CHM tests
311 on counts as opposed to ANOVA on square root transformed frequency estimates. The zinc
312 chloride experiment consisted of twelve replicate control versus zinc-selected comparisons, so
313 CHM is an appropriate test. The use of the CHM test for all three analytical approaches allows
314 for more intuitive comparisons between the study designs. For directly ascertained SNP scans
315 the CHM tests employed raw REF vs ALT counts tables, as a result the total counts in the
316 analysis at any given SNP is the total coverage at the SNP. In contrast the two imputation-
317 based designs use the raw counts and founder genotypes to impute haplotype or SNP
318 frequencies at 10kb steps or at each SNP in the genome respectively. SNP frequencies are
319 imputed by multiplying the closest set of haplotype frequency estimates by the founder states
320 (generally 0 or 1) for each haplotype. To convert per-sample imputed haplotype or SNP
321 frequencies to counts, those frequencies are multiplied by twice the number of individuals
322 contributing to that pooled DNA sample. The number of individuals in a pool is a genome-wide
323 constant per pool, as a result it will not generate variation between sites. Instead it will result in
324 inflated -log10(p-values)’s, since unlike the case of perfect genotyping, the counts are estimated
325 with error.
326
327 The simulations were carried out in R using the rbinom function, with an initial draw of size 2N to
328 obtain frequencies in the sample pool, followed by a second draw of size C to mimic sequence
329 coverage of that pool.

Data and Methods Availability:

The raw data, methods for calling SNPs, and haplotyping calling are described in the Hanson 2025. Briefly, DSPR RILs are available from the BDSC, raw FASTQ files are in the NCBI SRA under accession number PRJNA1127662, and all code to replicate the analyses is available on GitHub (https://github.com/Hanson19/Zinc-X-QTL). The analysis code, and code to reproduce the figures of this work are hosted on Github (https://github.com/tdlong/Compare_3_GWAS.git).

Acknowledgements:

This work was supported by NIH award R01-OD034064 (to SJM and ADL). The authors acknowledge the Bloomington Drosophila Stock Center (P40-OD018537) for supplying DSPR strains.

Figures

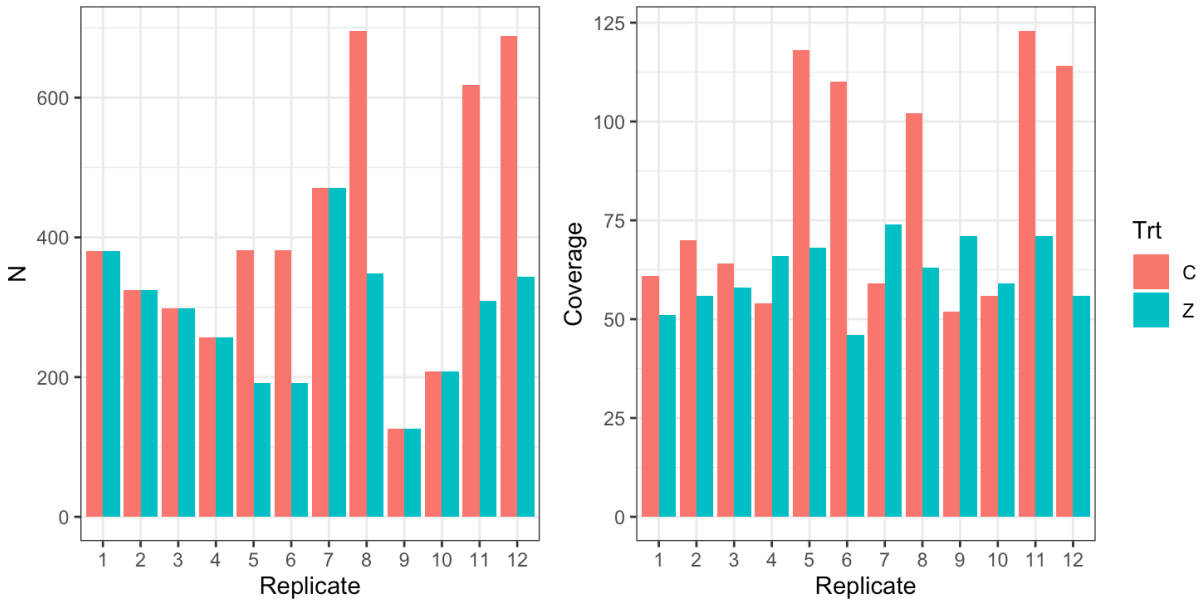


Figure 1: Sampling properties of an X-QTL experiment challenging a *Drosophila melanogaster* synthetic population with zinc chloride (Hanson *et al.* 2025). The experiment consisted of 12 paired replicates of Control versus Zinc-selected treatments (Trt). For each treatment and replicate N individuals were sampled to create DNA pools (left panel), and these were subsequently pool-sequenced to the coverage depicted in the right panel. Replicates 5, 6, 8, 11 and 12 had two separate, equally sized control pools.

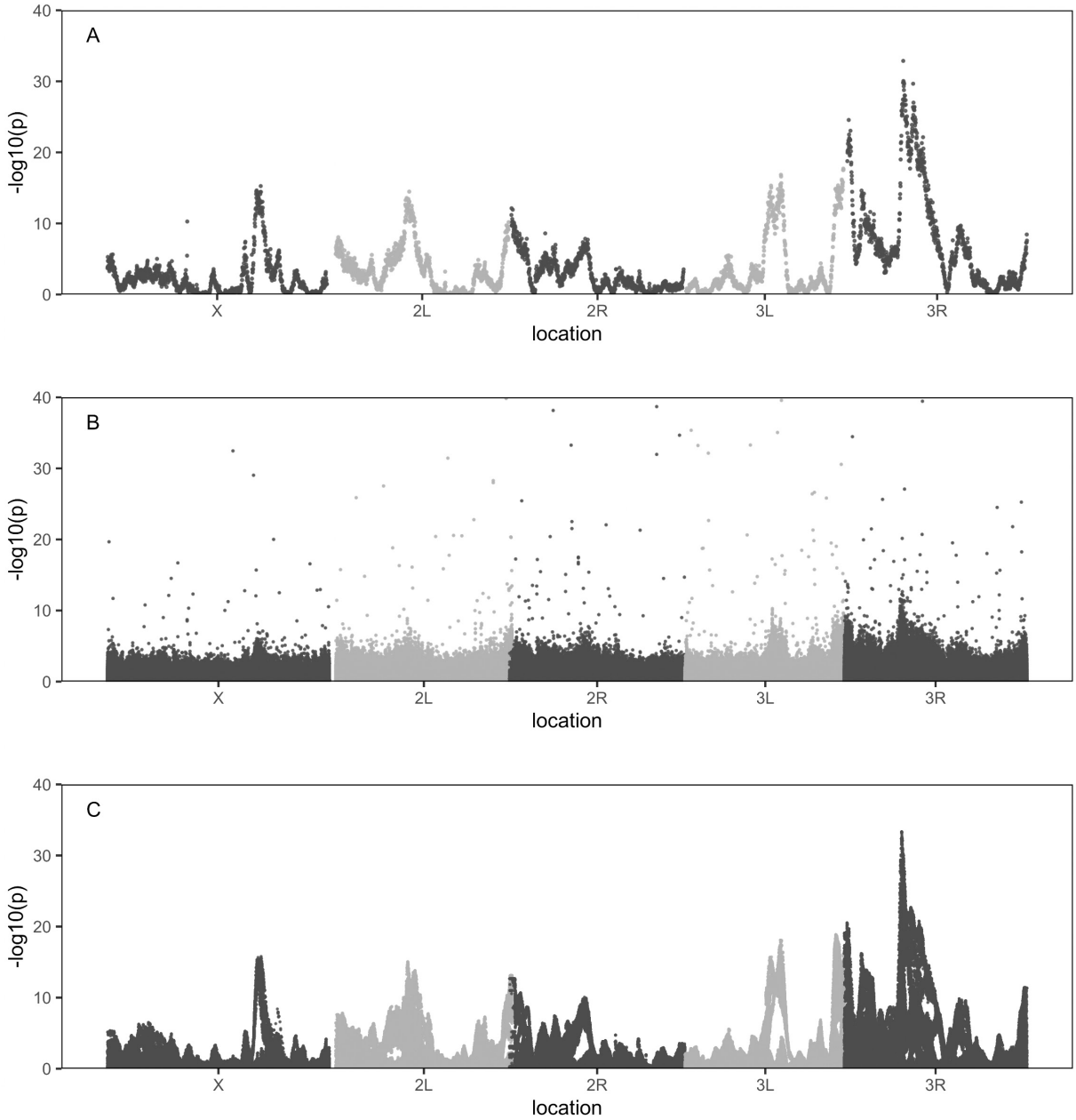


Figure 2: Zinc chloride developmental resistance Manhattan plots associated with three different approaches to statistical testing: (A) Imputed founder haplotype counts, (B) directly ascertained REF and ALT SNP counts, and (C) imputed SNP counts. X-axes are scaled in genetic as opposed to physical distances. *p*-values are obtained from Cochran–Mantel–Haenszel tests on counts over the 12 replicates of the experiment. For directly ascertained SNPs, the counts used are observed REF vs ALT counts. For the imputation-based tests, *pseudo*-counts are the product of haplotype (or SNP) frequency estimates and twice the number of individuals per DNA pool. These pseudo-counts will inflate -log10(p-values) by a constant multiplicative factor.

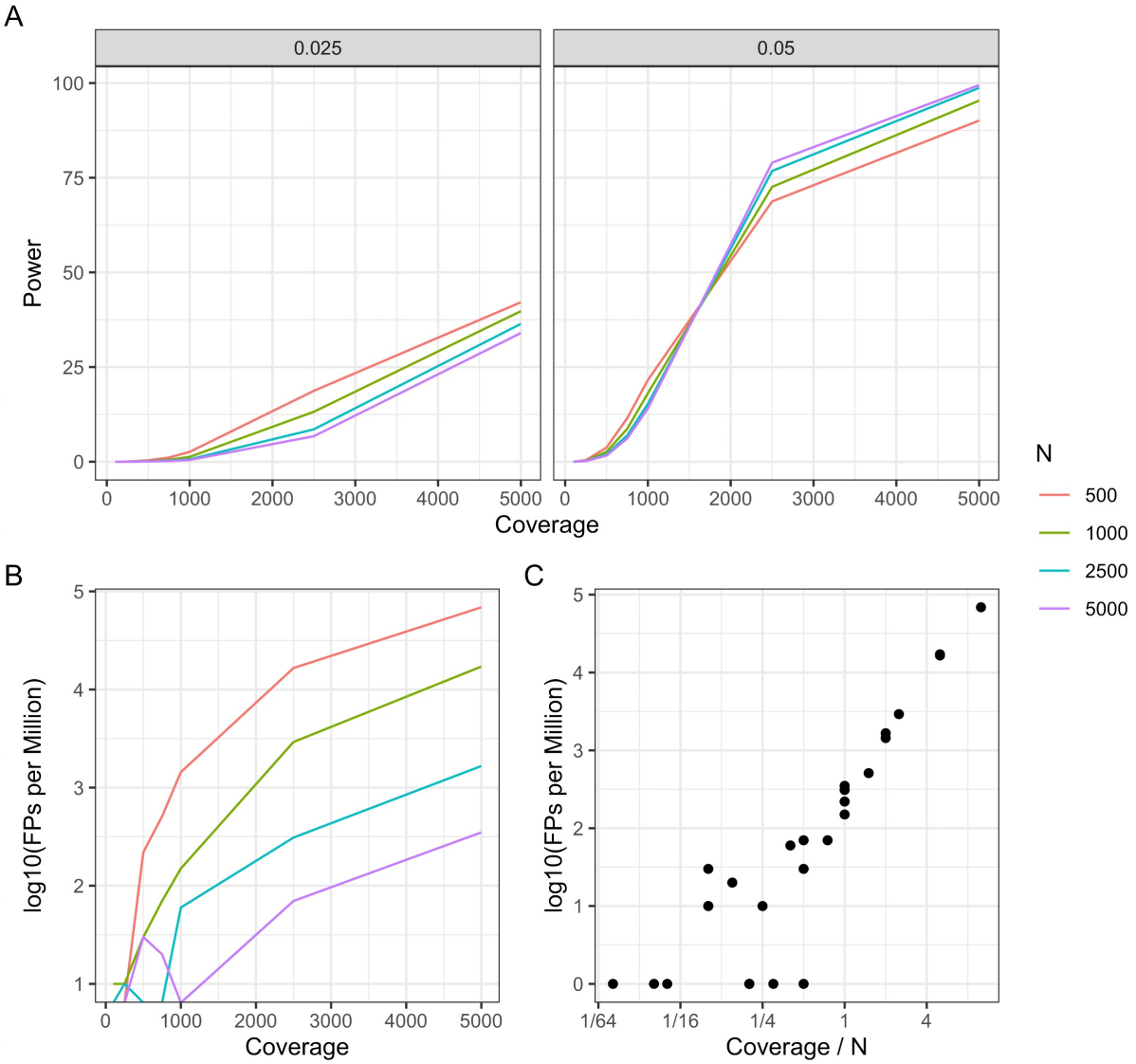
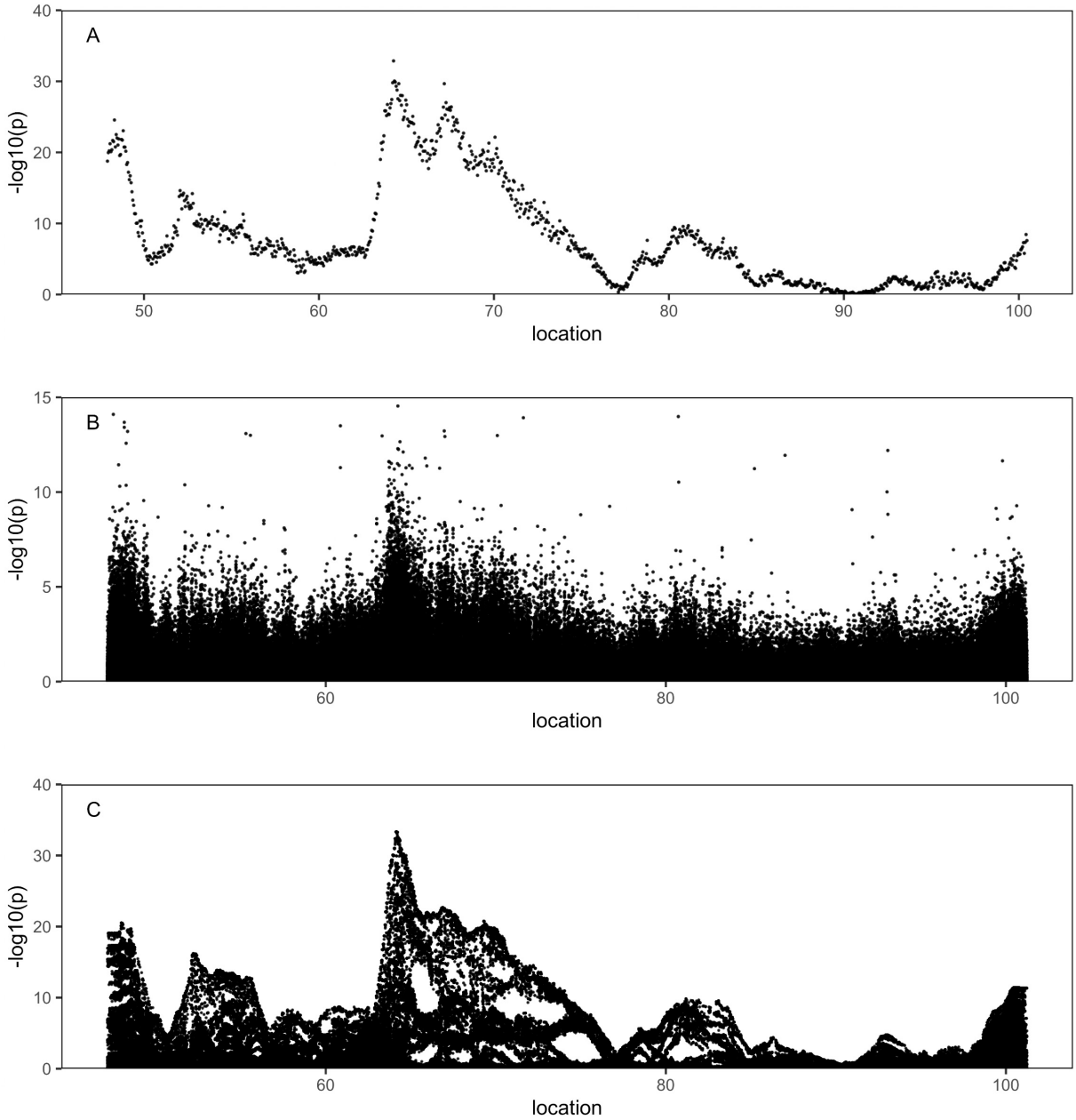


Figure 3: (A) Power of a poolseq case/control design under a directly-ascertained SNP framework as a function of the number of individuals (N) in each pool and sequence coverage for a difference in allele frequency of 2.5% (left) and 5.0% (right). (B) False positive hits per million tests as a function of short read coverage and number of individuals in each pool. (C) False positive hits per million tests as a function of short read coverage divided by the number of individuals in each pool. It is advisable for the sequence coverage of each pool to be much less than the number of individuals contributing to that pool.



Supplementary Figure 1: Zinc chloride developmental resistance Manhattan plots associated with three different approaches to statistical testing for chromosome 3R: (A) Imputed founder haplotype counts, (B) directly ascertained REF and ALT SNP counts, and (C) imputed SNP counts. These panels are otherwise the same as Figure 2, except for panel B where we limit the Y-axis to 15.

References

Abdellaoui A., L. Yengo, K. J. H. Verweij, and P. M. Visscher, 2023 15 years of GWAS discovery: Realizing the promise. Am. J. Hum. Genet. 110: 179–194.

Albert F. W., S. Treusch, A. H. Shockley, J. S. Bloom, and L. Kruglyak, 2014 Genetics of single-cell protein abundance variation in large yeast populations. Nature 506: 494–497.

Barghi N., R. Tobler, V. Nolte, A. M. Jakšić, F. Mallard, *et al.*, 2019 Genetic redundancy fuels polygenic adaptation in Drosophila. PLoS Biol. 17: e3000128.

Bastide H., A. Betancourt, V. Nolte, R. Tobler, P. Stöbe, *et al.*, 2013 A genome-wide, fine-scale map of natural pigmentation variation in Drosophila melanogaster. PLoS Genet. 9: e1003534.

Bastide H., J. D. Lange, J. B. Lack, A. Yassin, and J. E. Pool, 2016 A Variable Genetic Architecture of Melanic Evolution in Drosophila melanogaster. Genetics 204: 1307–1319.

Battlay P., P. B. Leblanc, L. Green, N. R. Garud, J. M. Schmidt, *et al.*, 2018 Structural Variants and Selective Sweep Foci Contribute to Insecticide Resistance in the Drosophila Genetic Reference Panel. G3 8: 3489–3497.

Bloom J. S., I. Kotenko, M. J. Sadhu, S. Treusch, F. W. Albert, *et al.*, 2015 Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nat. Commun. 6: 8712.

Boyle E. A., Y. I. Li, and J. K. Pritchard, 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169: 1177–1186.

Buniello A., J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, *et al.*, 2019 The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47: D1005–D1012.

Burke M. K., E. G. King, P. Shahrestani, M. R. Rose, and A. D. Long, 2014 Genome-wide association study of extreme longevity in Drosophila melanogaster. Genome Biol. Evol. 6: 1–11.

Clayton D. G., N. M. Walker, D. J. Smyth, R. Pask, J. D. Cooper, *et al.*, 2005 Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat. Genet. 37: 1243–1246.

Cogni R., C. Cao, J. P. Day, C. Bridson, and F. M. Jiggins, 2016 The genetic architecture of resistance to virus infection in Drosophila. Mol. Ecol. 25: 5228–5241.

Devlin B., and K. Roeder, 1999 Genomic control for association studies. Biometrics 55: 997–1004.

Ehrenreich I. M., J. Bloom, N. Torabi, X. Wang, Y. Jia, *et al.*, 2012 Genetic architecture of highly complex chemical resistance traits across four yeast strains. PLoS Genet. 8: e1002570.

Everman E. R., C. L. McNeil, J. L. Hackett, C. L. Bain, and S. J. Macdonald, 2019 Dissection of

Complex, Fitness-Related Traits in Multiple Drosophila Mapping Populations Offers Insight into the Genetic Control of Stress Resistance. Genetics 211: 1449–1467.

Everman E. R., K. M. Cloud-Richardson, and S. J. Macdonald, 2021 Characterizing the genetic basis of copper toxicity in Drosophila reveals a complex pattern of allelic, regulatory, and behavioral variation. Genetics 217: 1–20.

Everman E. R., S. J. Macdonald, and J. K. Kelly, 2023 The genetic basis of adaptation to copper pollution in Drosophila melanogaster. Front. Genet. 14: 1144221.

Gatti D. M., K. L. Svenson, A. Shabalin, L.-Y. Wu, W. Valdar, *et al.*, 2014 Quantitative trait locus mapping methods for diversity outbred mice. G3 4: 1623–1633.

Georgiopoulos G., and E. Evangelou, 2016 Power considerations for λ inflation factor in meta-analyses of genome-wide association studies. Genet. Res. 98: e9.

Gordon D., S. J. Finch, M. Nothnagel, and J. Ott, 2002 Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. Hum. Hered. 54: 22–33.

Hanson K. M., A. D. Long, and S. J. Macdonald, 2025 High-throughput genetic mapping discovers novel zinc toxicity response loci in Drosophila melanogaster. bioRxivorg 2025.06.12.659401.

Hardy C. M., M. K. Burke, L. J. Everett, M. V. Han, K. M. Lantz, *et al.*, 2018 Genome-wide analysis of starvation-selected Drosophila melanogaster-A genetic model of obesity. Mol. Biol. Evol. 35: 50–65.

Huang W., A. Massouras, Y. Inoue, J. Peiffer, M. Ràmia, *et al.*, 2014 Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 24: 1193–1208.

Huang W., T. Campbell, M. A. Carbone, W. E. Jones, D. Unselt, *et al.*, 2020 Context-dependent genetic architecture of Drosophila life span. PLoS Biol. 18: e3000645.

Kessner D., T. L. Turner, and J. Novembre, 2013 Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. Mol. Biol. Evol. 30: 1145–1158.

King E. G., C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen, *et al.*, 2012 Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource. Genome Res. 22: 1558–1566.

Koning D.-J. de, and L. M. McIntyre, 2017 Back to the Future: Multiparent Populations Provide the Key to Unlocking the Genetic Basis of Complex Traits. Genetics 206: 527–529.

Lander E. S., and D. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185–199.

Lander E. S., and N. J. Schork, 1994 Genetic dissection of complex traits. Science 265: 2037–2048.

Linder R. A., A. Majumder, M. Chakraborty, and A. Long, 2020 Two Synthetic 18-Way Outcrossed Populations of Diploid Budding Yeast with Utility for Complex Trait Dissection.

452 Genetics 215: 323–342.

453 Long Q., D. C. Jeffares, Q. Zhang, K. Ye, V. Nizhynska, *et al.*, 2011 PoolHap: inferring
454 haplotype frequencies from pooled samples by next generation sequencing. PLoS One 6:
455 e15292.

456 Long A. D., S. J. Macdonald, and E. G. King, 2014 Dissecting complex traits using the
457 Drosophila Synthetic Population Resource. Trends Genet. 30: 488–495.

458 Macdonald S. J., T. Pastinen, and A. D. Long, 2005 The effect of polymorphisms in the
459 enhancer of split gene complex on bristle number variation in a large wild-caught cohort of
460 Drosophila melanogaster. Genetics 171: 1741–1756.

461 Macdonald S. J., K. M. Cloud-Richardson, D. J. Sims-West, and A. D. Long, 2022 Powerful,
462 efficient QTL mapping in Drosophila melanogaster using bulked phenotyping and pooled
463 sequencing. Genetics 220. https://doi.org/10.1093/genetics/iyab238

464 Macdonald S. J., and A. D. Long, 2022 Discovery of malathion resistance QTL in Drosophila
465 melanogaster using a bulked phenotyping approach. G3 12.
466 https://doi.org/10.1093/g3journal/jkac279

467 Mackay T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, *et al.*, 2012 The
468 Drosophila melanogaster Genetic Reference Panel. Nature 482: 173–178.

469 Mackay T. F. C., and W. Huang, 2018 Charting the genotype-phenotype map: lessons from the
470 Drosophila melanogaster Genetic Reference Panel. Wiley Interdiscip. Rev. Dev. Biol. 7.
471 https://doi.org/10.1002/wdev.289

472 Magwire M. M., D. K. Fabian, H. Schweyen, C. Cao, B. Longdon, *et al.*, 2012 Genome-wide
473 association studies reveal a simple genetic basis of resistance to naturally coevolving
474 viruses in Drosophila melanogaster. PLoS Genet. 8: e1003057.

475 Manolio T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, *et al.*, 2009 Finding the
476 missing heritability of complex diseases. Nature 461: 747–753.

477 Morris R. W., and N. L. Kaplan, 2004 Testing for association with a case-parents design in the
478 presence of genotyping errors. Genet. Epidemiol. 26: 142–154.

479 Moskvina V., N. Craddock, P. Holmans, M. J. Owen, and M. C. O'Donovan, 2006 Effects of
480 differential genotyping error rate on the type I error probability of case-control studies. Hum.
481 Hered. 61: 55–64.

482 Najarro M. A., J. L. Hackett, B. R. Smith, C. A. Highfill, E. G. King, *et al.*, 2015 Identifying Loci
483 Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. PLoS Genet. 11:
484 e1005663.

485 Najarro M. A., J. L. Hackett, and S. J. Macdonald, 2017 Loci Contributing to Boric Acid Toxicity
486 in Two Reference Populations of Drosophila melanogaster. G3 7: 1631–1641.

487 Nguyen Ba A. N., K. R. Lawrence, A. Rego-Costa, S. Gopalakrishnan, D. Temko, *et al.*, 2022
488 Barcoded bulk QTL mapping reveals highly polygenic and epistatic architecture of complex
489 traits in yeast. Elife 11. https://doi.org/10.7554/eLife.73983

490 Pallares L. F., A. J. Lea, C. Han, E. V. Filippova, P. Andolfatto, *et al.*, 2023 Dietary stress
491 remodels the genetic architecture of lifespan variation in outbred Drosophila. Nat. Genet.
492 55: 123–129.

493 Scott M. F., O. Ladejobi, S. Amer, A. R. Bentley, J. Biernaskie, *et al.*, 2020 Multi-parent
494 populations in crops: a toolbox integrating genomics and genetic mapping with breeding.
495 Heredity 125: 396–416.

496 Tilk S., A. Bergland, A. Goodman, P. Schmidt, D. Petrov, *et al.*, 2019 Accurate Allele
497 Frequencies from Ultra-low Coverage Pool-Seq Samples in Evolve-and-Resequence
498 Experiments. G3 9: 4159–4168.

499 Uffelmann E., Q. Q. Huang, N. S. Munung, J. de Vries, Y. Okada, *et al.*, 2021 Genome-wide
500 association studies. Nature Reviews Methods Primers 1: 1–21.

501 Valdar W., L. C. Solberg, D. Gauguier, S. Burnett, P. Klenerman, *et al.*, 2006 Genome-wide
502 genetic association of complex traits in heterogeneous stock mice. Nat. Genet. 38: 879–
503 887.

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

506