

Pleiotropy increases parallel selection signatures during adaptation from standing genetic variation

Reviewed Preprint

v2 • March 6, 2025

Revised by authors

Reviewed Preprint

v1 • November 22, 2024

Wei-Yun Lai, Sheng-Kai Hsu, Andreas Futschik, Christian Schlötterer 

Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria • Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria • Department of Applied Statistics, Johannes Kepler University Linz, Linz, Austria

 https://en.wikipedia.org/wiki/Open_access
 Copyright information

eLife Assessment

This study makes the **important** finding that pleiotropy is positively associated with parallelism of evolutionary responses in gene expression. This finding, if true, runs counter to current expectations in the field. The analysis uses state-of-the art experimental evolution approach to study the genetic basis of adaptation of *Drosophila simulans* to a hot environment. Although the experimental results are **convincing**, the theoretical model is **incomplete**, due to several unusual assumptions. It remains to be seen whether the main conclusion can be replicated in other contexts.

<https://doi.org/10.7554/eLife.102321.2.sa4>

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.

In this study we employed gene expression analysis to investigate the interplay between polymorphism and pleiotropy on the degree of parallelism of adaptive responses. We assessed the selective response by measuring gene expression differences between populations from different environments, a widely employed approach [e. g. 34, 35-40]. Specifically, we compared gene expression in the ancestral *D. simulans* population with that of flies that had evolved for 100 generations in a novel hot temperature regime. To infer parallel evolution, the gene expression change was measured in 10 replicate populations. By focusing on the evolutionary response, we circumvent the challenge of connecting the phenotype to an unknown fitness landscape. Furthermore, we take advantage of well-established pleiotropy estimates [41-44] to evaluate the influence of pleiotropy on parallel gene expression evolution. The natural variation in gene expression was determined by examining the differences in expression levels among individual flies from the ancestral *D. simulans* population.

We observed significant interplays between pleiotropy, ancestral variation, and parallel selection responses among genes. The parallelism of gene expression evolution is positively associated with pleiotropy but negatively associated with the ancestral variation in gene expression. Given that pleiotropy was also negatively associated with ancestral variation, it is crucial to differentiate between statistical causality and correlation. Causal analysis revealed that pleiotropy affects parallel evolution both directly and indirectly via ancestral variation. It is likely that the indirect effects operate via the reduction of ancestral variation in pleiotropic genes. The direct effects of pleiotropy on parallelism are probably best explained by synergistic pleiotropy which results in a stronger selection response.

Methods

Estimating gene pleiotropy

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

τ indicates how specific the expression of a gene is across different tissues. τ was estimated for each gene using the adult male expression profiles on flyatlas2 [45] as:

$$\tau = \frac{\sum_i [1 - \frac{\text{gene expression}_i}{\text{gene expression}_{\max}}]}{N-1}, \text{ where } N \text{ is the number of tissues examined and } i \text{ indicates each of}$$

them [46]. If a gene is only expressed in one tissue, τ will equal to 1 while it equals 0 when a gene is expressed at the same level across all tissues. The relationship between τ and pleiotropy is

based on the idea that the genes expressed in many tissues are more likely to affect multiple traits than genes expressed in fewer tissues [41 [↗](#), 44 [↗](#), 47 [↗](#), 48 [↗](#)]. Hence, we used $1-\tau$ to indicate the pleiotropic effect of a gene.

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

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

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

Experimental evolution and common garden experiment

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

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

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

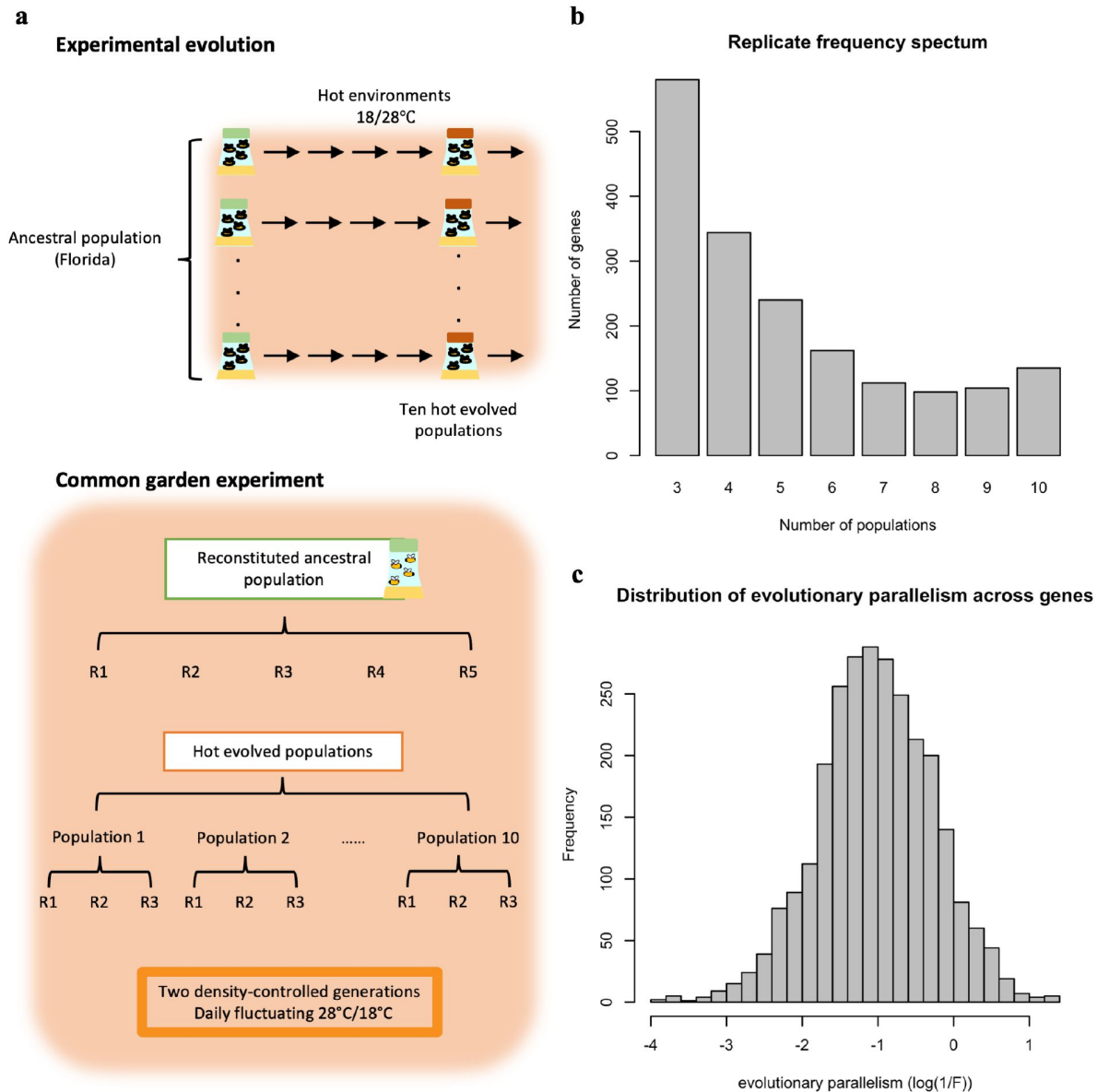


Figure 1.

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

Quantifying the evolutionary parallelism in gene expression across ten replicate populations

The raw RNA-Seq count table of five replicates of the reconstituted ancestral population and 10 independently evolved populations, each with three biological samples, was taken from (Jaksic et al., 2020). The raw read counts of each gene were normalized with the TMM method implemented in edgeR [58]. Only genes with at least 0.1 normalized counts per million base (CPM) across all samples were considered for further analysis. Because we are interested in the evolutionary response in each independently evolved population, we contrasted the three 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 ($\log_2FC$; fold change in expression between ancestral and evolved sample; see below) of each putatively adaptive gene ($n=1,775$) in each evolved sample.

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

where y_{evo} is the expression value (CPM) per gene per evolved sample, and $\bar{y}_{anc}$ 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 ($\log_2FC$), 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 $\log_2FC$ value for the j^{th} biological sample in the i^{th} evolution population. $i=1, 2, \dots, 10$, $j=1, 2, 3$.

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.

Estimating gene expression variation in an outbred ancestral population

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

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:

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

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

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

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

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

$p(Pl)$ 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:

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

Where $\mathcal{L}_i$ is the likelihood for model I-V, k_j 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:

$$y_{Pa} = \beta_1 x_A + \beta_2 x_{Pl} + \varepsilon \quad (9)$$

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:

$$y_A = \beta_3 x_{Pl} + \varepsilon \quad (10)$$

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 redundant 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 are redundant with in terms of their fitness effect during the experimental evolution, but they exhibit different levels of pleiotropy and thus have historically experienced different levels of purifying selection, leaving different levels of genetic variation at the start of the experiment. Using mimrEE2 [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 ($\log_2FC$) 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 ($1-\tau$ (tissue specificity) and network connectivity; see materials and methods) of each putatively adaptive gene ($n=1,775$) with the parallelism of its evolutionary responses ($1/F$). We observed a significant positive correlation between the parallelism and the degree of pleiotropy, ($\rho=0.26$, $p\text{-value}<2.2e-16$ for $1-\tau$ and $\rho=0.11$, $p\text{-value}<5.6e-07$ for network connectivity) (**Figure 2** and Supplementary figure 4a). This suggests that pleiotropy may enhance parallel changes in gene expression during evolution.

Pleiotropy constrains the ancestral (natural) variation in gene expression

The negative effect of pleiotropy on genetic variation over longer time scales, mostly between species, is well documented [13-16], but it not yet clear whether the same effect is seen for adaptation from standing genetic variation. We reconstituted the ancestral population from 189 isofemale lines [60] and estimated the expression variance (BCV^2 ; see materials and methods) of each putatively adaptive gene ($n=1,775$) in two replicate sets of 20 individuals. Since the variance in gene expression is positively correlated with genetic variation in cis-regulatory regions [65]

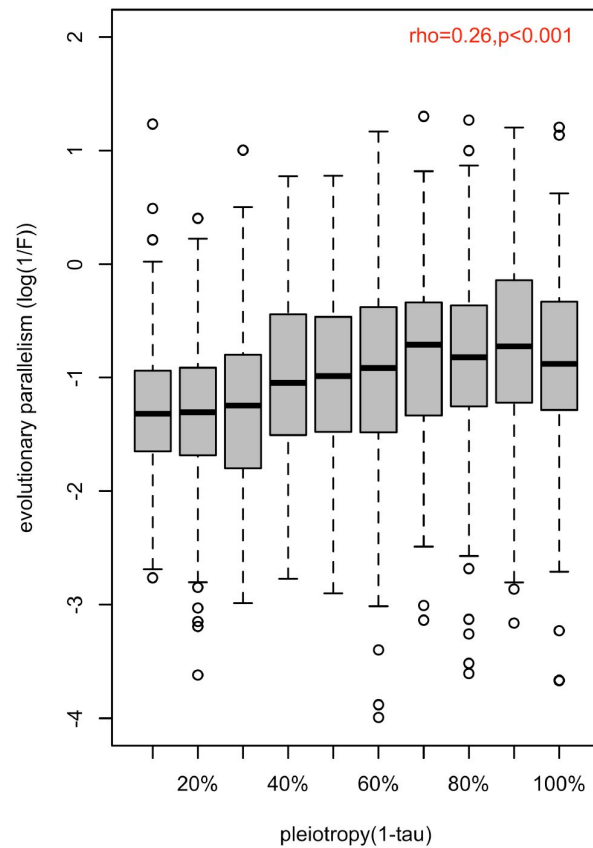


Figure 2.

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 ($1-\tau$). Overall, the strength of pleiotropy was positively correlated with evolution parallelism ($\rho=0.26$, $p\text{-value}<2.2\text{e-}16$).

and the heritability of gene expression is high [66], we assumed that the observed variation in gene expression has a genetic basis, rather than being noise (see discussion). We observed a highly significant negative correlation between the level of pleiotropy and the magnitude of ancestral variation in gene expression (**Figure 3a** and Supplementary figure 4b; $\rho = -0.34$, $p\text{-value} < 2.2 \times 10^{-16}$ for $1-\tau$ and $\rho = -0.37$, $p\text{-value} < 2.2 \times 10^{-16}$ for network connectivity). To exclude that this correlation was driven by our filtering for significant expression changes in at least three populations, we repeated these analyses with all genes ($n = 9,882$) and also observed a significant correlation ($\rho = -0.45$, $p\text{-value} < 2.2 \times 10^{-16}$ for $1-\tau$ and $\rho = -0.38$, $p\text{-value} < 2.2 \times 10^{-16}$ for network connectivity, data not shown). Hence, we suggest that the negative relationship between pleiotropy and polymorphism/diversity [32, 33] (i.e. reduced variants, lower frequencies and/or lower allelic effects) can be extended to gene expression.

Evolutionary parallelism is negatively associated with ancestral variation

For a better understanding of the factors determining parallel evolution, we examined the influence of ancestral variation in gene expression on the parallel evolution of gene expression. We observed a significant negative correlation between ancestral variation and parallelism of the evolutionary response ($\rho = -0.22$, $p\text{-value} < 2.2 \times 10^{-16}$; **Figure 3b**). We suggest that this pattern could be explained by genetic redundancy [5, 67], which describes the phenomenon that a polygenic trait can shift its phenotypic value through different combinations of contributing loci: When the diversity of the contributing loci is reduced (either through fewer loci, lower ancestral frequencies or smaller variation in effect size), fewer combinations of the contributing loci are available for adaptation, resulting in more parallel adaptive responses. To confirm this, we performed computer simulations of replicated polygenic adaptation and found that lower ancestral variation (i.e. fewer contributing loci) leads to more parallel trait evolution (**Figure 4**). Hence, that natural variation in the ancestral population could be one of factor determining the parallel adaptive response in replicate populations. Although we focused on the phenotypic response to match gene expression, it is interesting to note that a similar result was obtained for parallelism at the genetic level with different levels of standing genetic variation [68]. If pleiotropy reduces variation, the positive correlation between pleiotropy and parallelism may be an indirect (correlative) effect, rather than a direct consequence of pleiotropy.

Testing causality

Our previous analyses indicated that parallelism is affected by pleiotropy and ancestral variation. Because pleiotropy is negatively correlated with ancestral variation for both pleiotropy measures (**Figure 3a** and supplementary figure 4b), it is not clear whether pleiotropy affects parallelism directly or indirectly through the negative correlation with ancestral variation. We used causal analysis, which based on a Bayesian statistical framework [61], to disentangle the causal relationship between the three different parameters. Since the directionality of pleiotropic effects on ancestral variation has been previously demonstrated, and ancestral variation is unlikely to influence the pleiotropy of a gene, we considered five possible causal relationships between pleiotropic effects, ancestral variation, and parallelism (**Figure 5**). In the first model, parallelism is determined by ancestral variation, which is shaped by pleiotropic, but pleiotropy has no direct effect on parallelism. In model II, the pleiotropic effect affects ancestral variation and parallelism independently, but ancestral variation has no causal effect on parallelism. In model III, pleiotropy affects both ancestral variation and parallelism, and ancestral variation also determines parallelism. In model IV, parallelism is determined by ancestral variation and pleiotropy, but pleiotropic effects are assumed to have no effect on the ancestral variation. This serves as a null model to confirm the indirect effect of pleiotropy on parallelism via ancestral variation. We also add a global null model with no correlation between these three factors (model V, **Figure 5**). The analysis was performed independently for two measures of pleiotropy (tissue specificity and network connectivity).

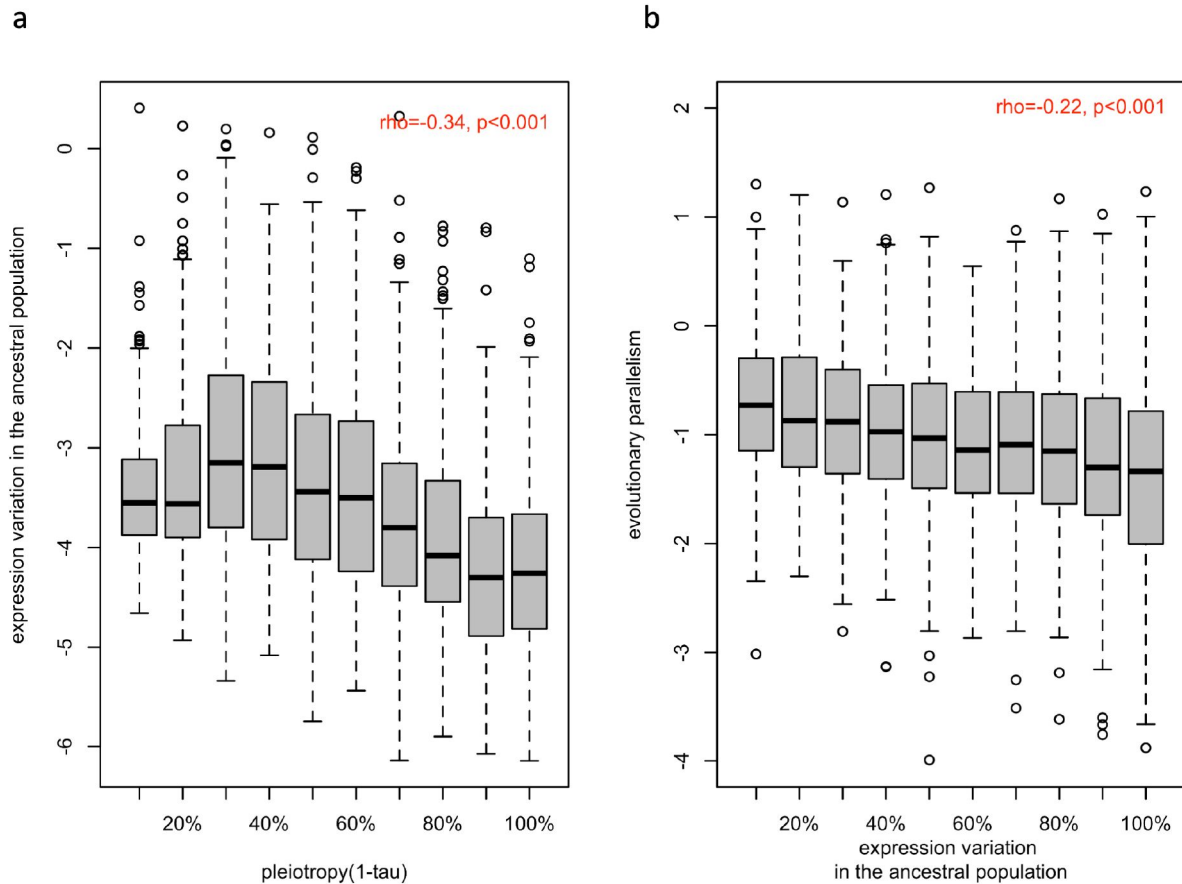


Figure 3.

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 ($\log(\text{BCV}^2)$) 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 ($\rho = -0.34$, $p\text{-value} < 2.2 \times 10^{-16}$). **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 ($\rho = -0.22$, $p\text{-value} < 2.2 \times 10^{-16}$).

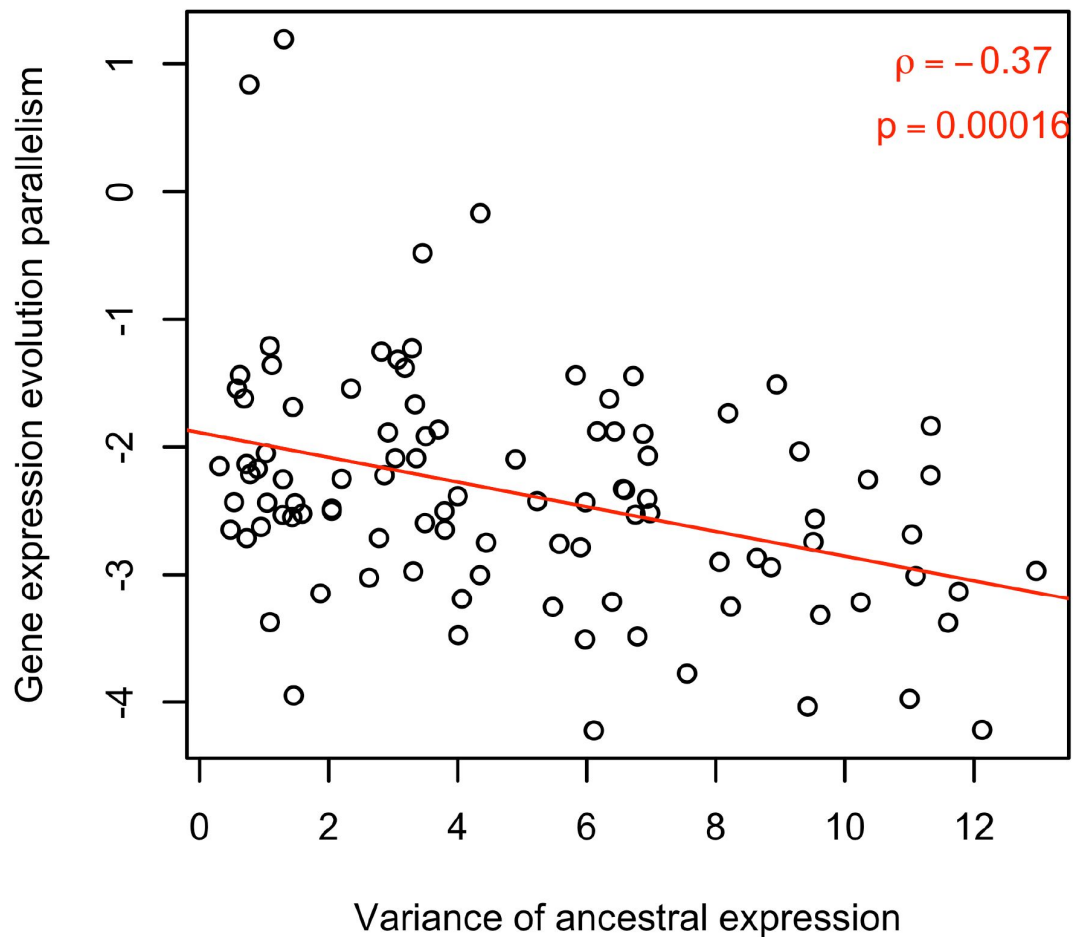


Figure 4.

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 ($\rho = -0.26$, $p\text{-value} < 1.77\text{e-}07$) suggests that the gene with less ancestral variation are resulting in more parallel responses.

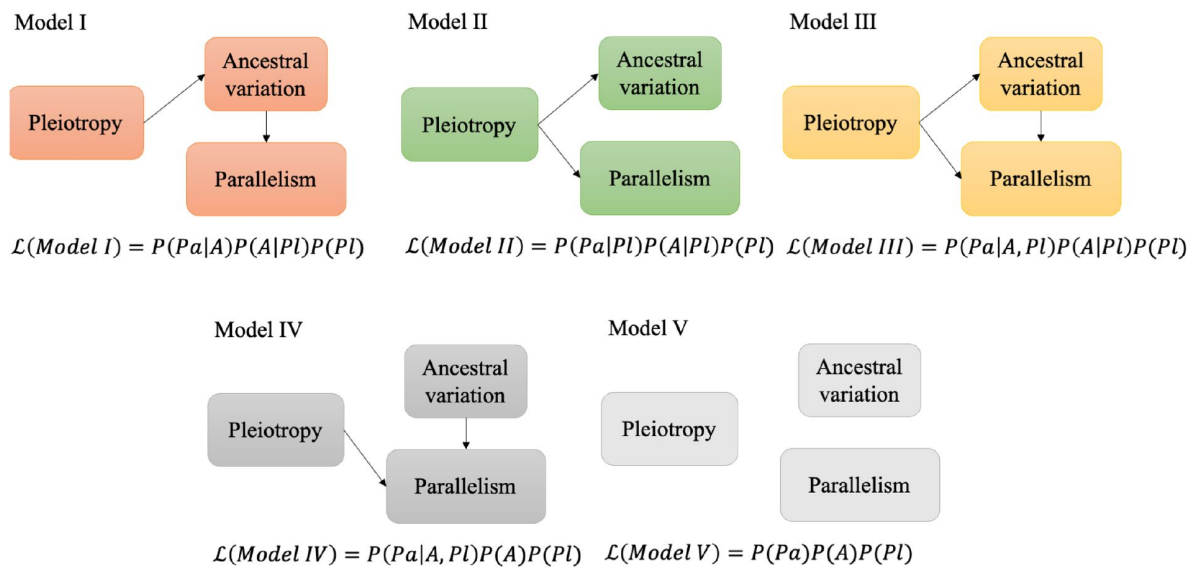


Figure 5.

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.

For tissue specificity, our data were better explained by model III (i.e., model III has the lowest 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.

Discussion

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

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

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

An important difference between adaptation from de novo mutations and standing genetic variation concerns the implications for parallel evolution. While pleiotropic genes are less likely to contribute to parallel evolution from de novo mutations [[7](#) [↗](#)], our results show that adaptation from standing genetic variation favors parallel selection responses of pleiotropic genes (**Figure 2** [↗](#), Supplementary Figure 4a). It is important to note that many previous studies focused on sequence variation, whereas we used expression variation. While it is possible that these two types of variation behave differently, we do not believe that this is the case as sequence polymorphism and expression variation are correlated [[69](#) [↗](#)]. Furthermore, a recent study focusing on genomic changes [[9](#) [↗](#)] reached similar conclusions.

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

Table 2.

The size of direct and indirect pleiotropic effects on the evolutionary parallelism.

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

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

A very important assumption in our interpretation is that differences in gene expression variation in the ancestral population reflect regulatory sequence polymorphism rather than noise. The high correlation in the variance of gene expression between two replicate groups of individuals indicates that our measures do not reflect technical noise but have a biological basis. However, stochastic fluctuations in gene expression have also been observed among cells in single-celled and multicellular organisms and the variance in gene expression differs among genes [70–72]. In this case, the differences in gene expression variance are biological in nature, but not associated with sequence variation. We do not believe that variation in the control of gene expression levels among genes can explain our pattern because *Drosophila* is a multicellular organism and we used whole body RNA-Seq data for our analysis. Therefore, even if expression is controlled to different extents among genes, by averaging across multiple cells, this does not affect our results. The only scenario we can imagine in which gene expression noise could propagate across the many cells of a multicellular organism is that during early development, differences in gene expression of a key regulator could affect downstream gene expression throughout the entire organism. However, we do not think this is a likely scenario, especially in the light of previous results showing that genetic variation in cis-regulatory regions correlates with expression variation [65]. Furthermore, whole-body gene expression in *Drosophila* is highly heritable, suggesting that gene expression is not stochastic [66]. Nevertheless, we would like to point out that our observations are valid irrespective of whether the correlation between parallelism and expression variation is the result of genetic polymorphism or genetic control of expression noise.

Based on the significant correlation results, the question arose as to whether the positive correlation between pleiotropy and parallelism, is causal (direct effect) or merely correlational (indirect effect). Regardless of the measure of pleiotropy, our causal analysis showed a consistent indirect effect of pleiotropy on parallelism (Table 1, 2). The estimated indirect effects were very similar for both measures of pleiotropy (Table 2). This indirect effect of pleiotropy could be explained by the reduced variation in the expression of more pleiotropic genes in the ancestral population leading to a more parallel selection response of gene expression. For the direct effect, however, the results differed between the two measures of 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 connectivity. 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.

We focused our analysis on genes that exhibited evidence of adaptive responses, defined as significant changes in the same direction across at least three populations. This approach allowed us to investigate the extent of parallel adaptive responses and how pleiotropy contributes to variation in parallelism among genes. Interestingly, when comparing pleiotropy between DE genes (those showing adaptive responses) and non-DE genes (those that did not change their response during evolution), we observed significantly higher pleiotropy in the non-DE gene group (Supplementary figure 5). This finding suggests that extremely high levels of pleiotropy may constrain evolutionary responses. Overall, the pattern of increased pleiotropy associated with evolutionary parallelism supports the idea that low to intermediate levels of pleiotropy are more favorable for adaptation [9, 73].

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.

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>).

Acknowledgements

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 research was funded in whole or in part by the Austrian Science Fund (FWF) [W1225, P32935] and the European Research Council (ERC) [ArchAdapt]. For open access PURPOSES, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission..

Additional information

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.

Additional files

Supplementary information

References

1. Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, et al. 2005) **Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles** *Science* **307**:1928–33 <https://doi.org/10.1126/science.1107239>
2. Westram AM, Faria R, Johannesson K, Butlin R. 2021) **Using replicate hybrid zones to understand the genomic basis of adaptive divergence** *Molecular ecology* **30**:3797–814 <https://doi.org/10.1111/mec.15861>
3. van der Zee MJ, Whiting JR, Paris JR, Bassar RD, Travis J, Weigel D, et al. 2022) **Rapid genomic convergent evolution in experimental populations of Trinidadian guppies (*Poecilia reticulata*)** *Evol Lett* **6**:149–61 <https://doi.org/10.1002/evl3.272>
4. Zong SB, Li YL, Liu JX 2021) **Genomic Architecture of Rapid Parallel Adaptation to Fresh Water in a Wild Fish** *Molecular biology and evolution* **38**:1317–29 <https://doi.org/10.1093/molbev/msaa290>
5. Barghi N, Tobler R, Nolte V, Jaksic AM, Mallard F, Otte KA, et al. 2019) **Genetic redundancy fuels polygenic adaptation in *Drosophila*** *PLoS biology* **17**:e3000128 <https://doi.org/10.1371/journal.pbio.3000128>
6. Barghi N, Hermisson J, Schlötterer C. 2020) **Polygenic adaptation: a unifying framework to understand positive selection** *Nature reviews Genetics* **21**:769–81 <https://doi.org/10.1038/s41576-020-0250-z>
7. Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE 2018) **(Non) Parallel Evolution** *Annu Rev Ecol Syst* **49**:303–30
8. Conte GL, Arnegard ME, Best J, Chan YF, Jones FC, Kingsley DM, et al. 2015) **Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback** *Genetics* **201**:1189–200 <https://doi.org/10.1534/genetics.115.182550>
9. Rennison DJ, Peichel CL 2022) **Pleiotropy facilitates parallel adaptation in sticklebacks** *Molecular ecology* **31**:1476–86 <https://doi.org/10.1111/mec.16335>
10. Stern DL 2013) **The genetic causes of convergent evolution** *Nature reviews Genetics* **14**:751–64 <https://doi.org/10.1038/nrg3483>
11. Fisher RA 1930) **The genetical theory of natural selection** Oxford: Clarendon Press
12. Orr HA 2000) **Adaptation and the cost of complexity** *Evolution; international journal of organic evolution* **54**:13–20
13. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW 2002) **Evolutionary rate in the protein interaction network** *Science* **296**:750–2 <https://doi.org/10.1126/science.1068696>
14. Hahn MW, Kern AD 2005) **Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks** *Molecular biology and evolution* **22**:803–6 <https://doi.org/10.1093/molbev/msi072>

15. Masalia RR, Bewick AJ, Burke JM (2017) **Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants** *PLoS one* **12**:e0182289 <https://doi.org/10.1371/journal.pone.0182289>
16. Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ (2017) **The Relationship between Selection, Network Connectivity, and Regulatory Variation within a Population of *Capsella grandiflora*** *Genome biology and evolution* **9**:1099–109 <https://doi.org/10.1093/gbe/evx068>
17. Hansen TF, Pii S0303-2647(02)00132-6 (2003) **Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability** *Biosystems* **69**:83–94
18. Blows MW (2007) **A tale of two matrices: multivariate approaches in evolutionary biology** *Journal of evolutionary biology* **20**:1–8 <https://doi.org/10.1111/j.1420-9101.2006.01164.x>
19. Lande R, Arnold SJ (1983) **The Measurement of Selection on Correlated Characters** *Evolution; international journal of organic evolution* **37**:1210–26 <https://doi.org/10.1111/j.1558-5646.1983.tb00236.x>
20. Lande R. (1979) **Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain:Body Size Allometry** *Evolution; international journal of organic evolution* **33**:402–16 <https://doi.org/10.1111/j.1558-5646.1979.tb04694.x>
21. Thorholludottir DAV, Nolte V, Schlötterer C. (2023) **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* <https://doi.org/10.1093/evolut/qpad132>
22. MacPherson A, Nuismer SL (2017) **The probability of parallel genetic evolution from standing genetic variation** *Journal of evolutionary biology* **30**:326–37 <https://doi.org/10.1111/jeb.13006>
23. Rivas-Sanchez DF, Melo-Florez L, Aragon A, Pardo-Diaz C, Salazar C, Montgomery SH, et al. (2023) **Parallel evolution of behavior, physiology, and life history associated with altitudinal shifts in forest type in *Heliconius* butterflies** *Evolution; international journal of organic evolution* **77**:1458–67 <https://doi.org/10.1093/evolut/qpad062>
24. Huang Y, Lack JB, Hoppel GT, Pool JE (2021) **Parallel and population-specific gene regulatory evolution in cold-adapted fly populations** *Genetics* **218** <https://doi.org/10.1093/genetics/iyab077>
25. Heckley AM, Pearce AE, Gotanda KM, Hendry AP, Oke KB (2022) **Compiling forty years of guppy research to investigate the factors contributing to (non)parallel evolution** *Journal of evolutionary biology* **35**:1414–31 <https://doi.org/10.1111/jeb.14086>
26. Fischer EK, Song Y, Hughes KA, Zhou W, Hoke KL (2021) **Nonparallel transcriptional divergence during parallel adaptation** *Molecular ecology* **30**:1516–30 <https://doi.org/10.1111/mec.15823>
27. Elmer KR, Fan S, Kusche H, Spreitzer ML, Kautt AF, Franchini P, et al. (2014) **Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes** *Nature communications* **5**:5168 <https://doi.org/10.1038/ncomms6168>
28. Butlin RK, Saura M, Charrier G, Jackson B, Andre C, Caballero A, et al. (2014) **Parallel evolution of local adaptation and reproductive isolation in the face of gene flow** *Evolution*;

29. Hsu SK, Belmouaden C, Nolte V, Schlötterer C. 2021) **Parallel gene expression evolution in natural and laboratory evolved populations** *Molecular ecology* **30**:884–94 <https://doi.org/10.1111/mec.15649>
30. Otte KA, Nolte V, Mallard F, Schlötterer C. 2021) **The genetic architecture of temperature adaptation is shaped by population ancestry and not by selection regime** *Genome biology* **22** <https://doi.org/10.1186/s13059-021-02425-9>
31. Papakostas S, Vollestad LA, Bruneaux M, Aykanat T, Vanoverbeke J, Ning M, et al. 2014) **Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions** *Nature communications* **5**:4071 <https://doi.org/10.1038/ncomms5071>
32. Lemos B, Meiklejohn CD, Hartl DL 2004) **Regulatory evolution across the protein interaction network** *Nature genetics* **36**:1059–60 <https://doi.org/10.1038/ng1427>
33. Mahler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR 2017) **Gene co-expression network connectivity is an important determinant of selective constraint** *PLoS genetics* **13** <https://doi.org/10.1371/journal.pgen.1006402>
34. Nourmohammad A, Rambeau J, Held T, Kovacova V, Berg J, Lassig M. 2017) **Adaptive Evolution of Gene Expression in Drosophila** *Cell Rep* **20**:1385–95 <https://doi.org/10.1016/j.celrep.2017.07.033>
35. Hsu SK, Jaksic AM, Nolte V, Lirakis M, Kofler R, Barghi N, et al. 2020) **Rapid sex-specific adaptation to high temperature in Drosophila** *eLife* **9** <https://doi.org/10.7554/eLife.53237>
36. Yu X, Wei P, Zhao S, Chen Z, Li X, Zhang W, et al. 2023) **Population transcriptomics uncover the relative roles of positive selection and differential expression in Batrachium bungei adaptation to the Qinghai-Tibetan plateau** *Plant Cell Rep* **42**:879–93 <https://doi.org/10.1007/s00299-023-03005-w>
37. Velotta JP, Wegrzyn JL, Ginzburg S, Kang L, Czesny S, O'Neill RJ, et al. 2017) **Transcriptomic imprints of adaptation to fresh water: parallel evolution of osmoregulatory gene expression in the Alewife** *Molecular ecology* **26**:831–48 <https://doi.org/10.1111/mec.13983>
38. Tang Q, Gu Y, Zhou X, Jin L, Guan J, Liu R, et al. 2017) **Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives** *Gigascience* **6**:1–9 <https://doi.org/10.1093/gigascience/gix105>
39. O'Neil ST, Dzurisin JD, Williams CM, Lobo NF, Higgins JK, Deines JM, et al. 2014) **Gene expression in closely related species mirrors local adaptation: consequences for responses to a warming world** *Molecular ecology* **23**:2686–98 <https://doi.org/10.1111/mec.12773>
40. Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Moriarty Lemmon E, et al. 2017) **Quantity, Not Quality: Rapid Adaptation in a Polygenic Trait Proceeded Exclusively through Expression Differentiation** *Molecular biology and evolution* **34**:3099–110 <https://doi.org/10.1093/molbev/msx231>
41. McShea DW 2000) **Functional complexity in organisms: Parts as proxies** *Biol Philos* **15**:641–68 <https://doi.org/10.1023/A:1006695908715>

42. Proulx SR, Promislow DE, Phillips PC (2005) **Network thinking in ecology and evolution** *Trends in ecology & evolution* **20**:345–53 <https://doi.org/10.1016/j.tree.2005.04.004>
43. He X, Zhang J. (2006) **Toward a molecular understanding of pleiotropy** *Genetics* **173**:1885–91 <https://doi.org/10.1534/genetics.106.060269>
44. Mank JE, Hultin-Rosenberg L, Zwahlen M, Ellegren H. (2008) **Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression** *Am Nat* **171**:35–43 <https://doi.org/10.1086/523954>
45. Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT (2018) **FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data** *Nucleic acids research* **46**:D809–D15 <https://doi.org/10.1093/nar/gkx976>
46. Dean R, Mank JE (2016) **Tissue Specificity and Sex-Specific Regulatory Variation Permit the Evolution of Sex-Biased Gene Expression** *Am Nat* **188**:E74–84 <https://doi.org/10.1086/687526>
47. Allen SL, Bonduriansky R, Chenoweth SF (2018) **Genetic constraints on microevolutionary divergence of sex-biased gene expression** *Philos Trans R Soc Lond B Biol Sci* **373** <https://doi.org/10.1098/rstb.2017.0427>
48. Watanabe K, Stringer S, Frei O, Mirkov MU, de Leeuw C, Polderman TJC, et al. (2020) **A global overview of pleiotropy and genetic architecture in complex traits (vol 51,1339, 2019)** *Nature genetics* **52**:353 <https://doi.org/10.1038/s41588-019-0571-z>
49. Hamala T, Gorton AJ, Moeller DA, Tiffin P. (2020) **Pleiotropy facilitates local adaptation to distant optima in common ragweed (*Ambrosia artemisiifolia*)** *PLoS genetics* **16**:e1008707 <https://doi.org/10.1371/journal.pgen.1008707>
50. Marbach D, Roy S, Ay F, Meyer PE, Candeias R, Kahveci T, et al. (2012) **Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks** *Genome research* **22**:1334–49 <https://doi.org/10.1101/gr.127191.111>
51. Mao YX, Li Y, Yang Z, Xu N, Zhang S, Wang X, et al. (2024) **Comparative transcriptome analysis between rhesus macaques (*Macaca mulatta*) and crab-eating macaques (*M. fascicularis*)** *Zool Res* **45**:299–310 <https://doi.org/10.24272/j.issn.2095-8137.2023.322>
52. Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu SH, et al. (2012) **Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution** *The Plant journal : for cell and molecular biology* **71**:492–502 <https://doi.org/10.1111/j.1365-3113X.2012.05005.x>
53. Julca I, Ferrari C, Flores-Tornero M, Proost S, Lindner AC, Hackenberg D, et al. (2021) **Comparative transcriptomic analysis reveals conserved programmes underpinning organogenesis and reproduction in land plants** *Nat Plants* **7**:1143–59 <https://doi.org/10.1038/s41477-021-00958-2>
54. Jaksic AM, Karner J, Nolte V, Hsu SK, Barghi N, Mallard F, et al. (2020) **Neuronal function and dopamine signaling evolve at high temperature in *Drosophila*** *Molecular biology and evolution* <https://doi.org/10.1093/molbev/msaa116>
55. Nouhaud P, Tobler R, Nolte V, Schlötterer C. (2016) **Ancestral population reconstitution from isofemale lines as a tool for experimental evolution** *Ecol Evol* **6**:7169–75 <https://doi.org/10.1002/ece3.2402>

56. Charlesworth B, Charlesworth D. 2010) **Elements of Evolutionary Genetics** Greenwood Village, Colorado: Roberts and Company Publishers
57. Nouhaud P, Mallard F, Poupardin R, Barghi N, Schlötterer C. 2018) **High-throughput fecundity measurements in *Drosophila*** *Sci Rep* **8**:4469
58. Robinson MD, McCarthy DJ, Smyth GK 2010) **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data** *Bioinformatics* **26**:139–40 <https://doi.org/10.1093/bioinformatics/btp616>
59. Benjamini Y, Hochberg Y. 1995) **Controlling the false discovery rate: a practical and powerful approach to multiple testing** *J R Stat Soc Ser B* **57**:289–300
60. Lai WY, Schlötterer C. 2022) **Evolution of phenotypic variance in response to a novel hot environment** *Molecular ecology* **31**:934–45 <https://doi.org/10.1111/mec.16274>
61. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, GuhaThakurta D, et al. 2005) **An integrative genomics approach to infer causal associations between gene expression and disease** *Nature genetics* **37**:710–7 <https://doi.org/10.1038/ng1589>
62. Wright S. 1920) **Correlation and causation Part I Method of path coefficients.** *J Agric Res* **20**:557–85
63. Vlachos C, Kofler R. 2018) **MimicrEE2: Genome-wide forward simulations of Evolve and Resequencing studies** *PLoS computational biology* **14**:e1006413 <https://doi.org/10.1371/journal.pcbi.1006413>
64. Bohutinska M, Vlcek J, Yair S, Laenen B, Konecna V, Fracassetti M, et al. 2021) **Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives** *Proceedings of the National Academy of Sciences of the United States of America* **118** <https://doi.org/10.1073/pnas.2022713118>
65. Hämälä T, Guiltinan MJ, Marden JH, Maximova SN, dePamphilis CW, Tiffin P. 2020) **Gene Expression Modularity Reveals Footprints of Polygenic Adaptation in *Theobroma cacao*** *Molecular biology and evolution* **37**:110–23 <https://doi.org/10.1093/molbev/msz206>
66. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, et al. 2009) **Systems genetics of complex traits in *Drosophila melanogaster*** *Nature genetics* **41**:299–307 <https://doi.org/10.1038/ng.332>
67. Laruson AJ, Yeaman S, Lotterhos KE 2020) **The Importance of Genetic Redundancy in Evolution** *Trends in ecology & evolution* <https://doi.org/10.1016/j.tree.2020.04.009>
68. Thompson KA, Osmond MM, Schluter D. 2019) **Parallel genetic evolution and speciation from standing variation** *Evol Lett* **3**:129–41 <https://doi.org/10.1002/evl3.106>
69. Lawniczak MK, Holloway AK, Begun DJ, Jones CD 2008) **Genomic analysis of the relationship between gene expression variation and DNA polymorphism in *Drosophila simulans*** *Genome biology* **9**:R125 <https://doi.org/10.1186/gb-2008-9-8-r125>
70. Swain PS, Elowitz MB, Siggia ED 2002) **Intrinsic and extrinsic contributions to stochasticity in gene expression** *Proceedings of the National Academy of Sciences of the United States of America* **99**:12795–800 <https://doi.org/10.1073/pnas.162041399>

71. Elowitz MB, Levine AJ, Siggia ED, Swain PS 2002) **Stochastic gene expression in a single cell** *Science* **297**:1183–6 <https://doi.org/10.1126/science.1070919>
72. Blake WJ M KA, Cantor CR, Collins JJ 2003) **Noise in eukaryotic gene expression** *Nature* **422**:633–7 <https://doi.org/10.1038/nature01546>
73. Frachon L, Libourel C, Villoutreix R, Carrere S, Glorieux C, Huard-Chauveau C, et al. 2017) **Intermediate degrees of synergistic pleiotropy drive adaptive evolution in ecological time** *Nat Ecol Evol* **1**:1551–61 <https://doi.org/10.1038/s41559-017-0297-1>

Author information

Wei-Yun Lai

Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria, Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria
ORCID iD: [0000-0002-5101-8695](https://orcid.org/0000-0002-5101-8695)

Sheng-Kai Hsu

Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria, Vienna Graduate School of Population Genetics, Vetmeduni Vienna, Vienna, Austria
ORCID iD: [0000-0002-6942-7163](https://orcid.org/0000-0002-6942-7163)

Andreas Futschik

Department of Applied Statistics, Johannes Kepler University Linz, Linz, Austria

Christian Schlötterer

Institut für Populationsgenetik, Vetmeduni Vienna, Vienna, Austria
ORCID iD: [0000-0003-4710-6526](https://orcid.org/0000-0003-4710-6526)

For correspondence: schlote@gmail.com

Editors

Reviewing Editor

Frank Chan

University of Groningen, Groningen, Netherlands

Senior Editor

Detlef Weigel

Max Planck Institute for Biology Tübingen, Tübingen, Germany

Reviewer #1 (Public review):

When different groups (populations, species) are presented with similar environmental pressures, how similar are the ultimate targets (genes, pathways)? This study sought to illuminate this broader question via experimental evolution in *D. simulans* and quantifying gene-expression changes, specifically in the context of standing genetic variation (and not de novo mutation). Ultimately, the authors showed pleiotropy and standing-genetic variation play a significant role in the "predictability" of evolution.

The results of this manuscript look at the interplay between pleiotropy, standing genetic variation and parallelism (i.e. predictability of evolution) in gene expression. Ultimately, their results suggest that (a) pleiotropic genes typically have a smaller range in variation/expression, and (b) adaptation to similar environments tends to favor changes in pleiotropic genes, which leads to parallelism in mechanisms (though not dramatically). However, it is still uncertain how much parallelism is directly due to pleiotropy, instead of a complex interplay between them and ancestral variation.

<https://doi.org/10.7554/eLife.102321.2.sa3>

Reviewer #2 (Public review):

Summary:

Lai and collaborators use a previously published RNAseq dataset derived from an experimental evolution set up to compare the pleiotropic properties of genes which expression evolved in response to fluctuating temperature for over 100 generations. The authors correlate gene pleiotropy with the degree of parallelisms in the experimental evolution set up to ask: are genes that evolved in multiple replicates more or less pleiotropic?

They find that, maybe counter to expectation, highly pleiotropic genes show more replicated evolution. And such effect seems to be driven by direct effects (which the authors can only speculate on) and indirect effect through low variance in pleiotropic genes (which the authors indirectly link to genetic variation underlying gene expression variance).

Weaknesses:

The results offer new insights into the evolution of gene expression and into the parameters that constrain such evolution, i.e., pleiotropy. Although the conclusions are supported by the data, I find the interpretation of the results a little bit complicated.

Major comment:

The major point I ask the authors to address is whether the connection between polygenic adaptation and parallelism can indeed be used to interpret gene expression parallelism. If the answer is not, please rephrase the introduction and discussion, if the answer is yes, please make it explicit in the text why it is so.

The authors argument: parallelism in gene expression is the same as parallelism in SNP allele frequency (AFC) (see L389-383 here they don't mention that this explanation is derived from SNP parallelism and not trait parallelism, and see Fig1 b). In previous publications the authors have explained the low level of AFC parallelism using a polygenic argument. Polygenic traits can reach a new trait optimum via multiple SNPs and therefore although the trait is parallel across replicates, the SNPs are not necessarily so.

In the current paper, they seem to be exchanging SNP AFC by gene expression, and to me, those are two levels that cannot be interchanged. Gene expression is a trait, not a SNP, and therefore the fact that a gene expression doesn't replicate cannot be explained by polygenic basis, because again the trait is gene expression itself. And, actually the results of the simulations show that high polygenicity = less trait parallelism (Fig4).

Now, if the authors focus on high parallel genes (present in e.g. 7 or more replicates) and they show that the eQTLs for those genes are many (highly polygenic) and the AFC of those eQTL are not parallel, then I would agree with the interpretation. But, given that here they just assess gene expression and not eQTL AFC, I do not think they can use the 'highly polygenic = low parallelism' explanation.

The interpretation of the results to me, should be limited to: genes with low variance and high pleiotropy tend to be more parallel, and the explanation might be synergistic pleiotropy.

Comments on revisions: The authors didn't really address any of the comments made by any of the reviewers - basically nothing was changed in the main text. Therefore, I leave my original review unchanged.

<https://doi.org/10.7554/eLife.102321.2.sa2>

Reviewer #3 (Public review):

The authors aim to understand how gene pleiotropy affects parallel evolutionary changes among independent replicates of adaptation to a new hot environment of a set of experimental lines of *Drosophila simulans* using experimental evolution. The flies were RNAsequenced after more than 100 generations of lab adaptation and the changes in average gene expression were obtained relative to ancestral expression levels from reconstructed ancestral lines. Parallelism of gene expression change among lines is evaluated as variance in differential gene expression among lines relative to error variance. Similarly, the authors ask how the standing variation in gene expression estimated from a handful of flies from a reconstructed outbred line affects parallelism. The main findings are that parallelism in gene expression responses is positively associated with pleiotropy and negatively associated with expression variation. Those results are in contradiction with theoretical predictions and empirical findings. To explain those seemingly contradictory results the authors invoke the role of synergistic pleiotropy and correlated selection, although they do not attempt to measure either.

Strengths:

The study uses highly replicated outbred laboratory lines of *Drosophila simulans* evolved in the lab under constant hot regime for over 100 generations. This allows for robust comparisons of evolutionary responses among lines.

The manuscript is well written and the hypotheses are clearly delineated at the onset.

The authors have run a causal analysis to understand the causal dependencies between pleiotropy and expression variation on parallelism.

The use of whole-body RNA extraction to study gene expression variation is well justified.

Weaknesses:

The accuracy of the estimate of ancestral phenotypic variation in gene expression is likely low because estimated from a small sample of 20 males from a reconstructed outbred line. It might not constitute a robust estimate of the genetic variation of the evolved lines under study.

There are no estimates of the standing genetic variation of expression levels of the genes under study, only estimates of their phenotypic variation. I wished the authors had been clear about that limitation and had refrained from equating phenotypic variation in expression level with standing genetic variation.

Moreover, since the phenotype studied is gene expression, its genetic basis extends beyond expressed sequences. The phenotypic variation of a gene's expression may thus likely misrepresent the genetic variation available for its evolution. The authors do not present evidence that sequence variation correlates with expression variation.

The authors have not attempted to estimate synergistic pleiotropy among genes, nor how selection acts on gene expression modules. It makes their conclusion regarding the role of synergistic pleiotropy rather speculative.

<https://doi.org/10.7554/eLife.102321.2.sa1>

Author response:

The following is the authors' response to the current reviews.

Reviewer #1 (Public review):

*When different groups (populations, species) are presented with similar environmental pressures, how similar are the ultimate targets (genes, pathways)? This study sought to illuminate this broader question via experimental evolution in *D. simulans* and quantifying gene-expression changes, specifically in the context of standing genetic variation (and not de novo mutation). Ultimately, the authors showed pleiotropy and standing-genetic variation play a significant role in the "predictability" of evolution.*

The results of this manuscript look at the interplay between pleiotropy, standing genetic variation and parallelism (i.e. predictability of evolution) in gene expression. Ultimately, their results suggest that (a) pleiotropic genes typically have a smaller range in variation/expression, and (b) adaptation to similar environments tends to favor changes in pleiotropic genes, which leads to parallelism in mechanisms (though not dramatically). However, it is still uncertain how much parallelism is directly due to pleiotropy, instead of a complex interplay between them and ancestral variation.

Yes, the reviewer is correct that our results for the direct effects of pleiotropy were not consistent for both measures of pleiotropy. We highlight this in the discussion: "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 connectivity. 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."

Reviewer #2 (Public review):

Summary:

Lai and collaborators use a previously published RNAseq dataset derived from an experimental evolution set up to compare the pleiotropic properties of genes which expression evolved in response to fluctuating temperature for over 100 generations. The authors correlate gene pleiotropy with the degree of parallelisms in the experimental evolution set up to ask: are genes that evolved in multiple replicates more or less pleiotropic?

They find that, maybe counter to expectation, highly pleiotropic genes show more replicated evolution. And such effect seems to be driven by direct effects (which the authors can only speculate on) and indirect effect through low variance in pleiotropic genes (which the authors indirectly link to genetic variation underlying gene expression variance).

Weaknesses:

The results offer new insights into the evolution of gene expression and into the parameters that constrain such evolution, i.e., pleiotropy. Although the conclusions are supported by the data, I find the interpretation of the results a little bit complicated.

We are very happy to read that the reviewer finds our conclusions to be supported by the data.

Major comment:

The major point I ask the authors to address is whether the connection between polygenic adaptation and parallelism can indeed be used to interpret gene expression parallelism. If the answer is not, please rephrase the introduction and discussion, if the answer is yes, please make it explicit in the text why it is so.

Yes, we think that gene expression parallelism can be explained by polygenic adaptation.

The authors argument: parallelism in gene expression is the same as parallelism in SNP allele frequency (AFC) (see L389-383 here they don't mention that this explanation is derived from SNP parallelism and not trait parallelism, and see Fig1 b). In previous publications the authors have explained the low level of AFC parallelism using a polygenic argument. Polygenic traits can reach a new trait optimum via multiple SNPs and therefore although the trait is parallel across replicates, the SNPs are not necessarily so.

In the current paper, they seem to be exchanging SNP AFC by gene expression, and to me, those are two levels that cannot be interchanged. Gene expression is a trait, not a SNP, and therefore the fact that a gene expression doesn't replicate cannot be explained by polygenic basis, because again the trait is gene expression itself. And, actually the results of the simulations show that high polygenicity = less trait parallelism (Fig4).

We agree with the reviewer that it is important to consider different hierarchies when talking about the implications of polygenic adaptation. The lowest hierarchical level is SNP variation and the highest level is fitness. In-between these extreme hierarchical levels is gene expression. While gene expression is a trait itself, as correctly pointed out by the reviewer, it is possible that selection is not favoring a specific trait value, because selection targets a trait on a higher hierarchical level. This implies that not only SNPs, but also intermediate traits such as gene expression can exhibit redundancy. Considering a simple example of one selected trait (e.g. body size), which is affected by the expression level of two genes A and B, each regulated by SNP A1, A2 and B1, B2. It is now possible to modulate the focal trait by allele frequency changes of A1, which in turn will only affect gene A. Alternatively, SNP B2 may change, modifying the expression of gene B, leading to the same change in body size. Hence, we could have redundancy both at the SNP level as well as on the gene expression level (although higher redundancy is expected on the SNP level). Most importantly, this redundancy at intermediate hierarchical levels is not pure theory, but it is supported by empirical evidence. We have shown that redundancy exists not only for gene expression (10.1111/mec.16274) but also for metabolite concentrations (10.1093/gbe/evad098).

Now, if the authors focus on high parallel genes (present in e.g. 7 or more replicates) and they show that the eQTLs for those genes are many (highly polygenic) and the AFC of those eQTL are not parallel, then I would agree with the interpretation. But, given that here they just assess gene expression and not eQTL AFC, I do not think they can use the 'highly polygenic = low parallelism' explanation.

This is clearly an interesting proposed research project, but we doubt that it would result in the expected outcome. Since most of the adaptive gene expression changes are not having a simple genetic basis (10.1093/gbe/evae077) and most expression variation is determined by trans-regulatory effects (10.1038/s41576-020-00304-w), eQTL mapping will most likely not identify all contributing loci. Large effect loci are more easily identified, but they are also expected to be more parallel.

The interpretation of the results to me, should be limited to: genes with low variance and high pleiotropy tend to be more parallel, and the explanation might be synergistic pleiotropy.

We thank the reviewer for the suggestion, but prefer to stick to our interpretation of the data.

Comments on revisions: The authors didn't really address any of the comments made by any of the reviewers - basically nothing was changed in the main text. Therefore, I leave my original review unchanged.

We modestly disagree, in our point to point reply, we respond to all reviewers' comments. Since, we did not identify any major problem in our manuscript, we only modified the wording in some parts where we felt that a clarification could resolve the misunderstanding of the reviewers. In response to the reviewers' comments, we added a new paragraph in the discussion and generated a new figure.

Reviewer #3 (Public review):

*The authors aim to understand how gene pleiotropy affects parallel evolutionary changes among independent replicates of adaptation to a new hot environment of a set of experimental lines of *Drosophila simulans* using experimental evolution. The flies were RNAsequenced after more than 100 generations of lab adaptation and the changes in average gene expression were obtained relative to ancestral expression levels from reconstructed ancestral lines. Parallelism of gene expression change among lines is evaluated as variance in differential gene expression among lines relative to error variance. Similarly, the authors ask how the standing variation in gene expression estimated from a handful of flies from a reconstructed outbred line affects parallelism. The main findings are that parallelism in gene expression responses is positively associated with pleiotropy and negatively associated with expression variation. Those results are in contradiction with theoretical predictions and empirical findings. To explain those seemingly contradictory results the authors invoke the role of synergistic pleiotropy and correlated selection, although they do not attempt to measure either.*

Strengths:

*The study uses highly replicated outbred laboratory lines of *Drosophila simulans* evolved in the lab under constant hot regime for over 100 generations. This allows for robust comparisons of evolutionary responses among lines.*

The manuscript is well written and the hypotheses are clearly delineated at the onset.

The authors have run a causal analysis to understand the causal dependencies between pleiotropy and expression variation on parallelism.

The use of whole-body RNA extraction to study gene expression variation is well justified.

Weaknesses:

The accuracy of the estimate of ancestral phenotypic variation in gene expression is likely low because estimated from a small sample of 20 males from a reconstructed outbred line. It might not constitute a robust estimate of the genetic variation of the evolved lines under study.

We agree with the reviewer that variation estimates based on 20 samples are not very precise. Nevertheless, we demonstrated that the estimated variance in gene expression was highly correlated between two independent samples from the same ancestral population. Furthermore, we identified a significant correlation of expression variance with evolutionary parallelism. In other words, the biological signal has been sufficiently strong despite the variance estimate has been noisy.

There are no estimates of the standing genetic variation of expression levels of the genes under study, only estimates of their phenotypic variation. I wished the authors had been clear about that limitation and had refrained from equating phenotypic variation in expression level with standing genetic variation.

The reviewer is right that we did not estimate genetic variation of gene expression, but use expression variation as a proxy for the standing genetic variation. There are two potential problems with this approach. First, a large expression variation could be caused by a single large effect variant segregating at intermediate frequency. Such large effect variants will exhibit a highly parallel selection response-contrary to our empirical results. Since we have shown previously (10.1093/gbe/evae077) that adaptive gene expression changes are mostly polygenic we do not consider this extreme scenario to be very relevant in our study. Rather, we would like to emphasize that neither a SNP analysis of the 5' region nor an eQTL study will provide an unbiased estimator of genetic variation of gene expression. The second problem arises if gene expression noise differs among genes, hence more noisy genes will appear to have more standing genetic variation than genes with less noise. Since, we average across many different cells and cell types, gene expression noise is expected to be levelled out- this aspect is discussed in detail in the manuscript.

In other words, despite these two potential limitations, we consider our approach superior to alternative approaches of estimating genetic variation in gene expression.

Moreover, since the phenotype studied is gene expression, its genetic basis extends beyond expressed sequences. The phenotypic variation of a gene's expression may thus likely misrepresent the genetic variation available for its evolution. The authors do not present evidence that sequence variation correlates with expression variation.

Gene expression is determined by the joint effects of cis-regulatory and trans-regulatory variation. Hence, recombination can create more extreme phenotypes than the one of the parental lines (in quantitative genetics this is called transgressive segregation). It is unclear to what extent this constitutes a problem for our analyses. Nevertheless, we would like to point out that eQTL mapping will miss many trans-acting variants and therefore we doubt that the requested empirical evidence for correlation between genetic variation (estimated by eQTL mapping) and observed expression variation is as straight forward as suggested by the reviewer.

Nevertheless, we reference an empirical study, which showed a positive correlation between expression variation and cis-regulatory variation.

The authors have not attempted to estimate synergistic pleiotropy among genes, nor how selection acts on gene expression modules. It makes their conclusion regarding the role of synergistic pleiotropy rather speculative.

The reviewer is correct that we did not demonstrate synergistic pleiotropy, but we discuss this as a possible explanation for the observed direct effects of pleiotropy.

The following is the authors' response to the original reviews.

Reviewer #1 (Public review):

The results of this manuscript look at the interplay between pleiotropy, standing genetic variation, and parallelism (i.e. predictability of evolution) in gene expression. Ultimately, their results suggest that (a) pleiotropic genes typically have a smaller range in variation/expression, and (b) adaptation to similar environments tends to favor changes in pleiotropic genes, which leads to parallelism in mechanisms (though not dramatically). However, it is still uncertain how much parallelism is directly due to pleiotropy, instead of a complex interplay between them and ancestral variation.

I have a few things that I was uncertain about. It may be these things are easily answered but require more discussion or clarity in the manuscript.

(1) The variation being talked about in this manuscript is expression levels, and not SNPs within coding regions (or elsewhere). The cause of any specific gene having a change in expression can obviously be varied - transcription factors, repressors, promoter region variation, etc. Is this taken into account within the "network connectivity" measurement? I understand the network connectivity is a proxy for pleiotropy - what I'm asking is, conceptually, what can be said about how/why those highly pleiotropic genes have a change (or not) in expression. This might be a question for another project/paper, but it feels like a next step worth mentioning somewhere.

In current study, we are only able to detect significant and repeatable expression changes but unable to identify the underlying causal variants. An eQTL study in the founder population in combination with genomic resequencing for both evolved and ancestral populations would be required to address this question.

(2) The authors do have a passing statement in line 361 about cis-regulatory regions. Is the assumption that genetic variation in promoter regions is the ultimate "mechanism" driving any change in expression? In the same vein, the authors bring up a potential confounding factor, though they dismiss it based on a specific citation (lines 476-481; citation 65). I'm of the mindset that in order to more confidently disregard this "issue" based on previous evidence, it requires more than one citation. Especially since the one citation is a plant. That specific point jumps out to me as needing a more careful rebuttal.

It was not our intention to claim that the expression changes in our experiment are caused by cis-regulatory variation only. We believe that the observed expression variation has both cis- and trans-genetic components, where as some studies tend to estimate much higher cisvariation for gene expression in *Drosophila* populations (e.g. [1, 2]). We mentioned the positive correlation between cis-regulatory polymorphism and expression variation to (1) highlight the genetic control of gene expression and (2) make the connection between polygenic adaptation and gene expression evolutionary parallelism.

(3) I feel like there isn't enough exploration of tissue specificity versus network connectivity. Tissue specificity was best explained by a model in which pleiotropy had both direct and indirect effects on parallelism; while network connectivity was best explained (by a small margin) via the model which was mostly pleiotropy having a direct effect on ancestral variation, that then had a direct effect on parallelism. When the

strengths of either direct/indirect effects were quantified, tissue specificity showed a stronger direct effect, while network connectivity had none (i.e. not significant). My confusion is with the last point - if network connectivity is explained by a direct effect in the best-supported model, how does this work, since the direct effect isn't significant? Perhaps I am misunderstanding something.

To clarify, for network connectivity, there's a significant "indirect" effect on parallelism (i.e. network connectivity affect ancestral gene expression and ancestral gene expression affect parallelism). Hence, in table 2, the direct effect of network connectivity on parallelism is weak and not significant while the indirect effect via ancestral variation is significant.

Also, network connectivity might favor the most pleiotropic genes being transcription factor hubs (or master regulators for various homeostasis pathways); while the tissue specificity metric perhaps is a kind of a space/time element. I get that a gene having expression across multiple tissues does fit the definition of pleiotropy in the broad sense, but I'm wondering if some important details are getting lost - I'm just thinking about the relative importance of what tissue specificity measurements say versus the network connectivity measurement.

We examined the statistical relationship between the two measures and found a moderate positive correlation on the basis of which we argued that the two measures may capture different aspects of pleiotropy. We appreciate the reviewer's suggestions about the biological basis of the two estimates of pleiotropy, but we think that without further experimental insights, an extended discussion of this topic is too premature to provide meaningful insights to the readership.

Reviewer #2 (Public review):

Summary:

Lai and collaborators use a previously published RNAseq dataset derived from an experimental evolution set up to compare the pleiotropic properties of genes whose expression evolved in response to fluctuating temperature for over 100 generations. The authors correlate gene pleiotropy with the degree of parallelisms in the experimental evolution set up to ask: are genes that evolved in multiple replicates more or less pleiotropic?

They find that, maybe counter to expectation, highly pleiotropic genes show more replicated evolution. Such an effect seems to be driven by direct effects (which the authors can only speculate on) and indirect effects through low variance in pleiotropic genes (which the authors indirectly link to genetic variation underlying gene expression variance).

Weaknesses:

The results offer new insights into the evolution of gene expression and into the parameters that constrain such evolution, i.e., pleiotropy. Although the conclusions are supported by the data, I find the interpretation of the results a little bit complicated.

Major comment:

The major point I ask the authors to address is whether the connection between polygenic adaptation and parallelism can indeed be used to interpret gene expression parallelism. If the answer is not, please rephrase the introduction and discussion, if the answer is yes, please make it explicit in the text why it is so.

Our answer is yes, we interpreted gene expression parallelism (high ancestral variance -> less parallelism) using the same framework that links polygenic adaptation and parallelism (high polygenicity = less trait parallelism). We believe that our response covers several of the reviewer's concerns.

The authors' argument: parallelism in gene expression is the same as parallelism in SNP allele frequency (AFC) (see L389-383 here they don't mention that this explanation is derived from SNP parallelism and not trait parallelism, and see Figure 1 b). In previous publications, the authors have explained the low level of AFC parallelism using a polygenic argument. Polygenic traits can reach a new trait optimum via multiple SNPs and therefore although the trait is parallel across replicates, the SNPs are not necessarily so.

Importantly, our rationale is based on the idea that gene expression is rarely the direct target of selection, but rather an intermediate trait [3]. Recently, we have specifically tested this assumption for gene expression and metabolite concentrations and our analysis showed that both traits were redundant [4], as previously shown for DNA sequences [5]. The important implication for this manuscript is that gene expression is also redundant, so that adaptation can be achieved by distinct changes in gene expression in replicate populations adapting to the same selection pressure. This implies that we can use the same simulation framework for gene expression as for sequencing data. In our case different SNP frequencies correspond to different expression levels (averaged across individuals from a population), which in turn increases fitness by modifying the selected trait. Importantly, the selected trait in our simulations is not gene expression, but a not defined high level phenotype. A key insight from our simulations is that with increasing polygenicity the expression of a gene is more variable in the ancestral population.

In the current paper, they seem to be exchanging SNP AFC by gene expression, and to me, those are two levels that cannot be interchanged. Gene expression is a trait, not an SNP, and therefore the fact that a gene expression doesn't replicate cannot be explained by a polygenic basis, because again the trait is gene expression itself. And, actually, the results of the simulations show that high polygenicity = less trait parallelism (Figure 4).

As detailed above, because adaptation can be reached by changes in gene expression at different sets of genes, redundancy is also operating on the expression level not just on the level of SNPs. To clarify, the x-axis of Fig. 4 is the expression variation in the ancestral population.

Now, if the authors focus on high parallel genes (present in e.g. 7 or more replicates) and they show that the eQTLs for those genes are many (highly polygenic) and the AFC of those eQTLs are not parallel, then I would agree with the interpretation. But, given that here they just assess gene expression and not eQTL AFC, I do not think they can use the 'highly polygenic = low parallelism' explanation.

The interpretation of the results to me, should be limited to: genes with low variance and high pleiotropy tend to be more parallel, and the explanation might be synergistic pleiotropy.

While we understand the desire to model the full hierarchy from eQTLs to gene expression and adaptive traits, we raise caution that this would be a very challenging task. eQTLs very often underestimate the contribution of trans-acting factors, hence the understanding of gene expression evolution based on eQTLs is very likely incomplete and cannot explain the redundancy of gene expression during adaptation. Hence, we think that the focus on

redundant gene expression is conceptually simpler and thus allows us to address the question of pleiotropy without the incorporation of allele frequency changes.

Reviewer #3 (Public review):

*The authors aim to understand how gene pleiotropy affects parallel evolutionary changes among independent replicates of adaptation to a new hot environment of a set of experimental lines of *Drosophila simulans* using experimental evolution. The flies were RNAsequenced after more than 100 generations of lab adaptation and the changes in average gene expression were obtained relative to ancestral expression levels from reconstructed ancestral lines. Parallelism of gene expression change among lines is evaluated as variance in differential gene expression among lines relative to error variance. Similarly, the authors ask how the standing variation in gene expression estimated from a handful of flies from a reconstructed outbred line affects parallelism. The main findings are that parallelism in gene expression responses is positively associated with pleiotropy and negatively associated with expression variation. Those results are in contradiction with theoretical predictions and empirical findings. To explain those seemingly contradictory results the authors invoke the role of synergistic pleiotropy and correlated selection, although they do not attempt to measure either.*

Strengths:

- (1) The study uses highly replicated outbred laboratory lines of *Drosophila simulans* evolved in the lab under a constant hot regime for over 100 generations. This allows for robust comparisons of evolutionary responses among lines.*
- (2) The manuscript is well written and the hypotheses are clearly delineated at the onset.*
- (3) The authors have run a causal analysis to understand the causal dependencies between pleiotropy and expression variation on parallelism.*
- (4) The use of whole-body RNA extraction to study gene expression variation is well justified.*

Weaknesses:

- (1) It is unclear how well phenotypic variation in gene expression of the evolved lines has been estimated by the sample of 20 males from a reconstructed outbred line not directly linked to the evolved lines under study. I see this as a general weakness of the experimental design.*

Our intention was not to measure the phenotypic variance of the evolved lines, but rather to estimate the phenotypic variance at the beginning of the experiment. Hence, we measured and investigated the variation of gene expression in the ancestral population since this was the beginning of the replicated experimental evolution. Furthermore, since the ancestral population represents the natural population in Florida, the gene expression variation reflects the history of selection history acting on it.

- (2) There are no estimates of standing genetic variation of expression levels of the genes under study, only phenotypic variation. I wished the authors had been clear about that limitation and had discussed the consequences of the analysis. This also constitutes a weakness of the study.*

The reviewer is correct that we do not aim to estimate the standing genetic variation, which is responsible for differences in gene expression. While we agree that it could be an interesting research question to use eQTL mapping to identify the genetic basis of gene expression, we caution that trans-effects are difficult to estimate and therefore an important

component of gene expression evolution will be difficult to estimate. Hence, we consider that our focus on variation in gene expression without explicit information about the genetic basis is simpler and sufficient to address the question about the role of pleiotropy.

(3) Moreover, since the phenotype studied is gene expression, its genetic basis extends beyond expressed sequences. The phenotypic variation of a gene's expression may thus likely misrepresent the genetic variation available for its evolution. The genetic variation of gene expression phenotypes could be estimated from a cross or pedigree information but since individuals were pool-sequenced (by batches of 50 males), this type of analysis is not possible in this study.

We agree with the reviewer that gene expression variation may also have a non-genetic basis, we discuss this in depth in the discussion of the manuscript.

(4) The authors have not attempted to estimate synergistic pleiotropy among genes, nor how selection acts on gene expression modules. It makes any conclusion regarding the role of synergistic pleiotropy highly speculative.

We mentioned synergistic pleiotropy as a possible explanation for our results. A positive correlation between the fitness effect of gene expression variation would predict more replicable evolutionary changes. A similar argument has been made by [6].

I don't understand the reason why the analysis would be restricted to significantly differentially expressed genes only. It is then unclear whether pleiotropy, parallelism, and expression variation do play a role in adaptation because the two groups of adaptive and non-adaptive genes have not been compared. I recommend performing those comparisons to help us better understand how "adaptive" genes differentially contribute to adaptation relative to "nonadaptive" genes relative to their difference in population and genetic properties.

We agree with the reviewer that the comparison between the pleiotropy of adaptive and nonadaptive genes is interesting. We performed the analysis but omitted from the current manuscript for simplicity. Similar to the results in [6], non-adaptive genes are more pleiotropic than the adaptive genes. For adaptive genes we find a positive correlation between the level of pleiotropy and evolutionary parallelism. Thus, high pleiotropy limits the evolvability of a gene, but moderate and potentially synergistic pleiotropy increases the repeatability of adaptive evolution. We included this result in the revised manuscript and discuss it.

There is a lack of theoretical groundings on the role of so-called synergistic pleiotropy for parallel genetic evolution. The Discussion does not address this particular prediction. It could be removed from the Introduction.

We modestly disagree with the reviewer, synergistic pleiotropy is covered by theory and empirical results also support the importance of synergistic pleiotropy.

References

- (1) Genissel A, McIntyre LM, Wayne ML, Nuzhdin SV. Cis and trans regulatory effects contribute to natural variation in transcriptome of *Drosophila melanogaster*. *Molecular biology and evolution*. 2008;25(1):101-10. Epub 20071112. doi: 10.1093/molbev/msm247. PubMed PMID: 17998255.
- (2) Osada N, Miyagi R, Takahashi A. Cis- and Trans-regulatory Effects on Gene Expression in a Natural Population of *Drosophila melanogaster*. *Genetics*. 2017;206(4):2139-48. Epub

20170614. doi: 10.1534/genetics.117.201459. PubMed PMID: 28615283; PubMed Central PMCID: PMC5560811.

(3) 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.

(4) Lai WY, Otte KA, Schlötterer C. Evolution of Metabolome and Transcriptome Supports a Hierarchical Organization of Adaptive Traits. *Genome biology and evolution*. 2023;15(6). Epub 2023/05/26. doi: 10.1093/gbe/evad098. PubMed PMID: 37232360; PubMed Central PMCID: PMC610246829.

(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) 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: PMC69306781.

<https://doi.org/10.7554/eLife.102321.2.sa0>