Experimental evidence that phenotypic evolution but not plasticity occurs along genetic lines of least resistance in homogeneous environments

Greg M Walter

Email: greg.walter@monash.edu

Address: School of Biological Sciences, Monash University, Melbourne 3800

ORCID: https://orcid.org/0000-0002-0883-3440

Abstract

Genetic correlations concentrate genetic variation in certain directions of the multivariate phenotype. Adaptation and, under some models, plasticity is expected to occur in the direction of the phenotype containing the greatest amount of genetic variation ($\boldsymbol{g}_{max}$). However, this may hinge upon environmental heterogeneity, which can affect patterns of genetic variation. I use experimental evolution to test whether plasticity and phenotypic evolution follow $\boldsymbol{g}_{max}$ during adaptation to environments that varied in environmental heterogeneity. For >25 generations, *Drosophila melanogaster* populations were exposed to six homogeneous or spatially and temporally heterogeneous treatments involving hot (25ºC) and cold (16ºC) temperatures. Five wing traits were assayed in both temperatures. Wing morphology diverged between populations evolving in homogeneous hot and cold temperatures in a direction of the phenotype containing a large proportion of genetic variance, and that aligned closely with $\boldsymbol{g}_{max}$ at 16ºC, but not 25ºC. Spatial heterogeneity produced an intermediate phenotype, which was associated with similar genetic variance across assay temperatures compared to all other treatments. Surprisingly, plasticity across assay temperatures evolved in a different direction to phenotypic evolution and aligned better with maternal variance than $\boldsymbol{g}_{max}$. Together, these results provide experimental evidence for evolution along genetic lines of least resistance in homogeneous environments, but no support for predicting plastic responses from the orientation of genetic variation. These results also suggest that spatial heterogeneity could maintain genetic variation that increases the stability of genetic variance across environments.

Keywords: Adaptation, additive genetic variance, environmental heterogeneity, experimental evolution, G-matrix, wing shape

Introduction

Traits rarely evolve in isolation. Instead, adaptation occurs as coordinated changes in multiple traits that leads to the evolution of a multivariate phenotype. This is because changes in any given trait will enact changes in other traits that are genetically correlated, which concentrates genetic variation in certain directions of the multivariate phenotype. For decades, evolutionary biologists have appreciated that genetic correlations can bias evolution towards certain multivariate phenotypes, which can slow adaptation if selection lies in a different direction (Arnold 1992; Chenoweth et al. 2010; Cheverud 1984; Lande 1979, 1980; McGuigan et al. 2005; Schluter 1996; Zeng 1988). If adaptation is determined by the distribution of genetic variation underlying a multivariate phenotype, rather than the direction of selection, then it is difficult to understand how rapid adaptive diversification is possible (Schluter 2000). We know, however, that genetic variation is not static and can evolve (Doroszuk et al. 2008; Jones et al. 2004; Roff and Mousseau 1999; Walter et al. 2018), or change in response to the environment (Wood and Brodie III 2015). Understanding how the environment affects genetic variation and influences phenotypic change remains a fundamental challenge in biology (Careau et al. 2015; Steppan et al. 2002).

Breeding designs are used to estimate the additive genetic component of the phenotype (i.e., variation created by different alleles), which is partitioned from the total phenotypic variance. The additive genetic variance and covariance among multiple traits, known as the G-matrix (**G**), captures the genetic architecture underlying a multivariate phenotype of the population (Lande 1979; Lynch and Walsh 1998; Walsh and Blows 2009). Genetic correlations among traits, created by pleiotropy or strong linkage, concentrate genetic variation into fewer axes (or directions in phenotypic space) than the number of traits originally measured (Arnold 1992; Lande 1980). The shape and orientation of **G** is expected to determine the direction of evolution, at least in the short term (Hansen and Houle 2008; Zu et al. 2020). The direction in phenotypic space that contains the greatest amount of genetic variation (i.e. the primary axis of **G**, $\boldsymbol{g}_{\max}$) is of particular interest because it is the direction that adaptation is predicted to follow (Marroig and Cheverud 2007; McGlothlin et al. 2018; McGuigan et al. 2005; Schluter 1996). Comparative studies have found both support for (Costa et al. 2020; Hangartner et al. 2020; McGlothlin et al. 2018; Schluter 1996), and against (Berner et al. 2010; Kimmel et al. 2012; McGuigan et al. 2005; Merilä and Björklund 1999), the theory that evolutionary trajectories are determined by the distribution and orientation of multivariate genetic variation, but we lack experimental manipulations that test whether phenotypic evolution is biased towards $\boldsymbol{g}_{\max}$. To understand how multivariate genetic variance influences evolution, we also need to identify how stable **G** is across environments, whether **G** is associated with plasticity, and how environmental heterogeneity influences genetic variance.

Although genetic correlations are expected to maintain stability in **G**, at least in the short-term (Zeng 1988), we know that **G** is not static (Arnold et al. 2008; Wood and Brodie III 2015). Changes in **G** have been observed for taxa that have adapted to contrasting habitats across relatively few generations

(Doroszuk et al. 2008; Eroukhmanoff and Svensson 2011). Within a generation, plasticity, the ability for a genotype to produce different phenotypes in response to environmental variation (Lind and Johansson 2007; Via et al. 1995), can affect genotypes differently. Genotypes that vary in plasticity (i.e. show genotype-by-environment interactions) are likely to produce changes in **G** when exposed to different environments (Sgrò and Hoffmann 2004; Walter et al. 2021; Wood and Brodie III 2015). If **G** changes rapidly in response environmental change, then it may be difficult to understand when adaptation will occur along $\boldsymbol{g}_{\max}$.

The relationship between plasticity and genetic variation during adaptation remains unclear. Under the 'plasticity-first' model, adaptation to a new environment occurs when plastic responses that are beneficial in the new environment become genetically incorporated (Levis and Pfennig 2016; West-Eberhard 2003). One mechanism that could allow an initially plastic response to become genetic is when plasticity lies in the direction of $\boldsymbol{g}_{\max}$, which means that adaptation can occur in the direction of both plasticity and genetic variation (Draghi and Whitlock 2012; Lind et al. 2015; Noble et al. 2019). However, a survey of studies concluded that the alignment between plasticity and genetic variation depended on the trait, species, and environment studied (Noble et al. 2019). Generalizing an alignment between plasticity and additive genetic variation may be difficult because the genetic basis of a trait can be different from that of its plasticity (Lafuente et al. 2018), and also because many other factors can determine plastic responses. For instance, maternal effects and epistasis were found to underlie differences in plasticity between populations of *Drosophila* from different latitudes (van Heerwaarden and Sgrò 2017). Understanding how plasticity relates to genetic variation to determine phenotypic responses to new environments therefore remains an ongoing challenge.

Environments are rarely constant and instead vary at different spatial and temporal scales that can determine how populations adapt. Environments with greater environmental heterogeneity experience more temporal or spatial variation in selection that moves the phenotypic optima, which can slow or prevent adaptation to new environments because phenotypic evolution towards any single optimum will be difficult (Jones et al. 2004; Steppan et al. 2002). As a consequence, however, greater environmental heterogeneity is expected to maintain more genetic variance in a given trait because spatial or temporal variation in selection prevents alleles from becoming fixed in the population (Bürger and Gimelfarb 2002; Gillespie and Turelli 1989; Mackay 1981; McDonald and Yeaman 2018; Via and Lande 1987). Adaptation towards a phenotypic optimum will therefore be more straightforward if conditions are relatively homogeneous, whereas populations experiencing greater heterogeneity should show smaller changes in phenotype but possess more genetic variation. By using experimental evolution with different environments (e.g., temperature) and different levels of environmental heterogeneity that represent more natural environments (i.e., fluctuations in temperature), it is possible to test whether evolution is more

predictable in homogeneous environments, and understand how genetic variance is affected by heterogeneous environments.

Due to the adaptive significance of wing size and shape in response to temperature in *Drosophila* (Cavicchi et al. 1991; Partridge et al. 1994; Santos et al. 2004), experimental evolution of wing shape is ideal to test whether changes in a multivariate phenotype can be predicted from patterns of genetic variation. Here, I use a previously published dataset to explore how phenotypic changes due to experimental evolution and plasticity are associated with additive genetic variance. For two years of experimental evolution, Yeaman et al. (2010a) exposed populations of *Drosophila melanogaster* to six treatments that differed in temperature (25ºC hot and 16ºC cold) and levels of environmental heterogeneity, created by moving flies between temperatures (**Fig. 1a**). Using a reciprocal transplant, they found that homogeneous hot and cold treatments showed greater performance in the environment in which they evolved, providing evidence for thermal adaptation. They then conducted experimental crosses to produce ~300 full-sibling families within each treatment and measured five wing traits (**Fig. 1b**) on the parents and multiple offspring to quantify mean phenotype and genetic variance in each wing trait. Wings were measured under both temperatures that capture plastic changes between the hot and cold treatments. Yeaman et al. (2010a) found that although wing size did not evolve in response to hot and cold treatments, wing angles did. In particular, wing angle 7-8-9 (**Fig. 1b**) was greater in the hot than cold treatments, which corresponds with thinner, longer wings under colder temperatures and suggests adaptive divergence in wing shape (Cavicchi et al. 1991; Debat et al. 2003; Partridge et al. 1994; Santos et al. 2004). Surprisingly, all treatments showed the same amount of genetic variance for all five wing traits, suggesting that greater environmental heterogeneity did not maintain higher levels of genetic variance in each univariate trait.

In this study, I use the data from Yeaman et al. (2010a) and extend the analyses by taking a multivariate approach to test whether evolved and plastic changes in wing shape align with $\boldsymbol{g}_{\max}$, and whether environmental heterogeneity influenced the outcome. I predicted that populations from hot and cold homogeneous environments would show: (1) Little difference in multivariate genetic variance (**Fig. 2a-b**); (2) Large changes in mean phenotype due to evolved differences between hot and cold treatments, as well as due to plasticity across assay temperatures (**Fig. 2a, c**); and (3) Changes in mean phenotype due to plasticity and evolution would align with $\boldsymbol{g}_{\max}$, the direction of the phenotype containing the greatest amount of genetic variance, assuming that selection has not depleted genetic variance too much (**Fig. 2a, d-e**). By contrast, I predicted that populations exposed to spatial and temporal environmental heterogeneity would show less of an evolved change in mean phenotype compared to populations from homogeneous environments, and instead show changes in the distribution or orientation of genetic variation (**Fig. 2a-b**).

Methods

Maintenance of experimental lines

Data for these analyses were obtained from Yeaman et al. (2010b). The maintenance of the experimental lines is described in detail in Yeaman et al. (2010a) and will only be reviewed briefly here. A mass-bred population was established from a large sample of *Drosophila melanogaster* collected in September 2005 from a permanent compost pile at Cawston (British Columbia, Canada). From the wild-caught flies, 298 isofemale lines were established over three generations at ambient temperature in the laboratory (20-23ºC), which were then used to create a mass-bred population that was maintained in 32 cages (8 bottles/cage; cage dimensions: 22cm×25cm×32cm) for six further generations. Populations for experimental evolution were established by allowing flies to lay eggs for 2 days in 400 bottles that were then distributed evenly among 50 cages, which were randomly assigned to six different temperature/heterogeneity treatments with five replicate populations (cages) per treatment (**Fig. 1a**). The six treatments included: Homogeneous (1) hot (25°C) and (2) cold (16°C) treatments, (3) spatial heterogeneity, where half the bottles were moved between hot and cold cages, (4) temporal heterogeneity created by moving the cages between hot and cold temperatures every four weeks, and homogeneous (5) hot and (6) cold treatments with limited migration, created by moving two mated females between hot and cold treatments every four weeks (**Fig. 1a**). Spatial heterogeneity was therefore created by moving half the population between cages, producing a migration rate of m~0.5 (proportion of individuals replaced per generation) to create a panmictic population experiencing two environments. By contrast, the homogeneous treatments with migration experienced a migration rate of around m~0.001, which simulates two spatially isolated subpopulations in different environments that experience a small amount of gene flow. After 116 weeks (29 vs 58 generations for the cold and hot treatments), the wing morphology of each treatment was assayed at 16°C and 25°C.

For each treatment (and replicate), 30 randomly picked eggs were placed in each of 20 vials. The vials were split across the two assay temperatures and maintained for three generations to prevent maternal environmental effects biasing the phenotype. To establish the parental generation, 60 full-sibling families were established for each replicate line by mating a male and female randomly selected from each vial (n=60 families per replicate, n=300 families per assay, n=600 families per treatment). Parents and four offspring (two males and two females) from each full-sibling family were measured (total N≈19,000 individuals). Five wing traits were measured from the landmarks of each wing that included wing size, three angles and a distance (**Fig. 1b**; Yeaman et al. 2010a). Wings were measured by the same person and families were processed in blocks of four from each of the six experimental treatments (24 families per block) so that any block effects would be shared across all treatments.

Estimation of G-matrices

Table 1 contains descriptions of all matrices and vectors. I estimated genetic, maternal and residual (co)variance matrices for the five wing traits by implementing multivariate linear mixed animal models (Kruuk 2004) within the R (R Core Team 2021) package *MCMCglmm* (Hadfield 2010) using

$$\boldsymbol{y_{ijkl}} = \mu + S + a_i + d_j + e_{l(ijk)}\,, \qquad (1)$$

where the global intercept for each trait (μ) and sex (S) were included as the only fixed effects. Preliminary analyses showed that analyzing male and female offspring separately did not change estimates of genetic variance or mean phenotype. The additive genetic merit of each individual is represented by a_i, while d_j represents the *j*th dam from the parental generation. $e_{l(ijk)}$ represents the residual variation. The five normally-distributed wing traits were included as a multivariate response variable ($\boldsymbol{y_{ijkl}}$). Prior to analysis, I standardized all traits by dividing them by their mean within each replicate cage, which calculates genetic variance relative to the mean phenotype. I chose this standardization because it best represents evolvability (Hansen and Houle 2008; Hansen et al. 2011). I used 60 separate implementations of equation 1 to estimate (co)variance matrices for each replicate cage (n=5) within each treatment (n=6) measured in each assay (n=2). With c.300 individuals measured per replicate cage that were derived from 60 families, estimates of genetic variance show similar levels of uncertainty to an analysis that incorporated all replicate cages in each treatment to estimate genetic variance (see **Methods S1**).

I applied equation 1 with a burn-in of 100,000 Monte Carlo Markov Chain (MCMC) iterations and a thinning interval of 1,000 iterations. I extracted the posterior distribution for a_i, d_j and $e_{l(ijk)}$ from each implementation of equation 1, which I used to construct genetic (**G**), maternal (**M**) and residual (**R**) covariance matrices. I checked model convergence by ensuring that autocorrelation did not exceed 0.05, that the effective sample size exceeded 85% of the number of saved samples and that the weakly informative prior did not have undue effect on the parameter estimates. Given that this analysis builds on Yeaman et al. (2010a), who found that all estimates of genetic variance were significant, I ensured that the animal model analysis produced similar estimates of genetic variance as previously reported.

For all analyses, I compared G-matrices among treatments and assays, which were calculated as the mean of the five replicate cages within each treatment (n=6) and for each assay (n=2). To identify how the replicate cages varied in their response to the treatments and assays, I incorporated all 60 G-matrices into each analysis. This allowed comparisons among treatments and assays while making sure that the effects were consistent across biological replicates. Comparing the 12 average G-matrices yielded identical results to an alternative analysis that estimated the 12 G-matrices but with an additional random effect in equation 1 to account for differences among the replicate cages (see **Methods S1**).

Prediction 1: Quantifying differences in genetic variance due to temperature and environmental heterogeneity

I then quantified overall differences in **G** among treatments and assays. Although a range of methods for comparing multiple matrices exists, the covariance tensor provides an elegant approach (Aguirre et al. 2014). By quantifying a matrix (**S**) that represents differences among all matrices and then decomposing **S** to describe the differences, it is possible to use a single framework to first quantify the extent of differences among the original matrices, and then identify how each original matrix contributes to the differences among all matrices (i.e. how each matrix contributes to the differences described by **S**).

For a detailed description of the covariance tensor approach, see Hine et al. (2009), Aguirre et al. (2014) or Figure 2 in Walter et al. (2018). First, **S** is constructed by quantifying the among-matrix variances and covariances for all the elements contained in the original matrices. This means that **S** represents the raw differences among the original $n \times n$ matrices (n = number of traits). Diagonal elements of **S** contain both the among-matrix variances in the variances of the original matrices, and the variance in the covariances. The off-diagonal elements contain the among-matrix covariances of the variances (of the original matrices), and the among-matrix covariances of the covariances. Next, **S** is decomposed, with the vectors describing axes of genetic variance that differ among the original matrices. The eigenvalues of **S** represent the magnitude of differences among the original matrices and are used to test whether differences captured by **S** are significantly greater than null expectations. The eigenvectors of **S** are then scaled and rearranged to calculate the eigentensors, the $n \times n$ matrices that represent orthogonal differences in the original matrices.

I used the covariance tensor for two purposes. First, to identify whether I detected biologically meaningful differences in genetic variance among treatments, I compared differences among the observed matrices to differences expected under a suitable null distribution. The null distribution was created by re-applying equation 1 to phenotypic data that was reconstructed to contain no differences in genetic variance among the treatments, as recommended by Morrissey et al. (2019) (see **Methods S2**). Second, to identify how each treatment contributed to the differences described by each eigentensor, I calculated the coordinates (Frobenius product) between each of the original matrices, and each significant eigentensor. Put simply, the coordinates quantify the overlap between an eigentensor and each original matrix. Differences in the coordinates for any given treatments suggest that they contribute differently to describing differences in genetic variance described by that particular eigentensor.

Prediction 2: Testing for differences in mean multivariate phenotype

I used a multivariate analysis of variance (MANOVA) to test for differences in multivariate phenotype among treatments and assays. I included the five traits as a multivariate response variable, with treatment, assay and their interaction as the main effects. I used replicate cages within treatment as the error term to test whether differences in mean phenotype were greater among treatments than among replicates within

treatments. I extracted the sums of squares and cross product matrices (SSCP) from the MANOVA and calculated mean-square matrices by dividing by their corresponding degrees of freedom. I then estimated **D**, the matrix representing differences in multivariate mean phenotype using

$$\mathbf{D} = \frac{(\mathbf{MS_T}+\mathbf{MS_A}+\mathbf{MS_{T\times A}})-\mathbf{MS_R}}{nf}, \quad (2)$$

where $\mathbf{MS_T}$, $\mathbf{MS_A}$ and $\mathbf{MS_{T\times A}}$ represent the mean-square matrices for treatment, assay and their interaction, respectively. $\mathbf{MS_R}$ represents mean-square matrices for replicate and therefore the variation within treatments. In an unbalanced design, nf represents the average number of observations for each replicate (here, the number of families per treatment), calculated using equation 9 in Martin et al. (2008). To visualize differences in multivariate phenotype, I used the first two eigenvectors of **D** to calculate the multivariate scores of all treatments and replicates.

Prediction 3, Test 1: Does genetic, maternal or residual variance align with changes in mean phenotype? To test whether genetic, maternal or residual variance align with changes in mean phenotype, I projected vectors representing axes of differences in mean phenotype, through **G** using

$$\mathrm{V}_{ij} = \frac{\boldsymbol{e}^T \mathbf{G}_{ij} \boldsymbol{e}}{Tr(\mathbf{G}_{ij})}, \quad (3)$$

where $\boldsymbol{e}$ represents a vector describing differences in mean phenotype, and $\mathbf{G}_{ij}$ is the genetic covariance matrix (for the *i*th MCMC iteration from the *j*th treatment) through which $\boldsymbol{e}$ is projected. T represents the transpose of a vector. Standardizing by the trace (*Tr*) of **G** (i.e. the total genetic variance in **G**) calculates the proportion of genetic variance along the vector $\boldsymbol{e}$. V_{ij} therefore quantifies the proportion of genetic variance in the direction of changes in mean phenotype.

I used equation 3 to project two different vectors ($\boldsymbol{e}$) through **G**: 1) the vector of plasticity, calculated as the difference in mean phenotype across assay temperatures, using $\Delta\bar{\boldsymbol{x}} = \bar{\boldsymbol{x}}_{cold} - \bar{\boldsymbol{x}}_{hot}$, where $\bar{\boldsymbol{x}}_{cold}$ and $\bar{\boldsymbol{x}}_{hot}$ are vectors of trait means in the cold and hot treatments, respectively. $\Delta\bar{\boldsymbol{x}}$ quantifies the direction of change in multivariate mean phenotype between assays (calculated for each treatment separately). 2) The vector representing divergence in mean phenotype across treatments (within each assay), which is the first eigenvector of **D**, calculated using equation 2, but for each assay temperature separately. Because plasticity can be determined by maternal and environmental effects, I repeated the analysis using equation 3 with the maternal (**M**) and residual (**R**) covariance matrices to test whether maternal or environmental variance aligned with plasticity (changes in mean across assay treatments) or evolved differences in mean phenotype among treatments. If changes in mean are associated with maternal or environmental variance (i.e. differences among individuals not due to additive genetic effects), then we would observe an alignment between changes in mean phenotype with either **M** or **R**.

To test whether plasticity and evolved changes in phenotype described significantly greater genetic variance than null expectations, I compared the observed projections with a null distribution created by quantifying the amount of genetic variation described by randomly generated vectors. Vectors were constructed by taking *n* samples (where *n* is the number of traits) from a uniform distribution between -1 and 1, and then normalizing the vector to unit length. I then used equation 3 to project the random vectors through the posterior distribution of **G**. Where the observed projections exceed the null distribution provides evidence that the observed vectors of plasticity or phenotypic evolution describe greater genetic variance than null expectations.

Prediction 3, Test 2: Did phenotypic evolution occur along genetic lines of least resistance?
To test whether evolved changes in mean phenotype were more closely aligned with the primary axis of **G** ($\boldsymbol{g}_{max}$) than expected under random sampling, I compared the observed alignment between $\boldsymbol{g}_{max}$ and the vectors representing differences in mean phenotype (i.e. plasticity and evolved differences) with a suitable null distribution. I constructed the null distribution by calculating the angle between 100,000 pairs of random vectors, which were generated as described above. If the observed distribution of angles between $\boldsymbol{g}_{max}$ and phenotypic vectors (plasticity and evolved differences) are smaller than distribution of the angles between the pairs of vectors generated at random, then there is evidence that changes in mean phenotype align better with $\boldsymbol{g}_{max}$ than expected under random sampling (McGlothlin et al. 2018).

Genetic variance can change due to drift, selection or environmental variation that can obscure an alignment between current estimates of genetic variation and phenotypic change. However, if phenotypic evolution occurred along an axis representing a genetic line of least resistance, then such an axis is likely to be conserved among G-matrices, even if differences in **G** are observed. By comparing conserved elements of **G** (here, among treatments) with changes in mean phenotype, it is possible to test whether phenotypic evolution has occurred along a line of genetic least resistance that was present in the common ancestor (McGlothlin et al. 2018). To quantify conserved elements of **G**, I used Krzanowski's common subspace analysis, which estimates the genetic subspace (**H**) common to the G-matrices of all treatments (Aguirre et al. 2014; Krzanowski 1979). **H** is constructed from a subset (*k*) of eigenvectors of **G**, where $k \leq \frac{n}{2}$ and *n* is the number of traits. I calculated **H** using two eigenvectors (*k*=2 for *n*=5 traits) from each G-matrix (within each assay), which captured >70% of total genetic variance in each treatment. An alignment between the subspace, **H**, with the vector representing evolutionary divergence therefore tests whether genetic variance common to all treatments aligns with phenotypic divergence (McGlothlin et al. 2018). The leading eigenvector of **H**, $\boldsymbol{h}_1$, is the axis describing the greatest amount of genetic subspace that is shared among **G** estimated in the six treatments. If the angle between $\boldsymbol{h}_1$ and phenotypic differences among treatments (within each assay) is close to zero and falls below the null distribution, then there is evidence that phenotypic evolution occurred along an axis of genetic variation that is conserved among treatments.

Results

Prediction 1: Environmental heterogeneity changed the orientation but not shape of genetic variance

The G-matrices for all treatments and assays are presented in **Table S1**, and all 60 G-matrices are included as supplementary material. As reported by Yeaman et al. (2010a), estimates of genetic variance for all traits were not significantly different across treatments (**Table S1**). However, for all wing angle traits, genetic variance was greater in the cold assay compared to the hot assay (**Fig. S1**). No change in multivariate evolvability in response to environmental heterogeneity for any of the metrics of evolvability was observed (**Table S2**), suggesting that environmental heterogeneity did not significantly affect the shape or size of **G**.

The proportion of genetic variation described by each eigenvector was similar for all treatments (**Table 2**), which provides further evidence that the shape of **G** has been conserved among treatments. Similar linear combinations of traits describe $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_2$ for all treatments in the cold assay, suggesting that the orientation of **G** was conserved (**Table 2a**). In the hot assay, however, different linear combinations of traits describe **G** in the different treatments (**Table 2b**), suggesting that environmental heterogeneity changed the orientation of **G**, but only at a warmer temperature. In the hot assay, the combinations of traits that describe $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_2$ were the same for all treatments, except the spatial heterogeneity treatment, which showed differences in the magnitude of the trait loadings compared to the other treatments (**Table 2b**). The pairwise angles between $\boldsymbol{g}_{max}$ of each treatment (and for $\boldsymbol{g}_2$) quantify how closely aligned the leading eigenvectors of **G** are for each pairwise comparison of treatments. For both $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_2$, the angles between treatments ranged between 2.5-36.7º in the cold assay and between 5.9-25º in the hot assay (**Table S3**), suggesting a similar orientation of **G** for all treatments. Only spatial heterogeneity in the hot assay showed a difference in the orientation of **G** to the other treatments, whereby $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_2$ were in a direction 40.2-59.8º different to the other treatments (**Table S3**). Therefore, G-matrices estimated for populations evolving under spatial heterogeneity showed differences in orientation to populations from homogeneous treatments, but only in the hot assay temperature.

To test for significant differences in **G** among treatments (within assay), I used a covariance tensor approach to identify axes (eigentensors) that describe differences in genetic variance among all treatments in a single analysis. No significant eigentensors were found in the cold assay, and only the first eigentensor in the hot assay describes greater differences in genetic variance than expected under a null distribution (**Fig. 3a**). Significant differences in genetic variance were therefore only found in the hot assay. The first eigentensor described 32% and 57% of the difference in genetic variance in the cold and hot assay, respectively (**Table S4**). To identify which treatments contributed to the differences in **G**, I calculated the coordinates (i.e. the correlation) between each of the original matrices and the eigentensors.

In both assays, differences in **G** were created by differences between the spatial and temporal heterogeneity treatments. In the cold assay, temporal heterogeneity showed the greatest difference from the other treatments, whereas spatial heterogeneity in the hot assay was significantly different to the remaining treatments (**Fig. 3b**). Significant differences in genetic variance in the hot assay were therefore created by differences in the orientation of **G** in the spatially heterogeneous treatment when compared to the homogeneous and temporally heterogeneous treatments.

Including the G-matrices of all treatments and both assays in a single tensor analysis showed that greater differences in genetic variance were created by assay temperature (hot vs. cold), which contrasts with smaller differences in genetic variance among treatments within each assay (**Fig. 3c**). Furthermore, while most treatments showed significant differences in **G** between assays, spatial heterogeneity showed no change in genetic variance (**Fig. 3c**). Given that one trait (wing angle 7-8-9) contributed the most to $\boldsymbol{g}_{max}$ in the cold assay for all treatments, and to $\boldsymbol{g}_{max}$ for spatial heterogeneity in the hot assay (**Table 2**), changes in genetic variance in this wing angle likely created changes in **G** across assays. While spatial heterogeneity maintained a similar level of genetic variance in angle 7-8-9 across assays, the other treatments showed significantly lower genetic variance in the hot assay, when compared to the cold assay. This result suggests that across assays, spatial heterogeneity maintained genetic variance in the trait that contributes the most to $\boldsymbol{g}_{max}$, which created greater stability in **G** across assay temperatures.

Prediction 2: Evolved changes in mean multivariate phenotype were nearly orthogonal to plasticity
Univariate trait means are presented in **Table S5**. Multivariate mean phenotype changed significantly between assays ($F_{1,48}$=862.39, P<0.0001) and among treatments ($F_{5,48}$=2.74, P<0.0001). The assay×treatment interaction was not significant ($F_{5,48}$=0.34, P=0.9988), suggesting that all treatments responded similarly to assay and that greater environmental heterogeneity did not affect plasticity differently to the homogeneous treatments (**Fig. 4**). Surprisingly, changes in mean phenotype due to plasticity (differences between assays) and evolved changes in phenotype among treatments were in different directions. The first axis of **D** ($\boldsymbol{d}_{max}$) described 98% of all differences in mean and captured large changes in mean phenotype due to plasticity between assay temperatures, but only small differences among treatments (**Fig. 4** and **Table S5**). By contrast, the second axis of **D** ($\boldsymbol{d}_2$) described 1.8% of all differences in mean phenotype and captured differences among treatments, but little difference between assays. Phenotypic differences due to plasticity (represented by $\boldsymbol{d}_{max}$) were created by smaller wings and larger wing angles in the hot assay compared to the cold assay. By contrast, evolved differences in phenotype among treatments (represented by blue and red lines in **Fig. 4**) were created by larger wing angles and wing veins in the hot homogeneous treatments compared to the cold treatments, but that were largely independent of wing size (**Table S5** and **Fig. 4**).

Prediction 3, Test 1: Genetic variance aligns with evolved differences in mean phenotype, but not plasticity across assays

Projection of vectors representing changes in mean phenotype due to plasticity (across assay temperatures) and evolved differences (between hot and cold treatments), through **G**, quantified how much genetic variation lies in the direction of each vector. The direction of evolved differences in mean phenotype among treatments (within assay temperature, $\boldsymbol{d}_2$ in **Fig. 4**) described a large proportion (c.40%) of total genetic variance (**Fig. 5**), which was greater than null expectations for all treatments (**Fig. S2**). Furthermore, the direction of phenotypic evolution described a similar amount of genetic variance to $\boldsymbol{g}_{max}$ for most treatments (**Fig. 5**). $\boldsymbol{g}_{max}$ only described more genetic variance than the direction of phenotypic evolution for the temporal treatment in the cold assay and three treatments in the hot assay (temporal heterogeneity and the migration treatments; **Fig. 5**). These results suggest that, as predicted, evolution occurred in a direction of the phenotype containing a large amount of genetic variation, which was close to $\boldsymbol{g}_{max}$. By contrast, and against the prediction that plasticity would also lie in a direction containing a large amount of genetic variance, plastic changes across assays only described c.10% of genetic variance (**Fig. 5**). Treatments showed little difference in the amount of genetic variance described for either vector, except in the hot assay, where phenotypic evolution described a greater proportion of genetic variance in the spatial heterogeneity treatment compared to most other treatments (**Fig. 5a**).

To understand why plasticity was in a different direction to evolved changes in mean phenotype and aligned poorly with genetic variance, I tested whether the vector of plasticity described a greater proportion of maternal and environmental variance. I found that the direction of plasticity across assays described a significantly greater proportion (c.22%) of maternal variance compared to genetic variance (c.10%) (**M** in **Fig. 5**). However, when compared to plasticity, evolved differences in mean phenotype described a similar proportion of maternal variance (c.20-30%). Only the spatial heterogeneity treatment showed greater maternal variance in the direction of plasticity when compared to the direction of evolved differences (**Fig. 5**). Residual (environmental) variance showed a similar pattern to genetic variance (**R** in **Fig. 5**) whereby the direction of phenotypic evolution described a greater proportion of residual variance (c.27%) than plasticity (c.10%). These results suggest that while maternal variance showed some alignment with the direction of plasticity, environmental variance better aligned with evolved changes in phenotype.

Prediction 3, Test 2: Phenotypic evolution occurred along genetic lines of least resistance in homogeneous treatments

The angles between $\boldsymbol{g}_{max}$ and the vector representing phenotypic evolution (i.e., $\boldsymbol{d}_2$ in **Fig. 4**) were less than 50° for ten (of 12) treatments (**Fig. 6**), suggesting that changes in mean phenotype were often in a direction similar to the major axis of genetic variance. This pattern is stronger in the cold assay where angles between $\boldsymbol{g}_{max}$ and phenotypic evolution were all less than 40°, which contrasted with the range of 25-50° observed in the hot assay (**Fig. 6**). In the cold assay, all treatments showed a significantly closer alignment between evolved differences in phenotype and $\boldsymbol{g}_{max}$ than null distribution (**Fig. 6**). In the hot

assay, only spatial heterogeneity and the homogeneous hot treatment showed a significantly closer alignment than null expectations (**Fig. 6**). For the treatments that showed a close alignment between $\boldsymbol{g}_{max}$ and phenotypic evolution there was consistent support across replicate cages with no more than one cage per treatment with an angle exceeding 45° from the direction of evolved differences in phenotype (**Fig. 6**). All angles between $\boldsymbol{g}_{max}$ and plasticity between assay temperatures were high (c.75°) and within the null expectation (**Fig. 6**), providing further evidence that plasticity did not align with genetic variance.

The spatial heterogeneity treatment was the only treatment where **G** did not change across assays (**Fig. 3b**), which was associated with the closest alignment between $\boldsymbol{g}_{max}$ and the direction of change in mean phenotype in both assays (23.7° and 27.1° in the cold and hot assays, **Fig. 6**). This means that although spatial heterogeneity did not affect mean phenotype (spatial heterogeneity is intermediate to the other treatments in **Fig. 4**), **G** exposed to spatial heterogeneity changed the least between the hot and cold assays, and maintained a close alignment between $\boldsymbol{g}_{max}$ and the direction of evolved differences in mean phenotype (**Fig. 6**). By contrast, the G-matrices estimated from the homogeneous and temporal heterogeneity treatments changed significantly across assays, which resulted in a weaker alignment between genetic variance and phenotypic evolution in the hot assay, when compared to the cold assay.

Although the hot assay was associated with changes in genetic variance that meant that $\boldsymbol{g}_{max}$ did not align very closely with phenotypic evolution, it is possible that conserved elements of genetic variance among treatments (representing genetic lines of least resistance in the common ancestor) could still align with phenotypic evolution. To test whether this was the case, I compared the genetic subspace (**H**) common to all treatments, with the vector representing evolved changes in mean phenotype among treatments (i.e., $\boldsymbol{d}_2$ in **Fig. 4**). Eigenvalues of **H** range from 0 to *m*, where *m* is the number of matrices. The closer eigenvalues are to *m*, the more of the original matrices (here, treatments) share the common subspace. I found that the first eigenvalue, which describes the direction of greatest shared genetic variance among treatments, $\boldsymbol{h}_1$, was 5.9 (5.78-5.98 90% HPD) in the cold assay and 5.85 (5.69-5.97 HPD) in the hot assay (**Table S6**), which is close to the maximum of 6. The angle between $\boldsymbol{h}_1$ and the subspace for each of the treatments were all less than 10° (**Table S6**), suggesting that $\boldsymbol{h}_1$ accurately represented genetic variance shared by all treatments. $\boldsymbol{h}_1$ also described a large proportion of genetic variance in all treatments (37-52%) that is similar to $\boldsymbol{g}_{max}$ and suggests that $\boldsymbol{h}_1$ represents a conserved axis of high evolvability. As evidence that phenotypic evolution occurred in the direction of the shared genetic subspace, $\boldsymbol{h}_1$ showed a close alignment with the vector describing evolved differences among treatments (i.e., $\boldsymbol{d}_2$ in **Fig. 4**) in both the cold (mean=25.7°; 20.7, 35.6 90% HPD) and hot (mean=26.1°; 18.5, 58.1 90% HPD) assays. The alignment was closer than expected under the null distribution for both assays (**Fig. S3**), providing evidence that phenotypic evolution occurred along genetic lines of least resistance. When the analysis was repeated using **G** for all replicate cages (i.e. by quantifying the shared genetic subspace for 30 G-

matrices in each assay), the results were consistent for both the cold (mean=27.4°; 21.8, 36.4 90% HPD) and hot (mean=24.5°; 15.5, 48.8 HPD) assays (**Fig. S3**).

Discussion

Using populations exposed to different temperatures and levels of environmental variation for >25 generations, these results represent some of the first experimental evidence supporting evolution along genetic lines of least resistance. The main results are visualized in **Fig. 7**. Populations evolving in homogeneous cold and hot temperatures diverged in their mean wing phenotype, which occurred in a direction of the phenotype that contained a large proportion (c.40%) of total genetic variance. The axis describing the largest amount of genetic variance, $\boldsymbol{g}_{max}$, aligned more closely with phenotypic evolution in the cold assay (22-37°) compared to the hot assay (27-53°). However, genetic subspace common to all treatments aligned with phenotypic divergence (c.26° in both assays), suggesting that phenotypic evolution followed genetic lines of least resistance, but that a closer alignment with $\boldsymbol{g}_{max}$ was obscured by environmental effects (i.e., plasticity) in the hot assay. Spatial heterogeneity created little change in mean phenotype and was associated with more stable genetic variance that aligned with evolved changes in phenotype in both assays. Surprisingly, plasticity changed the phenotype across assays in a direction different to that of evolved differences between hot and cold treatments. The direction of plasticity did not align with genetic or environmental variance and only showed a moderate alignment with maternal variance. Together, these results suggest that phenotypic evolution has occurred along genetic lines of least resistance, and that exposure to spatial environmental heterogeneity could stabilize genetic variance, whereas temporal heterogeneity was associated with more lability in **G**.

Replicate cages showed variation in the orientation of **G** and their alignment with phenotypic divergence. It is therefore possible that genetic drift could affect the orientation of genetic variation, which contrasts with the prediction that drift only changes the size of **G** (Jones et al. 2003). Although beyond the scope of this study, future work should identify whether changes in **G** due to drift influence adaptation and phenotypic evolution along genetic lines of least resistance.

Evolution along genetic lines of least resistance

Populations from homogeneous treatments showed phenotypic divergence along $\boldsymbol{g}_{max}$, suggesting that evolution along genetic lines of least resistance ($\boldsymbol{g}_{max}$) might only occur if environmental heterogeneity is sufficiently low (relative to selection). Adaptation in more spatially or temporally heterogeneous environments is likely to be more difficult than for homogeneous environments because the environment has a finer grain that allows high gene flow and hinders adaptation (Slatkin 1973, 1978). In the current study, homogeneous treatments with migration (two migrants each generation) showed less phenotypic divergence in the direction of the homogeneous treatments (without migration), suggesting that low levels

of migration are enough to erode adaptation along the genetic line of least resistance, even in response to a single environmental variable. Comparative evidence for evolution along lines of genetic least resistance (e.g. Costa et al. 2020; McGlothlin et al. 2018; McGuigan et al. 2005; Schluter 1996) could therefore be due to lower levels of environmental heterogeneity in the natural habitats in which the studied taxa evolved, which could be created, for example, by behavior that smooths environmental variation. Whether evolution along $\boldsymbol{g}_{max}$ only occurs below a threshold of environmental heterogeneity remains unexplored but could identify how ecological and genetic constraints together determine the rate of adaptation. Such data are especially important for understanding whether rapid adaptation to novel environments is possible.

Although the current study provides evidence for contemporary adaptation along genetic lines of least resistance, we still do not understand how and when genetic constraints to adaptation arise. Evidence for evolution along genetic lines of least resistance is derived from comparative studies that associate phenotypic divergence with axes of genetic variation. While comparative studies provide important insights about the current alignment between genetic variation and phenotypic evolution, they make it difficult to identify whether the common ancestor possessed similar patterns of genetic variance (but see McGlothlin et al. 2018 for a phylogenetic reconstruction of the common ancestor), and cannot test whether selection was also in the direction of $\boldsymbol{g}_{max}$, or in a different direction that would introduce genetic constraints to adaptation (Chenoweth et al. 2010). It is also important to note that comparative evidence for adaptation along genetic lines of least resistance rests on the assumption that selection did not deplete genetic variation during adaptation to the contrasting temperatures, although the debate about how selection depletes multivariate genetic variation is ongoing. Furthermore, it is possible to confuse causality using comparative approaches: If correlational selection was in the direction of $\boldsymbol{g}_{max}$, then we would expect genetic variation in the direction of $\boldsymbol{g}_{max}$ to be strengthened, which could make $\boldsymbol{g}_{max}$ the consequence of adaptation, rather than determining the direction of adaptation (Arnold et al. 2001; Roff and Fairbairn 2012). To understand how genetic correlations among traits determine evolutionary trajectories, future experimental evolution studies need to assay multivariate genetic variation and selection (ideally including correlational selection) in the common ancestor and then identify whether evolutionary trajectories can be predicted in multiple environments (e.g., Magalhães et al. 2007; Mallard et al. 2022; Zu et al. 2020).

Compared to the cold assay, phenotypic evolution in the hot assay showed a weaker association with $\boldsymbol{g}_{max}$, suggesting that evolutionary trajectories could be harder to predict in warmer environments. However, elements of **G** that were conserved among treatments (and described a large proportion of genetic variance) showed a strong association with phenotypic evolution, suggesting that evolution occurred along genetic lines of least resistance, even though initially obscured by environmental effects in the hot assay. It is therefore possible for evolution to occur along genetic lines of least resistance even when $\boldsymbol{g}_{max}$

and phenotypic divergence are not well aligned. This is because axes of high evolvability other than $\boldsymbol{g}_{max}$ exist that can allow rapid evolution (Bolstad et al. 2014; Hansen and Houle 2008; Hansen and Voje 2011). Results in the current study suggest that $\boldsymbol{g}_{max}$ (cold assay) and the common genetic subspace (in both assays) did not perfectly align with phenotypic divergence (angle of c.25°), which suggests that adaptation may occur along an axis of high evolvability that is slightly different to $\boldsymbol{g}_{max}$. While it is possible that changes in **G** due to plasticity in the hot assay could be a mechanism for coping with warmer environments but that still allows evolution along genetic lines of least resistance, it is also likely that such changes in **G** are incidental and are simply a byproduct of genotype-by-environment interactions inherent to the population. To better predict evolution, it is therefore important to understand how and why **G** changes across environments, and whether such changes will determine adaptation in the short and long-term.

The effect of environmental heterogeneity on genetic variance

Against predictions, genetic variance in more heterogeneous environments was not greater than in homogeneous environments (Yeaman et al. 2010a). This could be due to several factors: First, because selection in heterogeneous environments can favor different traits compared to selection in homogeneous environments, it is possible that the spatial and temporally heterogeneous environments could be associated with changes in other traits (e.g., early life history traits) that are more important than wing shape for adapting to more heterogeneous environments (Beaumont et al. 2009; Ketola and Saarinen 2015; Magalhães et al. 2014). We therefore need to understand how heterogeneous environments select for different traits and identify the consequences for adaptation. Second, given that selection can deplete genetic variation, it is possible that constant environments experienced during lab adaptation prior to experimental evolution could have reduced genetic variance available to the populations that were then exposed to temporal and spatial environmental heterogeneity. If so, then the genetic lines of least resistance ($\boldsymbol{g}_{max}$) found in this study may have been created by lab adaptation, which could have removed genetic variance in the common ancestor, and therefore reduced the ability to detect the maintenance of greater genetic variance in more heterogeneous environments.

Instead of maintaining greater genetic variance, environmental heterogeneity created differences in how **G** responded to assay temperature. In contrast with the other treatments, spatial heterogeneity was the only treatment that showed similar estimates of genetic variance in wing angle 7-8-9 across assays (**Fig. S1**). Spatial heterogeneity also showed the closest alignment between $\boldsymbol{g}_{max}$ and phenotypic divergence (**Fig. 6**). By contrast, temporal heterogeneity was associated with less stability in **G** and the greatest reduction in genetic variance for wing angle 7-8-9 in the hot assay (**Fig. S1**). While the potential for spatial and temporal environmental variation to maintain genetic variation remains poorly understood (Yeaman et al. 2010a), it is likely that spatial heterogeneity affects genetic variance (and **G**) differently to

temporal fluctuations, which could determine how plasticity evolves and how adaptation proceeds (Björklund and Gustafsson 2015).

Plastic and evolved responses of fly wings to temperature

Changes in wing shape and size under different temperatures and assays are likely functional (Cavicchi et al. 1991; Debat et al. 2003; Partridge et al. 1994; Santos et al. 2004). Consistent with Condon et al. (2014), plasticity produced larger wings in the cold assay compared to the hot assay, whereas significantly larger wings did not evolve in the cold homogeneous treatment compared to the hot treatment (**Table S5**). This contrasts with strong patterns of adaptive divergence in wing size in response to five years of experimental evolution in cold and hot temperatures (Partridge et al. 1994) and in response to a latitudinal cline that reflects a similar temperature gradient (James et al. 1997). By contrast, wing angles evolved large differences between hot and cold treatments. Wing angles were smaller in the cold assay temperatures and treatments, which produces narrower wings. Larger, narrower wings have been shown to improve flight in colder temperatures (Fraimout et al. 2018; Frazier et al. 2008), providing functional support for plastic and evolved changes in wing morphology in response to temperature.

Plastic changes in phenotype were in a different direction to evolved differences among treatments, suggesting that changes in mean phenotype due to plasticity and evolution could be determined by different processes. While plasticity created large changes in wing size and its negative correlation with wing angles, evolved differences between homogeneous hot and cold temperatures involved changes in wing angles independent of wing size. Wing size possessed genetic variation an order of magnitude smaller than the other wing traits (**Table S1**), suggesting that even if selection favored adaptive divergence in wing size, evolution is likely to be constrained by the low availability of genetic variation (Chenoweth et al. 2010). In responding to temperature, changes in wing size (and correlations with other traits) are therefore likely to be determined by plasticity rather than selection on genetic variation. Given that plasticity aligned better with maternal variance than genetic or environmental variance (**Figs. 5-6**) then, similar to other studies (Galloway et al. 2009; Lind et al. 2020; McAdam et al. 2014), it is possible that maternal genetic effects could play some role in plastic changes in wing morphology across temperature. If variation in maternal genotypes produce offspring phenotypes that differ in wing size (and the correlation with wing shape) along the direction of plasticity, then variation among maternal genotypes could, in part, contribute to plasticity. However, it is likely that plastic responses are largely the result of historical selection for specific developmental pathways that determine wing responses to temperature.

The poor alignment between genetic variance and plasticity contrasts with the conclusions of a recent meta-analysis that plasticity occurs in the direction of $\boldsymbol{g}_{max}$ (Noble et al. 2019). It is possible, however, that plasticity only occurs in the direction of genetic variation when there is high genetic variation in plasticity itself, which could occur when plasticity aligns with the direction of selection (Lind et al. 2015).

In this scenario, if the loci that are particularly environmentally sensitive contribute to genetic variation along $\boldsymbol{g}_{max}$, then an alignment between plasticity and $\boldsymbol{g}_{max}$ will be possible. However, plasticity is likely to be the result of selection on genetic variation in plasticity in the past, which could erode genetic variation when there is strong purifying selection for developmental pathways that are crucial for coping with environmental variation (Oostra et al. 2018). In the wild, *Drosophila* persist by developing in hot and cold temperatures that are often far from optimal. Given the importance of plasticity in wing morphology for responding to temperature variation, developmental pathways underlying plasticity could now possess little genetic variation as a consequence of strong historical purifying selection that eroded genetic variation (Oostra et al. 2018; Steward et al. 2022). It is therefore possible that plasticity could align with $\boldsymbol{g}_{max}$ during the early stages of adaptation when genetic variation for plasticity is greater, but as plasticity becomes canalized, genetic variance is reduced and plasticity becomes disassociated from $\boldsymbol{g}_{max}$. General conclusions about how $\boldsymbol{g}_{max}$ relates to plasticity remain difficult and require further knowledge on how plasticity and genetic variation are related during the early stages of adaptation.

Conclusions

Overall, these results reveal that evolution is likely to occur along genetic lines of least resistance in homogeneous environments. While spatial heterogeneity was associated with more stability in **G**, temporal heterogeneity showed greater lability in **G**, which suggests that we need to better understand how environmental heterogeneity imposes selection on different traits. Relating plasticity to multivariate genetic variation also remains a challenge. If strong purifying selection on developmental pathways erodes genetic variation in plasticity, then plasticity might only align with genetic variation during the early stages of adaptation. Future work needs to estimate selection and plasticity prior to experimental evolution to better understand how the environment and genetic variation impose constraints on adaptation.

Acknowledgements

I am grateful to Sam Yeaman for providing valuable feedback, and for encouraging me to develop this work. I thank Jennifer Lau, Sara Magalhães and three reviewers for providing extensive comments that helped me to improve and clarify this paper. I thank Keyne Munro for early discussions, and Vanessa Kellermann for comments on an earlier version. This work was supported by an Australian Research Council early career (DECRA) fellowship DE200101019.

Data Code and Accessibility

Data is located at https://doi.org/10.5061/dryad.1719, and also accompanies the R code located at DOI: 10.5281/zenodo.7160306.

References

Aguirre, J. D., E. Hine, K. McGuigan, and M. W. Blows. 2014. Comparing G: multivariate analysis of genetic variation in multiple populations. Heredity 112:21-29.

Arnold, S. J. 1992. Constraints on Phenotypic Evolution. The American Naturalist 140:S85-S107.

Arnold, S. J., R. Burger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the Evolution and Stability of the G-Matrix. Evolution 62:2451-2461.

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 112:9-32.

Beaumont, H. J., J. Gallie, C. Kost, G. C. Ferguson, and P. B. Rainey. 2009. Experimental evolution of bet hedging. Nature 462:90-93.

Berner, D., W. E. Stutz, and D. I. Bolnick. 2010. Foraging trait (co)variances in stickleback evolve deterministically and do not predict trajectories of adaptive radiation. Evolution 64:2265-2277.

Björklund, M., and L. Gustafsson. 2015. The stability of the G-matrix: The role of spatial heterogeneity. Evolution 69:1953-1958.

Bolstad, G. H., T. F. Hansen, C. Pelabon, M. Falahati-Anbaran, R. Perez-Barrales, and W. S. Armbruster. 2014. Genetic constraints predict evolutionary divergence in *Dalechampia* blossoms. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 369:20130255.

Bürger, R., and A. Gimelfarb. 2002. Fluctuating environments and the role of mutation in maintaining quantitative genetic variation. Genetics research 80:31-46.

Careau, V., M. E. Wolak, P. A. Carter, and T. Garland. 2015. Evolution of the additive genetic variance-covariance matrix under continuous directional selection on a complex behavioural phenotype. Proceedings of the Royal Society of London Series B-Biological Sciences 282.

Cavicchi, S., G. Giorgi, V. Natali, and D. Guerra. 1991. Temperature-related divergence in experimental populations of *Drosophila melanogaster*. III. Fourier and centroid analysis of wing shape and relationship between shape variation and fitness. Journal of Evolutionary Biology 4:141-159.

Chenoweth, S. F., H. D. Rundle, and M. W. Blows. 2010. The contribution of selection and genetic constraints to phenotypic divergence. The American Naturalist 175:186-196.

Cheverud, J. M. 1984. Quantitative Genetics and Developmental Constraints on Evolution by Selection. Journal of Theoretical Biology 110:155-171.

Condon, C., B. S. Cooper, S. Yeaman, and M. J. Angilletta, Jr. 2014. Temporal variation favors the evolution of generalists in experimental populations of *Drosophila melanogaster*. Evolution 68:720-728.

Costa, E. S. J., B. M. Potts, and P. A. Harrison. 2020. Population divergence along a genetic line of least resistance in the tree species *Eucalyptus globulus*. Genes 11:1095.

Debat, V., M. Bégin, H. Legout, and J. R. David. 2003. Allometric and nonallometric components of *Drosophila* wing shape respond differently to developmental temperature. Evolution 57:2773-2784.

Doroszuk, A., M. W. Wojewodzic, G. Gort, and J. E. Kammenga. 2008. Rapid divergence of genetic variance-covariance matrix within a natural population. The American Naturalist 171:291-304.

Draghi, J. A., and M. C. Whitlock. 2012. Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. Evolution 66:2891-2902.

Eroukhmanoff, F., and E. I. Svensson. 2011. Evolution and stability of the G-matrix during the colonization of a novel environment. Journal of Evolutionary Biology 24:1363-1373.

Fraimout, A., P. Jacquemart, B. Villarroel, D. J. Aponte, T. Decamps, A. Herrel, R. Cornette et al. 2018. Phenotypic plasticity of Drosophila suzukii wing to developmental temperature: implications for flight. Journal of Experimental Biology 221:jeb166868.

Frazier, M. R., J. F. Harrison, S. D. Kirkton, and S. P. Roberts. 2008. Cold rearing improves cold-flight performance in Drosophila via changes in wing morphology. Journal of Experimental Biology 211:2116-2122.

Galloway, L. F., J. R. Etterson, and J. W. McGlothlin. 2009. Contribution of direct and maternal genetic effects to life-history evolution. New Phytologist 183:826-838.

Gillespie, J. H., and M. Turelli. 1989. Genotype-environment interactions and the maintenance of polygenic variation. Genetics 121:129-138.

Hadfield, J. D. 2010. MCMC Methods for multi-response generalized linear mixed models: The MCMCglmm R package. Journal of Statistical Software 33:1-22.

Hangartner, S., C. Lasne, C. M. Sgrò, T. Connallon, and K. Monro. 2020. Genetic covariances promote climatic adaptation in Australian *Drosophila*. Evolution 74:326-337.

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. Journal of Evolutionary Biology 21:1201-1219.

Hansen, T. F., C. Pelabon, and D. Houle. 2011. Heritability is not evolvability. Evolutionary Biology 38:258-277.

Hansen, T. F., and K. L. Voje. 2011. Deviation from the line of least resistance does not exclude genetic constraints: a comment on Berner et al. (2010). Evolution 65:1821-1822.

Hine, E., S. F. Chenoweth, H. D. Rundle, and M. W. Blows. 2009. Characterizing the evolution of genetic variance using genetic covariance tensors. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 364:1567-1578.

James, A. C., R. B. Azevedo, and L. Partridge. 1997. Genetic and environmental responses to temperature of Drosophila melanogaster from a latitudinal cline. Genetics 146:881-890.

Jones, A. G., S. J. Arnold, and R. Burger. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution 58:1639-1654.

Jones, A. G., S. J. Arnold, and R. Bürger. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. Evolution 57:1747-1760.

Ketola, T., and K. Saarinen. 2015. Experimental evolution in fluctuating environments: tolerance measurements at constant temperatures incorrectly predict the ability to tolerate fluctuating temperatures. Journal of Evolutionary Biology 28:800-806.

Kimmel, C. B., W. A. Cresko, P. C. Phillips, B. Ullmann, M. Currey, F. von Hippel, B. K. Kristjánsson et al. 2012. Independent axes of genetic variation and parallel evolutionary divergence of opercle bone shape in threespine stickleback. Evolution 66:419-434.

Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 359:873-890.

Krzanowski, W. J. 1979. Between-Groups Comparison of Principal Components. Journal of the American Statistical Association 74:703-707.

Lafuente, E., D. Duneau, and P. Beldade. 2018. Genetic basis of thermal plasticity variation in *Drosophila melanogaster* body size. PLoS Genetics 14:e1007686.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. Evolution 33:402-416.

—. 1980. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. Genetics 94:203-215.

Levis, N. A., and D. W. Pfennig. 2016. Evaluating 'plasticity-first' evolution in nature: key criteria and empirical approaches. Trends in Ecology & Evolution 31:563-574.

Lind, M. I., and F. Johansson. 2007. The degree of adaptive phenotypic plasticity is correlated with the spatial environmental heterogeneity experienced by island populations of *Rana temporaria*. Journal of Evolutionary Biology 20:1288-1297.

Lind, M. I., K. Yarlett, J. Reger, M. J. Carter, and A. P. Beckerman. 2015. The alignment between phenotypic plasticity, the major axis of genetic variation and the response to selection. Proceedings of the Royal Society B-Biological Sciences 282:20151651.

Lind, M. I., M. K. Zwoinska, J. Andersson, H. Carlsson, T. Krieg, T. Larva, and A. A. Maklakov. 2020. Environmental variation mediates the evolution of anticipatory parental effects. Evolution Letters 4:371-381.

Lynch, M., and B. Walsh. 1998, Genetics and analysis of quantitative traits. Sunderland, Sinauer Associates, Inc.

Mackay, T. F. 1981. Genetic variation in varying environments. Genetics research 37:79–93.

Magalhães, S., A. Cailleau, E. Blanchet, and I. Olivieri. 2014. Do mites evolving in alternating host plants adapt to host switch? Journal of Evolutionary Biology 27:1956-1964.

Magalhães, S., J. Fayard, A. Janssen, D. Carbonell, and I. Olivieri. 2007. Adaptation in a spider mite population after long-term evolution ona single host plant. Journal of Evolutionary Biology 20:2016-2027.

Mallard, F., B. Afonso, and H. Teotónio. 2022. Selection and the direction of phenotypic evolution. bioRxiv https://doi.org/10.1101/2022.05.28.493855.

Marroig, G., and J. M. Cheverud. 2007. Size as a line of least evolutionary resistance: Diet and adaptive morphological radiation in new world monkeys. Evolution 59:1128-1142.

Martin, G., E. Chapuis, and J. Goudet. 2008. Multivariate Q_{st}–F_{st} Comparisons: A Neutrality Test for the Evolution of the G Matrix in Structured Populations. Genetics 180:2135-2149.

McAdam, A. G., D. Garant, and A. J. Wilson. 2014. The effects of others' genes: maternal and other indirect genetic effects *in* A. Charmantier, D. Garant, and L. E. B. Kruuk, eds. Quantitative genetics in the wild. Oxford, UK, Oxford University Press.

McDonald, T. K., and S. Yeaman. 2018. Effect of migration and environmental heterogeneity on the maintenance of quantitative genetic variation: a simulation study. Journal of Evolutionary Biology 31:1386-1399.

McGlothlin, J. W., M. E. Kobiela, H. V. Wright, D. L. Mahler, J. J. Kolbe, J. B. Losos, and E. D. Brodie, III. 2018. Adaptive radiation along a deeply conserved genetic line of least resistance in *Anolis* lizards. Evolution Letters 2:310-322.

McGuigan, K., S. F. Chenoweth, and M. W. Blows. 2005. Phenotypic divergence along lines of genetic variance. The American Naturalist 165:32-43.

Merilä, J., and M. Björklund. 1999. Population divergence and morphometric integration in the greenfinch (*Carduelis chloris*) - evolution against the trajectory of least resistance? Journal of Evolutionary Biology 12:103-112.

Morrissey, M. B., S. Hangartner, and K. Monro. 2019. A note on simulating null distributions for **G** matrix comparisons. Evolution 73:2512-2517.

Noble, D. W. A., R. Radersma, and T. Uller. 2019. Plastic responses to novel environments are biased towards phenotype dimensions with high additive genetic variation. Proceedings of the National Academy of Sciences, USA 116:13452-13461.

Oostra, V., M. Saastamoinen, B. J. Zwaan, and C. W. Wheat. 2018. Strong phenotypic plasticity limits potential for evolutionary responses to climate change. Nature Communications 9:1-11.

Partridge, L., B. Barrie, K. Fowler, and V. French. 1994. Evolution and development of body size and cell size in *Drosophila melanogaster* in response to temperature. Evolution 48:1269-1276.

R Core Team. 2021. R: A language and environment for statistical computing, version 4.0.5. R Foundation for Statistical Computing, Vienna, Austria.

Roff, D. A., and J. Fairbairn. 2012. A test of the hypothesis that correlational selection generates genetic correlations. Evolution 66:2953-2960.

Roff, D. A., and T. Mousseau. 1999. Does natural selection alter genetic architecture? An evaluation of quantitative genetic variation among populations of *Allonemobius socius* and *A. fasciatus*. Journal of Evolutionary Biology 12:361-369.

Santos, M., P. F. Iriarte, W. Céspedes, J. Balanyà, A. Fontdevila, and L. Serra. 2004. Swift laboratory thermal evolution of wing shape (but not size) in *Drosophila subobscura* and its relationship with chromosomal inversion polymorphism. Journal of Evolutionary Biology 17:841-855.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution 50:1766-1774.

—. 2000, The Ecology of Adaptive Radiation. Oxford, Oxford University Press.

Sgrò, C. M., and A. A. Hoffmann. 2004. Genetic correlations, tradeoffs and environmental variation. Heredity 93:241-248.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733-756.

—. 1978. Spatial patterns in the distributions of polygenic characters. Journal of Theoretical Biology 70:213-228.

Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends in Ecology & Evolution 17:320-327.

Steward, R. A., M. A. de Jong, V. Oostra, and C. W. Wheat. 2022. Alternative splicing in seasonal plasticity and the potential for adaptation to environmental change. Nature Communications 13:755.

van Heerwaarden, B., and C. M. Sgrò. 2017. The quantitative genetic basis of clinal divergence in phenotypic plasticity. Evolution 71:2618-2633.

Via, S., R. Gomulkiewicz, G. de Jong, S. M. Scheiner, C. D. Schlichting, and P. H. Van Tienderen. 1995. Adaptive phenotypic plasticity: consensus and controversy. Trends in Ecology & Evolution 10:212-217.

Via, S., and R. Lande. 1987. Evolution of genetic variability in a spatially heterogeneous environment: effects of genotype-environment interaction. Genetics research 49:147-156.

Walsh, B., and M. W. Blows. 2009. Abundant Genetic Variation plus Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. Annual Review of Ecology, Evolution, and Systematics 40:41-59.

Walter, G. M., J. D. Aguirre, M. W. Blows, and D. Ortiz-Barrientos. 2018. Evolution of genetic variance during adaptive radiation. The American Naturalist 191:E108–E128.

Walter, G. M., D. Terranova, J. Clark, S. Cozzolino, A. Cristaudo, S. J. Hiscock, and J. R. Bridle. 2021. Environmental effects on genetic variance and the alignment with plasticity and selection. bioRxiv https://doi.org/10.1101/2021.02.08.430333.

West-Eberhard, M. J. 2003, Developmental Plasticity and Evolution. New York, Oxford University Press.

Wood, C. W., and E. D. Brodie III. 2015. Environmental effects on the structure of the G-matrix. Evolution 69:2927-2940.

Yeaman, S., Y. K. Chen, and M. C. Whitlock. 2010a. No effect of environmental heterogeneity on the maintenance of genetic variation in wing shape in *Drosophila melanogaster*. Evolution 64:3398-3408.

Yeaman, S., C. Yukon, and M. C. Whitlock. 2010b. Data from: No effect of environmental heterogeneity on the maintenance of genetic variation in wing shape in *Drosophila melanogaster*. Dryad, Dataset, https://doi.org/10.5061/dryad.1719.

Zeng, Z. B. 1988. Long-Term Correlated Response, Interpopulation Covariation, and Interspecific Allometry. Evolution 42:363-374.

Zu, P., F. P. Schiestl, D. Gervasi, X. Li, D. Runcie, and F. Guillaume. 2020. Floral signals evolve in a predictable way under artificial and pollinator selection in *Brassica rapa*. BMC Evolutionary Biology 20:127.

Tables

Table 1 Glossary of quantitative genetics parameters estimated.

Term	Sym-bol	Calculation	Definition
D-matrix	**D**	Covariance of means for multiple traits (here, treatment and assay)	The (co)variance matrix that captures differences in mean phenotype across assays and treatments.
$\boldsymbol{d}_{max}$ / $\boldsymbol{d}_2$		The first two eigenvectors of **D**	The two axes that describe the direction of the greatest differences in mean multivariate phenotype.
G-matrix	**G**	The variance in breeding values for all individuals where wing traits were measured. Estimated for each cage within assay and treatment.	The additive genetic (co)variance matrix underlying a set of traits. Genetic variances on the diagonal and genetic covariances among traits off the diagonal.
$\boldsymbol{g}_{max}$		The first eigenvector of **G**.	The axis describing the direction containing the greatest amount of additive genetic variance.
M-matrix	**M**	Variance among the dams in the pedigree. Estimated for each cage within assay and treatment.	The maternal (co)variance matrix that represents the maternal contribution to the phenotype.
R-matrix	**R**	Variance among individual flies after removing the genetic and maternal effects. Estimated for each cage within assay and treatment.	The residual (co)variance matrix that represents the residual phenotype that includes differences among siblings not due to additive or maternal effects.
S-matrix	**S**	For each element of **G**, variances and covariances among matrices (treatment and assay) are calculated.	A symmetric matrix used for a tensor analysis that captures the element-by-element differences among the original G-matrices.
Eigentensor (matrix)	**E**	Eigenvectors of **S** describing axes of differences in genetic variance and that are scaled and rearranged to form the eigentensors.	Orthogonal axes describing differences among the original matrices.
Coordinates of an eigentensor		The correlation between the original matrices and each eigentensor.	Quantifies which matrices contribute to the differences among all matrices that are captured by an eigentensor.

Table 2 Eigendecomposition of G-matrices for all treatments estimated in the **(a)** cold assay, and **(b)** hot assay. Treatments are labelled along the top row (Hom. = Homogeneous). Eigenvalues have been multiplied by 1000 for ease of presentation. Lower and upper represent the 90% HPD credible intervals of each eigenvalue. Proportion is of the total genetic variance that each eigenvector describes. Trait loadings in bold are greater than 0.2 to aid interpretation. For simplicity only the first three eigenvectors of each matrix are presented, which describes >90% of total genetic variance in each matrix.

	Hom. Cold			**Hom. Hot**			**Migration Cold**			**Migration Hot**			**Spatial**			**Temporal**		
(a) cold assay	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3
Eigenvalues	2.90	1.96	1.45	2.65	1.83	1.34	2.59	1.87	1.38	2.50	1.73	1.17	2.44	1.86	1.26	2.86	2.07	1.38
HPD lower	2.52	1.67	1.18	2.28	1.49	1.08	2.23	1.54	1.08	2.11	1.41	0.91	2.12	1.48	1.01	2.53	1.74	1.09
HPD upper	3.39	2.36	1.66	3.06	2.19	1.54	3.05	2.20	1.53	2.97	2.09	1.33	2.90	2.16	1.43	3.25	2.34	1.56
Proportion	0.44	0.29	0.22	0.43	0.30	0.22	0.42	0.30	0.22	0.43	0.30	0.20	0.41	0.31	0.21	0.43	0.31	0.21
Trait loadings:																		
Centroid	0.00	0.00	-0.08	-0.04	-0.02	-0.09	0.01	0.01	-0.12	-0.02	0.04	-0.13	-0.03	-0.01	-0.09	-0.02	-0.04	-0.06
Line 9-10	**-0.36**	**0.88**	**-0.29**	**-0.62**	**-0.78**	-0.04	**-0.41**	**-0.83**	**-0.36**	**-0.32**	**-0.83**	**-0.45**	**-0.30**	**0.90**	**-0.30**	**-0.52**	**-0.84**	0.18
Angle 7-8-9	**-0.93**	**-0.36**	0.07	**-0.78**	**0.62**	0.02	**-0.91**	**0.41**	0.08	**-0.94**	**0.32**	0.10	**-0.95**	**-0.30**	0.06	**-0.85**	**0.53**	0.04
Angle 3-10-4	-0.04	**0.29**	**0.94**	-0.01	-0.04	**0.99**	-0.07	**-0.34**	**0.92**	-0.05	**-0.45**	**0.87**	-0.03	**0.30**	**0.94**	0.13	0.12	**0.97**
Angle 2-4-8	0.04	-0.10	-0.14	0.07	0.04	-0.08	0.09	0.15	-0.10	0.08	0.08	-0.11	0.07	-0.09	-0.13	0.03	-0.08	-0.15
(b) hot assay	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3	g_{max}	g_2	g_3
Eigenvalues	2.26	1.91	1.00	2.57	1.66	1.09	2.38	1.45	0.97	2.97	1.87	1.13	2.43	1.31	0.94	2.79	1.68	0.99
HPD lower	1.92	1.49	0.79	2.06	1.39	0.88	1.96	1.16	0.77	2.45	1.54	0.88	2.07	0.98	0.72	2.16	1.36	0.79
HPD upper	2.83	2.15	1.17	3.16	1.95	1.27	2.94	1.68	1.12	3.65	2.10	1.31	2.85	1.62	1.11	3.41	1.95	1.19
Proportion	0.41	0.35	0.18	0.46	0.30	0.19	0.46	0.28	0.19	0.47	0.30	0.18	0.49	0.26	0.19	0.49	0.29	0.17
Trait loadings:																		
Centroid	0.00	0.01	-0.16	0.03	0.03	0.09	-0.01	-0.04	-0.11	0.00	0.02	-0.09	-0.01	-0.01	-0.11	0.01	0.00	0.07
Line 9-10	**-0.93**	**0.32**	-0.14	**-0.86**	**0.43**	**0.27**	**-0.91**	**0.38**	-0.18	**-0.94**	0.09	**-0.32**	**-0.42**	**0.85**	**-0.28**	**-0.93**	0.09	**0.35**
Angle 7-8-9	**-0.32**	**-0.94**	-0.04	**-0.39**	**-0.90**	0.17	**-0.37**	**-0.92**	-0.05	-0.06	**-0.99**	-0.09	**-0.89**	**-0.45**	-0.01	-0.16	**-0.97**	-0.18
Angle 3-10-4	-0.15	0.01	**0.97**	**-0.31**	-0.03	**-0.94**	-0.19	0.02	**0.98**	**-0.32**	-0.05	**0.94**	-0.13	**0.24**	**0.95**	**-0.32**	**0.23**	**-0.91**
Angle 2-4-8	0.01	0.06	-0.04	0.11	0.06	0.05	0.10	0.05	-0.01	0.09	0.08	-0.10	0.11	-0.09	-0.06	0.09	0.03	0.05

Figures

(a)

Code	Name	Description
C	Cold homogeneous	Two cages at 16°C, 8 bottles per cage, 4 bottles migrate reciprocally every 4 weeks
H	Hot homogeneous	Two cages at 25°C, 8 bottles per cage, 4 bottles migrate reciprocally every 4 weeks
S	Spatially heterogeneous with panmixia	One cage at 25°C, one at 16°C, 8 bottles per cage, 4 bottles migrate reciprocally every 4 weeks
T	Temporally heterogeneous	Two cages maintained together, 8 bottles per cage, 4 bottles migrate reciprocally, cages move between 25°C and 16°C every 4 weeks
M	Spatially heterogeneous with limited migration	One cage at 25°C (MH), one at 16°C (MC), 8 bottles per cage, 2 mated females migrate reciprocally every 4 weeks

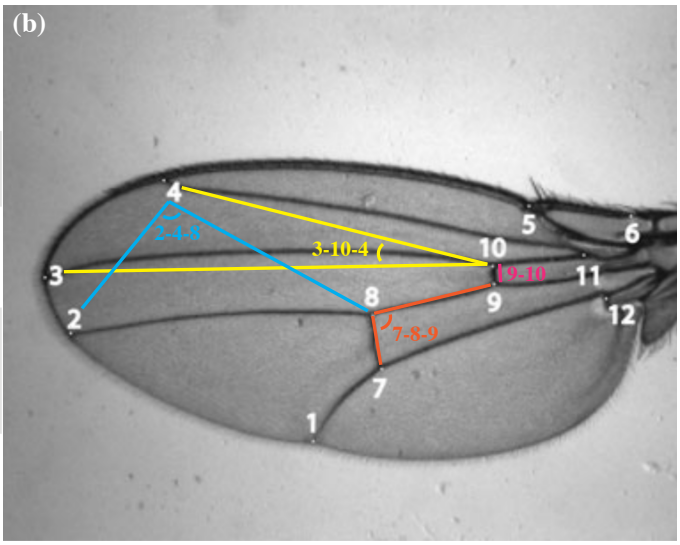

Fig. 1 (a) Treatments of environmental heterogeneity for the experimental evolution experiment from Yeaman et al. (2010a), each treatment was applied using five replicate populations with two cages per replicate. **(b)** The wing traits measured, with each trait represented by different colored lines (centroid size not pictured). Reproduced with permission.

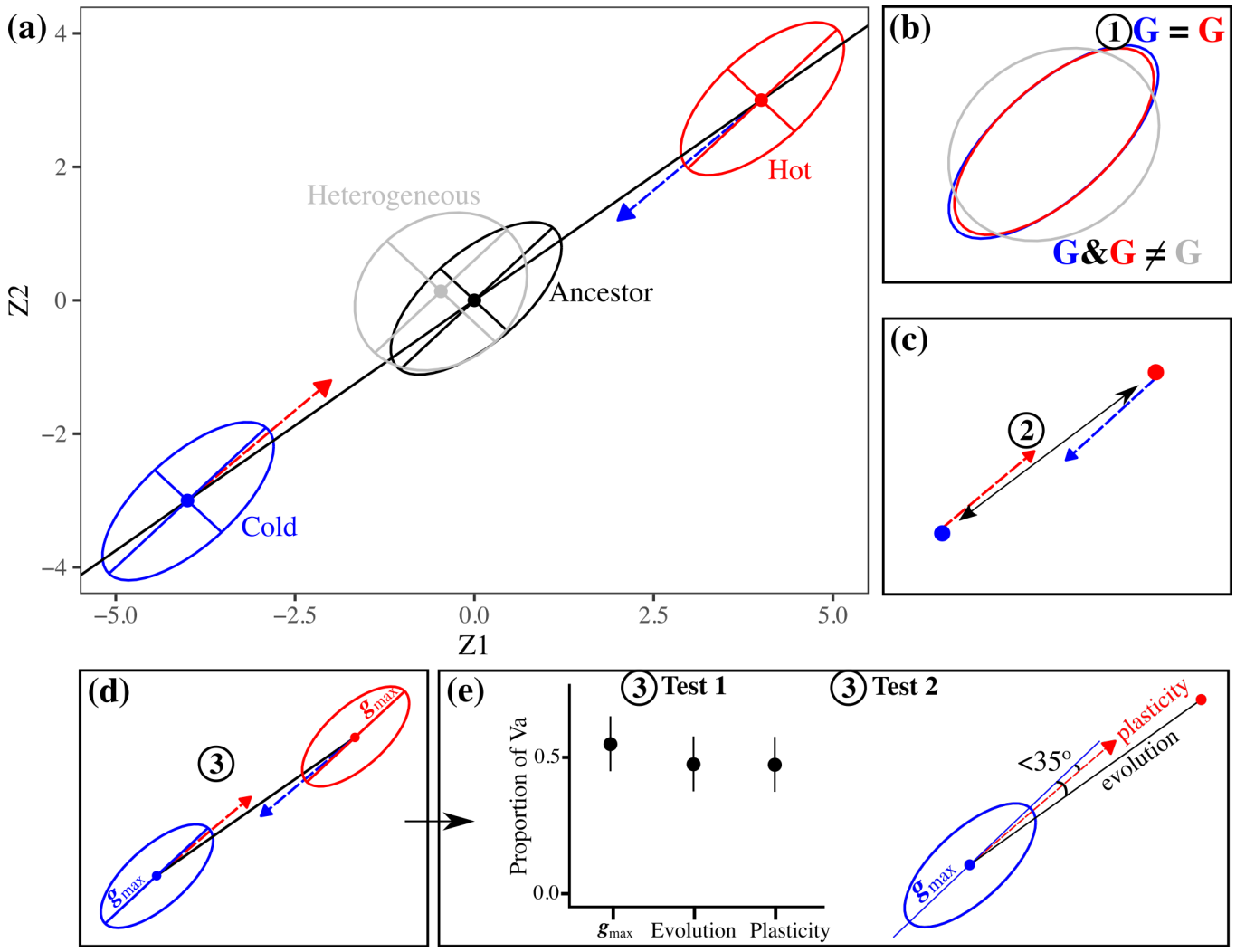

Fig. 2 (a) Conceptual diagram depicting how genetic variance in two traits (Z1 and Z2) that are genetically correlated would be associated with plasticity and phenotypic evolution if three predictions are true **(b-e)**. Mean phenotypes are represented by filled circles, and G-matrices are represented by the ellipses, where $\boldsymbol{g}_{max}$ is shown as the line along the long-axis of the ellipse, and $\boldsymbol{g}_2$ as the shorter axis that is perpendicular to $\boldsymbol{g}_{max}$. Populations evolving in homogeneous hot (red) and cold (blue) temperatures relative to their common ancestor (black), and a population exposed to alternating temperatures representing a heterogeneous environment (gray). Arrows represent plastic changes in mean phenotype when exposed to the other temperature. Numbers in panels **b-d** refer to the predictions. **(b)** Prediction 1: If **G** is stable in homogeneous environments, we expect to observe similar patterns of genetic variance, suggesting little change in **G** from the common ancestor. Conversely, if environmental heterogeneity influences genetic variance, we expect the distribution or orientation of **G** to change relative to the homogeneous treatments (gray ellipse is different to the hot and cold ellipses). **(c)** Prediction 2: Changes in mean phenotype due to plasticity between assays (colored arrows) and evolved changes between treatments (solid black arrow) would be significant and occur in similar directions. **(d)** Prediction 3: As evidence that the orientation of genetic variance determines the direction of evolution and plasticity, changes in mean phenotype (plasticity across assays and evolved differences among treatments) would

occur in a direction of high evolvability (i.e., changes in mean phenotype would align with genetic variance). There are two methods for testing an alignment between genetic variance and changes in mean phenotype **(e)**: Test 1 – Using matrix projection, vectors representing evolved and plastic changes in mean phenotype would describe a similar proportion of genetic variance (V_a) to that described by $\boldsymbol{g}_{max}$ (the direction of maximum V_a); Test 2 – $\boldsymbol{g}_{max}$ estimated in each treatment would align closely with both plastic and evolved change in mean phenotype with an angle less than 35º (closer than expected under random sampling).

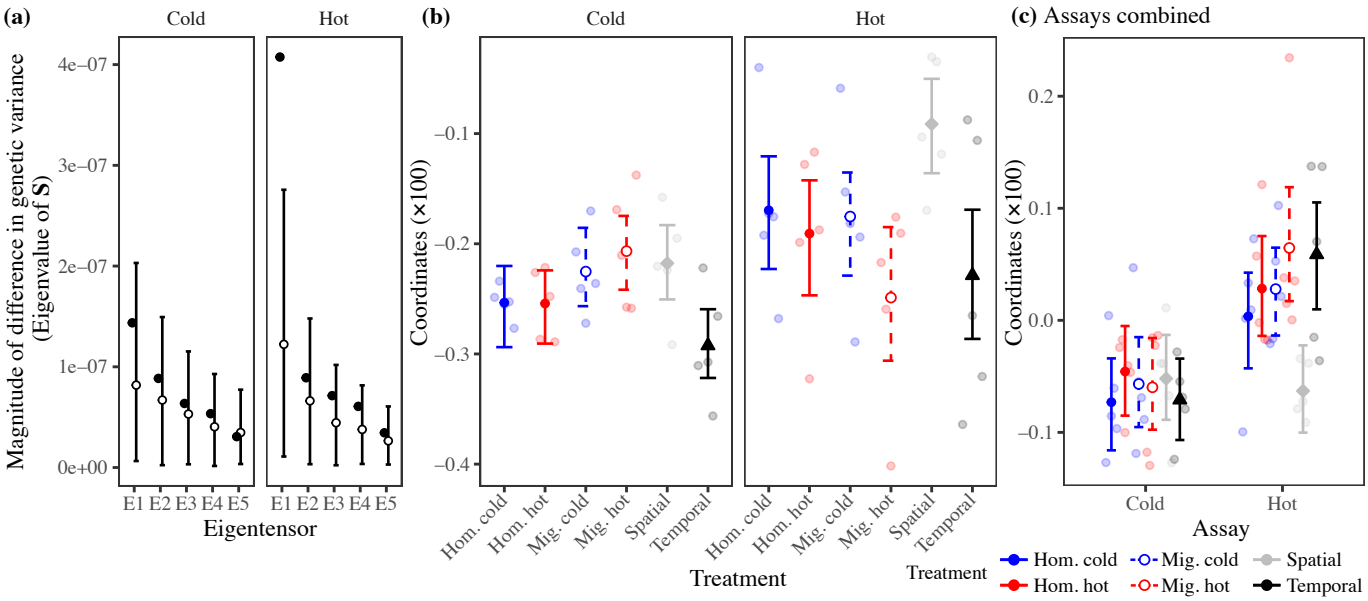

Fig. 3 Tensor analysis of **G** revealed significant differences in genetic variance among treatments and across assays. **(a)** Significant differences in genetic variance are represented by eigentensors that describe greater differences in observed **G** (filled circles) compared to the null distribution (open circles with 90% HPD intervals). The null distribution is created by calculating the mean from each of 500 separate implementations on data containing no expected differences in genetic variance among treatments (see **Methods S2**). Where the posterior mean from the observed estimate of **G** exceeds the distribution of posterior means (calculated from each of the 500 null models), there is evidence for significant difference in genetic variance. Only one significant eigentensor in the hot assay describes greater differences in genetic variance than expected under the null. **(b)** The coordinates for the first eigentensor in each assay describe how the original matrices (i.e. treatments) contributes to the differences described by each eigentensor. Hom. = Homogeneous, Mig. = Migration, Spat. = Spatial and Temp. = Temporal. In the cold assay, temporal heterogeneity shows small differences in genetic variance from spatial heterogeneity. In the hot assay, spatial heterogeneity shows the strongest differences in genetic variance from the other treatments. **(c)** Including all 12 G-matrices in the tensor analysis shows that assay temperature creates large changes in genetic variance for all treatments, except for spatial heterogeneity.

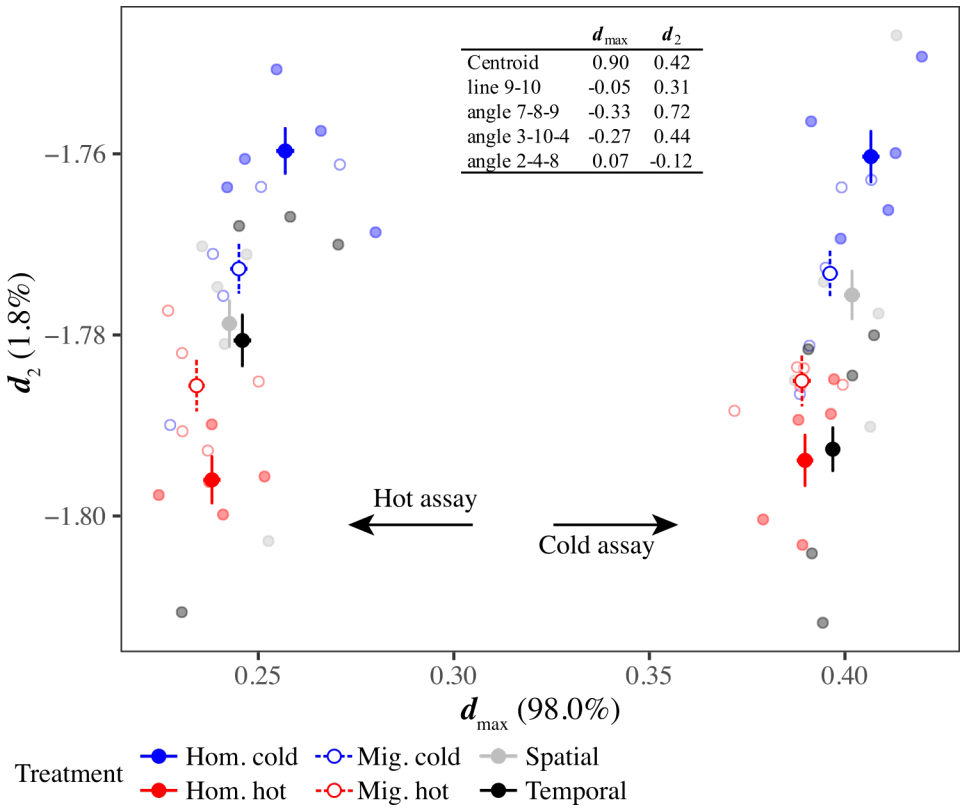

Fig. 4 Differences in multivariate mean phenotype separated the hot and cold assays in the first axis (d_{max}), and the six treatments along the second axis (d_2). Large circles with error bars (±1 SE) represent the mean for each treatment measured in each assay. Small circles represent the mean of each replicate cage. Inset table presents the trait loadings for each axis (loadings in bold are greater than 0.2 to aid interpretation). Red and blue lines represent the hot and cold d_{max}, which captures the difference in phenotype among treatments within each assay.

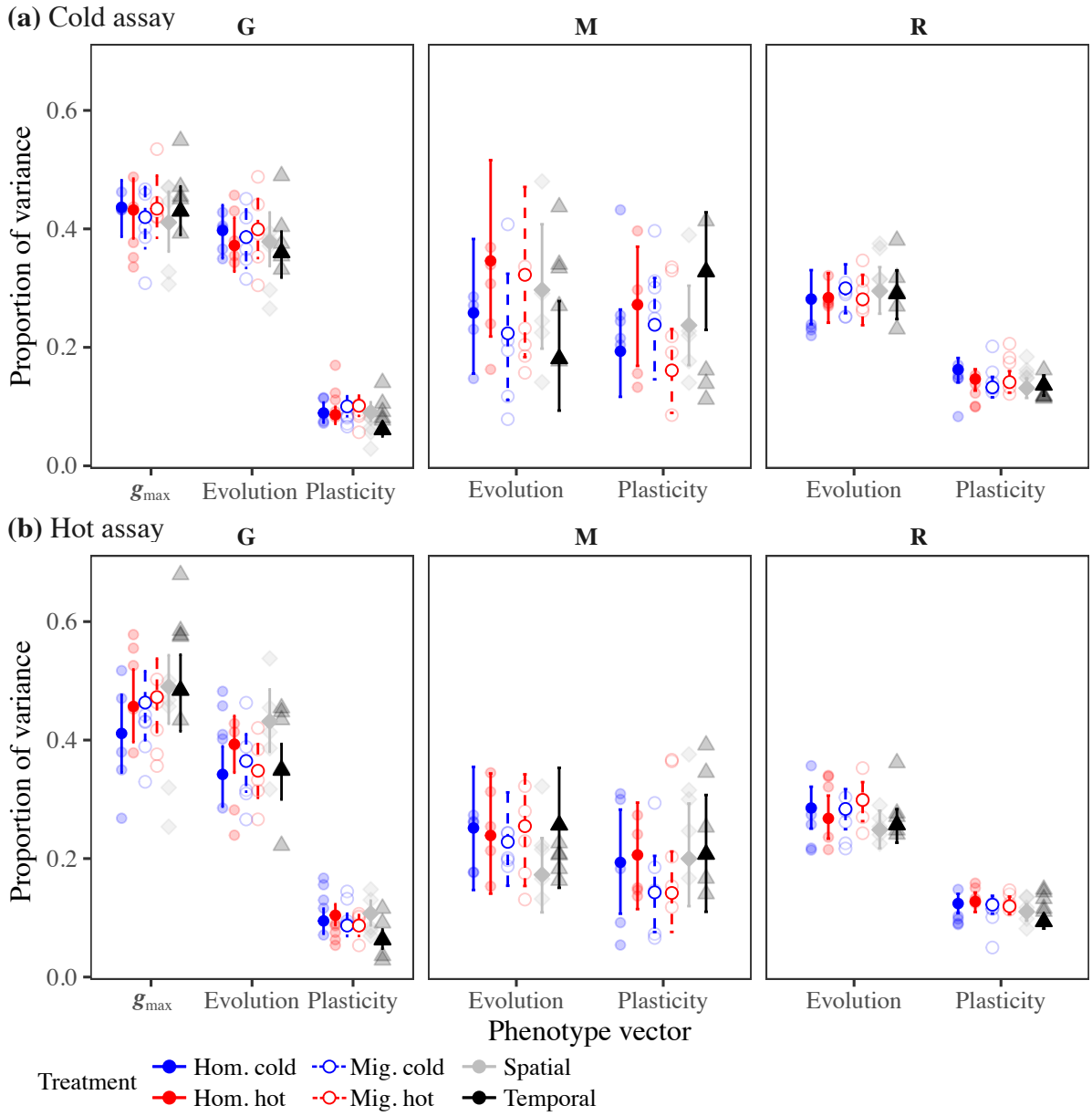

Fig. 5 For each treatment measured in the cold assay **(a)** and the hot assay **(b)**, the proportion of genetic (**G**), maternal (**M**) and environmental (**R**) variance in the direction of evolved differences in phenotype ($\boldsymbol{d}_2$ in **Fig. 4**) and plasticity associated with changes across assays ($\boldsymbol{d}_{max}$ in **Fig. 4**). Colored circles with credible intervals (90% HPD intervals) represent the amount of variance described by the average of the five replicate cages. Hom. = Homogeneous, Mig. = Migration. Lighter circles represent the posterior mean for each replicate cage. Evolved differences among treatments describes a greater proportion of genetic variance compared to plastic changes in phenotype.

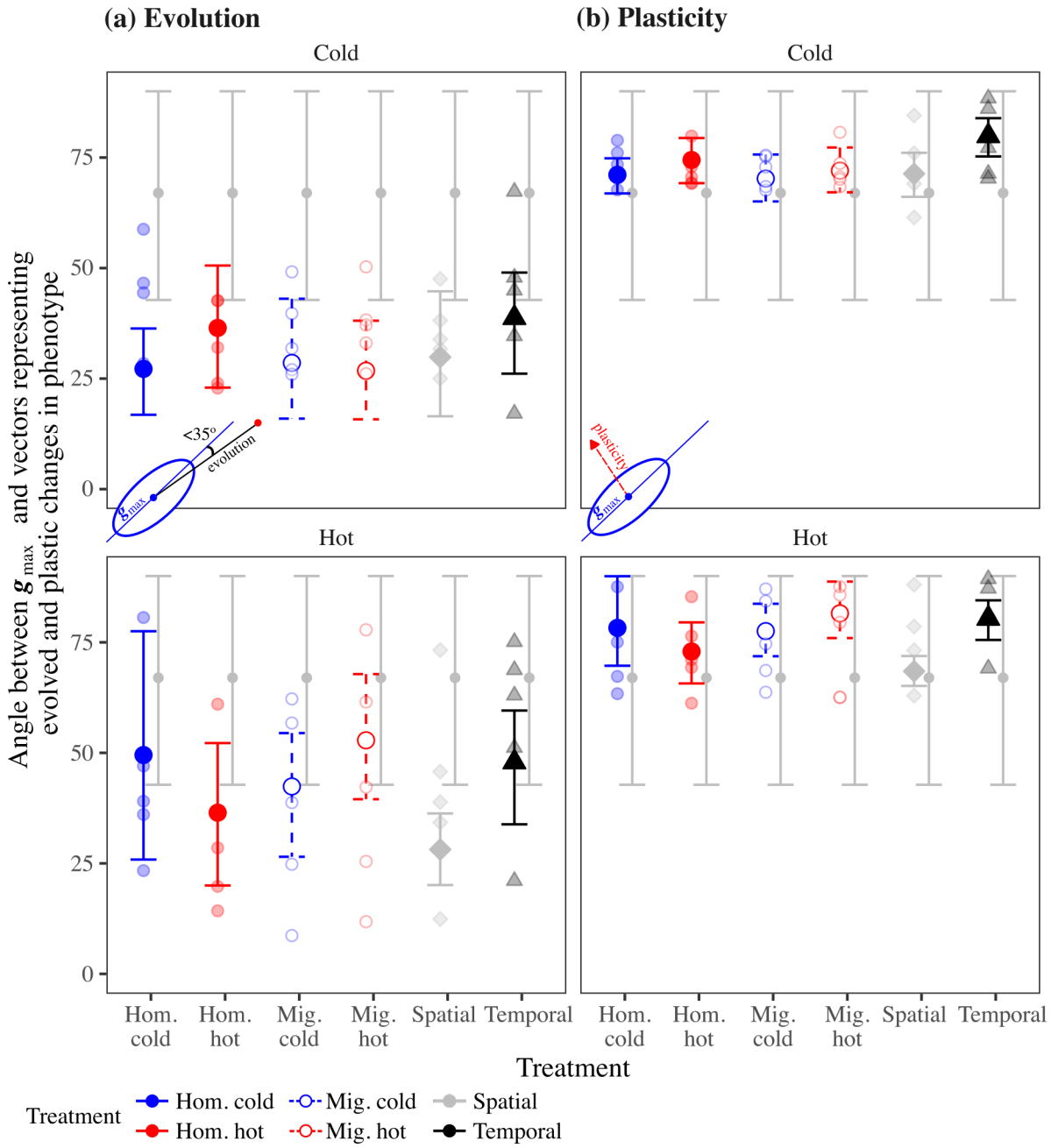

Fig. 6 For both assay temperatures, the angles between $\boldsymbol{g}_{max}$ and **(a)** evolution, and **(b)** plasticity. See inset diagrams for a summary of the result: Small angles in **(a)** suggest a close alignment between $\boldsymbol{g}_{max}$ and phenotypic divergence ($\boldsymbol{d}_2$ in **Fig. 4**), compared to large angles shown in **(b)** that suggest no alignment between $\boldsymbol{g}_{max}$ and plasticity ($\boldsymbol{d}_{max}$ in **Fig.4**). Large symbols with 90% HPD intervals represent the alignment for the posterior distribution of $\boldsymbol{g}_{max}$ calculated from mean **G**. Smaller symbols represent the alignment estimated for each replicate cage within treatments. Gray circles and credible intervals represent the null distribution.

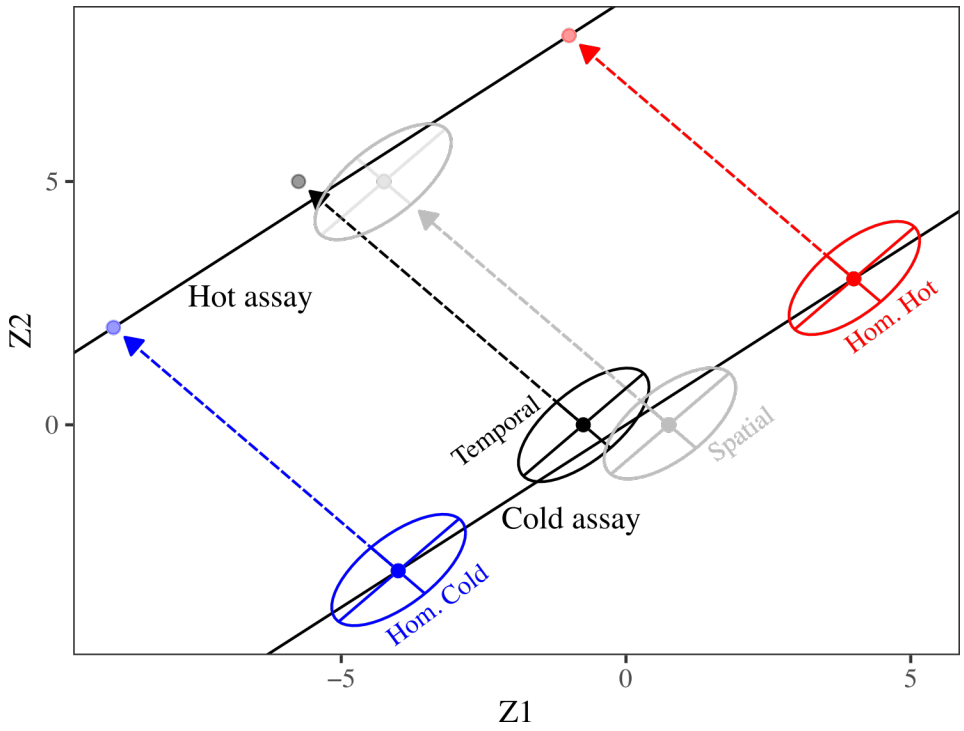

Fig. 7 Conceptual diagram summarizing the main results for two traits (Z1 and Z2) with circles representing the mean phenotype, ellipses representing **G** and dashed arrows representing plasticity across assay temperatures. Homogeneous hot and cold treatments diverged in mean phenotype along genetic lines of least resistance as both $\boldsymbol{g}_{max}$ (in the cold assay) and shared genetic subspace (for both assays). Plasticity changed the phenotype in a different direction to evolution. Temporal and spatial heterogeneity made little change in mean phenotype. **G** changed the most across assay temperatures for all treatments (not shown), except for spatial heterogeneity which showed a stable **G** across assays and a consistent alignment with phenotypic divergence.

Online Supplement for

Experimental evidence that phenotypic evolution but not plasticity occurs along genetic lines of least resistance in homogeneous environments

The American Naturalist

Greg M Walter
Email: greg.walter@monash.edu
Address: School of Biological Sciences, Monash University, Melbourne 3800

Contents

Table S1: G-matrices for each treatment assayed in **(a)** hot and **(b)** cold temperatures. Genetic variances are presented along the diagonal (gray shading), with genetic covariances above the diagonal. Variances and covariance are multiplied by 1,000 to aid interpretation. Genetic correlations are presented below the diagonal. Numbers in parentheses represent the 95% HPD intervals for each parameter estimated. Correlations in bold are significant and do not overlap zero.

		(a) Cold assay temperature					**(b) Hot assay temperature**				
Treatment		**Cent-roid**	**Line 9-10**	**Angle 7-8-9**	**Angle 3-10-4**	**Angle 2-4-8**	**Cent-roid**	**Line 9-10**	**Angle 7-8-9**	**Angle 3-10-4**	**Angle 2-4-8**
Cold homogen-eous (c)	**Centroid**	0.172 (0.098, 0.24)	0.036 (-0.066, 0.125)	0.001 (-0.112, 0.103)	-0.101 (-0.173, -0.017)	0.012 (-0.02, 0.039)	0.203 (0.126, 0.272)	0.016 (-0.102, 0.124)	-0.013 (-0.112, 0.081)	-0.144 (-0.228, -0.067)	0 (-0.028, 0.027)
	Line 9-10	0.06 (-0.09, 0.23)	2.11 (1.709, 2.505)	0.342 (0.081, 0.632)	0.134 (-0.061, 0.332)	-0.154 (-0.243, -0.083)	0.02 (-0.14, 0.19)	2.276 (1.728, 2.842)	0.073 (-0.226, 0.329)	0.196 (-0.03, 0.407)	0.026 (-0.061, 0.106)
	Angle 7-8-9	0 (-0.16, 0.16)	**0.14 (0.04, 0.26)**	2.869 (2.437, 3.311)	-0.006 (-0.224, 0.233)	-0.056 (-0.132, 0.042)	-0.02 (-0.16, 0.14)	0.03 (-0.1, 0.16)	1.999 (1.649, 2.378)	0.046 (-0.139, 0.238)	-0.109 (-0.183, -0.049)
	Angle 3-10-4	**-0.2 (-0.34, -0.05)**	0.08 (-0.03, 0.19)	0 (-0.11, 0.1)	1.505 (1.263, 1.778)	-0.232 (-0.302, -0.174)	**-0.31 (-0.47, -0.18)**	0.13 (-0.01, 0.26)	0.03 (-0.1, 0.15)	1.056 (0.836, 1.263)	-0.038 (-0.095, 0.018)
	Angle 2-4-8	0.06 (-0.08, 0.2)	**-0.21 (-0.33, -0.12)**	-0.07 (-0.16, 0.05)	**-0.38 (-0.47, -0.29)**	0.249 (0.212, 0.285)	0 (-0.15, 0.16)	0.04 (-0.1, 0.17)	**-0.19 (-0.31, -0.09)**	-0.09 (-0.2, 0.06)	0.167 (0.136, 0.196)
Hot homogen-eous (h)	**Centroid**	0.167 (0.106, 0.226)	0.09 (0.006, 0.173)	0.062 (-0.028, 0.148)	-0.104 (-0.186, -0.041)	-0.016 (-0.045, 0.008)	0.166 (0.103, 0.222)	-0.02 (-0.124, 0.074)	-0.057 (-0.132, 0.028)	-0.111 (-0.181, -0.039)	0.001 (-0.024, 0.025)
	Line 9-10	**0.15 (0.01, 0.3)**	2.22 (1.81, 2.683)	0.394 (0.124, 0.679)	0.016 (-0.174, 0.245)	-0.162 (-0.235, -0.07)	-0.03 (-0.21, 0.11)	2.414 (1.808, 2.966)	0.286 (0.034, 0.597)	0.381 (0.168, 0.63)	-0.185 (-0.273, -0.099)
	Angle 7-8-9	0.1 (-0.04, 0.23)	**0.17 (0.03, 0.28)**	2.409 (2.017, 2.827)	-0.005 (-0.181, 0.23)	-0.096 (-0.17, -0.002)	-0.1 (-0.25, 0.05)	**0.13 (0, 0.26)**	1.842 (1.546, 2.194)	0.182 (0.01, 0.346)	-0.182 (-0.246, -0.113)
	Angle 3-10-4	**-0.22 (-0.37, -0.08)**	0.01 (-0.12, 0.12)	0 (-0.1, 0.13)	1.357 (1.115, 1.563)	-0.097 (-0.165, -0.037)	**-0.24 (-0.38, -0.1)**	**0.22 (0.11, 0.35)**	**0.12 (0.01, 0.22)**	1.276 (1.06, 1.485)	-0.136 (-0.193, -0.076)
	Angle 2-4-8	-0.09 (-0.21, 0.07)	**-0.23 (-0.34, -0.12)**	**-0.13 (-0.23, 0)**	**-0.18 (-0.29, -0.07)**	0.216 (0.182, 0.254)	0 (-0.14, 0.14)	**-0.27 (-0.38, -0.14)**	**-0.31 (-0.41, -0.19)**	**-0.27 (-0.38, -0.17)**	0.191 (0.16, 0.227)
Cold spatially heterogen-eous with migration (mc)	**Centroid**	0.184 (0.129, 0.248)	0.03 (-0.046, 0.128)	-0.029 (-0.125, 0.066)	-0.156 (-0.234, -0.08)	0.011 (-0.015, 0.041)	0.215 (0.146, 0.276)	0.018 (-0.091, 0.123)	0.061 (-0.013, 0.145)	-0.092 (-0.164, -0.025)	0.008 (-0.02, 0.034)
	Line 9-10	0.05 (-0.1, 0.2)	1.951 (1.558, 2.368)	0.296 (0.006, 0.577)	0.158 (-0.057, 0.371)	-0.266 (-0.346, -0.178)	0.03 (-0.13, 0.18)	2.259 (1.743, 2.74)	0.302 (0.052, 0.549)	0.252 (0.06, 0.442)	-0.182 (-0.265, -0.091)
	Angle 7-8-9	-0.04 (-0.2, 0.08)	**0.13 (0.01, 0.26)**	2.553 (2.053, 2.992)	0.016 (-0.197, 0.27)	-0.094 (-0.192, -0.005)	0.11 (-0.02, 0.24)	**0.16 (0.03, 0.28)**	1.628 (1.336, 1.932)	0.097 (-0.053, 0.242)	-0.148 (-0.215, -0.085)
	Angle 3-10-4	**-0.3 (-0.43, -0.16)**	0.09 (-0.03, 0.22)	0.01 (-0.12, 0.13)	1.457 (1.211, 1.716)	-0.219 (-0.3, -0.154)	**-0.19 (-0.33, -0.05)**	**0.16 (0.04, 0.27)**	0.07 (-0.03, 0.19)	1.052 (0.856, 1.23)	-0.047 (-0.098, 0.003)
	Angle 2-4-8	0.05 (-0.07, 0.19)	**-0.38 (-0.47, -0.27)**	**-0.12 (-0.23, 0)**	**-0.36 (-0.45, -0.26)**	0.255 (0.21, 0.292)	0.04 (-0.11, 0.16)	**-0.29 (-0.4, -0.17)**	**-0.27 (-0.38, -0.15)**	**-0.11 (-0.22, 0)**	0.18 (0.15, 0.214)

Table S1 (continued):

Treatment		(a) Cold assay temperature					(b) Hot assay temperature				
		Cent-roid	**Line 9-10**	**Angle 7-8-9**	**Angle 3-10-4**	**Angle 2-4-8**	**Cent-roid**	**Line 9-10**	**Angle 7-8-9**	**Angle 3-10-4**	**Angle 2-4-8**
Hot spatially heterogeneous with migration (mh)	**Centroid**	0.192 (0.13, 0.26)	0.022 (-0.067, 0.107)	0.045 (-0.071, 0.15)	-0.142 (-0.218, -0.061)	0.023 (-0.011, 0.05)	0.181 (0.122, 0.245)	0.04 (-0.055, 0.153)	-0.026 (-0.099, 0.061)	-0.086 (-0.151, -0.01)	0.016 (-0.012, 0.042)
	Line 9-10	0.04 (-0.11, 0.2)	1.756 (1.382, 2.148)	0.255 (-0.004, 0.534)	0.232 (0.019, 0.429)	-0.118 (-0.205, -0.05)	0.06 (-0.09, 0.19)	2.87 (2.283, 3.508)	0.031 (-0.247, 0.299)	0.563 (0.293, 0.811)	-0.196 (-0.292, -0.097)
	Angle 7-8-9	0.06 (-0.1, 0.22)	0.12 (-0.01, 0.25)	2.49 (2.005, 2.93)	-0.009 (-0.219, 0.223)	-0.147 (-0.224, -0.05)	-0.04 (-0.17, 0.11)	0.01 (-0.1, 0.14)	1.896 (1.591, 2.21)	0.064 (-0.128, 0.234)	-0.15 (-0.219, -0.085)
	Angle 3-10-4	**-0.29 (-0.43, -0.13)**	**0.15 (0.03, 0.3)**	-0.01 (-0.12, 0.12)	1.307 (1.039, 1.537)	-0.163 (-0.232, -0.1)	**-0.17 (-0.29, -0.02)**	**0.28 (0.15, 0.39)**	0.04 (-0.08, 0.15)	1.368 (1.132, 1.638)	-0.18 (-0.249, -0.122)
	Angle 2-4-8	0.11 (-0.03, 0.25)	**-0.18 (-0.31, -0.08)**	**-0.19 (-0.31, -0.08)**	**-0.3 (-0.41, -0.19)**	0.231 (0.192, 0.271)	0.09 (-0.06, 0.23)	**-0.26 (-0.39, -0.14)**	**-0.25 (-0.35, -0.15)**	**-0.35 (-0.44, -0.24)**	0.198 (0.161, 0.229)
Spatially heterogeneous with panmixia (s)	**Centroid**	0.188 (0.126, 0.248)	0.037 (-0.054, 0.13)	0.077 (-0.02, 0.167)	-0.108 (-0.185, -0.035)	-0.021 (-0.05, 0.007)	0.158 (0.094, 0.225)	0.034 (-0.051, 0.117)	0.032 (-0.049, 0.117)	-0.089 (-0.152, -0.022)	-0.011 (-0.036, 0.011)
	Line 9-10	0.06 (-0.1, 0.2)	1.924 (1.534, 2.307)	0.172 (-0.099, 0.42)	0.169 (-0.003, 0.372)	-0.134 (-0.213, -0.057)	0.07 (-0.1, 0.25)	1.491 (1.023, 1.967)	0.419 (0.147, 0.699)	0.15 (-0.07, 0.32)	-0.191 (-0.261, -0.108)
	Angle 7-8-9	0.11 (-0.03, 0.24)	0.08 (-0.05, 0.19)	2.473 (2.032, 2.853)	-0.019 (-0.219, 0.193)	-0.106 (-0.185, -0.022)	0.05 (-0.1, 0.18)	**0.23 (0.09, 0.37)**	2.289 (1.953, 2.683)	0.126 (-0.046, 0.31)	-0.195 (-0.263, -0.131)
	Angle 3-10-4	**-0.22 (-0.34, -0.07)**	0.11 (-0.01, 0.22)	-0.01 (-0.12, 0.11)	1.327 (1.1, 1.566)	-0.189 (-0.257, -0.128)	**-0.22 (-0.37, -0.08)**	0.12 (-0.02, 0.29)	0.08 (-0.03, 0.2)	1.02 (0.806, 1.241)	-0.103 (-0.157, -0.049)
	Angle 2-4-8	-0.09 (-0.23, 0.03)	**-0.19 (-0.29, -0.08)**	**-0.13 (-0.23, -0.03)**	**-0.32 (-0.42, -0.23)**	0.265 (0.23, 0.302)	-0.07 (-0.21, 0.07)	**-0.35 (-0.49, -0.23)**	**-0.29 (-0.38, -0.19)**	**-0.23 (-0.33, -0.11)**	0.198 (0.169, 0.224)
Temporally heterogeneous (t)	**Centroid**	0.156 (0.101, 0.206)	0.083 (-0.006, 0.158)	0.014 (-0.072, 0.101)	-0.097 (-0.17, -0.028)	-0.011 (-0.039, 0.014)	0.159 (0.101, 0.231)	0.002 (-0.097, 0.111)	-0.017 (-0.098, 0.057)	-0.067 (-0.139, 0.006)	0.023 (-0.003, 0.046)
	Line 9-10	**0.14 (0.01, 0.28)**	2.342 (1.954, 2.695)	0.362 (0.113, 0.657)	-0.166 (-0.351, 0.044)	0.044 (-0.036, 0.12)	0 (-0.14, 0.19)	2.598 (2.075, 3.323)	0.2 (-0.082, 0.468)	0.544 (0.3, 0.791)	-0.211 (-0.306, -0.115)
	Angle 7-8-9	0.02 (-0.12, 0.15)	0.14 (0.05, 0.25)	2.714 (2.313, 3.077)	-0.131 (-0.323, 0.084)	-0.159 (-0.242, -0.079)	-0.03 (-0.19, 0.11)	0.09 (-0.03, 0.22)	1.749 (1.412, 2.067)	-0.07 (-0.256, 0.098)	-0.087 (-0.159, -0.023)
	Angle 3-10-4	**-0.2 (-0.35, -0.07)**	-0.09 (-0.19, 0.02)	-0.07 (-0.17, 0.03)	1.446 (1.234, 1.696)	-0.18 (-0.238, -0.113)	-0.15 (-0.29, 0.01)	0.3 (0.18, 0.42)	-0.05 (-0.17, 0.07)	1.237 (1.01, 1.461)	-0.108 (-0.169, -0.051)
	Angle 2-4-8	-0.06 (-0.19, 0.08)	0.06 (-0.03, 0.18)	**-0.19 (-0.29, -0.1)**	**-0.3 (-0.39, -0.2)**	0.253 (0.216, 0.29)	0.14 (-0.02, 0.26)	**-0.31 (-0.44, -0.18)**	**-0.15 (-0.28, -0.04)**	**-0.22 (-0.34, -0.12)**	0.186 (0.151, 0.221)

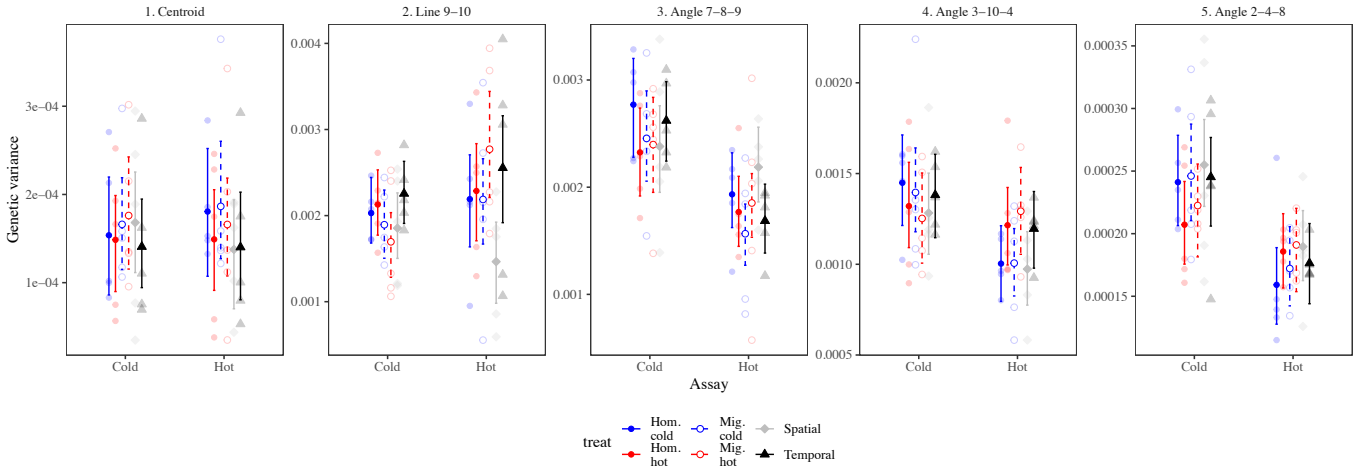

Fig. S1 Estimates of genetic variance for all traits. Dark shapes with 90% HPD intervals represent average **G** for each treatment. Smaller shapes represent the average for each cage within treatment.

Table S2 Metrics for comparing evolvability among treatments and assays. See **Methods S2** for the estimation procedure for all metrics. Presented are the posterior means for each metric with the 95% HPD (Highest Posterior Density intervals) in parentheses. Hom = Homogeneous, Mig = Migration.

Assay	Treatment	Dimensionality (n_D)	Maximum Evolvability ($\boldsymbol{e}_{max}$)	Total genetic variance (Trace of **G**)
Cold	Hom. cold	2.27 (2.01, 2.53)	0.0552 (0.0503, 0.0596)	0.0069 (0.0061, 0.0076)
	Hom. hot	2.29 (2.00, 2.55)	0.0528 (0.0484, 0.0572)	0.0064 (0.0057, 0.0072)
	Mig. cold	2.33 (2.03, 2.66)	0.0525 (0.0467, 0.0574)	0.0064 (0.0057, 0.0073)
	Mig. hot	2.27 (1.95, 2.57)	0.0514 (0.0467, 0.0571)	0.0060 (0.0052, 0.0068)
	Spatial	2.39 (2.04, 2.66)	0.0509 (0.0463, 0.0557)	0.0062 (0.0054, 0.0069)
	Temporal	2.30 (2.05, 2.57)	0.0548 (0.0504, 0.059)	0.0069 (0.0063, 0.0076)
Hot	Hom. cold	2.34 (2.05, 2.66)	0.0494 (0.0442, 0.0553)	0.0057 (0.0049, 0.0065)
	Hom. hot	2.16 (1.84, 2.48)	0.0523 (0.046, 0.0591)	0.0059 (0.0050, 0.0067)
	Mig. cold	2.15 (1.84, 2.46)	0.0499 (0.0439, 0.0558)	0.0053 (0.0047, 0.0061)
	Mig. hot	2.10 (1.81, 2.41)	0.0558 (0.049, 0.0623)	0.0065 (0.0055, 0.0073)
	Spatial	2.02 (1.75, 2.29)	0.0506 (0.0458, 0.0557)	0.0052 (0.0044, 0.0059)
	Temporal	2.07 (1.75, 2.41)	0.0536 (0.0462, 0.06)	0.0059 (0.0050, 0.0068)

Table S3 Matrix presenting the angles between $\boldsymbol{g}_{max}$ for each pairwise comparison of the treatments above the matrix diagonal (gray shading), and for $\boldsymbol{g}_2$ below the diagonal. Treatments are labelled within colored cells (Hom = Homogeneous, Mig = Migration). In the cold assay, $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_2$ are conserved among all treatments. In the hot assay, however, $\boldsymbol{g}_{max}$ and $\boldsymbol{g}_2$ are in similar directions for all treatments, except spatial heterogeneity, which shows larger angles with all other treatments for both vectors of **G**.

Assay		Hom. cold	Hom. hot	Mig. cold	Mig. hot	Spat-ial	Tem-poral
Cold	Hom. cold		17.3	4.1	3.3	4.6	14.0
	Hom. hot	21.9		15.1	19.6	21.0	10.7
	Mig. cold	6.1	22.7		5.3	7.6	14
	Mig. hot	10.3	30.0	9.2		2.5	16.5
	Spat-ial	4.1	25.3	9.1	9.8		16.8
	Tem-poral	27.7	13.1	31.0	36.7	29.6	
Hot	Hom. cold		12.1	6.4	18.3	45.3	13.9
	Hom. hot	7.0		7.6	19.7	40.2	13.8
	Mig. cold	4.3	6.1		19.3	41.7	14.0
	Mig. hot	14.3	20.3	18.0		59.8	5.9
	Spat-ial	45.5	40.5	42.1	59.7		53.9
	Tem-poral	18.6	25.1	20.7	16.8	55.2	

Table S4 Summaries of the covariance tensor analyses that captured differences among **G** in the cold and hot assay temperatures. λ represents the amount of difference in genetic variance described by each eigenvector of eigentensor. Eigenvalue of **S** represents the amount of difference in genetic variance described by each eigentensor, and the proportion represents the proportion of the total difference in genetic variance.

Assay	Eigen-tensor	Prop-ortion	Eigenvalue of S (90% HPD)	Eigenvector of eigentensor	λ	Cent-roid	Line 9-10	Angle 7-8-9	Angle 3-10-4	Angle 2-4-8
Cold	E1	0.32	1.54e-07 (5.06e-08, 2.62e-07)	$\boldsymbol{e}_{1,1}$	-0.952	-0.03	-0.84	-0.38	0.38	-0.1
				$\boldsymbol{e}_{1,2}$	-0.27	0.08	0.34	-0.91	-0.1	0.21
				$\boldsymbol{e}_{1,3}$	0.139	0.48	-0.39	0.08	-0.61	0.48
				$\boldsymbol{e}_{1,4}$	-0.03	0.35	0.15	0.16	0.68	0.61
				$\boldsymbol{e}_{1,5}$	0.006	0.8	0.1	-0.03	0.1	-0.58
	E2	0.2	9.5e-08 (1.26e-08, 1.74e-07)	$\boldsymbol{e}_{2,1}$	-0.904	0.04	-0.02	-1	-0.05	-0.01
				$\boldsymbol{e}_{2,2}$	0.363	-0.11	-0.95	0	0.31	-0.02
				$\boldsymbol{e}_{2,3}$	-0.222	0.16	-0.3	0.05	-0.83	0.44
				$\boldsymbol{e}_{2,4}$	0.034	-0.29	0.13	-0.04	0.36	0.88
				$\boldsymbol{e}_{2,5}$	-0.009	0.94	-0.02	0.02	0.29	0.19
	E3	0.14	6.53e-08 (8.36e-09, 1.24e-07)	$\boldsymbol{e}_{3,1}$	-0.885	0.06	-0.72	-0.44	-0.49	0.2
				$\boldsymbol{e}_{3,2}$	0.405	-0.06	0.49	-0.79	0.12	0.34
				$\boldsymbol{e}_{3,3}$	0.222	-0.02	-0.44	-0.31	0.76	-0.37
				$\boldsymbol{e}_{3,4}$	-0.049	-0.03	-0.21	0.29	0.4	0.84
				$\boldsymbol{e}_{3,5}$	-0.004	1	0.06	-0.02	0.07	0.03
Hot	E1	0.57	4.37e-07 (1.56e-07, 6.85e-07)	$\boldsymbol{e}_{1,1}$	-0.962	0	-0.94	0.13	-0.32	0.04
				$\boldsymbol{e}_{1,2}$	0.258	-0.12	-0.12	-0.98	-0.03	0.1
				$\boldsymbol{e}_{1,3}$	-0.092	-0.07	0.32	0.02	-0.91	0.24
				$\boldsymbol{e}_{1,4}$	-0.015	0.83	-0.01	-0.05	0.08	0.56
				$\boldsymbol{e}_{1,5}$	0.014	0.55	0.03	-0.15	-0.25	-0.79
	E2	0.12	9.58e-08 (1.4e-08, 1.72e-07)	$\boldsymbol{e}_{2,1}$	0.858	0	-0.68	-0.64	-0.25	0.24
				$\boldsymbol{e}_{2,2}$	-0.51	0.03	0.7	-0.7	0.04	0.16
				$\boldsymbol{e}_{2,3}$	-0.051	0.82	0.11	0.13	-0.54	0.08
				$\boldsymbol{e}_{2,4}$	0.025	0.56	-0.18	-0.1	0.8	0.05
				$\boldsymbol{e}_{2,5}$	-0.014	-0.11	0.05	0.27	0.06	0.95
	E3	0.1	7.69e-08 (1.41e-08, 1.38e-07)	$\boldsymbol{e}_{3,1}$	0.924	-0.06	-0.17	0.98	-0.06	-0.04
				$\boldsymbol{e}_{3,2}$	0.353	0.01	-0.46	-0.12	-0.84	0.26
				$\boldsymbol{e}_{3,3}$	-0.139	0.09	-0.78	-0.09	0.54	0.3
				$\boldsymbol{e}_{3,4}$	-0.043	0.99	0.1	0.07	-0.04	0.06
				$\boldsymbol{e}_{3,5}$	0.005	-0.1	0.38	0.1	0.06	0.91

Table S5 Univariate trait means measured for each population in both assays. Hom = Homogeneous, Mig = Migration. Numbers in parentheses represent one standard error.

Assay	Treatment	Centroid	line 9-10	angle 7-8-9	angle 3-10-4	angle 2-4-8
Cold	Hom. cold	320.95 (0.41)	0.0286 (0.00009)	89.17 (0.28)	15.53 (0.04)	84.78 (0.08)
	Hom. Hot	320.31 (0.38)	0.0293 (0.00008)	91.59 (0.28)	15.81 (0.04)	84.19 (0.08)
	Mig. cold	319.41 (0.36)	0.0291 (0.00008)	89.91 (0.25)	15.65 (0.03)	84.44 (0.08)
	Mig. hot	319.31 (0.42)	0.0288 (0.00008)	91.37 (0.29)	15.79 (0.04)	84.51 (0.09)
	Spatial	321.48 (0.41)	0.0289 (0.00008)	89.76 (0.27)	15.75 (0.04)	84.8 (0.08)
	Temporal	322.37 (0.35)	0.0288 (0.00007)	91.97 (0.26)	15.73 (0.03)	84.57 (0.08)
Hot	Hom. cold	280.26 (0.45)	0.0288 (0.00009)	93.76 (0.25)	16.15 (0.04)	84.03 (0.08)
	Hom. Hot	279.5 (0.41)	0.0295 (0.00009)	96.42 (0.27)	16.48 (0.04)	83.49 (0.07)
	Mig. cold	278.37 (0.41)	0.0293 (0.0001)	94.61 (0.24)	16.27 (0.04)	83.52 (0.08)
	Mig. hot	277.7 (0.35)	0.0289 (0.0001)	95.86 (0.26)	16.57 (0.04)	83.56 (0.07)
	Spatial	278.69 (0.4)	0.0292 (0.00009)	95.09 (0.24)	16.39 (0.04)	83.83 (0.07)
	Temporal	280.03 (0.39)	0.0289 (0.0001)	95.57 (0.26)	16.34 (0.04)	83.51 (0.08)

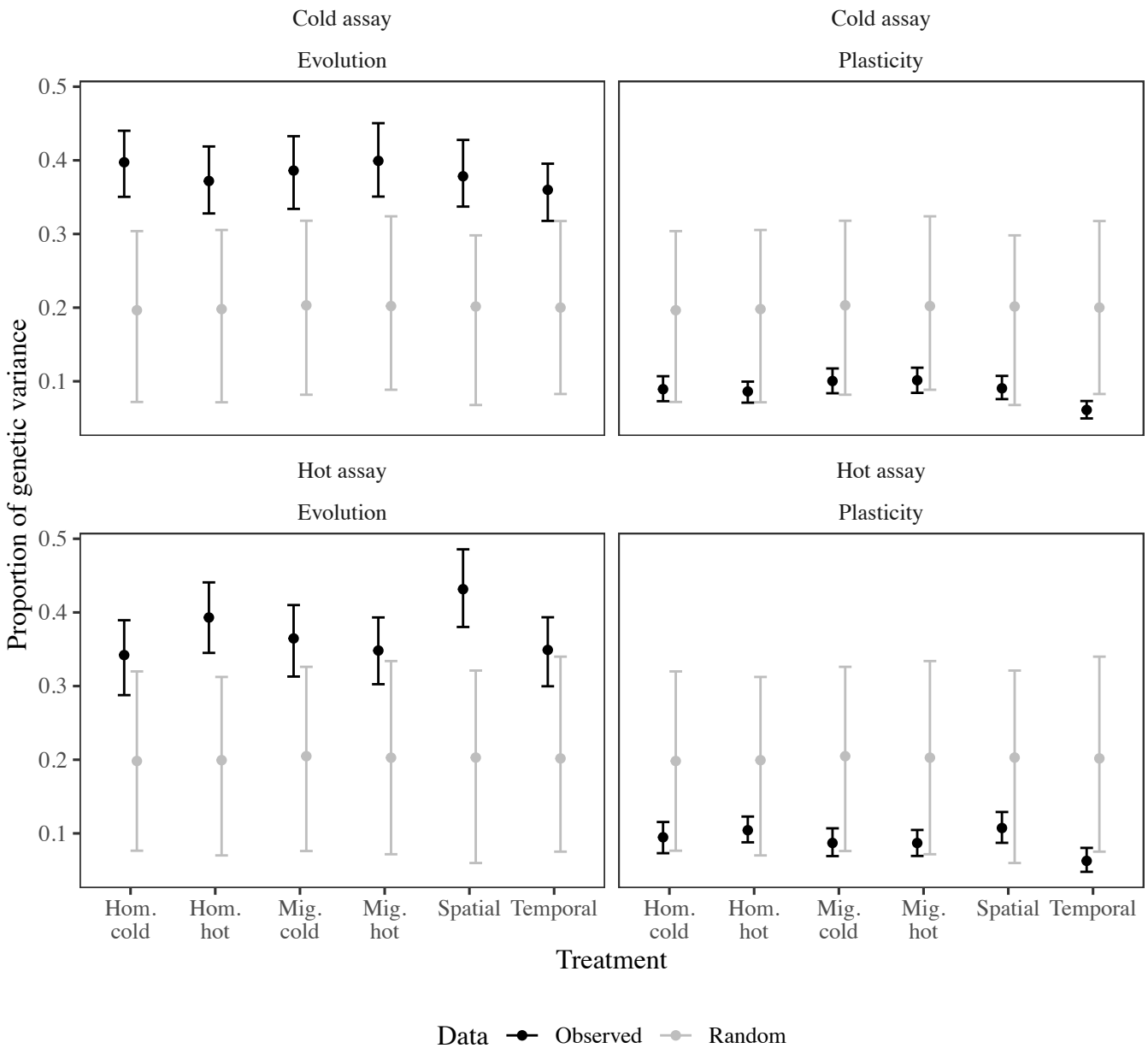

Fig. S2 The direction of evolved changes in phenotype described more genetic variance than expected under the null distribution. Circles and credible intervals represent 90% HPD intervals. The observed distribution (black) represents the proportion of genetic variance, for each MCMC sample of **G**, that is described by evolution and plasticity. The random distribution (gray) represents the proportion of genetic variation described by random vectors for each of MCMC iteration of observed **G**. Evolved differences described a significantly greater proportion of genetic variance than the null. For the hot assay, differences between the observed and random distribution are significant at 90% HPD. By contrast, the direction of plastic changes in phenotype described very small proportions of genetic variance, which for temporal heterogeneity, was lower than expected under random sampling.

Table S6 Summary of Krzanowski's common subspace analysis for the **(a)** cold and **(b)** hot assays. The top half of the table (this page) contains the analysis for mean **G** (the average calculated for each treatment), while the lower half (on the following page) contains the analysis for all 30 G-matrices (i.e. for all replicate cages). Eigenvectors of **H** ($\boldsymbol{h}_1$- $\boldsymbol{h}_5$) represent axes of shared genetic subspace between the matrices, with their associated eigenvalues representing how much the original matrices share the common subspace. Eigenvalues of H range from 0 to *m* (where *m* is the number of matrices), which means that the closer an eigenvalue of **H** is to *m*, the more the subspace is shared among the original matrices. For mean **G**, $\boldsymbol{h}_1$ represents the greatest amount of shared subspace, which is shared among all matrices because values of 5.9 (out of a possible 6 treatments), suggests that all treatments share the subspace identified by $\boldsymbol{h}_1$. Similarly, for **G** estimated on all 30 replicate cages, $\boldsymbol{h}_1$ has values of 26.3 (cold) and 24.8 (hot), suggesting that the subspace is common to almost of the matrices. Loadings represent how the linear combination of the original traits describe the subspace. θ is the angle between the subspace of each of the original treatments (listed for each row), and each of the eigenvectors describing the common subspace. Small angles show that $\boldsymbol{h}_1$ aligns closely with each of the treatments and almost all the replicate cages.

	(a) Cold assay					**(b)** Hot assay				
	Mean **G**					Mean **G**				
	$\boldsymbol{h}_1$	$\boldsymbol{h}_2$	$\boldsymbol{h}_3$	$\boldsymbol{h}_4$	$\boldsymbol{h}_5$	$\boldsymbol{h}_1$	$\boldsymbol{h}_2$	$\boldsymbol{h}_3$	$\boldsymbol{h}_4$	$\boldsymbol{h}_5$
Eigenvalue of **H**	5.9	5.31	0.76	0.02	0.01	5.85	5.66	0.46	0.02	0.01
HPD	5.78, 5.98	4.61, 5.85	0.18, 1.47	0.01, 0.03	0.00, 0.02	5.69, 5.97	5.16, 5.96	0.08, 1.09	0.01, 0.03	0.00, 0.03
Trait loadings										
Centroid	-0.02	0.00	0.11	-0.13	0.99	0.01	0.00	0.1	-0.57	0.81
Line 9-10	-0.29	-0.93	0.22	0.05	-0.02	-0.67	0.69	0.27	0.05	0
Angle 7-8-9	-0.96	0.28	-0.06	0.04	0.00	-0.7	-0.71	0.04	0.06	0.04
Angle 3-10-4	0.00	-0.21	-0.95	0.17	0.13	-0.21	0.17	-0.95	0.04	0.15
Angle 2-4-8	0.06	0.08	0.17	0.97	0.11	0.1	0.00	0.1	0.81	0.56
θ										
Hom. cold	5.41	14.05	74.65	87.75	88	7.96	8.41	79.29	85.84	87.26
Hom. hot	6.56	15.62	72.55	87.44	87.95	7.46	10.25	77.01	88.21	87.71
Mig. cold	6.78	18.18	70.55	86.23	87.77	6.15	10.84	77.8	87.32	87.07
Mig. hot	5.70	18.08	70.78	87.43	87.76	7.48	6.77	79.69	88.36	88.07
Spatial	5.82	10.90	77.3	88.26	87.73	8.83	17.39	70.41	86.66	87.51
Temporal	6.65	23.55	65.37	85.88	88.27	7.27	13.15	74.74	88.00.00	87.96

Table S6 (cont'd)

	(a) Cold assay **G** for the Replicate cages					**(b)** Hot assay **G** for the Replicate cages				
	$\boldsymbol{h}_1$	$\boldsymbol{h}_2$	$\boldsymbol{h}_3$	$\boldsymbol{h}_4$	$\boldsymbol{h}_5$	$\boldsymbol{h}_1$	$\boldsymbol{h}_2$	$\boldsymbol{h}_3$	$\boldsymbol{h}_4$	$\boldsymbol{h}_5$
Eigenvalue of **H**	26.27	20.48	12.44	0.46	0.36	24.76	23.02	11.27	0.55	0.40
HPD	24.71, 27.67	18.21, 23.07	9.48, 14.99	0.29, 0.65	0.22, 0.50	23.31, 26.39	20.86, 25.11	8.35, 13.82	0.30, 0.87	0.23, 0.63
Trait loadings										
Centroid	-0.02	0.00	-0.1	0.99	-0.11	0.01	0.00	-0.09	0.94	0.34
Line 9-10	-0.45	-0.87	-0.19	-0.04	-0.05	-0.68	-0.66	-0.32	0.00	-0.06
Angle 7-8-9	-0.89	0.45	0.05	-0.01	-0.04	-0.69	0.72	-0.01	0.03	-0.07
Angle 3-10-4	-0.03	-0.18	0.96	0.08	-0.17	-0.23	-0.21	0.94	0.11	-0.05
Angle 2-4-8	0.07	0.06	-0.14	-0.13	-0.98	0.11	0.00	-0.06	0.33	-0.94
θ										
Hom. cold 1	21.28	37.65	42.98	82.53	84.91	27.05	35.72	42.17	80.2	83.14
Hom. cold 2	18.79	24.52	57.43	82.61	84.95	10.96	16.51	73.02	82.83	82.97
Hom. cold 3	14.15	63.52	23.20	83.90	82.42	27.30	39.95	35.12	84.05	83.06
Hom. cold 4	9.13	27.48	60.66	85.36	86.78	25.74	14.83	61.38	82.43	82.71
Hom. cold 5	11.74	16.82	69.21	85.14	86.75	12.10	19.01	67.42	85.26	85.34
Hom. hot 1	17.54	35.25	48.06	83.35	84.42	14.15	42.79	43.21	83.88	84.67
Hom. hot 2	15.40	28.27	55.57	85.18	84.95	31.56	17.64	53.18	82.10	83.80
Hom. hot 3	12.01	32.62	55.67	83.01	84.52	18.63	30.97	49.94	87.05	85.25
Hom. hot 4	16.97	31.83	51.09	86.10	83.56	17.70	31.9	52.94	81.08	84.00
Hom. hot 5	16.73	28.52	56.35	81.58	85.67	17.83	12.89	68.31	84.04	86.04
Mig. cold 1	29.13	40.58	33.66	85.28	84.79	36.85	37.63	33.55	77.90	82.02
Mig. cold 2	7.89	39.76	49.56	85.51	85.58	11.24	16.87	71.41	82.84	85.47
Mig. cold 3	14.03	21.67	64.03	83.81	84.87	31.67	48.48	21.41	84.07	84.61
Mig. cold 4	35.64	30.37	37.86	80.04	83.28	7.62	13.00	78.86	85.59	81.53
Mig. cold 5	11.00	30.29	57.73	84.28	84.80	22.54	15.35	63.48	80.72	86.25
Mig. hot 1	16.37	30.38	56.86	81.82	82.00	25.97	8.98	62.53	85.77	85.36
Mig. hot 2	22.59	49.63	29.31	83.39	84.13	9.29	11.22	76.12	85.23	87.28
Mig. hot 3	17.68	22.97	59.22	83.00	86.07	43.23	35.34	23.61	84.55	85.78
Mig. hot 4	12.13	22.14	65.57	83.05	84.96	16.12	29.94	54.69	82.25	85.89
Mig. hot 5	24.95	33.13	44.48	82.37	83.35	45.23	21.23	34.91	83.26	85.52
Spatial 1	18.33	16.78	65.32	84.67	83.14	29.23	23.65	48.71	85.65	83.59
Spatial 2	19.93	61.28	22.07	83.18	79.71	9.13	20.14	68.88	83.61	86.15
Spatial 3	33.39	21.26	45.89	86.28	85.69	27.18	49.83	28.24	82.31	80.56
Spatial 4	8.33	10.31	78.26	84.46	86.74	19.39	30.20	54.61	83.15	81.02
Spatial 5	19.61	37.39	45.32	80.93	85.58	10.14	24.35	64.74	84.04	84.01
Temporal 1	8.04	41.28	48.44	85.05	84.48	27.45	16.65	57.03	85.54	86.92
Temporal 2	26.81	17.82	55.33	86.09	83.52	24.67	50.65	28.58	80.61	84.40
Temporal 3	13.25	39.92	46.19	83.96	85.71	15.47	20.75	63.00	85.79	84.48
Temporal 4	21.25	10.25	65.95	87.17	86.02	21.50	22.15	57.82	83.75	84.50
Temporal 5	17.02	52.02	32.23	83.61	84.83	10.97	16.19	70.83	86.19	85.87

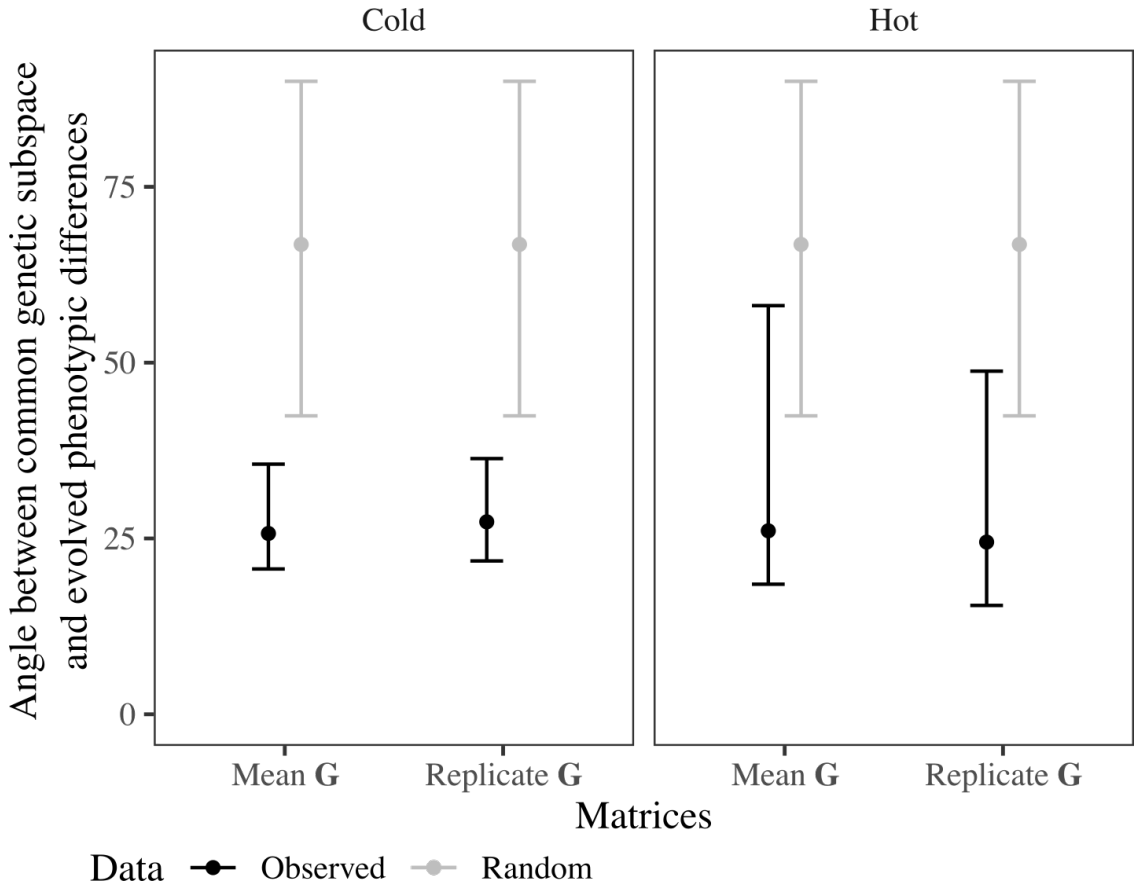

Fig. S3 For **G** estimated as the mean of each treatment, and for all 30 matrices in an assay, the observed angle (black) between shared genetic subspace and phenotypic divergence are smaller than expected by the null distribution (gray). Credible intervals represent 90% HPD intervals. In the hot assay, the observed distribution does not overlap the null distribution at 87% and 90%, respectively for mean **G** and replicate **G**.

Methods S1 Comparing **G** estimated as the mean for each treatment versus for each replicate cage

To ensure that the estimates of **G** accurately represent genetic variance for each treatment, I calculated mean **G** for each treatment using two methods and then compared them to estimates of **G** for all the replicate cages within each treatment. First, I calculated **G** for all replicate cages (n=60 matrices) using equation 1 (as described in the main text) and then I estimated the average of all the replicate cages within each treatment. This produced 12 G-matrices ($\mathbf{G_1}$). With the second approach, I estimated mean **G** (n=12 matrices) by including an additional random effect into equation 1, $\boldsymbol{b}_k$, which represents the kth replicate cage within each treatment ($\mathbf{G_2}$). To compare the two methods for estimating an average **G**, I first projected the observed eigenvectors of $\mathbf{G_1}$ and $\mathbf{G_2}$ through their posterior distributions using the matrix projection (from equation 3): $V_{ijk} = \frac{\boldsymbol{e}_{jk}^{T}\mathbf{G}_{ij}\,\boldsymbol{e}_{jk}}{Tr(\mathbf{G}_{ij})}$, where the kth eigenvector of **G** estimated in the jth treatment is projected through the ith MCMC iteration. Tr represents the trace of **G**, which means that the projection yields the proportion of total genetic variance. I then used the same method to project the same eigenvectors (from both estimates of mean **G**) through the G-matrices of all replicate cages. If mean **G** accurately reflects the G-matrices of the replicate cages, then each eigenvector of **G** should describe a similar amount of genetic variance for the posterior distribution of mean **G** and the posterior distribution of **G** estimated for each of the replicate cages. **Fig. S4** shows that the eigenvectors for both methods of estimating average **G** describe similar proportions of genetic variance for average **G**, and **G** estimated for all replicate cages.

To ensure that differences captured in mean **G** are the same for both methods of estimating average **G**, I then used the covariance tensor analysis on $\mathbf{G_1}$, $\mathbf{G}_2$ and then incorporated the five replicate cages for each treatment. **Fig. S5** shows that both methods for estimating an average **G** produce the same results, and that by incorporating **G** estimated for all the replicate cages, we can see that differences in **G** are driven by the spatially heterogeneous treatment in the hot assay (as reported in the main text), which is reflected by a consistent pattern shown by all replicate cages for this treatment (**Fig. S5**).

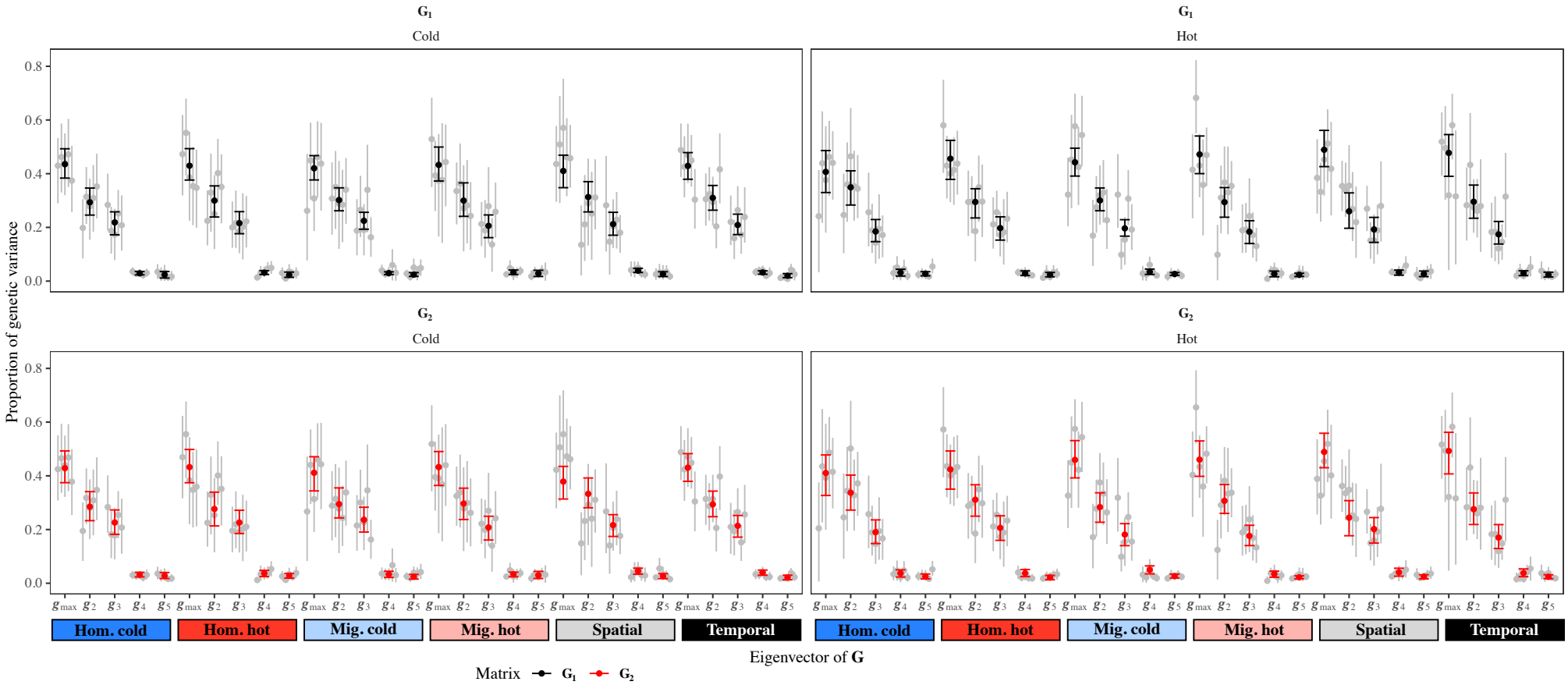

Fig. S4 Eigenvectors of average **G** ($\mathbf{G}_1$ and $\mathbf{G}_2$) describe a similar proportion of variance in the G-matrices estimated in the replicate cages.

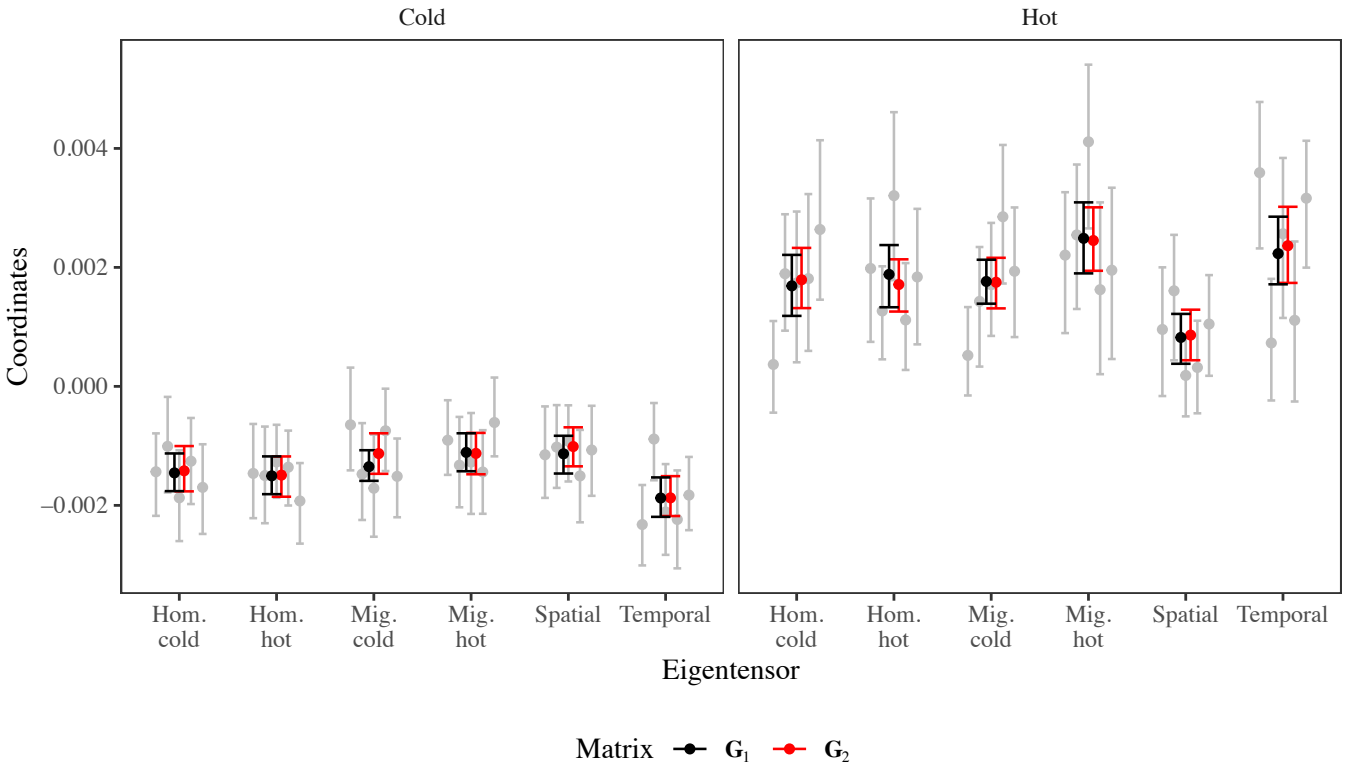

Fig. S5 Coordinates from the covariance tensor analysis that includes both estimates of average **G** (**G**$_1$ and **G**$_2$), and the replicate cages (in gray). Both methods for estimating an average **G** produced the same results, and replicate cages were consistent within each treatment.

Methods S2 Estimation of a suitable null distribution to test for significant differences in **G**.

To calculate a suitable null distribution for among-treatment differences in **G**, I used the approach outlined in the accompanying supplemental simulations and Morrissey et al. (2019). Each assay was analyzed separately. For each MCMC iteration of the observed array, I calculated the average **G** among treatments (representing no differences in **G**) and estimated breeding values for the entire pedigree. I then constructed the phenotypes around the breeding values by adding means and variance for each component of equation 1 (see supplementary code). I re-applied equation 1 on the newly constructed phenotypic data for each randomization. I used this approach because it estimates no differences in genetic variance among treatments, while keeping all other components the same as the observed models. This provides a suitable null distribution of no genetic differences among G-matrices.

Given the extensive time taken for each model, I constructed phenotype datasets for 500 MCMC iterations of the observed model. I also reduced the total number of iterations but kept the thinning interval and burn-in the same as the observed model. This reduces the computational requirements while ensuring convergence is the same as the observed model. I saved 100 iterations from each model and calculated the mean G-matrix for each model – this means that differences in the posterior mean of the observed model needs to be compared to the null distribution created by taking the posterior mean of the 500 models. I then constructed the null distribution for **G** as the difference among the treatments for the 500 implementations (quantified using the covariance tensor approach). If the posterior mean of the observed model described greater differences among **G** than the null distribution, then there is evidence for significant differences among **G** following experimental evolution. Overall, I applied equation 1 to 500 models for each of the observed G-matrices.

Methods S3 Estimating metrics of multivariate evolvability

To test whether **G** changed in response to the temperature and heterogeneity treatments, I first estimated three summary metrics of evolvability using the posterior distribution of **G** (Kirkpatrick 2009): 1) The effective number of dimensions (n_D), defined as the sum of the eigenvalues divided by the first eigenvalue; 2) Maximum evolvability ($\boldsymbol{e}_{max}$), represented by the square root of the leading eigenvalue, and; 3) Total genetic variance, estimated as the sum of the variances (the trace) of **G**. Differences among G-matrices in any of these metrics would suggest changes in the amount or distribution of genetic variance.

Supplementary references

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. Genetica 136:271-284.

Morrissey, M. B., S. Hangartner, and K. Monro. 2019. A note on simulating null distributions for **G** matrix comparisons. Evolution 73:2512-2517.