Evolution of phenotypic variance provides insights into the genetic basis of adaptation

Author list:

Wei-Yun Lai[1,2], Viola Nolte[1], Ana Marija Jakšić[1,2,3] and Christian Schlötterer[1*]

[1]Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria.

[2]Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria.

[3]Current affiliation: École polytechnique fédérale de Lausanne, Lausanne, Switzerland

[*]Correspondence: christian.schloetterer@vetmeduni.ac.at; Tel.: +43-1-25077-4300.

Keywords:

Phenotypic variance, temperature adaptation, *Drosophila simulans*, experimental evolution

Abstract

Most traits are polygenic and the contributing loci can be identified by GWAS. Their adaptive architecture is, however, difficult to characterize. Here, we propose to study the adaptive architecture of traits by monitoring the evolution of their phenotypic variance during adaptation to a new environment in well-defined laboratory conditions. Extensive computer simulations show that the evolution of phenotypic variance in a replicated experimental evolution setting provides power to distinguish between oligogenic and polygenic adaptive architectures. We compared gene expression variance in male *Drosophila simulans* before and after 100 generations of adaptation to a novel hot environment. The variance change in gene expression was indistinguishable for genes with and without a significant change in mean expression after 100 generations of evolution. We suggest that a majority of adaptive gene expression evolution can be explained by a polygenic architecture. We propose that tracking the evolution of phenotypic variance across generations can provide an approach to characterize the adaptive architecture.

Introduction

It is widely accepted that most complex traits have a polygenic basis (Ayroles *et al.*, 2009; Boyle *et al.*, 2017; Liu *et al.*, 2019). Nevertheless, it is difficult to predict which of these loci are responding to selection when a population is exposed to a new selection regime. If pleiotropic constraints are strong, only a small subset of the genes constituting are free to respond to selection. Hence, the genetic basis of the adaptive response of a complex trait (i.e. adaptive architecture (Barghi *et al.*, 2020)) may differ substantially from the genetic architecture. Characterizing the adaptive architecture by mapping selected loci is not easy, in particular when more than a handful of genes are involved. To circumvent this problem, we introduce a complementary approach, analogous to the Castle-Wright estimator (Castle, 1921), to infer the complexity of the adaptive architecture (i.e. simple with few contributing loci or with a polygenic basis). We propose to study the evolution of phenotypic variance, which may provide some insights into the key parameters of the adaptive architecture.

The phenotypic variance of a quantitative trait is a key determinant for its response to selection. It can be decomposed into genetic and environmental components (Falconer & Mackay, 1963). Over the past years, mathematical models have been developed to describe the expected genetic variance of a quantitative trait under selection and its maintenance in evolving populations (Bulmer, 1972; Chevalet, 1994; Kimura and Crow, 1964; Turelli, 1984). For infinitely large populations and traits controlled by many independent loci with infinitesimal effect, changes in trait optimum are not expected to affect the phenotypic variance (Lande, 1976). A much more complex picture is expected when the effect sizes are not equal, the population size is finite, or the traits have a simpler genetic basis (Barton & Turelli, 1987; Keightley & Hill, 1989; Barton & Keightley, 2002; Jain & Stephan, 2015; Franssen *et al.*, 2017; Hayward & Sella, 2019). For instance, for traits with oligogenic architecture, the genetic variance could drop dramatically during adaptation, while with polygenic architecture, only minor effects on the variance are expected (Barton et al., 2017; Franssen et al., 2017; Jain and Stephan , 2015). These studies suggest that a time-resolved analysis of phenotypic variance has the potential to shed light onto the complexity of the underlying adaptive architecture.

Despite its potential importance for the understanding of adaptation, we are faced with the situation that few empirical data are available for the evolution of phenotypic variance. The use of natural populations to study changes in phenotypes, and even more so phenotypic

variances, is limited as the environmental heterogeneity cannot be controlled. A complementary approach to study the evolution of phenotypic variance is experimental evolution (Kawecki *et al.*, 2012). With replicated populations starting from the same founders and evolving under tightly controlled environmental conditions, experimental evolution provides the opportunity to study the evolution of phenotypic variance.

Most experimental evolution studies in sexual populations focused on the evolution of phenotypic means, rather than variance (Chippindale *et al.*, 1996; Burke *et al.*, 2010; Mallard *et al.*, 2018; Jakšić *et al.*, 2020). Notable exceptions include a study which applied fluctuating, stabilizing and disruptive selection to a small number of wing shape-related traits (Pélabon *et al.*, 2010) and other studies tracking the variance evolution of behavioral and cranial traits under directional selection in mice (Careau *et al.*, 2015; Penna *et al.*, 2017). Instead of looking at a preselected subset of phenotypes which limits the generality, we will focus on gene expression, a set of molecular phenotypes, which can be easily quantified as microarrays and, more recently, RNA-Seq have become available. Importantly, the expression levels of genes exhibit the same properties (e.g.: continuality and normality) as other complex quantitative traits (Mackay *et al.*, 2009). Thus, gene expression has also been widely employed to study the adaptation of locally adapted populations (e.g.: Romero *et al.* 2012; Sork 2017; Signor and Nuzhdin 2018) or ancestral and evolved populations in the context of experimental evolution (e.g.: Lenski *et al.* 1994; Ferea *et al.* 1999; Huang and Agrawal 2016; Mallard *et al.* 2018).

In this study, we performed forward simulations that match not only essential design features of typical experimental evolution studies, but also incorporate realistic parameters of the genetic architecture. We recapitulate the classic expectations that even a moderately polygenic architecture is associated with a high stability of the phenotypic variance of selected traits across different phases of adaptation and evaluate the statistical power. Applying this insight to a recently published dataset (Lai & Schlötterer, 2022), we show that for putatively selected genes (DE genes), their average changes in expression variance were indistinguishable from genes without changes in mean expression. We propose that this pattern reflects that adaptive gene expression evolution generally has a polygenic basis.

Materials and Methods

Computer simulations

We performed forward simulations with MimicrEE2 (Vlachos & Kofler, 2018) using the qff mode to illustrate the influence of the genetic architecture on the evolution of phenotypic variance during the adaptation to a new trait optimum (Figure 1a). With 189 founder haplotypes (Barghi *et al.*, 2019), we simulated quantitative gene expression traits under the control of different numbers of loci (M = 5, 25, 50, 100, 200 and 1000) in a population with an effective population size of 300 (reflecting the estimated effective population size in an experimentally evolving population (Barghi *et al.*, 2019)). We used the empirical linkage map from *Drosophila simulans* (Howie *et al.*, 2019) to account for linkage. For each trait, we assume an additive model and a negative correlation (r = -0.7, reflecting the observed relationship from Otte *et al.* (2020)) between the effect size and starting frequency (Figure 1b). We used the *correlate()* function implemented in "fabricatr" R package (Blair *et al.*, 2019) to generate the effect sizes. The distribution of the effect sizes of loci controlling a trait can vary depending on the shape parameter of gamma sampling process (shape = 0.5, 2.5 and 100, Figure 1c). The mean of the effect sizes from 3 different shape parameters was standardized to 1/M. As the focus of our study is on expression traits, we used the distribution of heritabilities obtained from empirical gene expression data (Ayroles et al. 2009 and this study, see supplementary information) to assign the environmental/technical variance to a simulated trait (see supplementary information; Supplementary figure S1) (i.e. For each simulated trait, its heritability and associated environmental/technical variance is sampled from the empirical distribution). We do not simulate de novo mutations as they are not contributing to adaptation in this short time scale (Burke *et al.*, 2010). To simulate stabilizing selection with trait optimum shift, we provided the Gaussian fitness functions with mean of $\overline{X_{anc.}} + a\sqrt{V_{anc.}}$ and standard deviation of $b\sqrt{V_{anc.}}$, where $\overline{X_{anc.}}$ is the ancestral phenotypic mean and $V_{anc.}$ is the ancestral genetic variance (Figure 1a). Parameter "a" determines the distance of optimum shift, which is set to one (similar to the empirical case, Figure 4b) or three (Adopted from (Hayward & Sella, 2019)). Parameter "b" indicates how strong the phenotypic constraint would be when the trait optimum is reached. The value 3.6 for parameter "b" is adopted from Hayward and Sella (2019). In this study, we increase and decrease it by 50% to explore its impact (1.8 or 5.4). For the neutrality case, we assumed no variation in fitness among all individuals. For each scenario, 1000 traits were simulated, and each trait was affected by a different set of loci. We compared the phenotypic variance of selected and neutral traits for each genetic architecture. For each trait

124 under each scenario, the phenotypic variance was estimated at different generations and
125 compared to the ancestral phenotypic variance at generation 1 to illustrate the dynamic of
126 phenotypic variance during the evolution ($F = \sigma_x^2/\sigma_1^2$), where x stands for the number of
127 generations. We note that we do not assume that the ancestral population has reached an
128 equilibrium because the ancestral population in a typical experimental evolution study is often
129 phenotyped in the new environment.
130
131 Power analysis
132 We propose that the temporal dynamics of the phenotypic variance of selected traits can be
133 exploited to characterize the genetic basis of adaptation. This concept could be applied either
134 to a single trait under selection or to a group of selected traits. Depending on whether a single
135 trait or a group of traits is studied, different research objectives can be achieved. Hence, we
136 carefully examined the power and the required sample size for both types of inference based
137 on computer simulations.
138
139 *Power estimates for the analysis of a single trait:* For each simulated trait under different
140 genetic control, we tested whether the variance of this trait changed more than expected under
141 neutrality (F of 0.9 according to the neutral simulations) after 100 generations of selection.
142 First, we assumed that all simulated 300 individuals were phenotyped at both time points. For
143 the 1000 traits with each genetic architecture under each selection scenario, we calculated how
144 often a significant difference is detected. Since it is not possible to phenotype all individuals,
145 we investigated whether and how reduced sample sizes could compromise the power to infer
146 the adaptive architecture of a single trait. From the simulated population, we randomly selected
147 20 individuals for the variance estimation and tested whether the magnitude of variance change
148 for a given trait under selection is significantly different from neutral expectation. We
149 evaluated the power with the reduced sample sizes by calculating how often a significant
150 difference is detected among the 1000 simulated traits.
151
152 *Power estimates for the analysis on a group of traits:* We assessed the power to infer the
153 adaptive architecture for a group of selected traits under the assumption that most traits of
154 interest have similar level of complexity. In this case, a test on the average variance changes
155 (F) between a group of selected traits and another group of neutral traits is required. We
156 simulated 1,000 independent selected and 1,000 neutral traits for each genetic architecture and
157 selection regime, and estimated the variance before and after 100 generations of evolution for

158 the selected and neutral traits. For each trait, we measured the phenotypic variance for a sample
159 of 20 individuals. Next, we performed a t-test to compare the distribution of variance change
160 (F) between selected and neutral traits. The power was determined by 100 iterations of
161 sampling 20 individuals from a simulated population of 300 individuals.
162
163 Experimental evolution
164 The setup of populations and evolution experiment have been described by Barghi *et al.* (2019).
165 Briefly, ten outbred *Drosophila simulans* populations seeded from 202 isofemale lines were
166 exposed to a laboratory experiment at 28/18 °C with 12hr light/12hr dark photoperiod for more
167 than 100 generations. Each replicate consisted of 1000 to 1250 adults for each generation.
168
169 Common garden experiment
170 The collection of samples from the evolution experiment for RNA-Seq was preceded by two
171 generations of common garden (CGE). The common garden experiment was performed at
172 generation 103 of the evolution in the hot environment and this CGE has been described in
173 (Barghi *et al.*, 2019; Hsu *et al.*, 2019, 2020; Jakšić *et al.*, 2020; Lai & Schlötterer, 2022). In
174 brief, an ancestral population was reconstituted by pooling five mated females from 184
175 founder isofemale lines (Nouhaud *et al.*, 2016). No significant allele frequency differences are
176 expected between the reconstituted ancestral populations and the original ancestral populations
177 initiating the experiment (Nouhaud *et al.*, 2016). Because we evaluated phenotypes on the
178 population level, deleterious mutations will have a very limited impact. The reason is that
179 deleterious mutations are expected to occur at low frequencies in natural populations and novel
180 deleterious mutations occurring during the maintenance of the isofemale lines present in a
181 single line. In both cases, deleterious alleles are expected to represent a small fraction of the
182 total population. As described previously (Lai & Schlötterer, 2022), two replicates of the
183 reconstituted ancestral population and two independently evolved populations at generation
184 103 were reared for two generations with egg-density control (400 eggs/bottle) at the same
185 temperature regime as in the evolution experiment. Freshly eclosed flies were transferred onto
186 new food for mating. Sexes were separated under CO_2 anesthesia at day 3 after eclosure, left
187 to recover from CO_2 for two days, and at the age of five days whole-body mated flies of each
188 sex were snap-frozen at 2pm in liquid nitrogen and stored at -80°C until RNA extraction. More
189 than 30 individual male flies from two reconstituted ancestral populations (replicate no. 27 and
190 no. 28) and two evolved populations (replicate no. 4 and no. 9) were subjected to RNA-Seq.

191 The protocols of RNA extraction and library preparation are described in (Lai and Schlötterer
192 2022).
193
194 RNA-Seq data analysis for mean and variance evolution
195 The processed RNA-Seq data were obtained from Lai and Schlötterer (2022). To characterize
196 the adaptive architecture of gene expression evolution, we compared the evolutionary
197 dynamics of the variance between the genes with significant mean change and those without.
198 The underlying assumption is that genes with significant mean expression changes are under
199 selection and the rest of the transcriptome is not responding to the new environment (neutral).
200 The genes with/without significant mean evolution for two evolved populations were taken
201 from (Lai & Schlötterer, 2022).
202 For the quantification the variance changes during adaptation for each gene, we used the
203 variance estimates of the expression (lnCPM) of each gene in each population from (Lai &
204 Schlötterer, 2022). Due to the moderate sample size, we performed additional checks for the
205 uncertainty of variance estimates. Jackknifing was applied to measure the uncertainty of
206 estimator (Fukunaga & Hummels, 1989). The procedure was conducted independently on four
207 populations and we calculated the 95% confidence interval of the estimated variance
208 (Supplementary Figure S2). Additionally, the robustness of variance estimation was also
209 supported by the high correlation (rho = 0.79, Spearman's rank correlation) in variance
210 estimates between the two independently reconstituted ancestral populations. The change of
211 gene expression variance was determined by the F statistics calculated as the ratio between the
212 variance within the ancestral population and the variance within the evolved population of each
213 gene. To test whether selection on mean expression generally alters the expression variance,
214 we compared the F statistics of genes with significant changes in mean expression to the genes
215 without.
216
217 We note that the natural log transformation (as in Lai and Schlötterer 2022) does not attempt
218 to quantify and partition the noise due to the read sampling process at the gene level from the
219 across-individual variance. Since we were primarily interested in the latter, we evaluated the
220 effect of read sampling noise. We did an additional check by distinguishing the true biological
221 variation and the measurement error ([eq1], from User's Guide of edgeR) of each gene using
222 the statistical method implemented in edgeR, where the true variance across individuals
223 (biological coefficient of variation (BCV)) of each gene can be estimated (tag-wised
224 dispersion).

225 $$\text{T}otal\ CV^2 = Technical\ CV^2 + Biological\ CV^2\ [eq1]$$

226 Proportion of technical CV^2 in total CV^2 was calculated. We showed that the average
227 proportion of technical CV^2 gradually decreased with increasing sample size (Supplementary
228 Figure S3), and for our sample size (n = 22) it is 17%. The correlation between BCV^2 and the
229 variance estimates after log transformation is 0.98 (Spearman's rho), suggesting that the two
230 estimates are similar. As a sanity check, we repeated the comparisons of the variance changes
231 in selected (DE) and neutral (non-DE) genes using BCV^2. Similarly, the changes in variance
232 of putative adaptive genes are indistinguishable from the genes that do not change their mean
233 expression (t-test, p-value > 0.05; Supplementary Figure S3).

Results

The central idea of this study is that the complexity of adaptive traits can be inferred from the trajectory of the phenotypic variance during adaptation: the phenotypic variance remains relative stable for a trait with polygenic (infinitesimal) architecture while it changes across generations for a trait with oligogenic architecture. Although this prediction has been illustrated in multiple theoretical and simulation studies (e.g.: (Jain & Stephan, 2015; Barton *et al.*, 2017; Franssen *et al.*, 2017)), its application to empirical data has been limited. A recent experimental evolution study reporting the phenotypic variance of more than 10,000 gene expression traits before and after 100 generations of adaptation to a novel environment (Lai & Schlötterer, 2022) provides an opportunity to fill this gap. Since gene expression can be viewed as quantitative traits (Ayroles *et al.*, 2009; Hill *et al.*, 2021), we apply the theoretical predictions to the empirical data to infer the genetic basis of adaptive gene expression evolution. Hence, as the first step of this study, we explored to what extent theoretical predictions can be generalized to expression traits obtained from a typical experimental evolution setting by considering a broad parameter space and accounting for linkage. Using population genetic parameters (e.g. number of starting haplotypes, effective population size) estimated from the above-mentioned evolution experiment, we simulated 1000 neutral and 1000 selected expression traits. To account for experimental and environmental noise, we used an empirical distribution of heritabilities (estimated from gene expression measurements across F1 families; see supplementary information). The simulated selection regime consists of a mild/distant shift in trait optimum with weak/intermediate/strong stabilizing selection (Figure 1a and Supplementary Figure S4) (Hayward & Sella, 2019). We assume additivity and a negative correlation between the ancestral allele frequency and the effect size of contributing loci (Otte et al. 2020) (Figure 1b). With three different distributions of effect size (Figure 1c), we investigate how the number of contributing loci affects the evolution of gene expression variance with and without selection.

We monitored the change in phenotypic variance over 100 generations, which was sufficient to reach the trait optimum for most parameter combinations (Supplementary Figure S5). We compared the change in variance relative to the start of the experiment in populations with and without selection. First, we studied a mild (one standard deviation of the ancestral phenotypic distribution) shift in trait optimum. As expected for a founder population derived from a substantially larger natural population, we find that even under neutrality, the phenotypic variance does not remain constant, but gradually decreases during 100 generations of

experimental evolution (Figure 2). This pattern is unaffected by the genetic architecture of the neutral traits. We explain this loss of variance by the fixation of variants segregating in the founder population and the fact that we did not simulate new mutations, as they do not contribute to adaptation in such short time scales (Burke *et al.*, 2010). Although our simulations used moderate population sizes, they nicely recapitulate the patterns described for populations without drift (Jain & Stephan, 2015; Barton *et al.*, 2017; Franssen *et al.*, 2017). A pronounced drop in phenotypic variance is observed when a trait is approaching a new optimum with few contributing loci (Figure 2). When more loci (with smaller effects) are contributing to the selected phenotype, the difference to neutrality becomes very small (Figure 2). In addition to the number of contributing loci, the heterogeneity in effect size among loci and the shape of the fitness function have a major impact. The larger the difference in effect size is, the more pronounced was the influence of the number of contributing loci (Figure 2). The opposite effect was seen for the width of the fitness function – a wider fitness function decreased the influence of the number of contributing loci (Figure 2). Importantly, these patterns were not affected by the duration of the experiment-qualitatively identical patterns were seen at different time points until generation 200.

For a more distant trait optimum (three standard deviations of the ancestral phenotypic distribution), we noticed some interesting dynamics that were not apparent for a closer trait optimum (Supplementary Figure S6). The most striking one was the temporal heterogeneity of the phenotypic variance when few loci of unequal effects are contributing. During the early stage of adaptation, the variance increased and dropped later below the variance in the founder population. With an increasing number of contributing loci, this pattern disappeared and closely matched the neutral case (Supplementary Figure S6).

Overall, our simulations indicate that with a larger number of contributing loci the variance fitted the neutral pattern better. Modifying dominance did not change the overall patterns (Supplementary Figure S7). The large influence of key parameters of the adaptive architecture, in particular the number of contributing loci and their effect sizes, on the temporal phenotypic variance dynamics suggests that it should be possible to exploit this relationship to characterize the genetic basis of phenotypic adaptation.

A potentially interesting application of the relationship between number of contributing loci and evolution of variance is the inference of the adaptive architecture of a given trait. For a

single selected trait, a F test contrasting the reduction of variance after adaptation against neutral loss of variance (~0.9 after 100 generations in our simulations) can be used to infer the number of contributing loci behind its adaptive evolution. The lower bound for the number of contributing loci can be determined by the largest number of loci that results in a significant difference between a selected trait and the neutral expectation in all parameter combinations. Nevertheless, the power of this approach depends strongly on the sample size. When the entire population (N = 300) is phenotyped for the focal trait and a neutral trait, the power to detect a significant decrease in variance of a selected trait controlled by five contributing loci is only 44% (i.e. a lack of significant difference in variance after the adaptation cannot inform us whether the selected traits are controlled by five or more loci). With a more realistic sample size of 20 individuals, it is not possible to infer the adaptive architecture for a single trait when only two time points are available (Supplementary figure S8).

Alternatively, it is possible to study multiple, presumably independent, selected phenotypes (e.g.: transcriptomic data) together. Assuming a similar genetic architecture of the selected traits, we showed that the distribution of variance changes of 1000 selected traits with a simple genetic architecture (five loci) can be distinguished from neutral changes. Even with a sample size of 20 individuals, as in our experimental data (see Materials and Methods), significant differences from neutral expectations can be detected for some parameters with a power close to 100% (Figure 3 and Supplementary figure S9). This suggests that oligogenic and polygenic architectures of a group of selected traits can be distinguished experimentally even with moderate sample sizes.

As a case study, we investigated the evolution of gene expression variance in replicated populations evolving in a new hot temperature regime and inferred the adaptive architecture of gene expression traits. The evolved populations were derived from the same ancestral population but evolved independently for more than 100 generations in a novel temperature regime with daily temperature fluctuations between 18 and 28°C (Figure 4a). Rather than relying on pooled samples which only allow mean estimates, we quantified gene expression of 19-22 individuals from reconstituted ancestral populations (F0) and two evolved populations (F103) in a common garden setup (Lai & Schlötterer, 2021). We applied natural log transformation to the expression of each gene (counts per million (CPM)) to fit normal assumption for all subsequent analyses and make mean and variance independent from each other (see Material and Methods). Principal Component Analysis (PCA) indicated that 11.9%

336 of the variation in gene expression can be explained by the first PC which separates evolved
337 and ancestral populations, reflecting clear adaptive gene expression changes in response to the
338 novel hot temperature regime (Figure 4b). The means and variances of the expression of each
339 gene were estimated and compared between the two reconstituted ancestral populations and
340 the two evolved populations separately (see Materials and methods). Due to the usage of
341 different lot numbers for the RNA-Seq library preparation (Supplementary file 1), we only
342 contrasted ancestral and evolved samples generated with the same lot number (see Materials
343 and methods) to avoid any unnecessary confounding effects.
344
345 The comparison of ancestral and evolved populations identified 2,775 genes in the first
346 population and 2,677 genes in the second population which significantly changed mean
347 expression in the evolved flies (FDR<0.05, (Lai & Schlötterer, 2022)) (Figure 5a). The
348 concordance of both populations and the similarity to previous studies of the same evolution
349 experiment (Hsu *et al.*, 2020; Jakšić *et al.*, 2020; Lai & Schlötterer, 2022) suggests that most
350 of the altered expression means are mainly driven by selection, rather than by drift. We scaled
351 the gene expression change with the standard deviation in the ancestral population to
352 approximate the selection strength on each gene. The differentially expressed genes in both
353 populations showed a broad distribution of expression change, but the averaged mean
354 expression changed by one standard deviation (Figure 5b), which is unlikely to be reached by
355 drift based on our simulations (less than 1% of the neutral traits drifted by one standard
356 deviation of the ancestral phenotypic variance). Assuming that these putative adaptive gene
357 expression phenotypes reached the new trait optimum, this corresponds, on average, to a mild
358 shift in trait optimum in our computer simulations. Reasoning that genes with significant mean
359 expression changes are under selection and the rest of the transcriptome is not responding to
360 the new environment (neutral), we compared the evolutionary changes in expression variance
361 of genes with and without significant changes in expression mean (DE and non-DE genes), to
362 characterize the adaptive architecture of gene expression. The underlying hypothesis is that a
363 simple adaptive architecture will be associated with a significant drop in the expression
364 variance of the selected genes, while for polygenic gene expression evolution, the variance
365 remains similar to neutral genes. Our computer simulations showed that this expectation from
366 theoretical results (Jain & Stephan, 2015; Barton *et al.*, 2017; Franssen *et al.*, 2017) also holds
367 for moderately sized populations and linked loci (Figure 2).
368

369 The reduction in population size from a natural population to laboratory cultures would result
370 in a continuous reduction in phenotypic variance over time even under neutrality, as shown in
371 the simulations (Figure 2). Hence, a robust inference of the adaptive architecture of gene
372 expression relies on the comparison of putative selected (DE) and neutral (non-DE) genes. In
373 both replicates the expression variance of putative neutral (non-DE) genes dropped to 84% of
374 the ancestral variance, which is quite close to the neutral cases in our simulations. While it
375 would be desirable to test each putatively selected (DE) gene independently, the moderate
376 sample size (~20 per population) does not provide sufficient power to do so (Supplementary
377 figure S8). Rather, we consider all putatively selected (DE) genes jointly and compare their
378 variance changes to the ones of putatively neutral (non-DE) genes. Remarkably, we find that
379 the changes in variance of putatively selected genes with significant mean expression changes
380 are indistinguishable from the genes that do not change their mean expression (t test, p-value
381 > 0.05; Figure 5c). The magnitude of gene expression variance changes in both sets of genes
382 (median F-value = 0.85 and 0.84 for DE and non-DE genes, respectively, in replicate 1, and
383 0.82 and 0.84 in replicate 2) is quite close to the neutral conditions in our computer simulations.
384 This suggests that selection on mean expression does not significantly change the expression
385 variance during adaptation. This is further supported by the lack of correlation between the
386 evolutionary changes in gene expression variance and mean (Lai & Schlötterer, 2022).
387 Although the adaptive architecture may differ between genes, we suggest that variation in
388 expression of most putatively adaptive genes has polygenic basis (i.e. more than five
389 contributing loci).
390
391 Since we only explored two time points (F0 and F103) rather than a full time series, it may be
392 possible that an oligogenic basis could also result in a similar phenotypic variance as neutral
393 phenotypes like what is observed for a polygenic architecture (Supplementary Figure S9). This
394 can be seen in an intuitive case when a single/few major effect allele(s) starts at a low frequency
395 and becomes fixed (Yoo *et al.*, 1980). Because an oligogenic basis results in a highly parallel
396 genomic selection response (Supplementary Figure S10), it is possible to distinguish polygenic
397 and oligogenic architectures with phenotypes from two time points only when genomic data
398 are available. Because the genomic signature in the same experiment uncovered a highly
399 heterogeneous selection response (Barghi *et al.*, 2019), we can exclude the unlikely explanation
400 of an oligogenic architecture resulting in a similar expression variance as non-selected genes.
401 Rather, we suggest that the adaptive response in gene expression could be explained by a
402 polygenic architecture (i.e. more than five contributing loci).

403 **Discussion**

404 Population genetics has a long tradition of characterizing adaptation based on the genomic
405 signature of selected loci (Nielsen, 2005). Nevertheless, for selected phenotypes with a
406 polygenic architecture, the contribution of individual loci to phenotypic change may be too
407 subtle to be detected with classic population genetic methods (Pritchard *et al.*, 2010). Although
408 identification of many small-effect loci is challenging, it may be possible to determine the
409 number of contributing loci based on the dynamics of phenotypic variance.

410

411 Inspired by the Castle-Wright estimator (Castle, 1921) that estimates the number of loci
412 contributing to a quantitative trait from the phenotypic variance of the F2, we propose that the
413 temporal heterogeneity of the phenotypic variance can potentially be used to infer the number
414 of loci contributing to the adaptive response of a phenotype as well as other parameters of the
415 adaptive architecture. Reasoning that experimental evolution is probably the best approach to
416 obtain phenotypic time series, we performed computer simulations specifically tailored to
417 typical experimental evolution studies with *Drosophila*. We demonstrated that, in an
418 experimental evolution setup, the temporal dynamics of the phenotypic variance is strongly
419 affected by the number of contributing loci and other parameters of the adaptive architecture
420 such as the distribution of effect size.

421

422 We acknowledge that the proposed approach for the characterization adaptive architecture
423 critically depends on the power to distinguish the variance changes of selected traits from the
424 neutral expectations. Since the accurate inference of gene expression variance is key, we
425 partitioned the expression variance of each gene among three independent F1 families consist
426 of almost identical, heterozygous individuals (see Supplementary Information). We show that
427 on average, ~50% of expression variance can be attributed to random error - a combination of
428 technical (library preparation, RNA-extraction, sequencing) and biological (stochastic gene
429 expression differences among individuals) noise. We included this error in our simulations by
430 adopting the empirical heritability distribution (heritability of every gene inferred from this
431 experiment) such that random errors relative to the ancestral phenotypic variance were
432 introduced to the calculation of phenotypic values at each generation for each simulated trait.
433 The heritability estimates in our experiment may be too conservative (i.e. too high) because we
434 assumed no environmental heterogeneity. We also used heritability estimates from another
435 gene expression study on *Drosophila* which included more lines and accounted for

436 environmental heterogeneity (Ayroles et al., 2009). The results are qualitatively similar (Figure
437 3).
438
439 Given the limited power to infer the number of contributing loci for each expression phenotype,
440 we grouped all putatively selected expression traits and limited our inference of the adaptive
441 architecture to these genes jointly. This approach makes the implicit assumption that all
442 expression traits in this group are independent of each other and have similar level of
443 complexity in their adaptive architectures. This joint inference could potentially compromise
444 the detection of a minor subset of genes whose expression evolution is under simple genetic
445 control.
446
447 Our results may also provide some insights into the molecular basis of gene expression
448 evolution. Gene expression is modified by many trans-acting factors and some cis-regulatory
449 variation (Hill *et al.*, 2021). Adaptive gene expression changes can be either driven by
450 polymorphism in *cis*-regulatory elements or by *trans*-acting variants. While interspecific
451 differences in gene expression are predominantly caused by *cis*-regulation (Wittkopp *et al.*,
452 2004), intraspecific variation is mostly driven by trans regulatory changes (Wittkopp *et al.*,
453 2008; Suvorov *et al.*, 2013). Adaptive gene expression changes which are well-characterized
454 on the molecular level typically have a *cis*-regulatory basis that is not only frequently
455 associated with the insertion of a transposable element (e.g.: (Daborn *et al.*, 2002)) but also
456 sometimes with multiple regulatory variants (Endler *et al.*, 2018). Two lines of evidence
457 suggest that *cis*-regulatory variation cannot be the driver of adaptive gene expression changes
458 observed in this study. First, the mutational target size is too small to harbor a sufficiently large
459 number of alleles segregating in the founder population. Second, too few recombination events
460 occur during the experiment to uncouple regulatory variants located on a given haplotype such
461 that they could generate a signal of polygenic adaptation. More likely, the polygenic adaptive
462 architecture of gene expression change reflects the joint effects of many *trans*-acting variants.
463
464 Because we could only analyze phenotypic data from two time points, the founder population
465 and replicate populations evolved for 103 generations, we were not able to obtain a more
466 quantitative estimate of the number of contributing loci, in particular as other parameters of the
467 adaptive architecture are not known and need to be co-estimated. In addition, with only two
468 time points, for a few parameter combinations, an oligogenic response can also result in a
469 similar phenotypic variance change as a polygenic one but with a much higher parallel response

470 of genomic markers (Supplementary Figure S10). Hence, not only more time points describing
471 the phenotypic trajectory, but also some genomic data could contribute to inferring the adaptive
472 architecture in experimental evolution studies.
473
474 The extension of the concept proposed in this study to natural populations could face several
475 challenges that warrant extra caution and further investigation. First, the above-mentioned
476 challenge of sample size would be more pronounced in a natural setup. Second, phenotypic
477 time series over evolutionary relevant time scales are costly (but see (Clutton-Brock &
478 Pemberton, 2004)) and third, the distinction of environmental heterogeneity from genetic
479 changes is considerably more challenging than under controlled laboratory conditions.
480

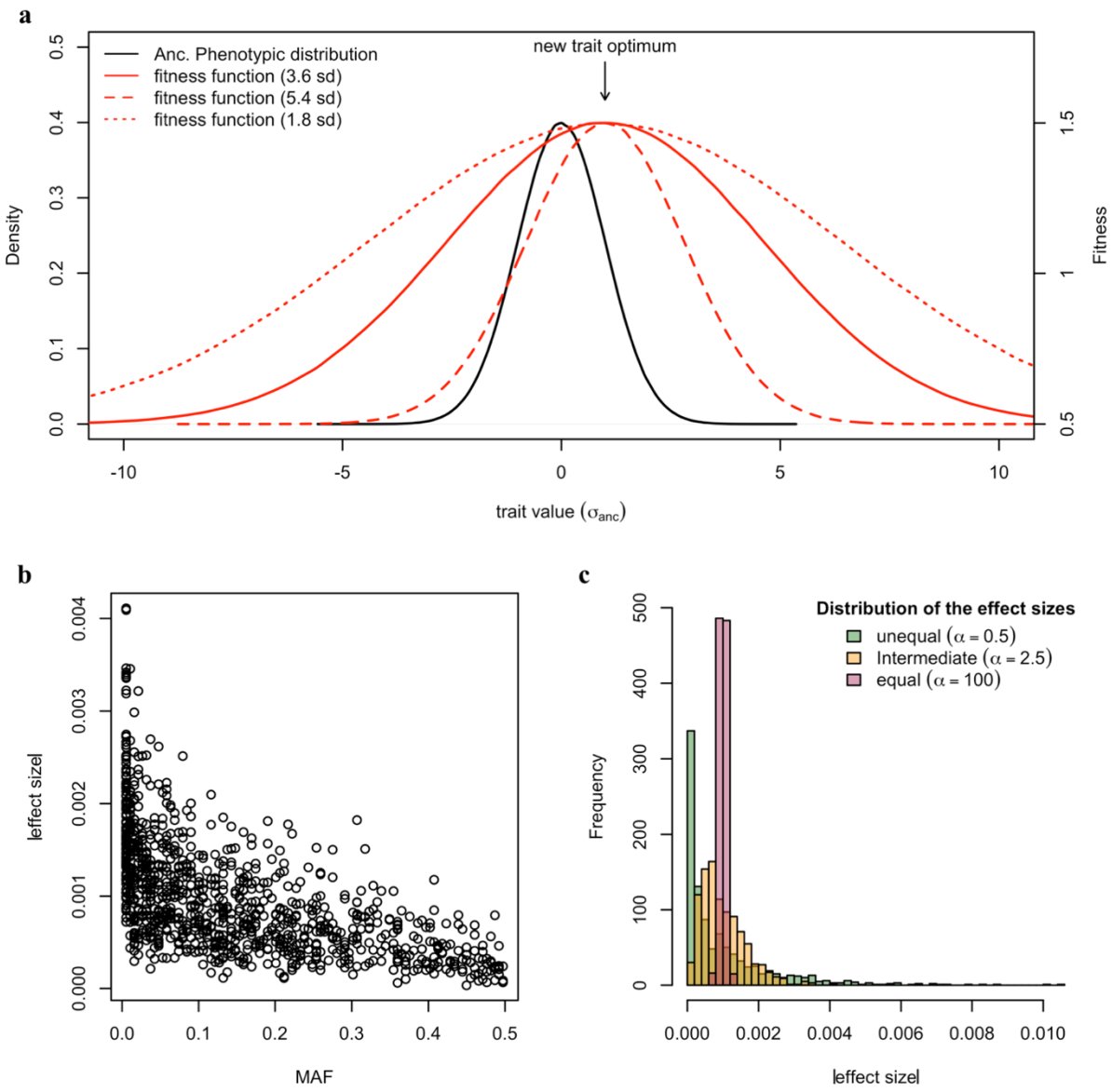


Figure 1. Simulating polygenic adaptation to a shift in trait optimum with different parameter combinations. a. For the computer simulations we consider a quantitative trait (in black) experiencing a sudden shift in trait optimum under stabilizing selection. The underlying fitness functions are illustrated in red. The new trait optimum is shifted from the ancestral trait mean by one/three standard deviation of the ancestral trait distribution. The strength of stabilizing selection is modified by changing the variance of the fitness function: 1.8, 3.6 and 5.4 standard deviations of the ancestral trait distribution. **b.** The negative correlation between the allele frequencies and the effect sizes (r = -0.7, estimated in Otte *et al.* (2020)). We consider such negative correlation when assigning the effect sizes to variants underlying a simulated trait. **c.** The distribution of effect sizes of the contributing loci is determined by the shape parameter (α) of gamma sampling process (α = 0.5, 2.5 and 100).

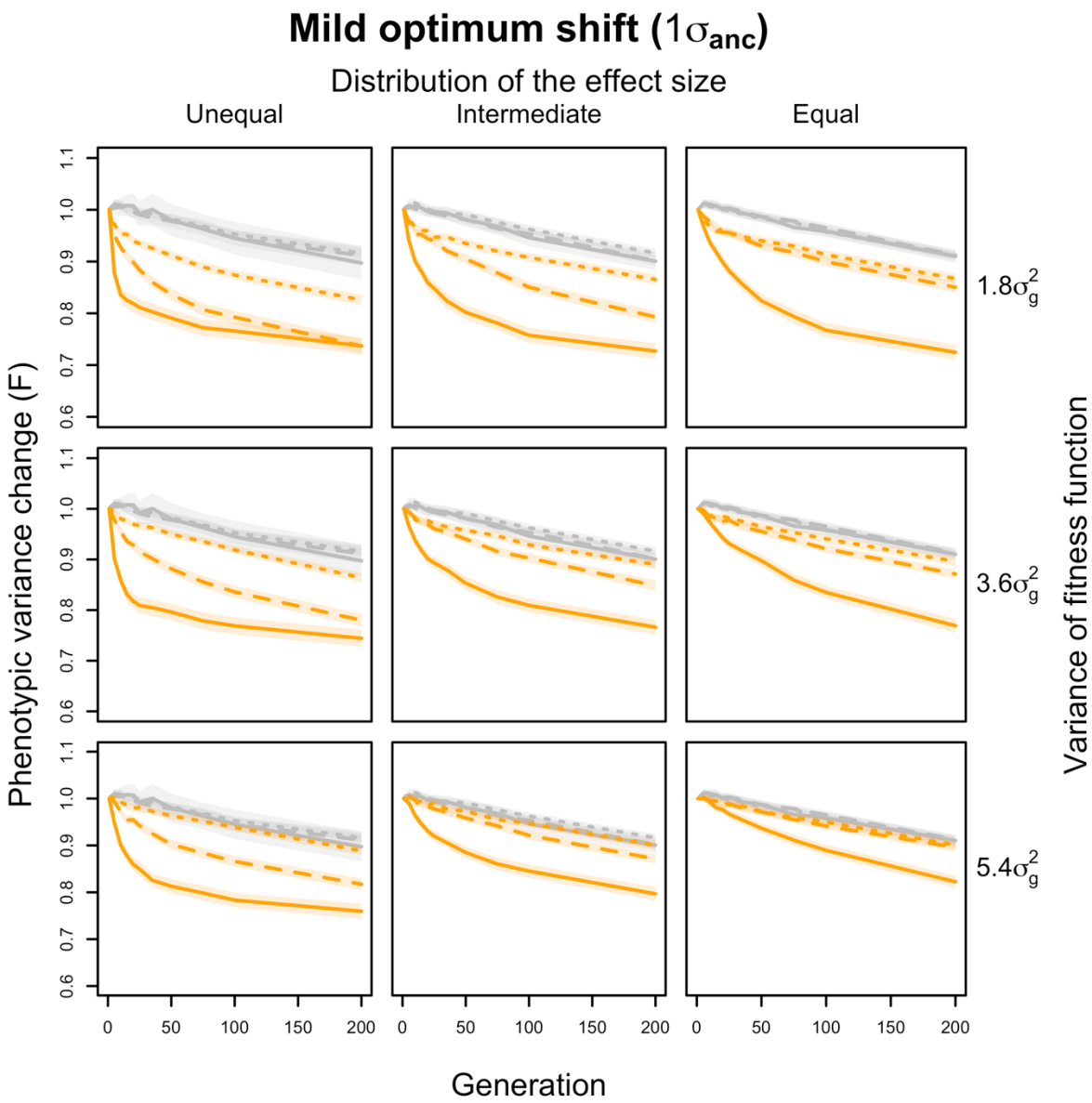


Figure 2. The trajectory of changes in phenotypic variance during adaptation to a mild optimum shift. The changes in phenotypic variance within 200 generations adapting to a moderate optimum shift (orange) are compared to the changes under neutrality (grey) on y axis. The average variance changes (F) of 1000 simulated traits is calculated as the ratio of phenotypic variance between each evolved time point (generation x) and the ancestral state (σ_x^2/σ_1^2). The translucent band indicates the 95% confidence interval for 1000 simulated traits. The simulations cover traits controlled by varying numbers of loci underlying the adaptation with three different distributions of effect sizes (columns) under different strength of stabilizing selection (rows). Only traits with the most (dotted lines, 1000 loci), intermediate (dash lines, 50 loci) and the least (solid lines, 5 loci) polygenic architectures are shown. In all scenarios, the variance of the trait decreases drastically when the adaptation is controlled by a small

number of loci (orange solid lines; 5 loci). While, for traits with extremely polygenic basis, the phenotypic variance stays stable over time (orange dotted lines).

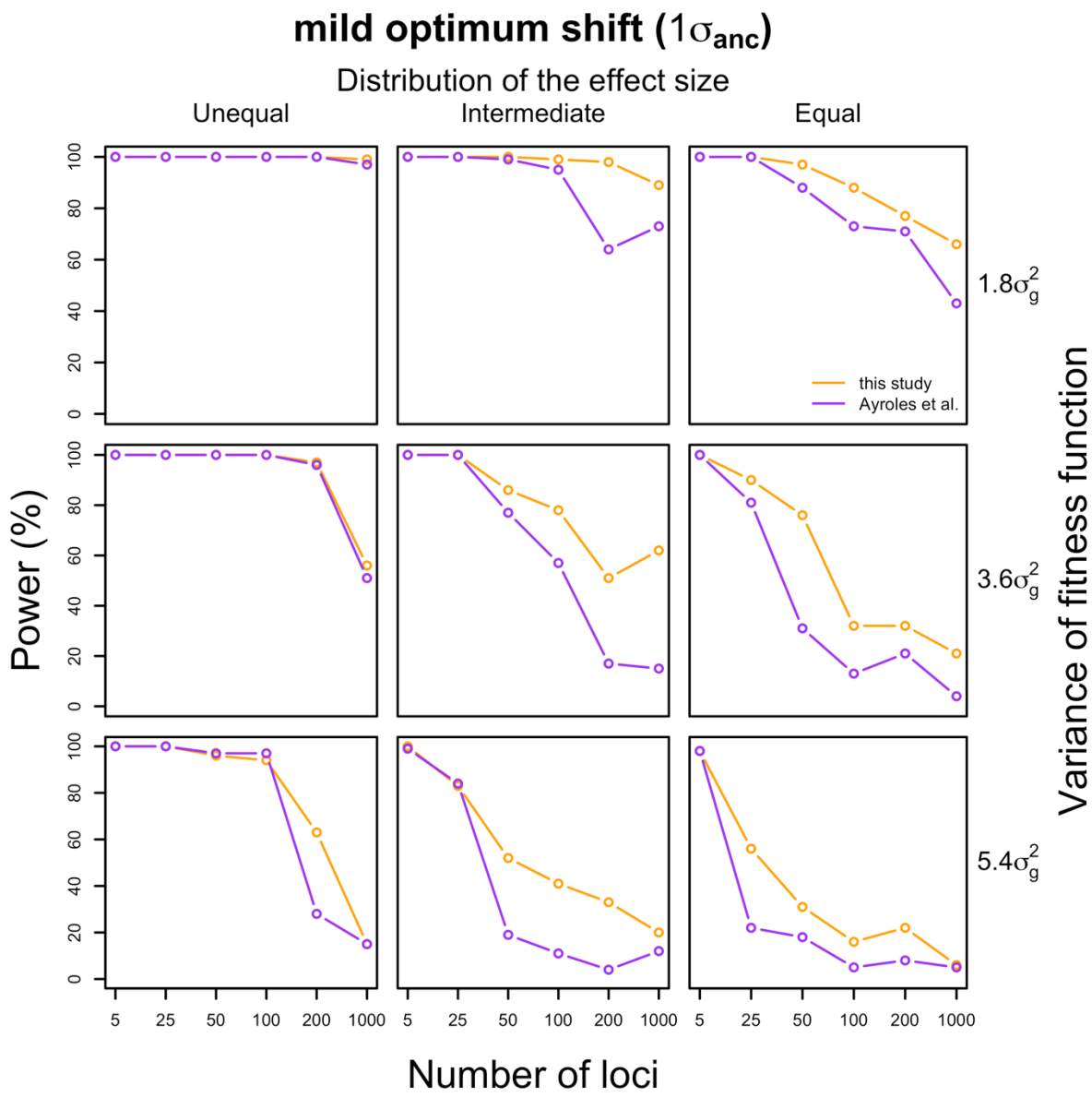


Figure 3. Power to detect significant differences in variance change between groups of selected and neutral traits. For each set of the 1000 traits controlled by different numbers of loci (x-axis) with varying effect sizes (columns) under each selection strength (rows), we calculated how often the variance change differs between 1000 neutral and 1000 selected traits after 100 generations (y-axis). The same genetic architecture is assumed for all 1000 selected traits with a sample size of 20. For each parameter combination, we have almost 100% power to detect a significant difference when all traits were under simple genetic control (5 loci). The power gradually decreases with an increasing number of loci (orange). A lack of significant difference indicates polygenic architecture of the adaptive traits. The results remain unaffected when we used heritability estimates from Ayroles et al., (2009) (purple).

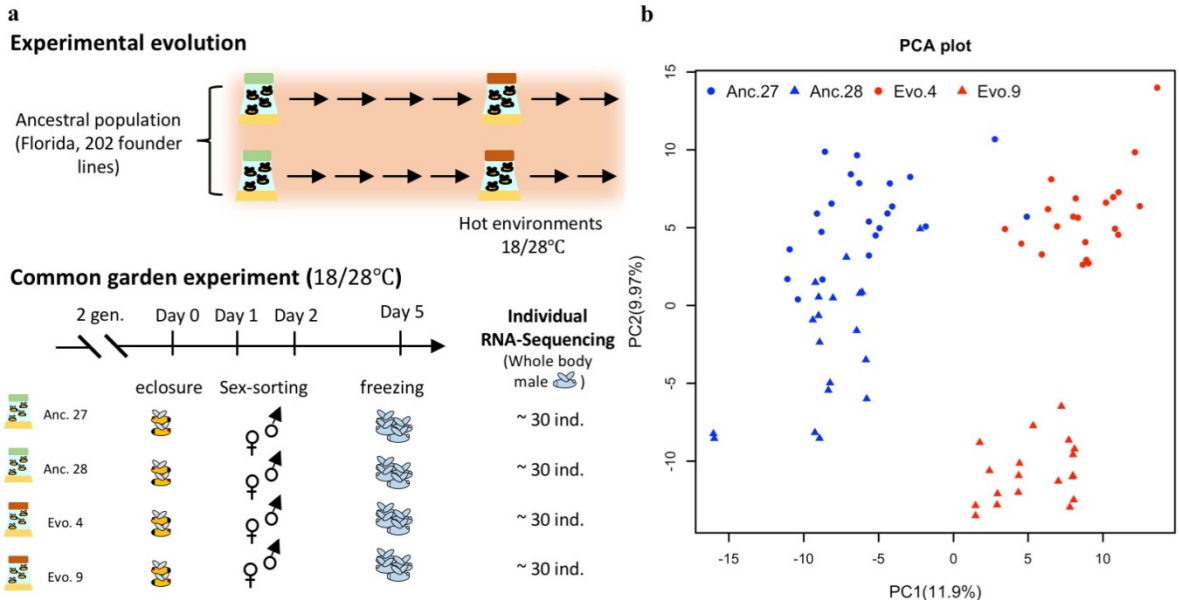
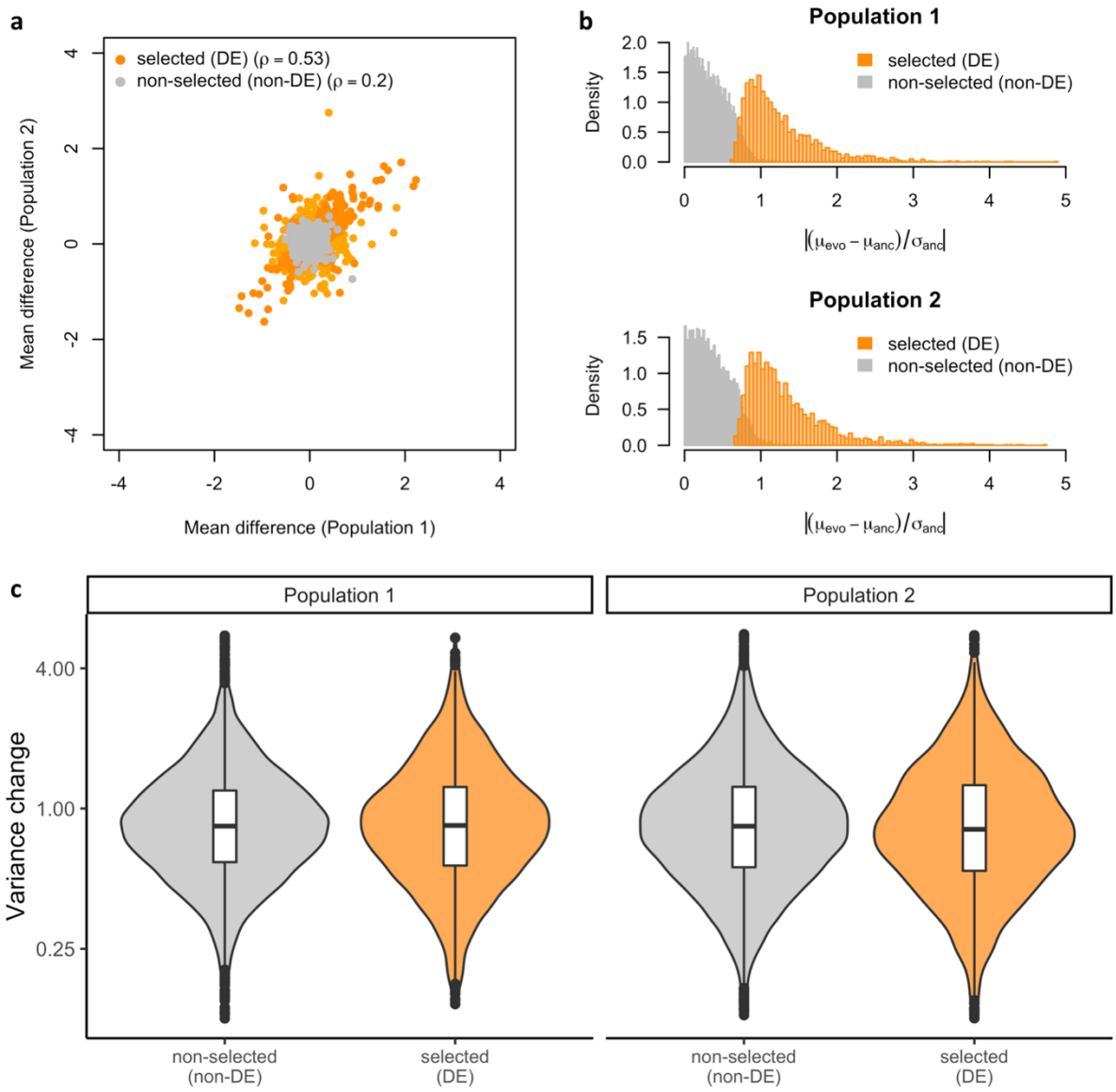


Figure 4. Schematic overview of the experimental procedures (a) and the divergence in gene expression during experimental evolution (b). **a.** Experimental evolution: starting from one common founder population, two replicate populations evolved for 100 generations in a hot laboratory environment fluctuating between 18 and 28°C. Common garden Experiment: after 100 generations, the two evolved replicate populations were maintained together with the reconstituted ancestral population for two generations in a hot laboratory environment fluctuating between 18 and 28°C. After this common garden procedure, about 30 males from each population were subjected to RNA-Seq. **b.** Principle Component Analysis (PCA) of the transcriptomic profiles of individuals from the ancestral population (blue) and the hot-evolved population (red). Circles indicate individuals of the first replicate (Anc. No. 27 and Evo. No. 4). Triangles represent individuals of the second replicate (Anc. No. 28 and Evo. No. 9). The two replicates were made with two different batches of library cards for RNA-Seq library preparation. The impact of the library card batch can be seen from the separation of the ancestral replicates which were reconstituted from the same iso-female lines.



537

538 **Figure 5. Evolution of phenotypic mean and variance over 100 generations of adaptation**
539 **in empirical data. a.** The evolution of gene expression means during adaptation in the two
540 replicates. For the genes with significant changes (DE, in orange), the changes are correlated
541 between replicates (Spearman's rho = 0.53). For the genes without significant changes (non-
542 DE, in grey), the correlation between replicates is much lower (Spearman's rho = 0.2). **b.** The
543 evolution of gene expression means scaled by the ancestral variation. For the DE genes (in
544 orange), the median change falls around 1 standard deviation of the ancestral expression value,
545 suggesting mild shift in trait optimum in the novel environment. For the non-DE genes (in
546 grey), the changes in expression are mostly negligible. The same pattern is seen in the second
547 replicate. **c.** The change in expression variance during adaptation for DE and non-DE genes. In
548 both replicates, the distribution of variance changes is indistinguishable between DE genes

549 (orange) and non-DE genes (grey) (t-test, $p > 0.05$ for both replicates). Almost identical pattern
550 is observed when the biological coefficient of variation (BCV; see materials and methods) is
551 used to estimate the expression variance (Supplementary Figure S4).
552

Acknowledgments

Special thanks to David Houle, who provided fantastic support during the collection and establishment of the isofemale lines in Florida. We thank all member of the Institut für Populationsgenetik for discussion. We are grateful to Reinhard Bürger, David Houle, Dagný Ásta Rúnarsdóttir and anonymous reviewers for helpful comments on earlier versions of the manuscript. Neda Barghi, François Mallard and Kathrin Otte contributed to the common garden experiment. Illumina sequencing was performed at the VBCF NGS Unit (www.vbcf.ac.at). This work was support by the Austrian Science Funds (FWF, W1225) and the European Research Council (ERC, ArchAdapt).

Author contribution

W.Y.L and C.S. conceived the study. V.N. prepared all RNA-Seq and supervised the maintenance of the evolution experiment. A.M.J supervised the common garden experiment. W.Y.L performed the simulation and data analysis. W.Y.L. and C.S. wrote the manuscript.

Data accessibility statement

All sequencing data will be available in European Nucleotide Archive (ENA) under the accession number PRJEB37011 upon publication.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to C.S.

References

Ayroles, J.F., Carbone, M.A., Stone, E.A., Jordan, K.W., Lyman, R.F., Magwire, M.M., *et al.* (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics*, **41**, 299–307.

Barghi, N., Hermisson, J. & Schlötterer, C. (2020) Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics*, **21**, 769–781.

Barghi, N., Tobler, R., Nolte, V., Jakšić, A.M., Mallard, F., Otte, K.A., *et al.* (2019) Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biology*, **17**, e3000128.

Barton, N.H., Etheridge, A.M. & Véber, A. (2017) The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, **118**, 50–73.

Barton, N.H. & Keightley, P.D. (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**, 11–21.

Barton, N.H. & Turelli, M. (1987) Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genetical Research*, **49**, 157–173.

Blair, G., Cooper, J., Coppock, A., Humphreys, M., Rudkin, A. & Fultz, N. (2019) fabricatr: imagine your data before you collect it.

Boyle, E.A., Li, Y.I. & Pritchard, J.K. (2017) An expanded view of complex traits: From polygenic to omnigenic. *Cell*, **169**, 1177–1186.

Bulmer, M.G. (1972) The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetical Research*, **19**, 17–25.

Burke, M.K., Dunham, J.P., Shahrestani, P., Thornton, K.R., Rose, M.R. & Long, A.D. (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature*, **467**, 587–590.

Careau, V., Wolak, M.E., Carter, P.A., Garland, T. & Jr. (2015) Evolution of the additive genetic variance-covariance matrix under continuous directional selection on a complex behavioural phenotype. *Proceedings. Biological sciences*, **282**.

Castle, W.E. (1921) An improved method of estimating the number of genetic factors concerned in the cases of blending inheritance. *Science (New York, N.Y.)*, **54**, 223.

Chevalet, C. (1994) *An approximate theory of selection assuming a finite number of quantitative trait loci*. *Genetics Selection Evolution*.

Chippindale, A.K., Chu, T.J.F. & Rose, M.R. (1996) Complex trade-offs and the evolution of starvation resistance in Drosophila melanogaster. *Evolution*, **50**, 753–766.

Clutton-Brock, T.H. & Pemberton, J.M. (Josephine M. (2004) *Soay sheep : population dynamics and selection on St. Kilda*. Cambridge University Press.

611 Daborn, P.J., Yen, J.L., Bogwitz, M.R., Goff, G. Le, Feil, E., Jeffers, S., *et al.* (2002) A single
612 p450 allele associated with insecticide resistance in *Drosophila*. *Science (New York, N.Y.)*, **297**,
613 2253–6.
614 Endler, L., Gibert, J.-M., Nolte, V. & Schlötterer, C. (2018) Pleiotropic effects of regulatory
615 variation in *tan* result in correlation of two pigmentation traits in *Drosophila melanogaster*.
616 *Molecular Ecology*, **27**, 3207–3218.
617 Falconer, D.S. & Mackay, T.F.C. (1963) *Introduction to quantitative genetics*. *Poultry Science*.
618 CreateSpace Independent Publishing Platform.
619 Ferea, T.L., Botstein, D., Brown, P.O. & Rosenzweig, R.F. (1999) Systematic changes in gene
620 expression patterns following adaptive evolution in yeast. *Proceedings of the National*
621 *Academy of Sciences of the United States of America*, **96**, 9721–6.
622 Franssen, S.U., Kofler, R. & Schlötterer, C. (2017) Uncovering the genetic signature of
623 quantitative trait evolution with replicated time series data. *Heredity*, **118**, 42–51.
624 Fukunaga, K. & Hummels, D.M. (1989) Leave-one-out procedures for nonparametric error
625 estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 421–423.
626 Hayward, L.K. & Sella, G. (2019) Polygenic adaptation after a sudden change in environment.
627 *bioRxiv*, 792952.
628 Hill, M.S., Zande, P. Vande & Wittkopp, P.J. (2021) Molecular and evolutionary processes
629 generating variation in gene expression. *Nature Reviews Genetics*, **22**, 203–215.
630 Howie, J.M., Mazzucco, R., Taus, T., Nolte, V. & Schlötterer, C. (2019) DNA motifs are not
631 general predictors of recombination in two drosophila sister species. *Genome Biology and*
632 *Evolution*, **11**, 1345–1357.
633 Hsu, S.-K., Jakšić, A.M., Nolte, V., Barghi, N., Mallard, F., Otte, K.A., *et al.* (2019) A 24 h
634 age difference causes twice as much gene expression divergence as 100 generations of
635 adaptation to a novel environment. *Genes*, **10**, 89.
636 Hsu, S.-K., Jakšić, A.M., Nolte, V., Lirakis, M., Kofler, R., Barghi, N., *et al.* (2020) Rapid sex-
637 specific adaptation to high temperature in *Drosophila*. *eLife*, **9**.
638 Huang, Y. & Agrawal, A.F. (2016) Experimental evolution of gene expression and plasticity
639 in alternative selective regimes. *PLOS Genetics*, **12**, e1006336.
640 Jain, K. & Stephan, W. (2015) Response of polygenic traits under stabilizing selection and
641 mutation when loci have unequal effects. *G3 (Bethesda, Md.)*, **5**, 1065–74.
642 Jakšić, A.M., Karner, J., Nolte, V., Hsu, S.-K., Barghi, N., Mallard, F., *et al.* (2020) Neuronal
643 function and dopamine signaling evolve at high temperature in *Drosophila*. *Molecular Biology*
644 *and Evolution*, **37**, 2630–2640.

645 Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I. & Whitlock, M.C. (2012)
646 Experimental evolution. *Trends in Ecology & Evolution*, **27**, 547–560.
647 Keightley, P.D. & Hill, W.G. (1989) Quantitative genetic variability maintained by mutation-
648 stabilizing selection balance: sampling variation and response to subsequent directional
649 selection. *Genetical Research*, **54**, 45–58.
650 Kimura, M. & Crow, J.F. (1964) The number of alleles that can be maintained in a finite
651 population. *Genetics*, **49**, 725–738.
652 Lai, W. & Schlötterer, C. (2022) Evolution of phenotypic variance in response to a novel hot
653 environment. *Molecular Ecology*, **31**, 934–945.
654 Lande, R. (1976) Natural selection and random genetic drift in phenotypic evoltion. *Evolution*,
655 **30**, 314–334.
656 Lenski, R.E., Travisano, M., Larison, B. & Moritz, C. (1994) Dynamics of adaptation and
657 diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the*
658 *National Academy of Sciences*, **91**, 6808–6814.
659 Liu, X., Li, Y.I. & Pritchard, J.K. (2019) Trans effects on gene expression can drive omnigenic
660 inheritance. *Cell*, **177**, 1022-1034.e6.
661 Mackay, T.F.C., Stone, E.A. & Ayroles, J.F. (2009) The genetics of quantitative traits:
662 challenges and prospects. *Nature Reviews Genetics*, **10**, 565–577.
663 Mallard, F., Nolte, V., Tobler, R., Kapun, M. & Schlötterer, C. (2018) A simple genetic basis
664 of adaptation to a novel thermal environment results in complex metabolic rewiring in
665 *Drosophila*. *Genome Biology*, **19**, 119.
666 Nielsen, R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**,
667 197–218.
668 Nouhaud, P., Tobler, R., Nolte, V. & Schlötterer, C. (2016) Ancestral population reconstitution
669 from isofemale lines as a tool for experimental evolution. *Ecology and Evolution*, **6**, 7169–
670 7175.
671 Otte, K.A., Nolte, V., Mallard, F. & Schlötterer, C. (2020) The adaptive architecture is shaped
672 by population ancestry and not by selection regime. *bioRxiv*, 2020.06.25.170878.
673 Pélabon, C., Hansen, T.F., Carter, A.J.R. & Houle, D. (2010) Evolution of variation and
674 variability under fluctuating, stabilizing, and disruptive selection. *Evolution*, **64**, 1912–25.
675 Penna, A., Melo, D., Bernardi, S., Oyarzabal, M.I. & Marroig, G. (2017) The evolution of
676 phenotypic integration: How directional selection reshapes covariation in mice. *Evolution*, **71**,
677 2370–2380.
678 Pritchard, J.K., Pickrell, J.K. & Coop, G. (2010) The genetics of human adaptation: Hard

679 sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–R215.
680 Romero, I.G., Ruvinsky, I. & Gilad, Y. (2012) Comparative studies of gene expression and the
681 evolution of gene regulation. *Nature Reviews Genetics*, **13**, 505–516.
682 Signor, S.A. & Nuzhdin, S. V. (2018) The evolution of gene expression in cis and trans. *Trends*
683 *in Genetics*, **34**, 532–544.
684 Sork, V.L. (2017) Genomic studies of local adaptation in natural plant populations. *Journal of*
685 *Heredity*, **109**, 3–15.
686 Suvorov, A., Nolte, V., Pandey, R.V., Franssen, S.U., Futschik, A. & Schlötterer, C. (2013)
687 Intra-specific regulatory variation in *Drosophila pseudoobscura*. *PLoS ONE*, **8**, e83547.
688 Turelli, M. (1984) Heritable genetic variation via mutation-selection balance: Lerch's zeta
689 meets the abdominal bristle. *Theoretical Population Biology*, **25**, 138–193.
690 Vlachos, C. & Kofler, R. (2018) MimicrEE2: Genome-wide forward simulations of Evolve
691 and Resequencing studies. *PLOS Computational Biology*, **14**, e1006413.
692 Wittkopp, P.J., Haerum, B.K. & Clark, A.G. (2004) Evolutionary changes in cis and trans gene
693 regulation. *Nature*, **430**, 85–88.
694 Wittkopp, P.J., Haerum, B.K. & Clark, A.G. (2008) Regulatory changes underlying expression
695 differences within and between *Drosophila* species. *Nature Genetics*, **40**, 346–350.
696 Yoo, B.H., Nicholas, F.W. & Rathie, K.A. (1980) Long-term selection for a quantitative
697 character in large replicate populations of *Drosophila melanogaster*. *Theoretical and Applied*
698 *Genetics*, **57**, 113–117.
699

Supplementary text

Estimating genetic variance in gene expression across F1 families

In this study, we included another experiment to evaluate how much of the expression variance can be explained by genetic variation by performing RNA-Seq on individual flies, with 3-4 individuals each from three iso-female lines maintained at the same density and culturing conditions (Supplementary file 1).

Six out of 184 founder iso-female lines from the evolution experiment and were maintained for one generation with controlled egg density (400 eggs/bottle) in the same environment as the main experiment (12h 28°C with light followed by 12h 18 °C with dark conditions). Using the offspring, we generated three crosses between two of the six lines each: FL 138 x FL 137, FL 157 x FL 112, FL 123 x FL 127: we combined 50 virgin females from one of the lines with 50 males from the other line, let them lay eggs under density control as above and maintained and froze their F1 offspring in the same way as in the main CGE: sexes were separated after mating at the age of three days and snap-frozen at the age of five days at 2pm. From each cross, we used four F1 males to prepare individual RNA-Seq libraries as described above.

Assuming no environmental heterogeneity, we decomposed the total variance of the expression of each gene measured in these individuals into the genetic difference among three different F1 families and environmental/technical error. The data were analyzed as follows:

Natural log-transformation was applied to CPMs of all genes to improve data normality (Rocke and Durbin 2003). We tested for the genetic variance of each gene separately using analysis of variance (ANOVA):

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

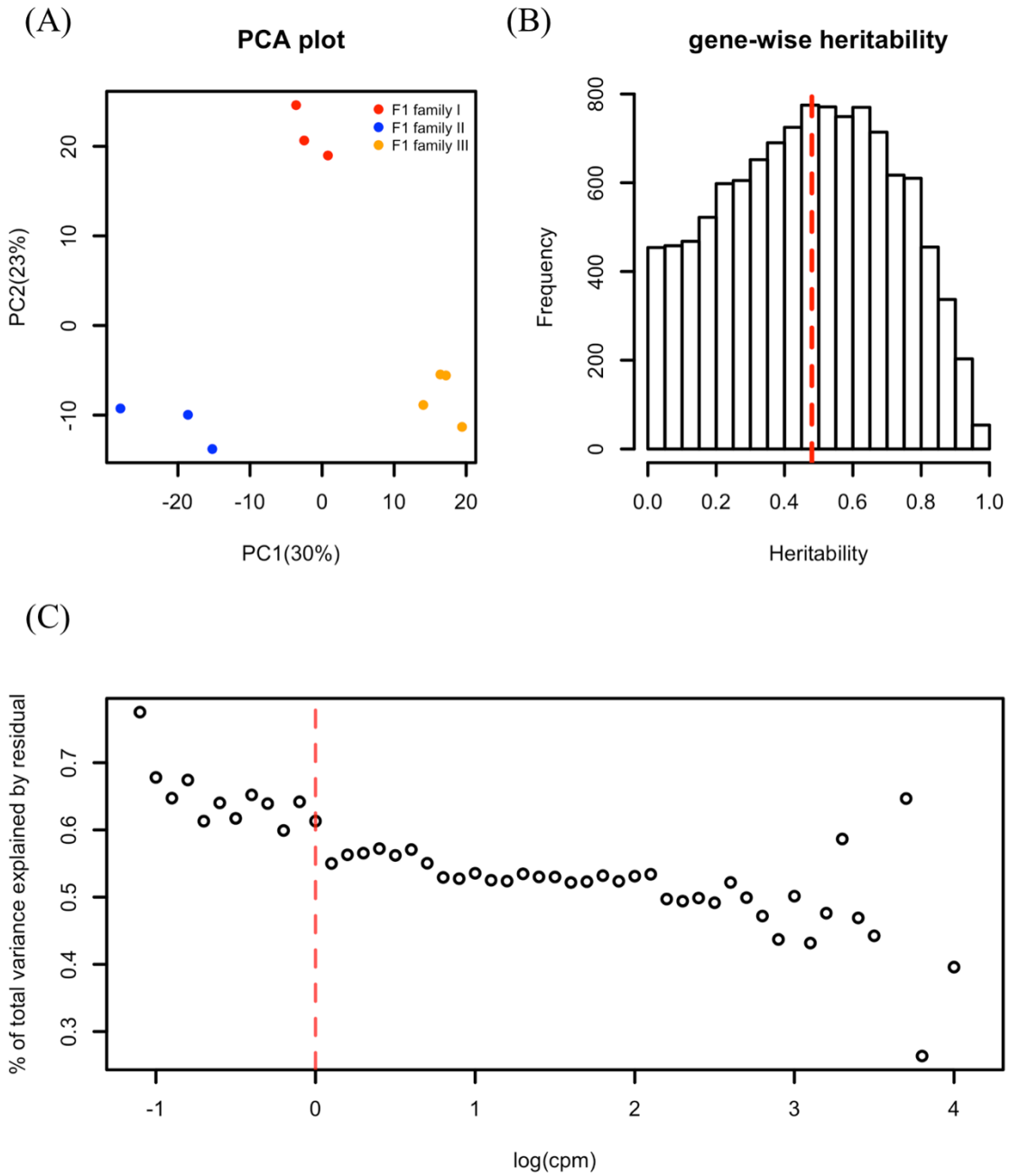
Where i=1, 2, 3 (the ith families); j =1, 2, 3, 4 (the jth individuals in each cross). y_{ij} is the observed expression level of a gene in a given sample, μ is the overall mean; τ_i is the effect of genetic background and ε_{ij} is the random noise. We calculated the proportion of total variance explained by random error using the following equation:

$$variance\ explained\ by\ random\ error = \frac{sum\ squares\ of\ error\ (SSE)}{sum\ squares\ of\ total\ (SST)}$$

We found that on average, around 50% of the gene expression variance can be explained by the genetic difference among the three F1 families (Supplementary Figure S1). This implies that a substantial within-population (between individuals) gene expression variance in our focal experiment is contributed by the genetic components. Because we only used offspring from single vials, we may have overestimated the heritability if the environment in the vials differs. Nevertheless, since our heritability estimates are very similar to previous ones (Ayroles *et al.* 2009), we consider our estimates reliable.

The technical error will go into residual in this analysis. Since variance of lowly expressed genes are mainly dominated by technical error, we defined a cut-off of expression value for a more reliable variance estimate. Genes were binned based on their average expression value (lnCPM) which ranged from -0.8 to 4.1, by bin size of 0.1. The average proportion of variance explained by random error of each bin was calculated. The expression variance of genes with less than 1 count per million bases (CPM) is dominated by technical errors (Supplementary Figure S9). Thus, genes with less than 1 count per million base (CPM) were excluded for subsequent analysis.

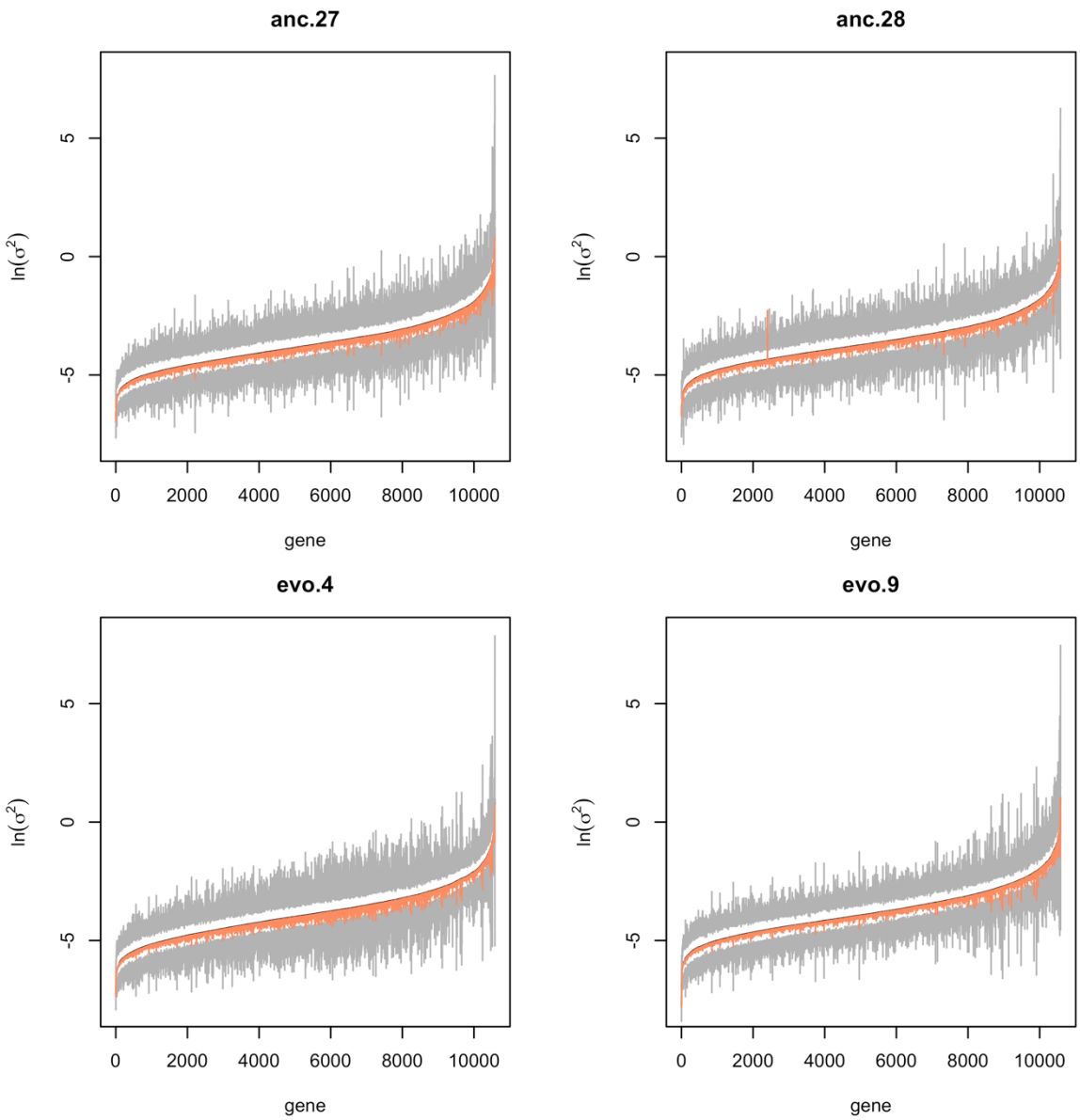
49



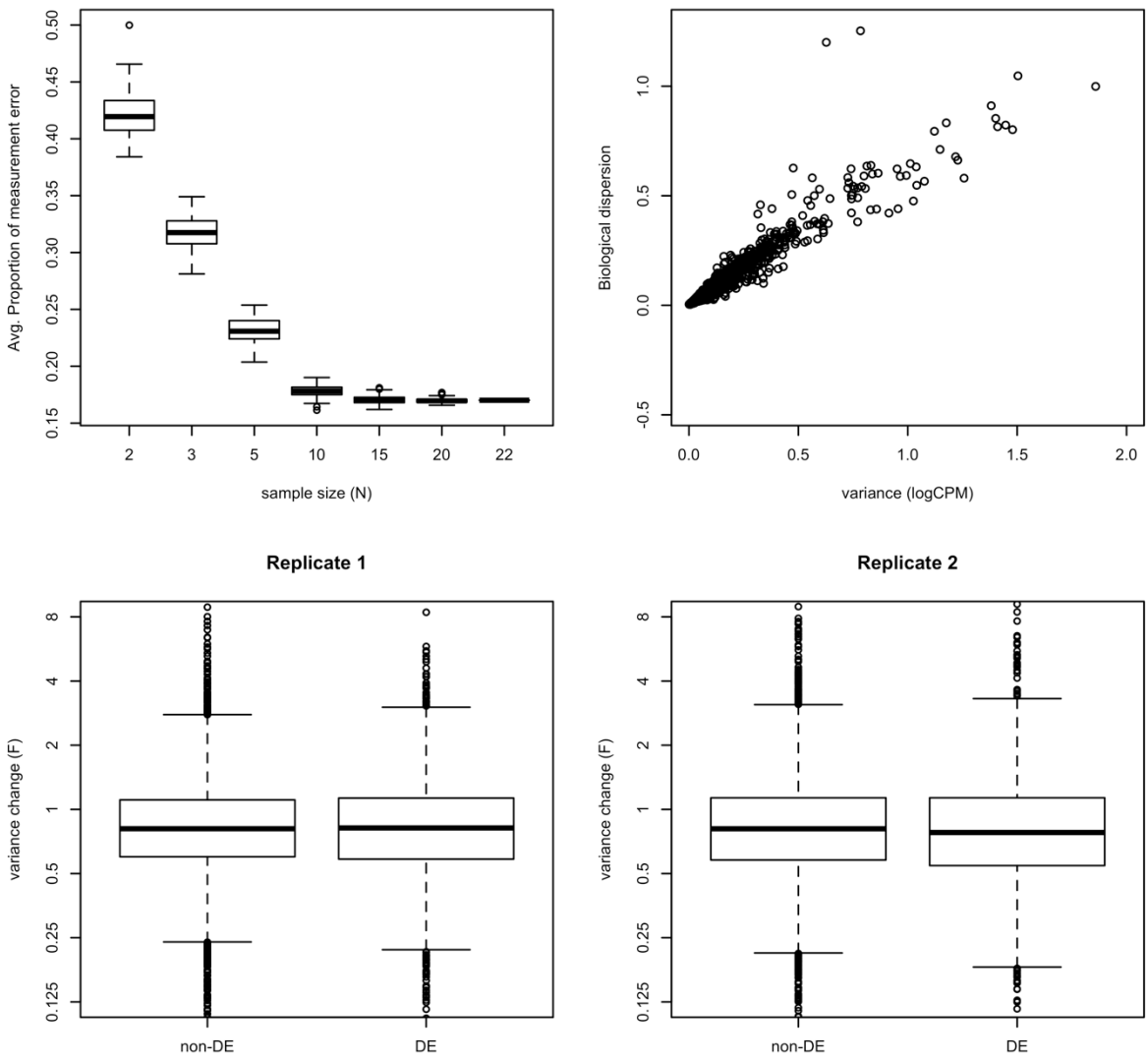
50

51 **Supplementary Figure S1. Genetic variance in gene expression across F1 families. (A)**
52 Principal component analysis (PCA) on the transcriptomes of F1 individuals from three
53 different crosses between the founder iso-female lines. Individuals from different families
54 clustered nicely based on the first two PCs. **(B)** Gene-wise heritability on the transcriptomes
55 of F1 individuals. On average, 48.8% of the total variance in gene expression was explained
56 by the genetic difference between the individuals. **(C)** Genes were binned based on their
57 average expression value (lnCPM) which ranged from -0.8 to 4.1, by bin size of 0.1. The
58 average proportion of variance explained by random error of each bin was visualized. The

59 expression variance of genes with less than 1 count per million bases (CPM) is dominated by
60 residuals.
61

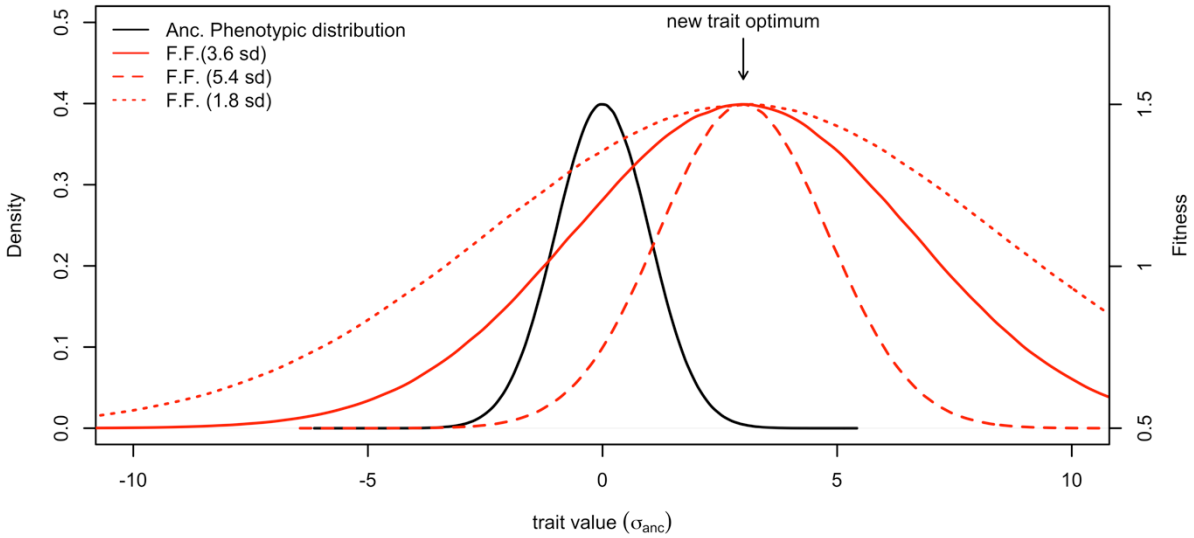


Supplementary Figure S2. Robustness of the variance estimation using individual sequencing data. Jackknife method was applied to measure the uncertainty of variance estimation within each population. Given a sample size of K, the procedure is to estimate the variance of each gene for K times, each time leaving one sample out. The procedure was conducted independently on 4 populations (anc.27, anc.28, evo.4 and evo.9). In each panel, we visualize Jackknife approximated 95% confidence interval for the variance estimates of each gene. The genes are ordered based on the average variance estimates (black dash line) on the x-axis. The upper and lower limits of the 95% confidence interval are indicated with grey curves. The salmon line denotes the observed value of the variance estimates. In most cases, the estimates lie in the confidence interval, suggesting robust estimation.

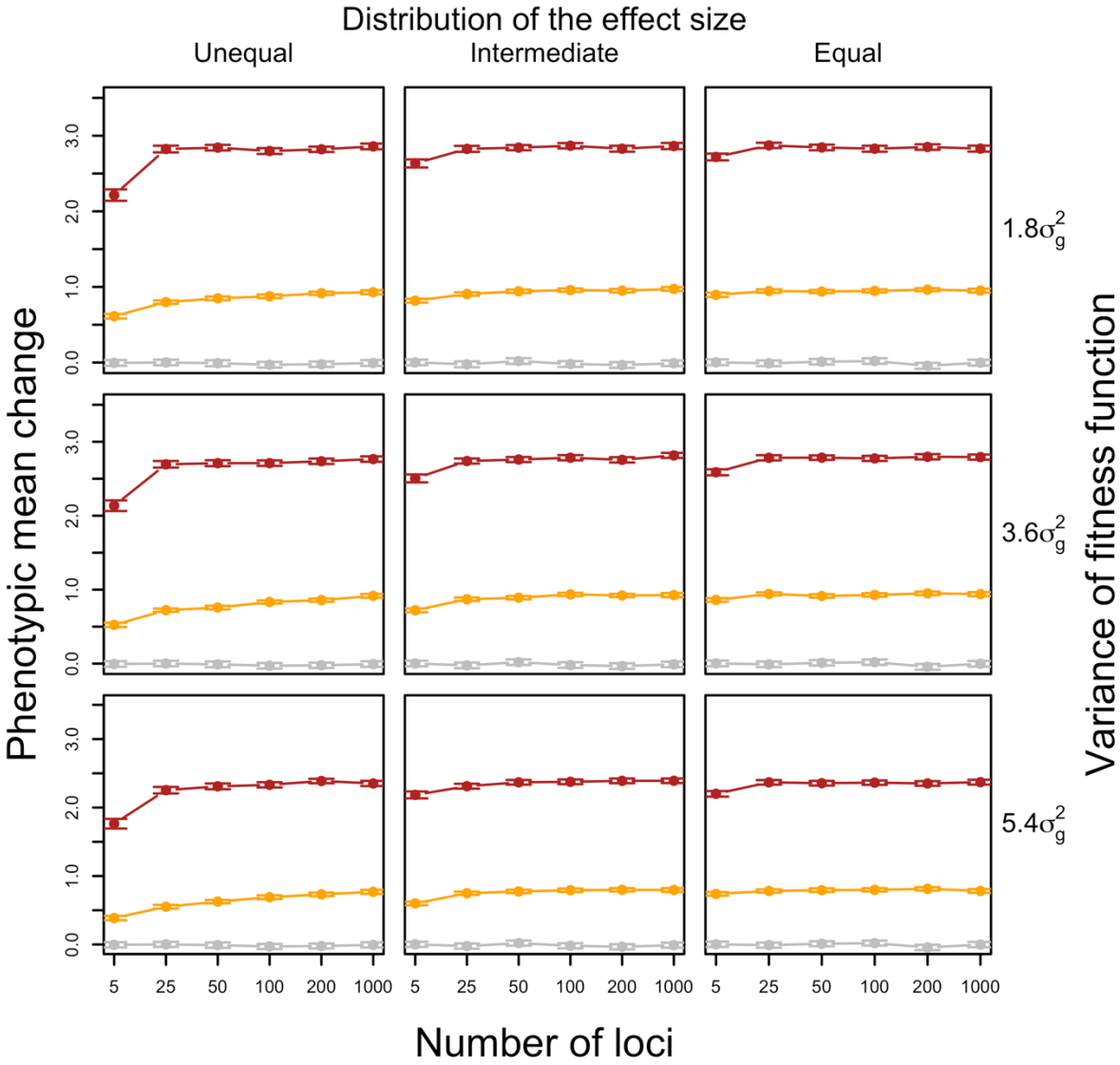


Supplementary Figure S3. The decomposition of measurement error in RNA-Seq data and its implication for the tests of variance evolution. (A) The relationship between the average proportion of measurement error among the RNA-Seq samples and the sample sizes is visualized. As the sample size increases, the measurement error of RNA-Seq data is well controlled. The results are consistent across replicate populations. For simplicity, only the result in the first replicate of ancestral population is shown. **(B)** We observed a strong correlation between the variance estimates after log-transformation and the biological coefficient of variation (BCV) estimated from edgeR, which claims to exclude the measurement error. Such strong correlation suggests that measurement error in RNA-Seq data wouldn't have strong impact on the test of variance of evolution. The results are consistent across replicate populations. For simplicity, only the result in the first replicate of ancestral

population is shown. **(C)** Using the BCV estimates, we repeated the comparison between the variance changes ($F = \frac{BCV_{evo}^2}{BCV_{anc}^2}$) of the genes with and without significant evolution in mean expression. In both replicates, the distribution of variance changes is indistinguishable between DE genes and non-DE genes (t-test, $p > 0.05$ for both replicates).

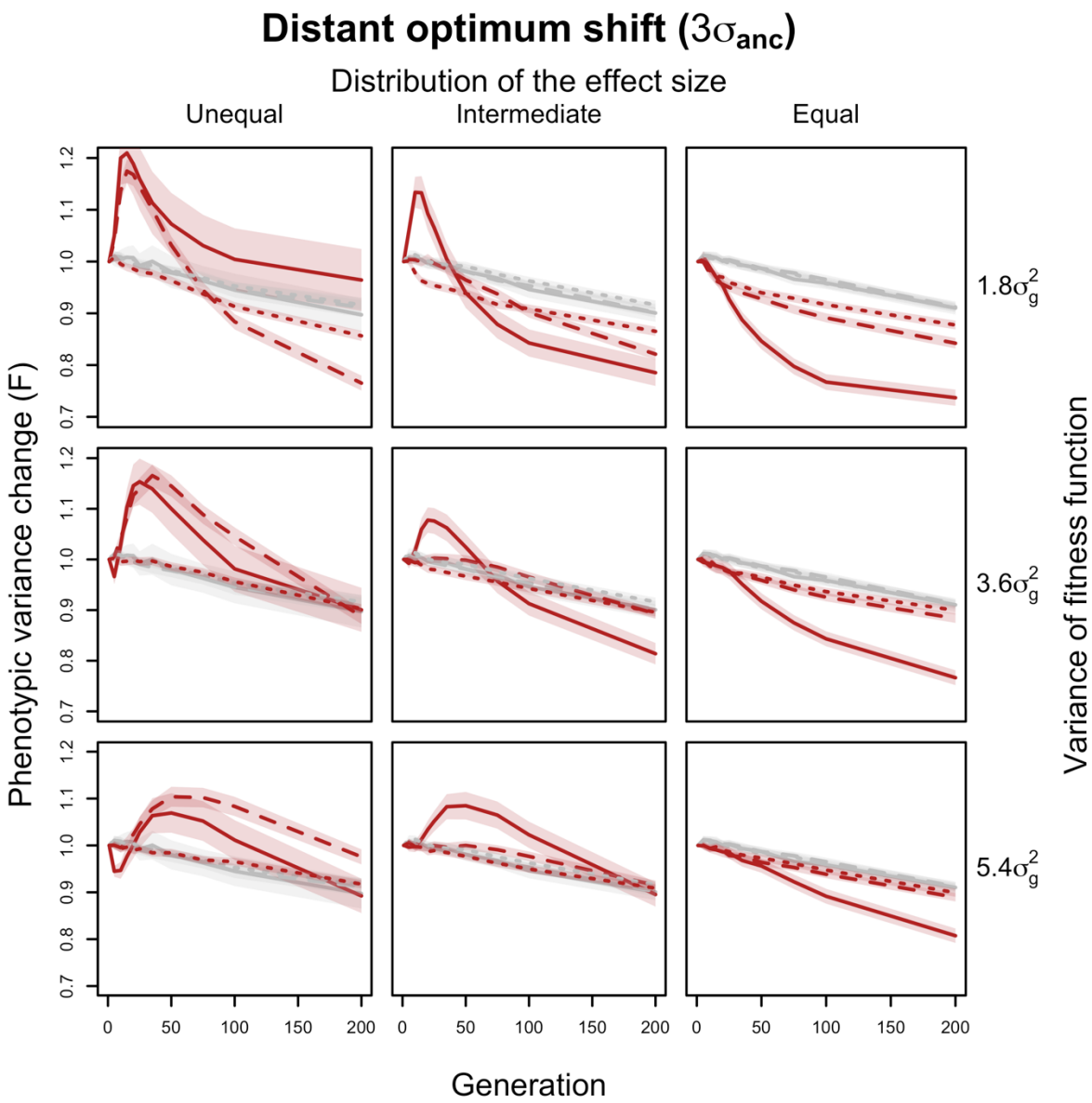


Supplementary Figure S4. The evolutionary scenario for distant optimum shift. We consider the case when a quantitative trait (in black) experiences a sudden shift in trait optimum under stabilizing selection. The imposed fitness functions (F.F.) are illustrated in red. The new trait optimum is set away from the ancestral trait mean by three standard deviation of the ancestral trait distribution for distal shift. To vary the strength of stabilizing selection, the variance of the fitness function is set as 1.8, 3.6 and 5.4 standard deviation of the ancestral trait distribution.



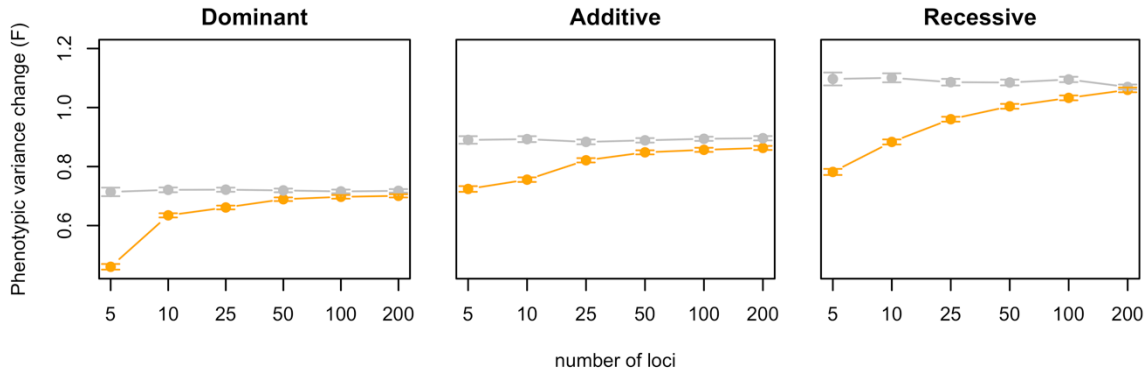
Supplementary Figure S5. The changes in phenotypic mean when adapting to optimum shift. The changes in phenotypic mean after 100 generations adapting to a mild/distant optimum shift (orange/red) are compared to the changes under neutrality (grey) on y axis. The changes in phenotypic mean are scaled by the standard deviation of the ancestral trait distribution. The simulations cover traits controlled by varying numbers of loci underlying the adaptation (x axes) with three different distributions of effect sizes (columns) and under different strength of stabilizing selection (rows). For each scenario, 1000 traits have been performed. In most cases, the traits under selection (orange/red) shift their means by one/three standard deviation of the ancestral trait distribution (i.e. reaching the new trait optimum) while

the neutral traits (grey) stay unchanged. The error bar indicates the 95% confidence interval for 1000 simulated traits.

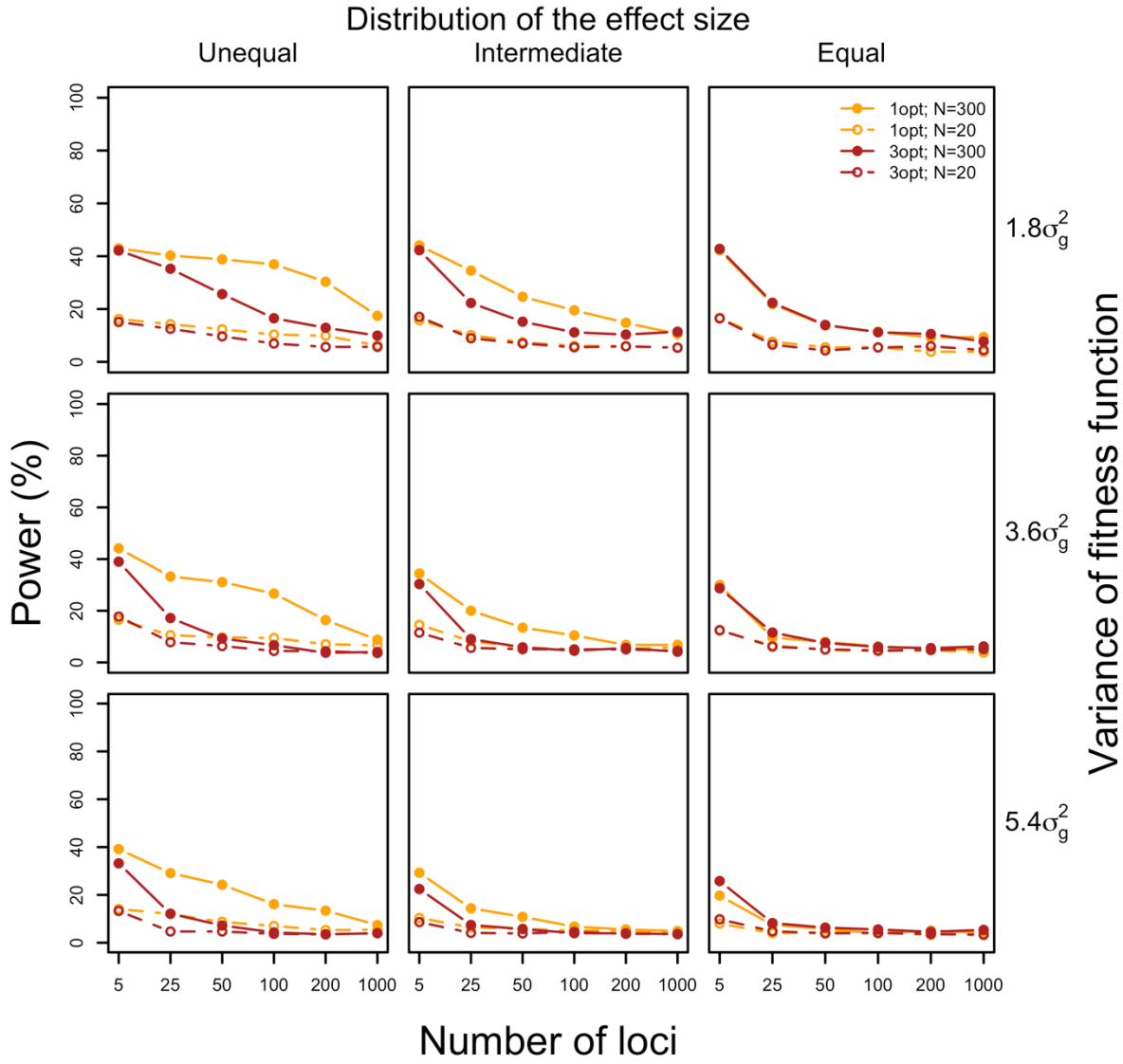


Supplementary Figure S6. The trajectory of expected changes in phenotypic variance when adapting to a distant optimum shift. The changes in phenotypic variance within 200 generations adapting to a distant optimum shift (red) are compared to the changes under neutrality (grey) on y axis. The average change in variance among 1000 traits (F) is calculated as the ratio of phenotypic variance between each evolved time point (generation x) and the ancestral state (σ_x^2/σ_1^2). The translucent band indicates the 95% confidence interval for 1000 simulated traits. The simulations cover traits controlled by varying numbers of loci underlying the adaptation with three different distributions of effect sizes (columns) and under different strength of stabilizing selection (rows). For each scenario, we simulated 1000 traits. Only traits with the most (dotted lines, 1000 loci), intermediate (dash lines, 50 loci) and the least (solid lines, 5 loci) polygenic architectures are shown. Unlike the continuous decreasing pattern in

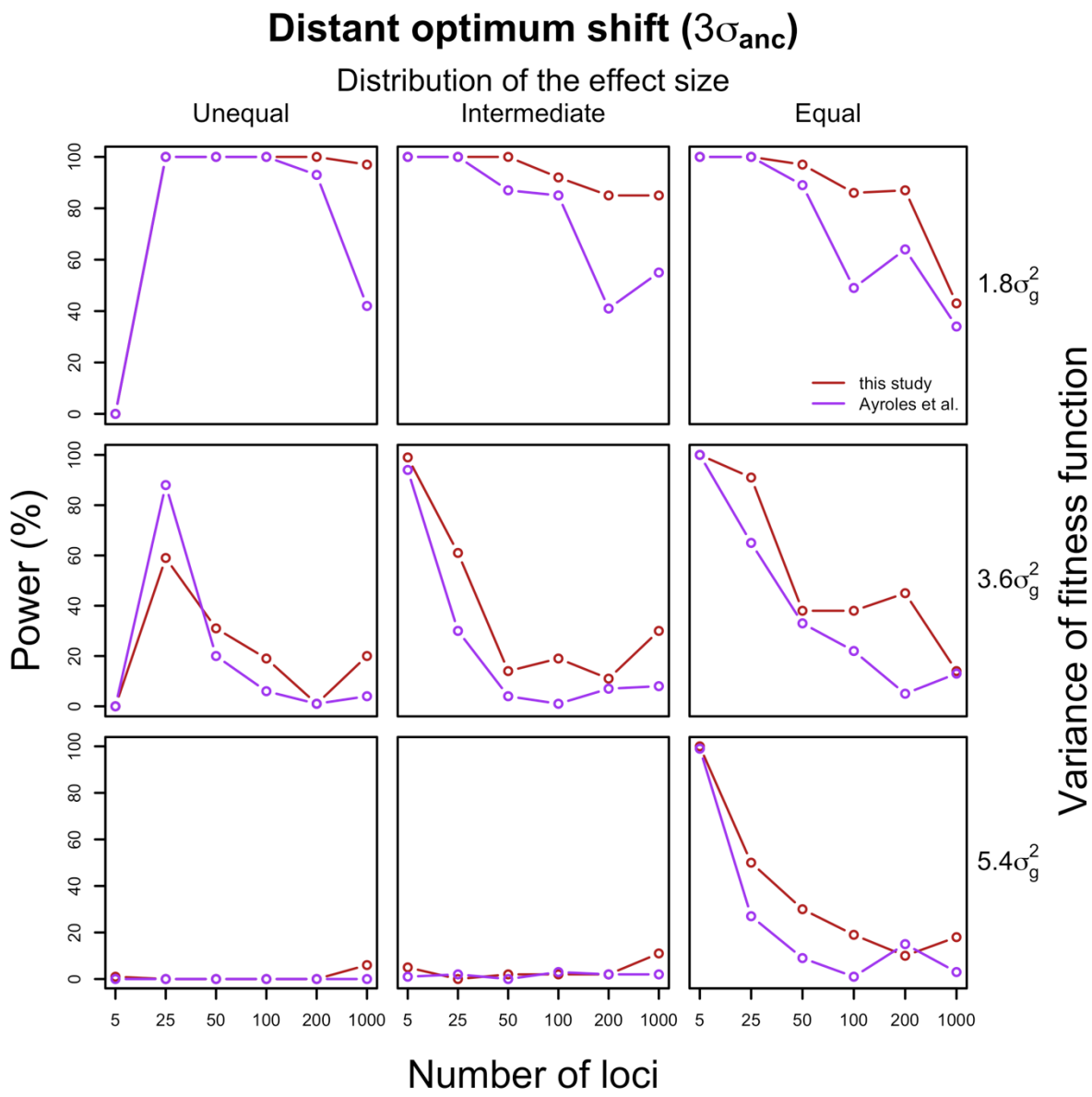
the cases with mild optimum shifts, the variance of traits controlled by a few loci (5 loci) with largely dispersed effects would increase first and then decrease when the effect sizes of contributing loci are dispersed (red solid lines). Nevertheless, for traits with extremely polygenic basis, the phenotypic variance always stays stable over time (red dotted lines).



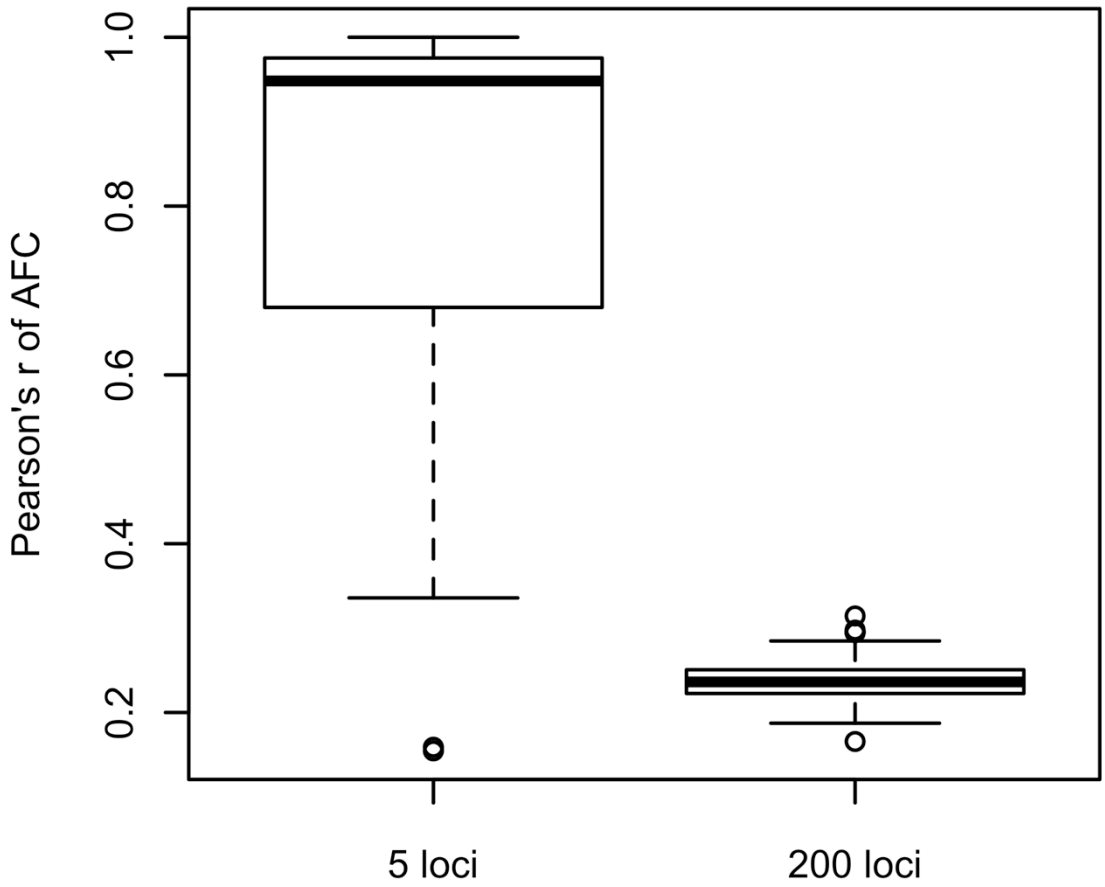
Supplementary Figure S7. The expected changes in phenotypic variance for traits controlled by dominant and recessive alleles. The changes in phenotypic variance after 100 generations adapting to a mild optimum shift (orange) are compared to the changes under neutrality (grey) on y axis. The change in variance (F) is calculated as the ratio between the evolved and ancestral phenotypic variance ($\sigma_{100}^2/\sigma_0^2$). This simulation covers traits controlled by varying numbers of loci underlying the adaptation (x axes) with recessive, additive and dominant effects. For each scenario, we simulated 1000 traits. No matter how the dominance varies, the variance of the trait decreases drastically when the adaptation is controlled by a small number of loci. As the number of contributing loci increases, the phenotypic variance becomes more stable. The error bar indicates the 95% confidence interval for 1000 simulated traits.



Supplementary Figure S8. Power for detecting significant variance changes for a single trait. With each genetic architecture (x-axis) under each selection scenario, we calculated how often we can detect a significant difference between the variance change of a single trait after 100 generations of selection and the neutral expectation (F of 0.9 according to the neutral simulations) (y-axis). The power is largely limited by the sample size. When all 300 individuals were considered, only 40% of power is observed under simple genetic control (5 loci). When the sampling processes of 20 individuals from the 300 simulated individuals was considered, there is nearly no power for the inference.



Supplementary Figure S9. Power for detecting significant different variance change between a group of selected and neutral traits under distant optimum shift. For each set of the 1000 traits controlled by different numbers of loci (x-axis) with varying effect sizes (columns) under each selection strength (rows), we calculated how often the variance changes of these 1000 selected traits differs from 1000 neutral traits after 100 generations (y-axis). The same genetic architecture is assumed for all 1000 selected traits with with a sample size of 20. In five out of the nine parameter combinations, we have nearly 100% power to detect a significant difference when all traits were under simple genetic control (5 loci). The power gradually decreases when the number of loci increased. Nevertheless, in the other parameter combinations, exception can be found for the traits controlled by a few loci (5 loci) with largely dispersed effects (see Discussion).



166

167 **Supplementary Figure S10. Parallelism in genomic response across 10 evolved replicates**
168 **for traits under the control of five loci and 100 loci.** For the loci with allele frequency change
169 of at least 10% in 100 generations, the average Pearson's correlation coefficient of the
170 frequency change between all pairs of two evolved replicates was calculated to describe the
171 parallelism of the evolution at these loci. An average across loci is used to obtain a general
172 parallelism. 100 traits with 10 evolution replicates have been performed for each scenario. With
173 five contributing loci, the genomic evolution of the contributing loci is more parallel across the
174 10 evolved replicates compared to the case when with 100 contributing loci.
175

Titles and legend for supplementary files

Supplementary file 1. Library information of the sample in this study. This file provides a list of all sequenced samples and the library information.

References

Ayroles J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, *et al.*, 2009 Systems genetics of complex traits in Drosophila melanogaster. Nat. Genet. 41: 299–307. https://doi.org/10.1038/ng.332

Rocke D. M., and B. Durbin, 2003 Approximate variance-stabilizing transformations for gene-expression microarray data. Bioinformatics 19: 966–72.