Pleiotropy increases parallel selection signatures during adaptation from standing genetic variation

Wei-Yun Lai[1,2], Sheng-Kai Hsu[1,2], Andreas Futschik[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]Department of Applied Statistics, Johannes Kepler University Linz, Linz, Austria.
[*]Correspondence: christian.schloetterer@vetmeduni.ac.at; Tel.: +43-1-25077-4300.

Abstract

The phenomenon of parallel evolution, whereby similar genomic and phenotypic changes occur across replicated pairs of population or species, is widely studied. Nevertheless, the determining factors of parallel evolution remain poorly understood. Theoretical studies have proposed that pleiotropy, the influence of a single gene on multiple traits, is an important factor. In order to gain a deeper insight into the role of pleiotropy for parallel evolution from standing genetic variation, we characterized the interplay between parallelism, polymorphism and pleiotropy. The present study examined the parallel gene expression evolution in 10 replicated populations of *Drosophila simulans*, which were adapted from standing variation to the same new temperature regime. The data demonstrate that parallel evolution of gene expression from standing genetic variation is positively correlated with the strength of pleiotropic effects. The ancestral variation in gene expression is, however, negatively correlated with parallelism. Given that pleiotropy is also negatively correlated with gene expression variation, we conducted a causal analysis to distinguish cause and correlation and evaluate the role of pleiotropy. The causal analysis indicated that both direct (causative) and indirect (correlational) effects of pleiotropy contribute to parallel evolution. The indirect effect is mediated by historic selective constraint in response to pleiotropy. This results in parallel selection responses due to the reduced standing variation of pleiotropic genes. The direct effect of pleiotropy is likely to reflect a genetic correlation among adaptive traits, which in turn gives rise to synergistic effects and higher parallelism.

Background

Distinguishing selection from stochastic changes continues to be a major challenge in evolutionary research. The evidence for selection mostly comes from parallel evolution, where similar genes, traits or functions are identified in replicate populations exposed to similar environments [1-4]. Given that non-parallel evolution does not preclude the possibility of selection [5, 6], there has been a sustained interest in elucidating the factors that influence the extent of genetic parallelism (i.e., how similar is the response of the same gene to selection in multiple populations from the same/similar ecological niche(s)) [7].

It has been proposed that the degree of pleiotropy, the number of traits affected by a single gene, may be a potential factor associated with parallelism [7-10]. The majority of hypotheses regarding the influence of pleiotropy on parallel selection responses are (implicitly) related to the ideas of Fisher [11], later modified by Orr [12], focusing on the cost of pleiotropy. This concept posits that any new mutation in a pleiotropic gene results in a shift in trait values away from their optimum across multiple traits. New mutations in highly pleiotropic genes are more likely to have negative fitness effects than mutations in less pleiotropic genes as more traits are displaced from their optima. This "cost of pleiotropy" model [12] is evidenced by the lack of sequence divergence observed in genes with a higher level of pleiotropy [13-16]. Consequently, it can be postulated that less pleiotropic genes are more likely to contribute to adaptation over long evolutionary timescales, given that they have a higher probability of accumulating beneficial genetic variants. This results in a more parallel selection response in different populations/species [17].

Fisher's model of pleiotropy is simplistic in its assumption of uncorrelated traits and uncorrelated selection. It has been proposed that the influence of pleiotropy on the selection response strongly depends on the genetic correlation of the affected traits and the underlying fitness function which connects phenotypic values with fitness [18-20]. If the fitness effects of the affected traits are positively correlated (synergistic pleiotropy), this may result in stronger net selection [21] and consequently in more parallel selection signatures [22]. This implies that the correlation of fitness effects determines whether pleiotropy increases or decreases parallel selection signatures.

The impact of pleiotropy on parallel evolution has typically been considered in the context of long-term adaptation via de novo mutations. However, many studies of parallel evolution have

focused on short-term adaptation [e. g. 23, 24-30], where standing genetic variation plays a more significant role for selection responses than de novo mutations. Nevertheless, it remains unclear whether a similar impact of pleiotropy on evolutionary parallelism is to be expected for standing genetic variation.

Indeed, the empirical evidence regarding the role of pleiotropy for intra-specific adaptive responses is somewhat inconclusive. A gene expression study in European grayling populations adapted to different temperature regimes provided compelling evidence for the constraining effect of pleiotropy [31], which supports the cost of complexity concept. A comparison of gene expression changes in natural *Drosophila* populations along a temperature cline and populations evolved in the laboratory to different temperature regimes demonstrated the importance of correlated fitness effects [21]. A direct test of the influence of pleiotropy on parallel evolution was conducted on 19 independent contrasts of stickleback populations [9]. The authors discovered that pleiotropic genes were more likely to exhibit parallel adaptation signatures [9], thereby supporting the importance of synergistic pleiotropy.

Previous studies on pleiotropy and parallelism did not account for polymorphism (standing genetic variation), which plays a central role for selection responses during short-term adaptation. Although natural variation provides the basis for adaptive responses, it is not well-documented how ancestral variation (polymorphism) determines the level of genetic parallelism during short-term adaptation in replicate populations. Moreover, polymorphism of a gene is also affected by its pleiotropic level while the expectations regarding the influence of pleiotropy on the extent of polymorphism varies. The cost of complexity predicts that pleiotropic genes with uncorrelated fitness effects will be subject to stronger purifying selection, which will result in reduced polymorphism. In the case of multiple selected traits with aligned fitness effect, synergistic pleiotropy may facilitate the occurrence of selective sweeps and thereby reducing variation in pleiotropic genes. Alternatively, it may be possible that once pleiotropic alleles have been established in a population, their multivariate fitness effects lead to a balanced polymorphism, which in turn maintains higher levels of variation. Empirical studies have observed a negative correlation between the level of pleiotropy and sequence/expression polymorphism [32, 33]. This suggests that pleiotropy may preferentially reduce variation, either by purifying selection, synergistic effects, or a combination of both.

102 In this study we employed gene expression analysis to investigate the interplay between
103 polymorphism and pleiotropy on the degree of parallelism of adaptive responses. We assessed
104 the selective response by measuring gene expression differences between populations from
105 different environments, a widely employed approach [e. g. 34, 35-40]. Specifically, we
106 compared gene expression in the ancestral *D. simulans* population with that of flies that had
107 evolved for 100 generations in a novel hot temperature regime. To infer parallel evolution, the
108 gene expression change was measured in 10 replicate populations. By focusing on the
109 evolutionary response, we circumvent the challenge of connecting the phenotype to an
110 unknown fitness landscape. Furthermore, we take advantage of well-established pleiotropy
111 estimates [41-44] to evaluate the influence of pleiotropy on parallel gene expression evolution.
112 The natural variation in gene expression was determined by examining the differences in
113 expression levels among individual flies from the ancestral *D. simulans* population.
114
115 We observed significant interplays between pleiotropy, ancestral variation, and parallel
116 selection responses among genes. The parallelism of gene expression evolution is positively
117 associated with pleiotropy but negatively associated with the ancestral variation in gene
118 expression. Given that pleiotropy was also negatively associated with ancestral variation, it is
119 crucial to differentiate between statistical causality and correlation. Causal analysis revealed
120 that pleiotropy affects parallel evolution both directly and indirectly via ancestral variation. It
121 is likely that the indirect effects operate via the reduction of ancestral variation in pleiotropic
122 genes. The direct effects of pleiotropy on parallelism are probably best explained by synergistic
123 pleiotropy which results in a stronger selection response.
124

125 **Methods**

126 Estimating gene pleiotropy

127 We approximated the pleiotropy of each gene with two alternative estimators, network
128 connectivity and tissue specificity (τ).

129

130 τ indicates how specific the expression of a gene is across different tissues. τ was estimated for
131 each gene using the adult male expression profiles on flyatlas2 [45] as: $\tau =$
132 $\frac{\sum_i [1-\frac{gene\ expression_i}{gene\ expression_{max}}]}{N-1}$, where *N* is the number of tissues examined and *i* indicates each of them
133 [46]. If a gene is only expressed in one tissue, τ will equal to 1 while it equals 0 when a gene
134 is expressed at the same level across all tissues. The relationship between τ and pleiotropy is
135 based on the idea that the genes expressed in many tissues are more likely to affect multiple
136 traits than genes expressed in fewer tissues [41, 44, 47, 48]. Hence, we used 1-τ to indicate the
137 pleiotropic effect of a gene.

138

139 Network connectivity has also been used as proxy for pleiotropy, as more connected genes are
140 likely to be more pleiotropic [16, 42, 43, 49]. We used published information about the
141 transcriptional regulatory network of *Drosophila*, which was inferred from several sources,
142 including genome-wide chromatin immuno-precipitation, conserved transcription factor
143 binding motifs, gene expression profiles across different development stages, and chromatin
144 modification profiles among several cell types [50]. The connectivity for each gene was
145 estimated as the sum of adjacencies between the focal gene and other genes in the network.

146

147 The significant positive correlation (rho=0.54) between 1-τ and network connectivity suggests
148 that both estimates capture a similar, but not identical, signature of the pleiotropic properties
149 among the genes (Supplementary figure 1).

150

151 We note that approximating pleiotropic effects in *D. simulans* by using data-rich estimates
152 from the close relative *Drosophila melanogaster* could potentially compromise the
153 interpretability of our results. However, comparative analyses of tissue-specific gene
154 expression have shown an extremely high consistency across species in animals [51] and plants
155 [52, 53]. Hence, these data suggest that at least for tissue specificity, one of the pleiotropy
156 measures used, estimates from one species can be safely used for close relatives.

157

158 Experimental evolution and common garden experiment

159 The procedures of the evolution and common garden experiments were described in [5, 35,
160 54]. Briefly, 202 isofemale lines collected from Florida, USA were used to constitute 10
161 outbred *Drosophila simulans* populations, which have been exposed for more than 100
162 generations to a laboratory experiment at 28/18 °C with 12hr light/12hr dark photoperiod. The
163 census population size of each replicate population is 1000 to 1250 adult flies.

164 Each founder isofemale line was maintained at a small population size (typically less than 50
165 individuals) at 18 °C on standard laboratory food. Adaptation to the laboratory environment
166 with the residual heterogeneity or de novo mutation is unlikely due to the small effective
167 population size during the maintenance of each line [5]. This is supported by the observation
168 that 50 generations of maintenance as isofemale lines does not result in significant genetic
169 differentiation from the ancestral population [55]. Furthermore, mutations that occur in the
170 laboratory are mostly recessive [56], and the effects are likely to be masked because after two
171 generations of random mating during common garden maintenance, most individuals will be
172 heterozygous for isofemale line-specific variants.

173 The collection of samples for RNA-Seq was preceded by two generations of common garden
174 experiment (CGE). As described previously [5, 35, 54], at generation 103 of the evolution
175 experiment, five biological replicates of reconstituted ancestral populations were generated by
176 pooling five mated females from 184 founder isofemale lines [57] and reared together with 10
177 independently evolved populations for two generations with controlled egg density (400
178 eggs/bottle) at the same temperature regime as in the evolution experiment. Three biological
179 replicates were included for each independently evolved population (each biological replicate
180 was cultivated separately for the two generations of common garden). From each biological
181 replicate, a sample of 50 five-day-old adult males were collected (Figure 1a).

182

183 Quantifying the evolutionary parallelism in gene expression across ten replicate populations

184 The raw RNA-Seq count table of five replicates of the reconstituted ancestral population and
185 10 independently evolved populations, each with three biological samples, was taken from
186 (Jaksic et al., 2020). The raw read counts of each gene were normalized with the TMM method
187 implemented in edgeR [58]. Only genes with at least 0.1 normalized counts per million base
188 (CPM) across all samples were considered for further analysis. Because we are interested in
189 the evolutionary response in each independently evolved population, we contrasted the three
190 biological samples from each evolved population to the five biological samples from the

ancestral population. The differential expression (DE) analysis was done separately in each of the ten evolved populations. For DE analysis, we utilized negative binomial generalized linear modeling implemented in edgeR to fit the expression to the model $Y = E + \varepsilon$, in which Y stands for gene expression, E is the effect of evolution and ε is the random error. Likelihood ratio tests were performed to test the effect of evolution. P-values was adjusted using Benjamini-Hochberg false discovery rate (FDR) correction [59]. After we identified DE genes in each evolved population, we constructed a replicate-frequency spectrum (RFS; [5]). We considered genes with significant changes in the same direction in at least three evolved populations as putatively adaptive genes.

To quantify the degree of parallel gene expression evolution, we calculated the expression changes in response to evolution ($\log_2$FC; fold change in expression between ancestral and evolved sample; see below) of each putatively adaptive gene (n=1,775) in each evolved sample.

$$log_2FC = \log\left(\frac{y_{evo}}{\overline{y_{anc}}}\right), \quad (1)$$

where y_{evo} is the expression value (CPM) per gene per evolved sample, and $\overline{y_{anc}}$ is the overall mean expression value (CPM) per gene across five ancestral samples.

We calculated the ratio of between-evolution population variation (MS_{pop}) and residual (MS_e) among the expression changes ($\log_2$FC), which is denoted as F.

$$F = \frac{MS_{pop}}{MS_e} \quad (2)$$

$$MS_{pop} = \frac{\sum_{i=1}^{10}(\bar{X}_i - \bar{X})^2}{10 - 1} \quad (3)$$

$$MS_e = \frac{\sum_{i=1}^{10}\sum_{j=1}^{3}(X_{ij} - \bar{X}_i)^2}{30 - 10} \quad (4)$$

Where X_{ij} represent the $\log_2$FC value for the jth biological sample in the ith evolution population. i=1, 2, …,10, j=1, 2, 3.

As F reflects the heterogeneity of the evolutionary response for a gene across ten evolutionary populations, we used the reciprocal of F (i.e., 1/F) to quantify the degree of parallelism in gene expression evolution.

<u>Estimating gene expression variation in an outbred ancestral population</u>

223 The variation in expression of each gene within an outbred ancestral population was estimated
224 from 20 adult males sampled from an outbred ancestral population [60]. We distinguished
225 biological variation and measurement error of each gene using the statistical method
226 implemented in edgeR [58], where the biological variance across individual (biological
227 coefficient of variation; BCV; defined in [58]) of each gene can be estimated (tag-wised
228 dispersion). BCV^2 was used to represent the expression variance relative to the mean of each
229 gene within the population. To illustrate that the gene expression variation is not explained by
230 random measurement error, we took the data from two ancestral population replicates, which
231 are considered to be genetically identical and calculated BCV^2 of expression levels across
232 individuals for each population replicates separately. We found that both mean and variance in
233 expression for each gene was similar in two ancestral population replicates (mean: rho>0.99;
234 variance: rho=0.8, p-value<2.2e-16), suggesting that sequencing noise does not mask the
235 biological variability.
236

Causal analysis

To infer the causal relationship between three factors: pleiotropy (Pl), ancestral variation (A) and parallelism (Pa), we applied causal analysis. The causal analysis is built upon a previously published statistical framework [61]. We consider five possible causal relationships between pleiotropic effects, ancestral variation and parallelism (Figure 5). Our goal is to determine which of the five models is best supported by the data. The selection is based on Bayesian information criteria (BIC) which are calculated from maximum likelihood of the respective models. The likelihood of each model is represented as:

$$\mathcal{L}(\text{Model I}) = p(Pa|A)p(A|Pl)p(Pl) \tag{5}$$

$$\mathcal{L}(\text{Model II}) = p(Pa|Pl)p(A|Pl)p(Pl) \tag{6}$$

$$\mathcal{L}(\text{Model III}) = p(Pa|A, Pl)p(A|Pl)p(Pl) \tag{7}$$

$$\mathcal{L}(\text{Model IV}) = p(Pa|A, Pl)p(A)p(Pl) \tag{8}$$

$$\mathcal{L}(\text{Model V}) = p(Pa)p(A)p(Pl) \tag{9}$$

p(Pl) is the probability distribution of the pleiotropic effect. To utilize normal probability density function in the derivation of the likelihoods for each joint probability distributions, we transformed the random variables Pa and A to be normally distributed (natural log transformation on 1/F and BCV^2). The exact forms of these likelihoods are given in supplementary information.

For each model, we maximized the likelihood and estimated the parameters using standard maximum likelihood methods. Based on the maximum likelihood, we computed the BICs for each model; the model with the smallest BIC is the one best support by the data.

BIC is calculated as:

$$\text{BIC} = -2\ln(\mathcal{L}_i) + k_i ln(n) \tag{8}$$

Where $\mathcal{L}_i$ is the likelihood for model I-V, k_i is the corresponding number of free parameters and n is total number of putatively selected genes. The analysis was performed independently on two measures of pleiotropy (tissue specificity and network connectivity).

Estimating the sizes of direct and indirect pleiotropic effects on parallelism

We used path analysis to estimate the effect sizes of direct and indirect pleiotropic effects on parallelism [62]. We standardized all three transformed variables (Parallelism (Pa), Pleiotropy (Pl), Ancestral variation (A); see supplementary information) by subtracting the mean and dividing by the standard deviation for the following analysis.

First, we fit the regression model for all three variables across all putatively adaptive genes:

$$y_{Pa} = \beta_1 x_A + \beta_2 x_{Pl} + \varepsilon \tag{9}$$

where y_{Pa} stands for parallelism, x_A is ancestral variation, x_{Pl} is pleiotropy measure and ε is random error. β_1 and β_2 are the regression coefficients corresponding to x_A and x_{Pl}, respectively.

Second, we fit another regression model for ancestral variation and pleiotropy measure across all putatively adaptive genes:

$$y_A = \beta_3 x_{Pl} + \varepsilon \tag{10}$$

where y_A stands for ancestral variation, x_{Pl} is pleiotropy measure and ε is random error. β_3 is the regression coefficient corresponding to x_{pl}.

The sizes of the direct effect of pleiotropy on parallelism will be β_2. The size of the indirect effect of pleiotropy on parallelism via ancestral variation will then be the product of $\beta_1 \times \beta_3$.

The analysis was performed independently on two different measures of pleiotropy.

Computer simulations

In our interpretation of the indirect effect of pleiotropy on parallelism, we assumed that ancestral variation in gene expression is negatively correlated with its parallel evolution. Because this has not been previously demonstrated, we used computer simulations to illustrate how the level of standing genetic variation impacts the parallelism of adaptive responses after a shift in trait optimum. For simplicity, we considered four genes equally contributing to a fitness-associated trait. The phenotype (expression level) of each gene is controlled by a different number of genetic loci (5, 15, 30 and 50) with equal effect. The underlying assumption is that these four genes exhibit different levels of pleiotropy and thus experience varying strength of purifying selection, retaining different levels of genetic variation. Using mimicrEE2

[63], we evolved 10 replicate populations (N=300) derived from the same set of 189 natural *Drosophila* haplotypes [5] in same selection regime with a shift in trait optimum of one standard deviation relative to the ancestral phenotypic distribution. After 100 generations, we determined the evolutionary parallelism of the four genes using the same approach as for the empirical data. Each simulation was repeated 100 times and one out of four genes was randomly picked from each simulation run to generate Figure 4.

Results

The parallelism of expression evolution differs among genes

We investigated the degree of parallelism in gene expression evolution, using an RNA-Seq dataset from an experimental evolution study in which 10 replicated populations with the same genetic variation were exposed to the same environment for more than 100 generations [29, 54, 60] (Figure 1a). While previous studies have focused on the shared parallel response across replicated populations [29, 54, 60], we re-analyzed the RNA-Seq data to quantify the degree of evolutionary parallelism across different genes. We identified 1,775 putatively selected genes with a significant change in gene expression in the same direction in response to the novel environment in at least three populations. This conditioning on multiple replicate populations with a significant change in the same direction reduces false positives due to drift [9, 64]. Relaxing the threshold of responding populations provides qualitatively similar results.

Barghi et al. (2019) introduced the replicate frequency spectrum to describe the parallel genomic selection response in replicate populations. The replicate frequency spectrum of DE genes shows considerable heterogeneity among populations, with many genes displaying significant gene expression changes only in a few populations (Figure 1b). The replicate frequency spectrum is strongly influenced by the significance threshold used and provides an incomplete quantitative measure of parallelism. To obtain a more quantitative measure of parallelism, we evaluated the between-population variation in the evolutionary response ($\log_2$FC) of each putatively adaptive gene (n=1,775) while controlling for measurement error (F). The parallelism of evolution for each gene is then quantified as 1/F (Figure 1b; see Materials and Methods). A higher value of 1/F indicates a more parallel evolutionary response among replicated populations (Supplementary figure 2). Importantly, the measure of parallelism varies between genes (Figure 1c) and is positively correlated with the number of populations in which a significant change in gene expression is observed for each gene (Supplementary figure 3).

Pleiotropy is positively associated with the degree of parallelism in gene expression evolution

Pleiotropy is one of the factors that influence the degree of parallel evolution [7-10]. Therefore, we tested whether and how the pleiotropic effect of a gene influences the degree of parallelism of its expression evolution in a new environment. We correlated two different measures of pleiotropy (1-τ (tissue specificity) and network connectivity; see materials and methods) of

349 each putatively adaptive gene (n=1,775) with the parallelism of its evolutionary responses
350 (1/F). We observed a significant positive correlation between the parallelism and the degree of
351 pleiotropy, (rho=0.26, p-value<2.2e-16 for 1-τ and rho=0.11, p-value<5.6e-07 for network
352 connectivity) (Figure 2 and Supplementary figure 4a). This suggests that pleiotropy may
353 enhance parallel changes in gene expression during evolution.
354
355 ***Pleiotropy constrains the ancestral (natural) variation in gene expression***
356 The negative effect of pleiotropy on genetic variation over longer time scales, mostly between
357 species, is well documented [13-16], but it not yet clear whether the same effect is seen for
358 adaptation from standing genetic variation. We reconstituted the ancestral population from 189
359 isofemale lines [60] and estimated the expression variance (BCV^2; see materials and methods)
360 of each putatively adaptive gene (n=1,775) in two replicate sets of 20 individuals. Since the
361 variance in gene expression is positively correlated with genetic variation in cis-regulatory
362 regions [65] and the heritability of gene expression is high [66], we assumed that the observed
363 variation in gene expression has a genetic basis, rather than being noise (see discussion). We
364 observed a highly significant negative correlation between the level of pleiotropy and the
365 magnitude of ancestral variation in gene expression (Figure 3a and Supplementary figure 4b;
366 rho=-0.34, p-value<2.2e-16 for 1-τ and rho=-0.37, p-value<2.2e-16 for network connectivity).
367 To exclude that this correlation was driven by our filtering for significant expression changes
368 in at least three populations, we repeated these analyses with all genes (n=9,882) and also
369 observed a significant correlation (rho=-0.45, p-value<2.2e-16 for 1-τ and rho=-0.38, p-
370 value<2.2e-16 for network connectivity, data not shown). Hence, we suggest that the negative
371 relationship between pleiotropy and polymorphism/diversity [32, 33] (i.e. reduced variants,
372 lower frequencies and/or lower allelic effects) can be extended to gene expression.
373
374 ***Evolutionary parallelism is negatively associated with ancestral variation***
375 For a better understanding of the factors determining parallel evolution, we examined the
376 influence of ancestral variation in gene expression on the parallel evolution of gene expression.
377 We observed a significant negative correlation between ancestral variation and parallelism of
378 the evolutionary response (rho=-0.22, p-value<2.2e-16; Figure 3b). We suggest that this pattern
379 could be explained by genetic redundancy [5, 67], which describes the phenomenon that a
380 polygenic trait can shift its phenotypic value through different combinations of contributing
381 loci: When the diversity of the contributing loci is reduced (either through fewer loci, lower

382 ancestral frequencies or smaller variation in effect size), fewer combinations of the contributing
383 loci are available for adaptation, resulting in more parallel adaptive responses. To confirm this,
384 we performed computer simulations of replicated polygenic adaptation and found that lower
385 ancestral variation (i.e. fewer contributing loci) leads to more parallel trait evolution (Figure
386 4). Hence, that natural variation in the ancestral population could be one of factor determining
387 the parallel adaptive response in replicate populations. Although we focused on the phenotypic
388 response to match gene expression, it is interesting to note that a similar result was obtained
389 for parallelism at the genetic level with different levels of standing genetic variation [68]. If
390 pleiotropy reduces variation, the positive correlation between pleiotropy and parallelism may
391 be an indirect (correlative) effect, rather than a direct consequence of pleiotropy.
392
393 ***Testing causality***
394 Our previous analyses indicated that parallelism is affected by pleiotropy and ancestral
395 variation. Because pleiotropy is negatively correlated with ancestral variation for both
396 pleiotropy measures (Figure 3a and supplementary figure 4b), it is not clear whether pleiotropy
397 affects parallelism directly or indirectly through the negative correlation with ancestral
398 variation. We used causal analysis, which based on a Bayesian statistical framework [61], to
399 disentangle the causal relationship between the three different parameters. Since the
400 directionality of pleiotropic effects on ancestral variation has been previously demonstrated,
401 and ancestral variation is unlikely to influence the pleiotropy of a gene, we considered five
402 possible causal relationships between pleiotropic effects, ancestral variation, and parallelism
403 (Figure 5). In the first model, parallelism is determined by ancestral variation, which is shaped
404 by pleiotropic, but pleiotropy has no direct effect on parallelism. In model II, the pleiotropic
405 effect affects ancestral variation and parallelism independently, but ancestral variation has no
406 causal effect on parallelism. In model III, pleiotropy affects both ancestral variation and
407 parallelism, and ancestral variation also determines parallelism. In model IV, parallelism is
408 determined by ancestral variation and pleiotropy, but pleiotropic effects are assumed to have
409 no effect on the ancestral variation. This serves as a null model to confirm the indirect effect
410 of pleiotropy on parallelism via ancestral variation. We also add a global null model with no
411 correlation between these three factors (model V, Figure 5). The analysis was performed
412 independently for two measures of pleiotropy (tissue specificity and network connectivity).
413
414 For tissue specificity, our data were better explained by model III (i.e., model III has the lowest
415 BIC, Table 1), suggesting that both direct and indirect effects of pleiotropy determine the

degree of parallelism of adaptive gene expression evolution. On the other hand, when pleiotropy was estimated by network connectivity, model I had a slightly lower BIC than model III (Table 1). Thus, our causal analysis does not provide evidence for a direct effect of network connectivity. We further quantified the strength of the direct and indirect effects of pleiotropy on parallelism using a path analysis (see materials and methods). The path analysis confirmed the significant direct effect of tissue specificity, but network connectivity has no statistically significant direct effect on the parallelism of gene expression evolution (Table 2 and Table S1). Although the relevance of direct effects of pleiotropy on parallel gene expression evolution differs for the two pleiotropy estimates, the similarity of the indirect effect sizes for both pleiotropy measures is striking (Table 2). Taken together, these results suggest that pleiotropy may enhance the parallel evolutionary response of gene expression directly and indirectly through its influence on ancestral variation.

429 **Discussion**

430 This study examined the influence of pleiotropy on the parallelism of the adaptive response
431 from standing genetic variation. By examining the role of natural variation, which is important
432 for parallelism of short-term evolution, we extended the scope of previous studies. Because
433 differences in genetic background and selection strength among replicated populations cannot
434 be controlled in natural populations and may affect the degree of parallelism, we relied on
435 experimental evolution for our study. We took advantage of ten replicated populations that
436 have evolved from the same genetic background in the same hot environment for more than
437 100 generations [5].

438

439 The adaptive response during adaptation to the new hot temperature regime was measured by
440 changes in gene expression. This approach has the advantage that pleiotropic effects and their
441 associated fitness function, which are typically not known, are already included in the adaptive
442 response. Furthermore, we were able to rely on two well-established measures of pleiotropy to
443 study the interplay with parallelism and natural variation.

444

445 A key finding of our study is that that adaptation from standing genetic variation results in
446 complex interplays between pleiotropy, ancestral variation, and parallelism. Of particular
447 interest is the observed negative correlation between pleiotropy and ancestral variation in gene
448 expression, which confirms previous observations [32, 33]. This suggests that the primary
449 effect of pleiotropy is not to maintain variation. Rather, stronger purifying selection removes
450 pleiotropic variants from the population, resulting in a reduced standing genetic variation. The
451 reduced variation in more pleiotropic genes could reflect historical hard sweeps or purifying
452 selection. A similar pattern is seen over longer evolutionary timescales: less sequence
453 divergence in pleiotropic genes [13-16].

454

455 An important difference between adaptation from de novo mutations and standing genetic
456 variation concerns the implications for parallel evolution. While pleiotropic genes are less
457 likely to contribute to parallel evolution from de novo mutations [7], our results show that
458 adaptation from standing genetic variation favors parallel selection responses of pleiotropic
459 genes (Figure 2, Supplementary Figure 4a). It is important to note that many previous studies
460 focused on sequence variation, whereas we used expression variation. While it is possible that
461 these two types of variation behave differently, we do not believe that this is the case as

462 sequence polymorphism and expression variation are correlated [69]. Furthermore, a recent
463 study focusing on genomic changes [9] reached similar conclusions.
464
465 A very important assumption in our interpretation is that differences in gene expression
466 variation in the ancestral population reflect regulatory sequence polymorphism rather than
467 noise. The high correlation in the variance of gene expression between two replicate groups of
468 individuals indicates that our measures do not reflect technical noise but have a biological
469 basis. However, stochastic fluctuations in gene expression have also been observed among
470 cells in single-celled and multicellular organisms and the variance in gene expression differs
471 among genes [70-72]. In this case, the differences in gene expression variance are biological
472 in nature, but not associated with sequence variation. We do not believe that variation in the
473 control of gene expression levels among genes can explain our pattern because *Drosophila* is
474 a multicellular organism and we used whole body RNA-Seq data for our analysis. Therefore,
475 even if expression is controlled to different extents among genes, by averaging across multiple
476 cells, this does not affect our results. The only scenario we can imagine in which gene
477 expression noise could propagate across the many cells of a multicellular organism is that
478 during early development, differences in gene expression of a key regulator could affect
479 downstream gene expression throughout the entire organism. However, we do not think this is
480 a likely scenario, especially in the light of previous results showing that genetic variation in
481 cis-regulatory regions correlates with expression variation [65]. Furthermore, whole-body gene
482 expression in *Drosophila* is highly heritable, suggesting that gene expression is not stochastic
483 [66]. Nevertheless, we would like to point out that our observations are valid irrespective of
484 whether the correlation between parallelism and expression variation is the result of genetic
485 polymorphism or genetic control of expression noise.
486
487 Based on the significant correlation results, the question arose as to whether the positive
488 correlation between pleiotropy and parallelism, is causal (direct effect) or merely correlational
489 (indirect effect). Regardless of the measure of pleiotropy, our causal analysis showed a
490 consistent indirect effect of pleiotropy on parallelism (Table 1, 2). The estimated indirect
491 effects were very similar for both measures of pleiotropy (Table 2). This indirect effect of
492 pleiotropy could be explained by the reduced variation in the expression of more pleiotropic
493 genes in the ancestral population leading to a more parallel selection response of gene
494 expression. For the direct effect, however, the results differed between the two measures of
495 pleiotropy. Only tissue specificity had a significant direct effect, which was even larger than

the indirect effect (Table 2). No significant direct effect was found for network connectivitiy. The discrepancy between the two measures of pleiotropy is particularly interesting given their significant correlation (Supplementary Figure 1). This suggests that both measures capture aspects of pleiotropy that differ in their biological implications. The direct effects of pleiotropy, most likely arise from the positively correlated fitness effects of non-focal traits (synergistic pleiotropy). The synergistic pleiotropic effects result in a stronger fitness advantage, which translates into a more parallel selection response - as previously shown [22]. Given that temperature is one of the most important abiotic factors, it is expected that positive genetic correlations between temperature-adaptive traits have been selected over long evolutionary timescales, probably prior to the speciation of *D. simulans*. Thus, it is conceivable that positive pleiotropy for temperature adaptation related traits is common and thus determined the positive contribution of pleiotropy to the parallel evolution observed here.

In a concluding remark, we would like to emphasize that although pleiotropy and ancestral variation were found to be significantly associated with the degree of parallelism of gene expression evolution, it is likely that additional factors also contribute. For example, diminishing return epistasis, population size (drift), and the contribution of gene expression to the selected traits may also affect the parallelism of gene evolution. Further investigation of other potential factors will continue to advance our understanding of parallelism and the predictability of evolution.

References

1. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Jr., Dickson M, Grimwood J, et al. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science. 2005;307(5717):1928-33. Epub 2005/03/26. doi: 10.1126/science.1107239. PubMed PMID: 15790847.
2. Westram AM, Faria R, Johannesson K, Butlin R. Using replicate hybrid zones to understand the genomic basis of adaptive divergence. Molecular ecology. 2021;30(15):3797-814. Epub 2021/02/28. doi: 10.1111/mec.15861. PubMed PMID: 33638231.
3. van der Zee MJ, Whiting JR, Paris JR, Bassar RD, Travis J, Weigel D, et al. Rapid genomic convergent evolution in experimental populations of Trinidadian guppies (Poecilia reticulata). Evol Lett. 2022;6(2):149-61. Epub 2022/04/08. doi: 10.1002/evl3.272. PubMed PMID: 35386829; PubMed Central PMCID: PMCPMC8966473.
4. Zong SB, Li YL, Liu JX. Genomic Architecture of Rapid Parallel Adaptation to Fresh Water in a Wild Fish. Molecular biology and evolution. 2021;38(4):1317-29. Epub 2020/11/05. doi: 10.1093/molbev/msaa290. PubMed PMID: 33146383; PubMed Central PMCID: PMCPMC8480189.
5. Barghi N, Tobler R, Nolte V, Jaksic AM, Mallard F, Otte KA, et al. Genetic redundancy fuels polygenic adaptation in *Drosophila*. PLoS biology. 2019;17(2):e3000128. Epub 2019/02/05. doi: 10.1371/journal.pbio.3000128. PubMed PMID: 30716062.
6. Barghi N, Hermisson J, Schlötterer C. Polygenic adaptation: a unifying framework to understand positive selection. Nature reviews Genetics. 2020;21(12):769-81. Epub 2020/07/01. doi: 10.1038/s41576-020-0250-z. PubMed PMID: 32601318.
7. Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. (Non) Parallel Evolution. Annu Rev Ecol Syst. 2018;49:303-30.
8. Conte GL, Arnegard ME, Best J, Chan YF, Jones FC, Kingsley DM, et al. Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. Genetics. 2015;201(3):1189-200. Epub 2015/09/19. doi: 10.1534/genetics.115.182550. PubMed PMID: 26384359; PubMed Central PMCID: PMCPMC4649644.
9. Rennison DJ, Peichel CL. Pleiotropy facilitates parallel adaptation in sticklebacks. Molecular ecology. 2022;31(5):1476-86. Epub 2022/01/09. doi: 10.1111/mec.16335. PubMed PMID: 34997980; PubMed Central PMCID: PMCPMC9306781.
10. Stern DL. The genetic causes of convergent evolution. Nature reviews Genetics. 2013;14(11):751-64. Epub 2013/10/10. doi: 10.1038/nrg3483. PubMed PMID: 24105273.
11. Fisher RA. The genetical theory of natural selection. Oxford: Clarendon Press; 1930.
12. Orr HA. Adaptation and the cost of complexity. Evolution; international journal of organic evolution. 2000;54(1):13-20. Epub 2000/08/11. PubMed PMID: 10937178.
13. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. Science. 2002;296(5568):750-2. Epub 2002/04/27. doi: 10.1126/science.1068696. PubMed PMID: 11976460.
14. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Molecular biology and evolution.

2005;22(4):803-6. Epub 2004/12/24. doi: 10.1093/molbev/msi072. PubMed PMID: 15616139.
15. Masalia RR, Bewick AJ, Burke JM. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PloS one. 2017;12(7):e0182289. Epub 2017/08/02. doi: 10.1371/journal.pone.0182289. PubMed PMID: 28759647; PubMed Central PMCID: PMCPMC5536297.
16. Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ. The Relationship between Selection, Network Connectivity, and Regulatory Variation within a Population of Capsella grandiflora. Genome biology and evolution. 2017;9(4):1099-109. doi: 10.1093/gbe/evx068. PubMed PMID: WOS:000406755800024.
17. Hansen TF. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. Biosystems. 2003;69(2-3):83-94. doi: Pii S0303-2647(02)00132-6
Doi 10.1016/S0303-2647(02)00132-6. PubMed PMID: WOS:000182602700002.
18. Blows MW. A tale of two matrices: multivariate approaches in evolutionary biology. Journal of evolutionary biology. 2007;20(1):1-8. Epub 2007/01/11. doi: 10.1111/j.1420-9101.2006.01164.x. PubMed PMID: 17209986.
19. Lande R, Arnold SJ. The Measurement of Selection on Correlated Characters. Evolution; international journal of organic evolution. 1983;37(6):1210-26. Epub 1983/11/01. doi: 10.1111/j.1558-5646.1983.tb00236.x. PubMed PMID: 28556011.
20. Lande R. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain:Body Size Allometry. Evolution; international journal of organic evolution. 1979;33(1Part2):402-16. Epub 1979/03/01. doi: 10.1111/j.1558-5646.1979.tb04694.x. PubMed PMID: 28568194.
21. Thorholludottir DAV, Nolte V, Schlötterer C. Temperature driven gene expression evolution in natural and laboratory populations highlights the crucial role of correlated fitness effects for polygenic adaptation. Evolution; international journal of organic evolution. 2023. Epub 2023/07/17. doi: 10.1093/evolut/qpad132. PubMed PMID: 37455661.
22. MacPherson A, Nuismer SL. The probability of parallel genetic evolution from standing genetic variation. Journal of evolutionary biology. 2017;30(2):326-37. Epub 2016/11/02. doi: 10.1111/jeb.13006. PubMed PMID: 27801996.
23. Rivas-Sanchez DF, Melo-Florez L, Aragon A, Pardo-Diaz C, Salazar C, Montgomery SH, et al. Parallel evolution of behavior, physiology, and life history associated with altitudinal shifts in forest type in Heliconius butterflies. Evolution; international journal of organic evolution. 2023;77(6):1458-67. Epub 2023/04/19. doi: 10.1093/evolut/qpad062. PubMed PMID: 37075171.
24. Huang Y, Lack JB, Hoppel GT, Pool JE. Parallel and population-specific gene regulatory evolution in cold-adapted fly populations. Genetics. 2021;218(3). Epub 2021/05/15. doi: 10.1093/genetics/iyab077. PubMed PMID: 33989401; PubMed Central PMCID: PMCPMC8864734.
25. Heckley AM, Pearce AE, Gotanda KM, Hendry AP, Oke KB. Compiling forty years of guppy research to investigate the factors contributing to (non)parallel evolution. Journal of evolutionary biology. 2022;35(11):1414-31. Epub 2022/09/14. doi: 10.1111/jeb.14086. PubMed PMID: 36098479.
26. Fischer EK, Song Y, Hughes KA, Zhou W, Hoke KL. Nonparallel transcriptional divergence during parallel adaptation. Molecular ecology. 2021;30(6):1516-30. Epub 2021/02/02. doi: 10.1111/mec.15823. PubMed PMID: 33522041.

615 27. Elmer KR, Fan S, Kusche H, Spreitzer ML, Kautt AF, Franchini P, et al. Parallel
616 evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nature
617 communications. 2014;5:5168. Epub 2014/10/28. doi: 10.1038/ncomms6168. PubMed
618 PMID: 25346277.
619 28. Butlin RK, Saura M, Charrier G, Jackson B, Andre C, Caballero A, et al. Parallel
620 evolution of local adaptation and reproductive isolation in the face of gene flow.
621 Evolution; international journal of organic evolution. 2014;68(4):935-49. Epub
622 2013/12/05. doi: 10.1111/evo.12329. PubMed PMID: 24299519; PubMed Central
623 PMCID: PMCPMC4261988.
624 29. Hsu SK, Belmouaden C, Nolte V, Schlötterer C. Parallel gene expression evolution
625 in natural and laboratory evolved populations. Molecular ecology. 2021;30(4):884-94.
626 Epub 2020/09/27. doi: 10.1111/mec.15649. PubMed PMID: 32979867; PubMed
627 Central PMCID: PMCPMC7891358.
628 30. Otte KA, Nolte V, Mallard F, Schlötterer C. The genetic architecture of
629 temperature adaptation is shaped by population ancestry and not by selection regime.
630 Genome biology. 2021;22(1). doi: ARTN 211
631 10.1186/s13059-021-02425-9. PubMed PMID: WOS:000673395800001.
632 31. Papakostas S, Vollestad LA, Bruneaux M, Aykanat T, Vanoverbeke J, Ning M, et al.
633 Gene pleiotropy constrains gene expression changes in fish adapted to different thermal
634 conditions. Nature communications. 2014;5:4071. Epub 2014/06/04. doi:
635 10.1038/ncomms5071. PubMed PMID: 24892934; PubMed Central PMCID:
636 PMCPMC4059932.
637 32. Lemos B, Meiklejohn CD, Hartl DL. Regulatory evolution across the protein
638 interaction network. Nature genetics. 2004;36(10):1059-60. Epub 2004/09/21. doi:
639 10.1038/ng1427. PubMed PMID: 15378060.
640 33. Mahler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR. Gene
641 co-expression network connectivity is an important determinant of selective constraint.
642 PLoS genetics. 2017;13(4). doi: ARTN e1006402
643 10.1371/journal.pgen.1006402. PubMed PMID: WOS:000402549200001.
644 34. Nourmohammad A, Rambeau J, Held T, Kovacova V, Berg J, Lassig M. Adaptive
645 Evolution of Gene Expression in Drosophila. Cell Rep. 2017;20(6):1385-95. doi:
646 10.1016/j.celrep.2017.07.033. PubMed PMID: 28793262.
647 35. Hsu SK, Jaksic AM, Nolte V, Lirakis M, Kofler R, Barghi N, et al. Rapid sex-specific
648 adaptation to high temperature in *Drosophila*. Elife. 2020;9. Epub 2020/02/23. doi:
649 10.7554/eLife.53237. PubMed PMID: 32083552; PubMed Central PMCID:
650 PMCPMC7034977.
651 36. Yu X, Wei P, Zhao S, Chen Z, Li X, Zhang W, et al. Population transcriptomics
652 uncover the relative roles of positive selection and differential expression in
653 Batrachium bungei adaptation to the Qinghai-Tibetan plateau. Plant Cell Rep.
654 2023;42(5):879-93. Epub 2023/03/28. doi: 10.1007/s00299-023-03005-w. PubMed
655 PMID: 36973418.
656 37. Velotta JP, Wegrzyn JL, Ginzburg S, Kang L, Czesny S, O'Neill RJ, et al.
657 Transcriptomic imprints of adaptation to fresh water: parallel evolution of
658 osmoregulatory gene expression in the Alewife. Molecular ecology. 2017;26(3):831-48.
659 Epub 2016/12/25. doi: 10.1111/mec.13983. PubMed PMID: 28012221.
660 38. Tang Q, Gu Y, Zhou X, Jin L, Guan J, Liu R, et al. Comparative transcriptomics of 5
661 high-altitude vertebrates and their low-altitude relatives. Gigascience. 2017;6(12):1-9.
662 Epub 2017/11/18. doi: 10.1093/gigascience/gix105. PubMed PMID: 29149296;
663 PubMed Central PMCID: PMCPMC5729692.

664 39. O'Neil ST, Dzurisin JD, Williams CM, Lobo NF, Higgins JK, Deines JM, et al. Gene
665 expression in closely related species mirrors local adaptation: consequences for
666 responses to a warming world. Molecular ecology. 2014;23(11):2686-98. Epub
667 2014/04/29. doi: 10.1111/mec.12773. PubMed PMID: 24766086.
668 40. Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Moriarty Lemmon E,
669 et al. Quantity, Not Quality: Rapid Adaptation in a Polygenic Trait Proceeded Exclusively
670 through Expression Differentiation. Molecular biology and evolution.
671 2017;34(12):3099-110. Epub 2017/09/30. doi: 10.1093/molbev/msx231. PubMed
672 PMID: 28962003.
673 41. McShea DW. Functional complexity in organisms: Parts as proxies. Biol Philos.
674 2000;15(5):641-68. doi: Doi 10.1023/A:1006695908715. PubMed PMID:
675 WOS:000165088500002.
676 42. Proulx SR, Promislow DE, Phillips PC. Network thinking in ecology and evolution.
677 Trends in ecology & evolution. 2005;20(6):345-53. Epub 2006/05/17. doi:
678 10.1016/j.tree.2005.04.004. PubMed PMID: 16701391.
679 43. He X, Zhang J. Toward a molecular understanding of pleiotropy. Genetics.
680 2006;173(4):1885-91. Epub 2006/05/17. doi: 10.1534/genetics.106.060269. PubMed
681 PMID: 16702416; PubMed Central PMCID: PMCPMC1569710.
682 44. Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. Pleiotropic constraint
683 hampers the resolution of sexual antagonism in vertebrate gene expression. Am Nat.
684 2008;171(1):35-43. Epub 2008/01/04. doi: 10.1086/523954. PubMed PMID:
685 18171149.
686 45. Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. FlyAtlas 2: a new version of
687 the Drosophila melanogaster expression atlas with RNA-Seq, miRNA-Seq and sex-
688 specific data. Nucleic acids research. 2018;46(D1):D809-D15. Epub 2017/10/27. doi:
689 10.1093/nar/gkx976. PubMed PMID: 29069479; PubMed Central PMCID:
690 PMCPMC5753349.
691 46. Dean R, Mank JE. Tissue Specificity and Sex-Specific Regulatory Variation Permit
692 the Evolution of Sex-Biased Gene Expression. Am Nat. 2016;188(3):E74-84. Epub
693 2016/08/09. doi: 10.1086/687526. PubMed PMID: 27501094.
694 47. Allen SL, Bonduriansky R, Chenoweth SF. Genetic constraints on
695 microevolutionary divergence of sex-biased gene expression. Philos Trans R Soc Lond B
696 Biol Sci. 2018;373(1757). Epub 2018/08/29. doi: 10.1098/rstb.2017.0427. PubMed
697 PMID: 30150225; PubMed Central PMCID: PMCPMC6125734.
698 48. Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJC, et al. A
699 global overview of pleiotropy and genetic architecture in complex traits (vol 51,1339,
700 2019). Nature genetics. 2020;52(3):353-. doi: 10.1038/s41588-019-0571-z. PubMed
701 PMID: WOS:000511537400002.
702 49. Hamala T, Gorton AJ, Moeller DA, Tiffin P. Pleiotropy facilitates local adaptation
703 to distant optima in common ragweed (Ambrosia artemisiifolia). PLoS genetics.
704 2020;16(3):e1008707. Epub 2020/03/27. doi: 10.1371/journal.pgen.1008707. PubMed
705 PMID: 32210431; PubMed Central PMCID: PMCPMC7135370.
706 50. Marbach D, Roy S, Ay F, Meyer PE, Candeias R, Kahveci T, et al. Predictive
707 regulatory models in Drosophila melanogaster by integrative inference of
708 transcriptional networks. Genome research. 2012;22(7):1334-49. Epub 2012/03/30.
709 doi: 10.1101/gr.127191.111. PubMed PMID: 22456606; PubMed Central PMCID:
710 PMCPMC3396374.
711 51. Mao YX, Li Y, Yang Z, Xu N, Zhang S, Wang X, et al. Comparative transcriptome
712 analysis between rhesus macaques (Macaca mulatta) and crab-eating macaques (M.

713 fascicularis). Zool Res. 2024;45(2):299-310. Epub 2024/03/15. doi:
714 10.24272/j.issn.2095-8137.2023.322. PubMed PMID: 38485500; PubMed Central
715 PMCID: PMCPMC11017088.
716 52. Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SH, et al.
717 Comparative transcriptomics of three Poaceae species reveals patterns of gene
718 expression evolution. The Plant journal : for cell and molecular biology.
719 2012;71(3):492-502. Epub 2012/03/27. doi: 10.1111/j.1365-313X.2012.05005.x.
720 PubMed PMID: 22443345.
721 53. Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner AC, Hackenberg D, et al.
722 Comparative transcriptomic analysis reveals conserved programmes underpinning
723 organogenesis and reproduction in land plants. Nat Plants. 2021;7(8):1143-59. Epub
724 2021/07/14. doi: 10.1038/s41477-021-00958-2. PubMed PMID: 34253868.
725 54. Jaksic AM, Karner J, Nolte V, Hsu SK, Barghi N, Mallard F, et al. Neuronal function
726 and dopamine signaling evolve at high temperature in Drosophila. Molecular biology
727 and evolution. 2020. Epub 2020/05/14. doi: 10.1093/molbev/msaa116. PubMed PMID:
728 32402077.
729 55. Nouhaud P, Tobler R, Nolte V, Schlötterer C. Ancestral population reconstitution
730 from isofemale lines as a tool for experimental evolution. Ecol Evol. 2016;6(20):7169-
731 75. doi: 10.1002/ece3.2402. PubMed PMID: 27895897; PubMed Central PMCID:
732 PMCPMC5114691.
733 56. Charlesworth B, Charlesworth D. Elements of Evolutionary Genetics. Greenwood
734 Village, Colorado: Roberts and Company Publishers; 2010.
735 57. Nouhaud P, Mallard F, Poupardin R, Barghi N, Schlötterer C. High-throughput
736 fecundity measurements in *Drosophila*. Sci Rep. 2018;8:4469.
737 58. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for
738 differential expression analysis of digital gene expression data. Bioinformatics.
739 2010;26(1):139-40. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308;
740 PubMed Central PMCID: PMC2796818.
741 59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and
742 powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57:289-300.
743 60. Lai WY, Schlötterer C. Evolution of phenotypic variance in response to a novel
744 hot environment. Molecular ecology. 2022;31(3):934-45. Epub 2021/11/15. doi:
745 10.1111/mec.16274. PubMed PMID: 34775658.
746 61. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, GuhaThakurta D, et al. An
747 integrative genomics approach to infer causal associations between gene expression
748 and disease. Nature genetics. 2005;37(7):710-7. doi: 10.1038/ng1589. PubMed PMID:
749 WOS:000230196400017.
750 62. Wright S. Correlation and causation Part I. Method of path coefficients. J Agric
751 Res. 1920;20:0557-85. PubMed PMID: WOS:000188347400034.
752 63. Vlachos C, Kofler R. MimicrEE2: Genome-wide forward simulations of Evolve and
753 Resequencing studies. PLoS computational biology. 2018;14(8):e1006413. Epub
754 2018/08/17. doi: 10.1371/journal.pcbi.1006413. PubMed PMID: 30114186; PubMed
755 Central PMCID: PMCPMC6112681.
756 64. Bohutinska M, Vlcek J, Yair S, Laenen B, Konecna V, Fracassetti M, et al. Genomic
757 basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives.
758 Proceedings of the National Academy of Sciences of the United States of America.
759 2021;118(21). Epub 2021/05/19. doi: 10.1073/pnas.2022713118. PubMed PMID:
760 34001609; PubMed Central PMCID: PMCPMC8166048.

65. Hämälä T, Guiltinan MJ, Marden JH, Maximova SN, dePamphilis CW, Tiffin P. Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in *Theobroma cacao*. Molecular biology and evolution. 2020;37(1):110-23. Epub 2019/09/11. doi: 10.1093/molbev/msz206. PubMed PMID: 31501906.
66. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, et al. Systems genetics of complex traits in *Drosophila melanogaster*. Nature genetics. 2009;41(3):299-307. doi: 10.1038/ng.332. PubMed PMID: WOS:000263640200010.
67. Laruson AJ, Yeaman S, Lotterhos KE. The Importance of Genetic Redundancy in Evolution. Trends in ecology & evolution. 2020. Epub 2020/05/23. doi: 10.1016/j.tree.2020.04.009. PubMed PMID: 32439075.
68. Thompson KA, Osmond MM, Schluter D. Parallel genetic evolution and speciation from standing variation. Evol Lett. 2019;3(2):129-41. Epub 2019/07/11. doi: 10.1002/evl3.106. PubMed PMID: 31289688; PubMed Central PMCID: PMCPMC6591551.
69. Lawniczak MK, Holloway AK, Begun DJ, Jones CD. Genomic analysis of the relationship between gene expression variation and DNA polymorphism in Drosophila simulans. Genome biology. 2008;9(8):R125. Epub 2008/08/14. doi: 10.1186/gb-2008-9-8-r125. PubMed PMID: 18700012; PubMed Central PMCID: PMCPMC2575515.
70. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(20):12795-800. Epub 2002/09/19. doi: 10.1073/pnas.162041399. PubMed PMID: 12237400; PubMed Central PMCID: PMCPMC130539.
71. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297(5584):1183-6. Epub 2002/08/17. doi: 10.1126/science.1070919. PubMed PMID: 12183631.
72. Blake WJ, M KA, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422(6932):633-7. Epub 2003/04/11. doi: 10.1038/nature01546. PubMed PMID: 12687005.

Acknowledgments

We thank Viola Nolte for preparing all RNA-Seq libraries and supervising the maintenance of the evolution experiment. We thank Dagný Ásta Rúnarsdóttir for the suggestions on the early version of the manuscript and all member of the Institut für Populationsgenetik for discussion. Ana Marija Jakšić, Neda Barghi, François Mallard and Kathrin Otte performed the common garden experiment. Illumina sequencing was done at the VBCF NGS Unit (www.vbcf.ac.at). This work was support by the Austrian Science Funds (FWF, W1225, P32935) and the European Research Council (ERC, ArchAdapt).

Author contribution

W.Y.L, S.K.H. and C.S. conceived the study. A.F. contributed statistical advice, in particular for the causal analysis. W.Y.L. performed the data analysis. S.K.H. performed the simulations. W.Y.L. and C.S. wrote the manuscript. W.Y.L., S.K.H., A.F. and C.S. revised and edited the final manuscript.

Competing interests

The authors declare no competing interests.

Data accessibility statement

All sequencing data are available in European Nucleotide Archive (ENA) under the accession number PRJEB35504, PRJEB35506 and PRJEB37011. Other raw data and scripts are available on github (https://github.com/cloudweather34/pleiotropy).

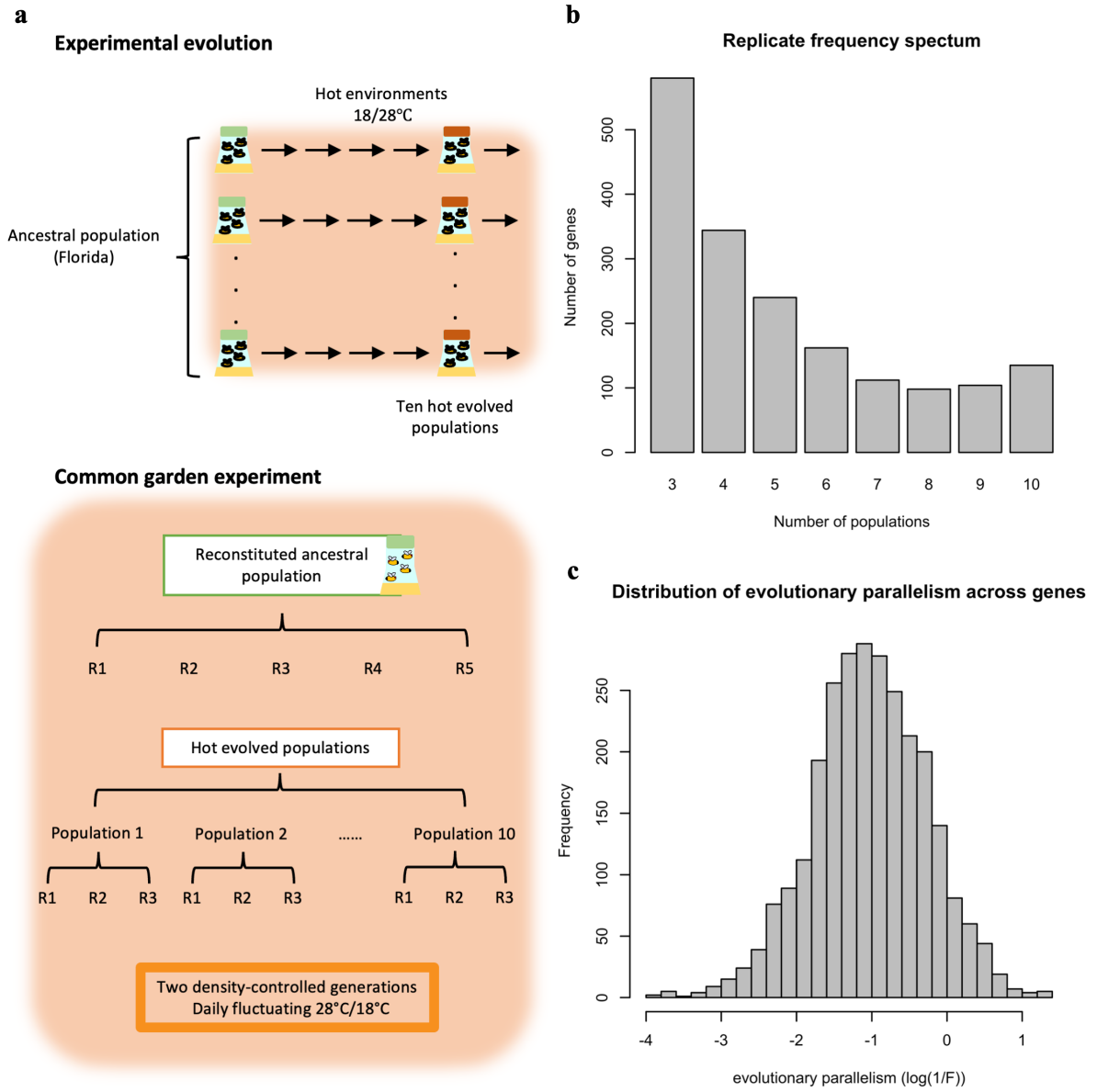

Figure 1. Overview of the experimental procedures (a) and the variation in parallel gene expression evolution across genes (b,c). **a.** Experimental Evolution: ten replicated populations seeded from one common founder population have been evolving for >100 generations in a hot laboratory environment at 18 and 28°C. Common Garden Experiment: at the 103rd generation of the evolution experiment, the ten evolved populations (each with three biological replicates) were maintained together with the reconstituted ancestral population (with five biological replicates) for two generations in the same environment as in the evolution experiment. After two generations in common garden, 50 males from each biological replicates were pooled and subjected to RNA-Seq. **b**. Replicate frequency spectrum. Number of populations (x-axis) in which a given gene experienced a significant change in gene expression. The y-axis indicates the number of genes in each category. Most of the genes experienced a significant change in gene expression in few evolved populations while much fewer genes were

827 significant in all 10 populations. This pattern suggests that the parallelism in gene expression
828 evolution differs across genes. **c.** The distribution of gene expression evolution parallelism
829 (log(1/F)) across genes. Larger log(1/F) values indicate more parallel evolution of a gene. The
830 exhibit variation suggests that genes varied in their parallelism of its expression evolution.
831

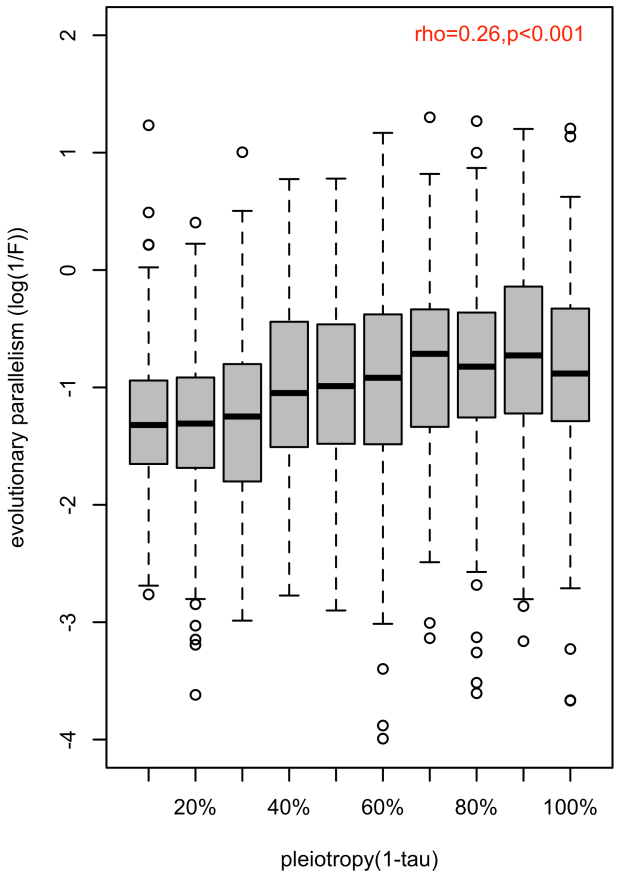

Figure 2. Association between the magnitude of evolution parallelism and the strength of pleiotropy. The distribution of evolution parallelism (log(1/F)) of different genes is shown in boxplots binned by their strength of pleiotropy (1-τ). Overall, the strength of pleiotropy was positively correlated with evolution parallelism (ρ=0.26, p-value<2.2e-16).

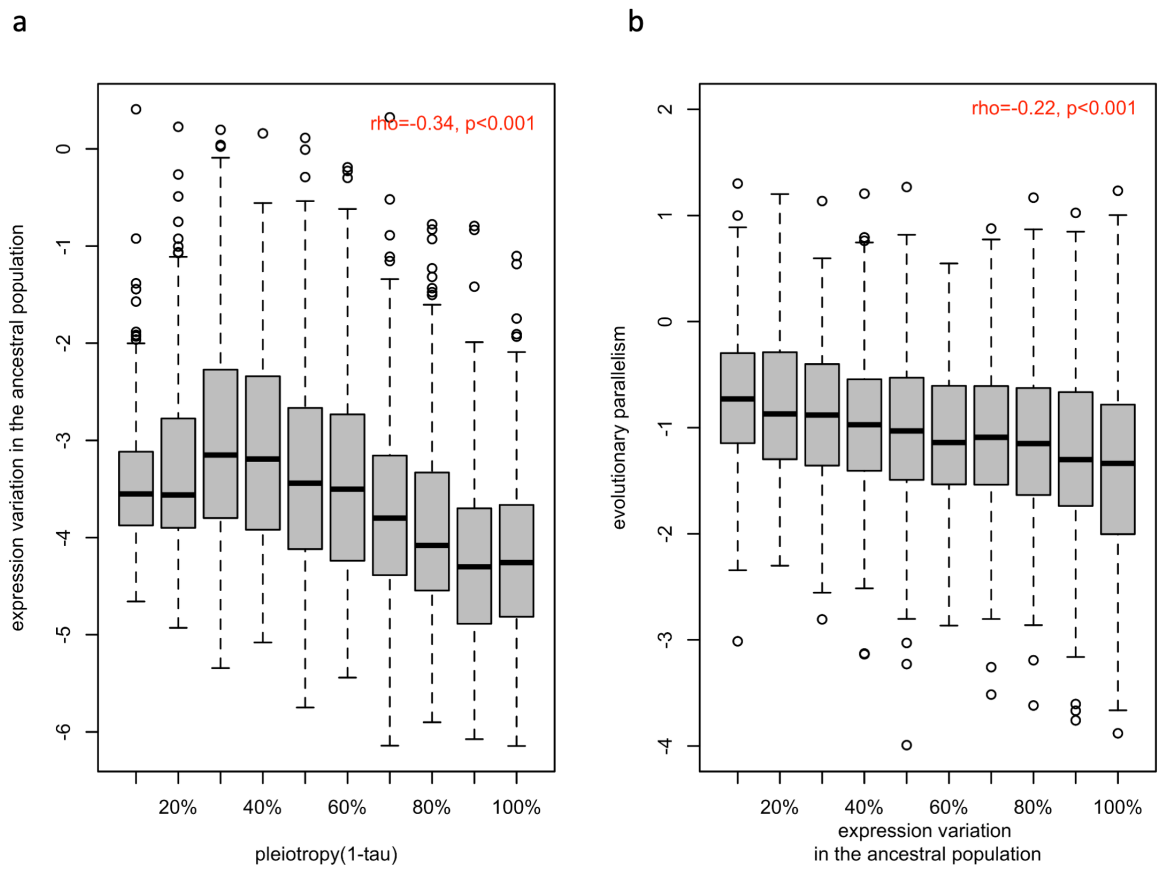

Figure 3. Association between ancestral variation in gene expression with the strength of pleiotropy (1- τ) (a) and with the evolution parallelism (b). a. The distribution of gene expression variation in the ancestral population (log (BCV^2)) of different genes is shown in boxplots binned by their strength of pleiotropy (1-τ). Overall, the strength of pleiotropy was negatively correlated with ancestral variation in gene expression (ρ=-0.34, p-value<2.2e-16). **b.** The distribution of evolutionary parallelism (log(1/F)) of different genes is shown in boxplots binned by their strength of ancestral variation. The strength of parallelism is negatively correlated with the ancestral variation in gene expression (ρ=-0.22, p-value<2.2e-16).

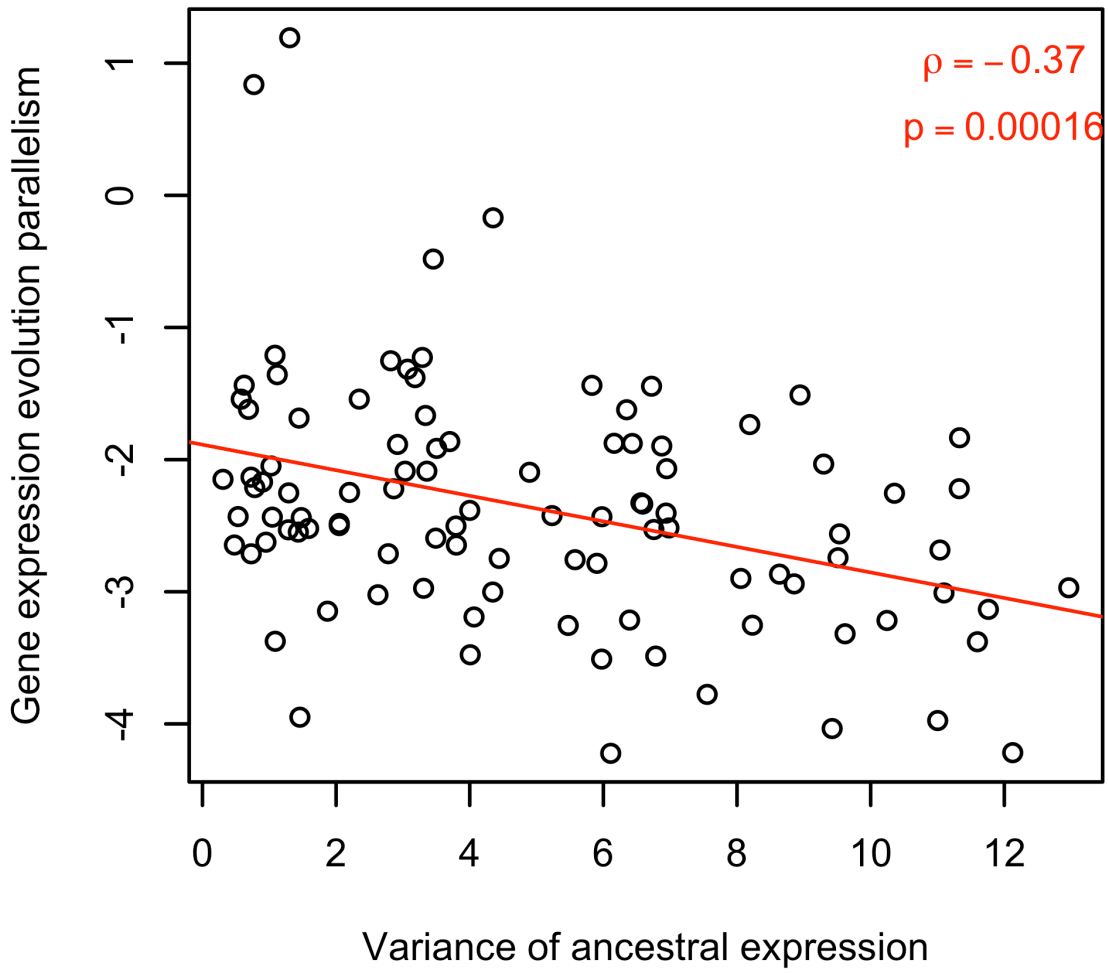

Figure 4. Computer simulations illustrate the negative effect of ancestral variation on parallel evolution. Computer simulations assume a fitness related trait determined by the expression of four genes. The expression of these four genes is determined by a different number of genetic loci, reflecting the influence of pleiotropy on ancestral genetic variation. A shift in trait optimum was used to illustrate the connection between ancestral variation in expression (x-axis) and parallelism of expression change across ten evolved populations (y-axis). The negative correlation between them (rho = -0.26, p-value < 1.77e-07) suggests that the gene with less ancestral variation are resulting in more parallel responses.

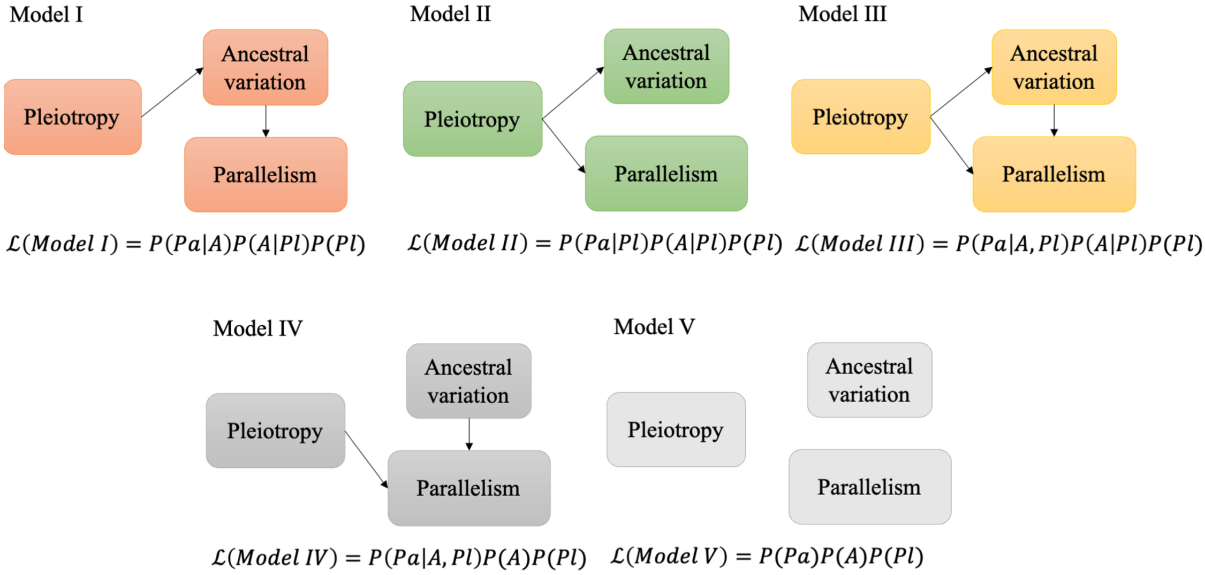

Figure 5. Schematic figure illustrating the causal models evaluated. Five possible relationships between pleiotropic effects, ancestral variation, and evolutionary parallelism. $\mathcal{L}$ denotes the likelihood of each model given the data. P is the probability or conditional probability of the measurements for each gene; *Pa* is the parallelism level and *A* is the level of ancestral variation in gene expression. *Pl* is the pleiotropic effects. See materials and methods for a more detailed description.

Table 1. BIC value for Model I-V. The model with the smallest BIC is the one best support by the data.

Model	BIC value (Network connectivity)	BIC value (Tissue specificity)
I	9258.395	8958.125
II	9316.852	8972.626
III	9264.769	8939.342
IV	9490.017	9162.090
V	9550.309	9247.540

874 **Table 2. The size of direct and indirect pleiotropic effects on the evolutionary parallelism.**
875

	Direct effect	**Indirect effect**
1-τ	0.127	0.055 (-0.181, -0.332) *
Network connectivity	0.026	0.068 (-0.210, -0.337) *

876 * Coefficient estimates (β_2, β_3) whose product provides the indirect effect.
877