PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Careau V, Wolak ME, Carter PA, Garland Jr T. 2015 Evolution of the additive genetic variance–covariance matrix under continuous directional selection on a complex behavioural phenotype. *Proc. R. Soc. B* **282**: 20151119.
http://dx.doi.org/10.1098/rspb.2015.1119

Received: 13 May 2015
Accepted: 20 October 2015

Subject Areas:
evolution, behaviour

Keywords:
Bulmer effect, experimental evolution, genetic covariance tensor, **G**-matrix, selection limit, wheel running

Author for correspondence:
Vincent Careau
e-mail: vcareau@uottawa.ca

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2015.1119 or via http://rspb.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Evolution of the additive genetic variance–covariance matrix under continuous directional selection on a complex behavioural phenotype

Vincent Careau[1], Matthew E. Wolak[2], Patrick A. Carter[3] and Theodore Garland Jr[4]

[1]Canada Research Chair in Functional Ecology, Department of Biology, University of Ottawa, Ottawa, Ontario, Canada
[2]School of Biological Sciences, University of Aberdeen, Aberdeen, UK
[3]School of Biological Sciences, Washington State University, Pullman, WA, USA
[4]Department of Biology, University of California, Riverside, CA, USA

Given the pace at which human-induced environmental changes occur, a pressing challenge is to determine the speed with which selection can drive evolutionary change. A key determinant of adaptive response to multivariate phenotypic selection is the additive genetic variance–covariance matrix (**G**). Yet knowledge of **G** in a population experiencing new or altered selection is not sufficient to predict selection response because **G** itself evolves in ways that are poorly understood. We experimentally evaluated changes in **G** when closely related behavioural traits experience continuous directional selection. We applied the genetic covariance tensor approach to a large dataset ($n = 17\,328$ individuals) from a replicated, 31-generation artificial selection experiment that bred mice for voluntary wheel running on days 5 and 6 of a 6-day test. Selection on this subset of **G** induced proportional changes across the matrix for all 6 days of running behaviour within the first four generations. The changes in **G** induced by selection resulted in a fourfold slower-than-predicted rate of response to selection. Thus, selection exacerbated constraints within **G** and limited future adaptive response, a phenomenon that could have profound consequences for populations facing rapid environmental change.

1. Introduction

Given the pace at which human-induced environmental changes occur, a pressing challenge is to determine the speed with which selection can drive evolutionary change and adaptation (as opposed to extinction [1–3]). Prediction of evolutionary change for multiple traits is obtained from the product of the linear selection gradient vector (β), which indicates how selection is acting on individual traits after adjusting for trait correlations, and the additive genetic variance–covariance matrix of those traits (**G**) [4]. Most typically, **G** is viewed as affecting the efficacy of selection [5]; crucially, however, **G** can evolve in response to selection because of changes in allele frequencies and/or the build up of linkage disequilibrium (LD) among important loci [6,7]. Thus, evolutionary processes can shape genetic variation in ways that may either facilitate or constrain future evolutionary changes [5,8].

Simulation-based studies point to the hypothetical conditions under which **G** will evolve to the greatest extent (e.g. small population, weak correlational selection, low mutational correlation among traits [9,10]). However, general answers from simulations seem unlikely to emerge because it is extremely complex to consider all of the relevant parameters involved across the whole range of relevant parameter values (e.g. population size, number of loci, number of alleles, distribution of allelic effects, dominance, epistasis, mutation rate, strength of selection

[11]). Therefore, if and how selection influences **G** remains largely an empirical question [11–13]. Most empirical studies to date have been modest in scale, biased towards certain taxa (mainly a few insect and plant species) or types of characters (mostly morphological and life-history traits), and varied in both statistical and experimental approaches [11]. As a result, how long the **G** matrix remains stable under multiple generations of selection (and/or random genetic drift) remains a contentious issue [2].

Replicated artificial selection experiments evaluating the stability of **G** under selection are rare [14,15], but bring crucial strengths to this key endeavour (e.g. replication, potentially large sample size, known selection, constant environment). Interestingly, selection experiments under laboratory conditions commonly reach an evolutionary limit (i.e. a plateau in the response to selection) even when significant additive genetic variance (V_A) remains for the trait under selection [16]. This result is paradoxical from a univariate perspective, in which the response to selection is determined strictly by the product of the directional selection differential and the narrow-sense heritability (h^2) of a trait. The paradox can sometimes be resolved by adopting a multivariate perspective on additive genetic variance and selection [5]. Even in the presence of V_A in all traits, an absolute genetic constraint can occur if there are one or more directions within **G** for which no V_A exists (i.e. **G** is semi-positive definite and singular because one or more of its eigenvectors have zero eigenvalues). Thus, an evolutionary limit to selection can result from a multivariate constraint within **G**, manifested as a lack of V_A in the multivariate direction favoured by selection [15]. (Alternatively, genetic covariances between traits can also accelerate evolution if they allow V_A in the direction favoured by selection [5].)

Testing the possibility that the evolution of **G** may result in genetic limits to adaptive evolution is crucial for accurately predicting the phenotypic outcomes of selection because response to selection in wild populations may be more limited than is currently believed [17]. Moreover, in a rapidly changing environment, multivariate constraints could increase risk of local extinctions because populations may not be able to adapt quickly enough to keep pace with environmental change [2]. Here, we experimentally test if and how continuous directional selection induces changes in the form of **G** in a way that limits future adaptation. We use data from a long-term, fourfold replicated, artificial selection experiment on voluntary wheel-running behaviour in house mice [18]. Considering wheel running expressed over 6 successive days as a series of closely related traits, our objective is to evaluate if and how directional selection on a subset of the daily behaviours induced changes in **G** overall.

2. Material and methods

(a) Animals, housing and selection protocol

The experiment started from outbred Hsd : ICR mice, the origins of which are described fully elsewhere [18]. A total of 112 males and 112 females, each from a different family, were purchased from Harlan Sprague Dawley (designated generation −2) and paired randomly (details in electronic supplementary material, appendix S1). The resulting offspring, designated generation −1, were assigned to eight closed lines. To establish lines, one male and one female were chosen randomly from each litter. These individuals were then paired randomly except that full-sibling matings were disallowed. Lines were randomly assigned into four non-selected control (C) lines and four selected 'high-runner' (HR) lines. Their offspring were designated generation 0 and selection started at this generation in HR lines. Each line was maintained with at least 10 families per generation, routinely housed in same-sex groups of four per cage except during breeding (one pair per cage) and wheel-running measurements (one individual per cage). Mice were maintained on a 12 L : 12 D cycle (lights on 07.00 h), which was also maintained during the wheel-running trials.

Each generation, approximately 600 mice six to eight weeks of age were monitored for wheel revolutions for 6 consecutive days. Voluntary wheel running was measured on stainless steel and Plexiglas, Wahman-type activity wheels (circumference = 112 cm, diameter = 35.7 cm and width = 10 cm; Lafayette Instruments, Lafayette, IN, USA) attached to standard home cages. Three batches of approximately 200 mice each were measured during three successive weeks. Mice from a given batch were weighed and placed on randomly assigned wheels during the morning of the first day; data collection was started at approximately 13.00 h. Data were downloaded every 23.5 h, at which time wheels were checked to remove any food pellets or wood shavings and to ensure freedom of rotation. On the sixth day, mice were removed from the wheels and weighed.

Starting at generation 0, selection was based on the residual average number of wheel revolutions run on days 5 and 6 (transformed as necessary to improve normality of residuals) from a multiple regression model used to control for several biological and nuisance variables, including measurement block (batches 1–3 and rooms 1–2), sex (in interaction with selection history) and family [18]. Within-family selection was used to increase the effective population size, reduce the rate of inbreeding and help to eliminate the possibly confounding influences of some maternal effects. To produce offspring in the next generation for each of the four HR lines, the highest running male and the highest running female from each family were chosen to breed and paired randomly, with the provision of no full-sibling mating. Over 31 generations, the selection differential averaged 0.92 (s.d. = 0.23) phenotypic standard deviations in the HR lines and approximated zero in the C lines [16]. In C lines, a male and female from within each family were chosen randomly to obtain breeders, again with no full-sibling matings allowed.

(b) Data and pedigree

We used the same data and pedigree for wheel running on days 5 and 6 as compiled and checked in a previous study [16]. Here, we added data on days 1–4 and excluded an additional 572 individuals for which wheel running on a given day was obviously abnormal compared with wheel running on the other days (this often resulted from wheel problems that were detected and corrected during the first 4 days). This resulted in a sample of 17 328 individuals with wheel running measured on the 6 days from generation 0–31.

(c) Animal models for quantitative genetic analyses

Falconer [19] and many authors since have recognized that, conceptually, a given element of the phenotype (e.g. body mass) measured in several environments or serially at several ages or life stages can be considered as a set of traits that are expected to be genetically correlated. We adopted such a 'character-state' approach and considered wheel running expressed over the 6 successive days as a series of closely related traits. Our choice is supported by previous research showing that the genetic architecture for running distance varies somewhat across successive days [20]. For example, some detected quantitative trait loci have an effect only during the initial 3 days of exposure to wheels, which are believed to be related to anxiety or fear-related

behaviours upon initial exposure to a new environment (e.g. individual housing, introduction to a running wheel [20]). Another option would have been to adopt a 'function-valued' approach and use random regression to model wheel running as a function of day for each individual. However, in this case, many of the logistical and statistical advantages of using the function-valued approach are not relevant, and the low number of index values at which measures were taken (six) makes it likely that the function-valued and character-state approaches will result in models that are equivalent [21]. A major disadvantage of using the character-state approach here is the increased number of model parameters (i.e. covariances among all traits); however, this is not a practical problem because of the relatively low number of separate traits (i.e. 6) and relatively large sample size.

We quantified **G** for wheel running on each of the 6 days (referred to below as 'traits' for simplicity) using Bayesian multivariate animal models in the MCMCglmm package of R [22]. We chose to pool the data for replicate lines within C and HR, and divide the 32 generations into roughly three equal time periods: generations 0–10, 11–20 and 21–31. This offered the best trade-off between having large sample size within each time period (electronic supplementary material, table S1) to allow estimation of **G** with adequate precision while at the same time examining temporal change in **G** over periods before (generations 0–10), during (generations 11–20) and mostly after (generations 21–31) the selection plateaus had been reached [16]. For each time period, we ran two separate models (one for C and one for HR mice) restricted to the pedigree and phenotypic data from that time period, yielding a total of six **G** matrices. We followed a heuristic approach for estimating **G** for the specific time periods in the selection experiment [23] where the pedigree used for each time period treated animals in the first generation of the period (i.e. 0, 11 or 21) as the descendants of a hypothetical unrelated, non-selected and non-inbred population, and included pedigree information for all individuals in that time period [23]. Thus, the animal model inferred **G** back to the hypothetical parental population of the time period given the data provided. Obviously, in finite populations (with drift) and especially in populations under selection (the HR lines) in which allele frequencies change [24], some assumptions of the infinitesimal model do not hold. In this case, the animal model may not yield a strict estimate of **G** in the base parental population, but one that is 'biased' towards subsequent generations for which data are included. Further research is required to better understand exactly how animal models perform in such situations [25]. To ensure the differences found were attributable to selection *per se* and not genetic drift, we also ran MCMC models on a line-by-line basis, which yielded a total of 24 **G** matrices for comparison (i.e. one for each of the eight lines per generation block).

Each model included the six traits as response variables and a series of fixed effects (age, sex, measurement batch and room, and inbreeding coefficient) fitted separately for each trait. All traits were standardized (mean = 0, variance = 1) for each line in each generation. Choice of scale can have a substantial influence on both univariate and multivariate analyses [26]; we thus wanted to account for the large fluctuations in phenotypic means across generations and the increase in the total phenotypic variance across generations in the HR lines [16] (see also electronic supplementary material, figure S1). Moreover, by standardizing trait variances, we made sure that no individual trait dominates the eigenstructure of individual **G** matrices [27].

Each model included unstructured (co)variance matrices for random effects of maternal identity (common environmental (co)variances), animal identity linked to the pedigree (additive genetic (co)variance; **G**) and the residuals. Analyses were run using weakly informative inverse-Wishart priors with the scale parameter defined as a diagonal matrix containing values of one-third of the phenotypic variance and distribution parameters set to 0.001 for the degrees of freedom [22]. Posterior distributions were estimated from 6 500 000 MCMC iterations sampled at 5000 iteration intervals following an initial burn-in period of 1 500 000 iterations. Autocorrelations were less than 0.1 for all (co)variance components (electronic supplementary material, figure S2), which yielded effective samples sizes close to 1000 for all (co)variances and fixed effects. We visually inspected plots of the traces and posterior distributions, and calculated the autocorrelation between samples to make sure that all models properly converged.

(d) Covariance tensor within a Bayesian framework

A genetic covariance tensor analysis [27] captures all of the variation among multiple **G** matrices and subsequently allows identification of the trait combinations that differ most in V_A among the **G** matrices. The covariance tensor analysis yields matrices of loadings that each represent independent aspects of how the original **G** matrices differ and can be interpreted in a fashion similar to the original **G** matrices. The covariance tensor approach has recently been integrated within a Bayesian framework [12], meaning that uncertainty in estimates of each **G** is subsequently incorporated into the comparisons among **G** matrices. Consequently, this also provides estimates of confidence to be placed on the estimation of differences among **G** matrices. Moreover, one can conduct a covariance tensor analysis on **G** matrices created under a null model (e.g. by randomization), thereby allowing tests of whether the identified differences among observed **G** are more pronounced than predicted by a particular null hypothesis. Indeed, it is possible to use the posterior predictive distribution of breeding values obtained from a Bayesian animal model to generate a set of **G** matrices where any differences among these are assumed to be driven by random sampling variation. Thus, applying the covariance tensor analysis to both the observed and randomized sets of **G** matrices provides a test of the hypothesis that the observed **G** matrices differ more than expected by random sampling *per se* (i.e. of individuals from a population for analysis). We briefly describe the implementation below, but refer interested readers to an example in the electronic supplementary material, appendix S2, and recommend Hine *et al*. [27] and Aguirre *et al*. [12] for more details about the methods.

Tensors from multilinear algebra extend the notion of vectors (which are first-order tensors) and matrices (second-order tensors) to higher-order structures that can be used to characterize variation in these lower-order variables (i.e. vectors and matrices). The variation among multiple **G** matrices can be characterized by a fourth-order genetic covariance tensor, Σ. The independent aspects of variation among multiple **G** matrices are described by second-order eigentensors (**E**) of Σ, and the amount by which an **E** contributes to the total variation among **G** matrices is quantified by the size of the eigenvalue corresponding to an **E**. The eigentensors of Σ are obtained by first mapping Σ onto the symmetric matrix **S** [27]. The elements in the eigenvectors of the matrix **S** are then scaled and arranged to form the second-order eigentensors (**E**). Therefore, the **E** matrices (second-order eigentensors) contain information regarding independent directions in which the original **G** matrices differ and can be read, similar to how one interprets a **G** matrix. For example, if an **E** contains large positive values along the diagonal and values near zero on the off-diagonals then the conclusion would follow that the original **G** matrices differ in the V_A of individual traits (with the relative difference among **G** for each trait reflected by the relative magnitude of the diagonal elements of **E**; see electronic supplementary material, appendix S2). Further, we can determine the genetically orthogonal linear combinations of traits that describe independent changes among **G** matrices by obtaining the eigenvectors (***e***) and eigenvalues of each **E**. These eigenvectors (***e***) can be interpreted in a way similar to eigenvectors of a **G** matrix. For example, if the largest eigenvalue of an ***e*** is close to 1, then the detected

change in covariance structure, described by the eigentensor, can be attributed to the change in V_A for a single combination of traits.

We next applied the covariance tensor method in a Bayesian framework. For the *i*th MCMC sample of the set of **G**, we determined the matrix representation of the tensor, $\mathbf{S}_i$. We then calculated the elements of $\bar{\mathbf{S}}$ from the corresponding posterior means of the elements of the set of $\mathbf{S}_i$ ($i = 1$–1000). We projected the *j*th eigenvector of $\bar{\mathbf{S}}$ onto $\mathbf{S}_i$ (equivalent to projecting of $\overline{\mathbf{E}_j}$ onto Σ) to determine α_{ij}, the variance among the *i*th MCMC sample of the **G** matrices for the aspect of covariance structure specified by $\overline{\mathbf{E}_j}$. Projecting an eigentensor onto a tensor is analogous to projecting a vector onto a **G** matrix to determine how much variance is present in a particular direction. This allowed us to calculate the V_A along the axis of greatest variation among the six **G** of interest by projecting the leading eigenvector of $\mathbf{E}_1$ (i.e. $\boldsymbol{e}_{11}$) onto the observed **G** matrices. The posterior distribution of α_j summarizes the uncertainty in the variance in covariance structure represented by $\overline{\mathbf{E}_j}$.

To test whether the observed differences among **G** matrices were statistically significant, we used randomizations to generate a null model in which we assumed the differences among **G** were driven by random sampling variation alone (see electronic supplementary material, appendix S2, for an application of the Bayesian genetic covariance tensor on simulated **G** matrices to help interpret the results obtained from this approach). Note that for a single trait, the variance in true breeding values is V_A and similarly the covariance in true breeding values between two traits is the additive genetic covariance (COV_A). Thus, we used the marginal posterior distribution of each estimated **G** to generate breeding values and assigned them randomly to one of the six combinations of selection group (i.e. C or HR) and generation block (i.e. 0–10, 11–20 or 21–31) under the null hypothesis that sampling variation is the only process explaining the divergence among **G** [12]. Thus, we calculated posterior predictive breeding values for each trait in every individual by taking draws from a multivariate normal distribution with a mean of zero and covariance according to Kronecker product between the *i*th MCMC sample of the *j*th **G** and the pedigree-derived numerator relationship matrix. We then calculated randomized **G** matrices as the variances within and covariances between the simulated vectors of breeding values assigned to each hypothetical population. Subsequently, we applied the covariance tensor method to both the observed and randomized sets of **G** to test the hypothesis that the observed **G** matrices differ significantly more than by sampling variation alone. Specifically, we compared the distribution of α_j obtained from the genetic covariance tensor applied to the observed versus randomized sets of **G**. Although the expectation of α_j for the randomized set of **G** is approximately 0, the upper bound of its 95% highest posterior density (HPD) interval provides an estimation of the variation among **G** matrices that can be obtained by chance given the dataset and structure of the pedigrees.

(e) The multivariate breeder's equation

We used the multivariate breeder's equation [4], $\mathbf{R} = \mathbf{G}\beta$, to illustrate the extent to which changes in **G** induced by continuous directional selection influence subsequent response to selection (**R**). To do this, we used the β estimated in HR mice (see below) to predict **R** in both C and HR mice for each generation block, which demonstrates how the changes in **G** induced by selection altered evolutionary responses in units of phenotypic standard deviations. We calculated relative fitness as the number of pups produced by a given HR mouse divided by the average number of pups in HR mice for different blocks of generations (i.e. 0–10, 11–20 and 21–31). This analysis included all HR mice for which we had wheel-running data, including the non-selected individuals (who were attributed a pup production of zero). We then used MCMCglmm and the data for each generation block to run multiple linear regressions of relative fitness as a function of wheel running on days 1–6 (each day standardized to a mean of 0 and a variance of 1) with the same fixed effects as used in the multivariate animal models (see above). The partial regression coefficients from these multiple linear regressions provided the vectors of selection gradients (β), which represents the strength of directional selection on wheel running for each of days 1–6, corrected for correlations with all other traits in the model and adjusting for other fixed effects. Interpretation of partial regression coefficients can in some cases be counterintuitive to univariate examination of mean differences between successful and unsuccessful individuals. In particular, it is not uncommon that a trait that displays a very low selection differential when examined from a univariate perspective appears as very important in the multiple regression by having a large partial regression coefficient, or *vice versa*.

Very high correlations among independent variables are well known to cause instability in the estimates of partial regression coefficients and their standard errors, which is possible here because the raw phenotypic correlations between wheel running on different days ranged from 0.53 to 0.85 in selected mice. The generalized variance inflation factors were below 5 for all wheel-running variables, however, suggesting problems with multicollinearity were not serious [28]. Most importantly, we incorporated the uncertainty in both **G** and β to obtain the posterior distribution of **R** [29]. For a given generation block, we took the *i*th sample of the posterior distribution of **G** from the multivariate animal model and post-multiplied it by the *i*th sample of the posterior distribution of the vector of β from the multiple regression model to obtain the *i*th sample of the posterior distribution for **R**. We could then calculate the posterior mode and 95% HPD intervals for the predicted response to selection for wheel running on days 1–6 in each generation block, separately for C and HR mice.

3. Results

(a) Correlated responses to selection

Over 31 generations, we measured 17 328 young adult mice for wheel revolutions on 6 consecutive days, across which wheel running increases monotonically in every generation (figure 1*a*). Substantial response to selection occurred over the first 20 generations, at which point HR mice were running approximately 2.5-fold more revolutions per day compared with C (figure 1*b*) and after which no further increase in wheel running has occurred, despite continued directional selection and the presence of V_A in the selected trait (see also [16]). Although the first 4 days of wheel running were intended primarily to allow familiarization with the testing apparatus (e.g. reduce or eliminate neophobia) and not part of the selection criterion, substantial correlated responses occurred for each of those days (figure 1*a*). On an absolute basis, HR mice show a greater daily increase than C mice in wheel running across the 6-day trial (figure 1*c*), which may require co-adaptational changes in training ability and other subordinate traits that support or cause wheel running [30]. The differential increase in wheel running is less apparent on a relative basis, especially after generation 24 (figure 1*b*).

(b) Additive genetic (co)variances matrices

Throughout the experiment, the six behavioural traits showed higher V_A in C mice (range 0.167 to 0.307) than HR

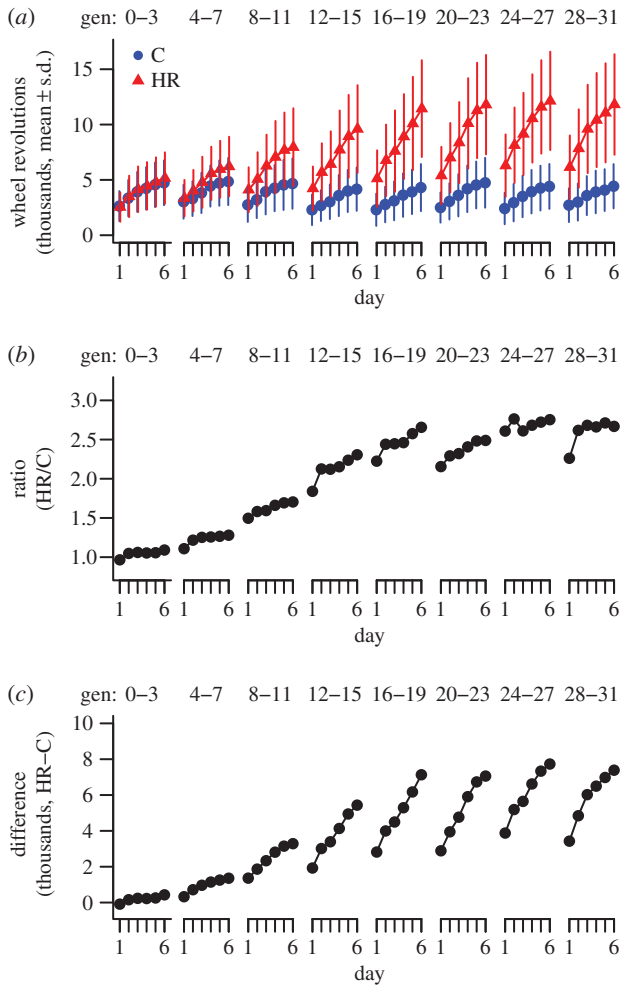

Figure 1. (*a*) Daily average number of wheel revolutions run (pooled means ± s.d. for four replicate lines in each selection group) over 6 days of wheel access in control (C) and selected (HR) mice, averaged over blocks of four generations (see also electronic supplementary material, figure S1). (*b*) Relative and (*c*) absolute difference in wheel running in C versus HR mice. (Online version in colour.)

mice (range 0.073 to 0.118) (electronic supplementary material, figures S3 and S4). All of the COV_A were positive, but they were consistently higher in C (range 0.112 to 0.219) than HR mice (range 0.026 to 0.068), even over the first 10 generations (electronic supplementary material, figures S3 and S4). For sake of completeness, we also calculated scaled individual components of **G** (i.e. h^2 and r_A) which yielded similar differences between C and HR mice (see electronic supplementary material, figure S4). That the estimated COV_A were all positive is reflected in the correlated increases observed for days 1–4 as a result of selection only on days 5–6 (see above).

(c) Genetic covariance tensor

The eigentensor $\mathbf{E}_1$ described 91.9% of the variation among the six observed **G** matrices (figure 2*a* and table 1). $\mathbf{E}_1$ reveals that the variation across the six **G** matrices was substantial for all elements (figure 2*b*). Note that there seems to be more variability among **G** matrices for elements in the lower right corner of the matrix, which corresponds to V_A and COV_A for days 4–6 (figure 2*b*). The leading eigenvector $\boldsymbol{e}_{11}$ explained 99.8% of the variation captured by $\mathbf{E}_1$ (table 1). Thus, almost all of the variation among the six **G** matrices is captured by a single combination of traits. The loadings of $\boldsymbol{e}_{11}$ range over a rather restricted range of −0.37 to −0.44 (table 1), suggesting that the six traits contribute roughly equally to the major axis of variation among **G** matrices. Calculating the V_A along this axis of variation shows that the changes captured in $\boldsymbol{e}_{11}$ were driven by a steady difference in V_A between C and HR mice across generation blocks (figure 2*c*). Note that V_A along $\boldsymbol{e}_{11}$ did not differ statistically within HR mice across the generation blocks (figure 2*c*).

Applying the tensor analysis to the 24 **G** matrices estimated for each line and generation block yielded results that are consistent with the pooled analysis (electronic supplementary material, figure S5 and table S2). $\mathbf{E}_1$ captured a large part (75.6%) of the variance among **G** matrices and the comparison of observed versus null **G** indicated that this eigentensor explained significantly more variance than expected by chance (electronic supplementary material, figure S5*a*,*b* and table S2). The leading eigenvector $\boldsymbol{e}_{11}$ explained 99.6% of the variation captured by $\mathbf{E}_1$ (electronic supplementary material, table S2), with loadings ranging from −0.35 to −0.45 only (electronic supplementary material, table S2). Compared with C lines, all four HR lines consistently showed lower V_A along $\boldsymbol{e}_{11}$ (electronic supplementary material, figure S5*c*). These results show that the effect of selection on **G** was repeatable at the level of replicate lines and that differences among **G** were not affected by genetic drift among the lines.

(d) Predicted response to selection

Multiple regressions of relative fitness as a function of wheel running on days 1 to 6 revealed that the selection gradients were positive for days 5 and 6, but some were negative and significantly different from zero for days 1–4 (electronic supplementary material, table S3). We used these partial regression coefficients estimated in HR mice as our β in the multivariate breeder's equation to predict the between-generation changes for days 1–6 if selection were to be applied to both C and HR lines separately for each generation block (i.e. using the six separate **G**-matrices estimated above). For all six behavioural traits, HR mice show an approximately four-fold lower predicted **R** than do C mice (figure 3; electronic supplementary material, table S4). Because the predicted responses for days 5 and 6 were still positive and statistically significant in generations 21–31, however, we can confidently rule out the possibility that changes in the form of **G** and/or β caused the selection plateaus that occurred in each of the four HR lines during generations 21–31, which is consistent with the conclusions of Careau *et al*. [16].

4. Discussion

Our results address two major questions in evolutionary biology. First, does the multivariate quantitative genetic architecture typically cause some trait combinations to have far more V_A than others? Second, is V_A typically low in the direction of past prevailing selection? These questions are related by the possibility that selection itself can contribute to the generation of semi-positive definite and singular **G** matrices by eroding V_A for the trait combinations that are under strong selection, thus potentially causing selection plateaus [5]. More generally, if and how the elements within **G** (V_A and COV_A) change under selection are long-standing questions in

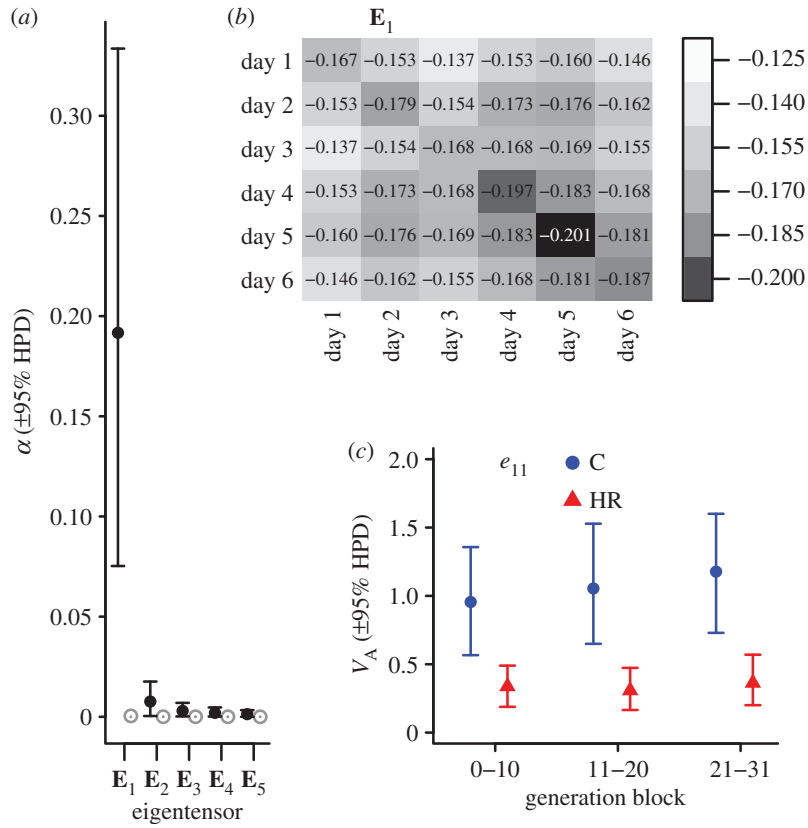

Figure 2. (*a*) Variance (α; $\pm$95% HPD intervals) accounted for by each eigentensor ($\mathbf{E}_n$) for the observed (black dots) and randomized (grey dots) sets of **G**. Because the 95% HPD intervals of the observed versus randomized sets of **G** did not overlap for $\mathbf{E}_1$ and $\mathbf{E}_2$, these eigentensors described significantly more variation among the observed **G** than by chance. (*b*) 'Heat map' displaying the pattern of greatest variation among **G** matrices as captured by $\mathbf{E}_1$ (darker shading indicates greater variation among **G** matrices as measured by elements of $\mathbf{E}_1$, which reflect variance of the (co)variances among the six **G** matrices). Hence, variability among **G** matrices was distributed throughout the entire matrix, but was slightly more intense for trait combinations involving days 4–6 compared with those earlier in the day sequence. (*c*) The additive genetic variance (V_A) in control (C) and selected (HR) mice along the direction of the first eigenvector of $\mathbf{E}_1$ (i.e. $\boldsymbol{e}_{11}$) across generation blocks. (Online version in colour.)

evolutionary biology [7,31]. We used a Bayesian framework [12] to directly incorporate the uncertainty associated with estimating **G** and applied the genetic covariance tensor approach [27], where we identified a single trait combination that captured most of the variation among **G** matrices from a single selection experiment. These methods enabled us to determine that selection induced statistically significant changes in the quantitative genetic architecture of behavioural traits. Most importantly, we found that directional selection significantly reduced V_A present in the dominant direction of variation within **G**, which resulted in a fourfold slower predicted response to subsequent selection.

Experimental studies evaluating the effect of selection on **G** in a constant environment are rare [14,15]. In a selection experiment on thorax length in *Drosophila melanogaster*, the **G** matrix of the control population did not seem to differ from that of the populations selected for small or large thorax [14]. In *D. serrata*, Hine *et al.* [15] showed that an evolutionary limit to selection for an increase in male attractiveness was caused by a lack of V_A in the direction favoured by sexual selection. Hine *et al.*'s study exemplifies a key feature of the multivariate response to selection: V_A can exist for all traits within **G**, yet little or no V_A may exist along the direction selection is trying to move the population [5]. This raises the possibility that a change in the orientation of **G**, induced by selection, could have caused the selection plateaus that occurred in HR mice during generations 21–31 [16]. We can confidently rule out this possibility, as the level of V_A captured by the leading eigentensor ($\mathbf{E}_1$) did not differ significantly in HR mice before versus after the observed selection plateaus (figure 2c), and the multivariate breeder's equation predicted positive and significant response in the selected traits (days 5 and 6) in generations 21–31 (figure 3). Such discrepancy between the predicted and observed responses to selection may be related to additional constraints on the evolution of locomotor behaviour arising from unmeasured behavioural, neurobiological or physiological traits.

Changes in **G** under selection were repeatable across the four replicate HR lines (electronic supplementary material, figure S5), suggesting a negligible effect of genetic drift on differences in the estimates of **G** between C and HR. Yet one could argue that the observed differences among **G** matrices are attributable to founder effects that were simply maintained throughout the experiment. This scenario is highly improbable for several reasons. First, the possibility for any large difference in V_A among lines at generation 0 is very unlikely because all eight lines were established by randomly choosing individuals from a randomly bred population who all shared the same grandparents (see also the paper by Garland *et al.* [32] regarding the mini-muscle allele). Second, each line was assigned by coin toss to a selection group (i.e. C or HR), meaning there is an extremely low probability of the observed pattern occurring by chance. The probability that four lines with relatively lower **G** were all randomly assigned to the selection group is $p = 0.014$ (details in electronic supplementary material, appendix S1). Third, we have good evidence that h^2 of the selected trait

Table 1. Summary of the genetic covariance tensor. Shown are the eigenvectors (***e***) of the two leading eigentensors ($\mathbf{E}_1$ and $\mathbf{E}_2$), their eigenvalues and the percentage of variation each eigenvector explains within their respective **E**, and loadings on each trait (wheel running on days 1–6). Only shown are the eigenvectors describing more than 45% of the variance within their respective eigentensor.

E	eigenvalues of E	variation explained	*e*	eigenvalues of *e*	variation explained in E	trait loadings					
						day 1	day 2	day 3	day 4	day 5	day 6
$\mathbf{E}_1$	0.1917	91.9%	e_{11}	−0.999	99.8%	−0.374	−0.408	−0.389	−0.427	−0.439	−0.409
$\mathbf{E}_2$	0.0076	3.7%	e_{21}	0.725	52.6%	0.820	0.461	0.151	0.262	0.135	−0.070
			e_{22}	−0.687	47.2%	0.277	−0.149	−0.386	−0.325	−0.455	−0.663

Figure 3. Predicted response to selection (**R**; in units of standard deviation) for wheel running on days 1–6 in control (C) and selected (HR) mice over different blocks of generations. The posterior modes and 95% HPD intervals incorporate the uncertainty from estimating both **G** (estimated separately for C and HR lines) and β (as estimated in HR lines). This figure shows the consequences of changes to **G** induced by directional selection (i.e. a large reduction in **R** in HR lines). However, for wheel running on days 5 and 6 (the selected traits), values of **R** were still significantly higher than zero even at the selection plateau (i.e. generations 21–31). (Online version in colour.)

(average wheel running on days 5 and 6) was very similar in C and HR mice at generation 0, but rapidly decreased in HR mice (electronic supplementary material, figure S6). Finally, a complete reanalysis of the data, based on smaller blocks of four generations, also suggests that C and HR lines did not initially differ in V_A along e_{11} (electronic supplementary material, figure S7). Thus, we are confident that the differences among **G** matrices in C versus HR mice were caused by selection *per se*.

Some of our observations could suggest that the changes in **G** were caused by selection generating negative LD among the loci that affect the six behavioural traits [6]. Indeed, the differences in **G** between C and HR mice emerged within the first generation block (i.e. 0–10), after which the differences remained relatively stable (see electronic supplementary material, figure S8, for a fine-scale analysis based on smaller generation blocks). However, because the selected trait has relatively low h^2 and selection was practised within families, the change attributable to the combined effects of inbreeding and LD should be relatively small (see also [25]). According to calculations based on the infinitesimal model, the predicted reduction in each element of **G** is only approximately 11% after 10 generations of selection (electronic supplementary material, figure S9). This clearly does not account for the 50–80% reduction we observed throughout the entire matrix.

Under the infinitesimal model, selection is not expected to significantly change allele frequencies at any particular locus, because this (null) model assumes that quantitative traits are determined by an infinite number of genes, each of them with an infinitesimally small effect. However, changes in allele frequency under directional selection have been shown in a few livestock populations [24] and in the current experiment. Specifically, we previously showed, in two of the four HR lines, dramatic changes in the frequency of a Mendelian recessive 'mini-muscle' allele that causes a 50% reduction in hind limb muscle mass, as well as many pleiotropic effects that are apparently conducive to supporting high levels of

endurance exercise [32]. It is likely that allele frequencies were also altered at many additional loci.

In addition, previous research has shown that wheel running results for days 1–6 generally share some common quantitative trait loci [20], which could explain why changes in **G** were proportional (i.e. distributed throughout the entire matrix). Unique genomic regions were also identified to specifically affect wheel running during the initial exposure (days 1–3) to wheels [20]. These differences among days may be related to anxiety or fear-related behaviours having an effect on wheel running only during the initial days following exposure to a wheel and altered housing conditions [20]. Previous research failed to detect behavioural differences in C versus HR mice in a novel open-field test [33], suggesting that anxiety or fear-related traits were not indirectly selected in the HR lines. Generally, trait combinations under weak selection should display higher V_A than those under strong selection [31,34]. Consistent with this expectation, we observed fewer differences among **G** matrices for trait combinations involving wheel running on days 1–3 than for elements of **G** corresponding to later in the sequence of days (figure 2*b*; see also trait loadings on e_{11}, e_{21} and e_{22}, table 1), which could be explained by differential alteration of allele frequencies for days 1–3 versus 4–6.

An important feature of this experiment is that selection was applied on behaviour, a type of trait for which we have little information regarding **G**. Most previous comparisons of **G** matrices among populations and species have concentrated on relationships among morphological or life-history traits [11]. As this study is the first robust empirical test of the stability of **G** under selection involving behaviour, it broadens our understanding of the stability of **G** to a new type of quantitative trait. Compared with morphological traits, behavioural traits generally are more closely associated with Darwinian fitness [35] and generally have lower h^2 [5]. Future research is needed to determine whether the evolutionary dynamics of **G** under selection are generally different among types of characters, and whether the stability of **G** relates to how closely the character types are related to fitness.

More research is also needed to determine the extent to which our results apply to wild populations experiencing natural rates of immigration and emigration (gene flow) and spatio-temporal variation in selection gradients. To the extent that the evolutionary dynamics of **G** are similar in laboratory versus wild populations, then our results have important implications for populations facing a consistent and directional rapid environmental change. The sensitivity of a population to global change will be mediated by both its resilience (i.e. ability to withstand and recover from a perturbation) and its adaptive capacity (i.e. both phenotypic plasticity and evolutionary potential) [1]. Locomotor capacity may be a key component of the resilience of a population [3], because it strongly affects the ability of individuals to disperse within and across increasingly fragmented habitats, and hence the ability of the population to expand its range rapidly following disturbance. Moreover, locomotor capacity can itself evolve towards higher levels if global change favours higher dispersal ability or broader daily ranging patterns. For example, individuals with high locomotor capacity might have higher Darwinian fitness if they can successfully disperse to and take advantage of habitats that have become newly available or favourable [2]. Thus, evolution towards higher locomotor capacity may help populations to track the preferred climate space [3]. However, if directional selection further exacerbates constraints on the evolution of locomotor behaviour, as we observed here, then evolution may not occur quickly enough to allow evolutionary adaptation to rapidly shifting environmental conditions [6].

Data accessibility. Dryad: http://dx.doi.org/10.5061/dryad.k6m50.

Authors' contributions. T.G. and P.A.C designed and performed the selection experiment. P.A.C. and V.C. compiled the data and pedigree. V.C. and M.E.W. analysed the data. All authors were involved in interpreting the results and contributed to writing the manuscript.

Competing interests. We have no competing interests.

Funding. This project was supported by grants from NSF to T.G. (IOS-1121273) and to P.A.C. (EF-0328594).

Acknowledgements. We thank people who collected the data over the generations, and J. G. Kingsolver, J. Endler, A. Charmantier and three anonymous reviewers for comments on various versions of the manuscript.

References

1. Williams SE, Shoo LP, Isaac JL, Hoffmann AA, Langham G. 2008 Towards an integrated framework for assessing the vulnerability of species to climate change. *PLoS Biol*. **6**, 2621–2626. (doi:10.1371/journal.pbio.0060325)
2. Hoffmann AA, Sgro CM. 2011 Climate change and evolutionary adaptation. *Nature* **470**, 479–485. (doi:10.1038/nature09670)
3. Feder ME, Garland Jr T, Marden JH, Zera AJ. 2010 Locomotion in response to shifting climate zone: not so fast. *Annu. Rev. Physiol*. **72**, 167–190. (doi:10.1146/annurev-physiol-021909-135804)
4. Lande R. 1979 Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–416. (doi:10.2307/2407630)
5. Walsh B, Blows MW. 2009 Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annu. Rev. Ecol. Syst*. **40**, 41–59. (doi:10.1146/annurev.ecolsys.110308.120232)
6. Blows MW, Walsh B. 2009 Spherical cows grazing in flatland: constraints to selection and adaptation. In *Adaptation and fitness in animal populations* (eds J van der Werf, H-U Graser, R Frankham, C Gondro), pp. 83–101. Dordrecht, The Netherlands: Springer.
7. Barton NH, Turelli M. 1987 Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genet. Res*. **49**, 157–173. (doi:10.1017/S0016672300026951)
8. Etterson JR, Shaw RG. 2001 Constraint to adaptive evolution in response to global warming. *Science* **294**, 151–154. (doi:10.1126/science.1063656)
9. Revell LJ. 2007 The G matrix under fluctuating correlational mutation and selection. *Evolution* **61**, 1857–1872. (doi:10.1111/j.1558-5646.2007.00161.x)
10. Jones AG, Arnold SJ, Bürger R. 2007 The mutation matrix and the evolution of evolvability. *Evolution* **61**, 727–745. (doi:10.1111/j.1558-5646.2007.00071.x)
11. Arnold SJ, Bürger R, Hohenlohe PA, AJie BC, Jones AG. 2008 Understanding the evolution and stability of the **G**-matrix. *Evolution* **62**, 2451–2461. (doi:10.1111/j.1558-5646.2008.00472.x)
12. Aguirre JD, Hine E, McGuigan K, Blows MW. 2014 Comparing G: multivariate analysis of genetic variation in multiple populations. *Heredity* **112**, 21–29. (doi:10.1038/hdy.2013.12)
13. Turelli M. 1988 Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**, 1342–1347. (doi:10.2307/2409017)
14. Shaw FH, Shaw RG, Wilkinson GS, Turelli M. 1995 Changes in genetic variances and covariances: G whiz! *Evolution* **49**, 1260–1267. (doi:10.2307/2410450)

15. Hine E, McGuigan K, Blows MW. 2011 Natural selection stops the evolution of male attractiveness. *Proc. Natl Acad. Sci. USA* **108**, 3659–3664. (doi:10.1073/pnas.1011876108)
16. Careau V, Wolak ME, Carter PA, Garland TJr. 2013 Limits to behavioral evolution: the quantitative genetics of a complex trait under directional selection. *Evolution* **67**, 3102–3119. (doi:10.1111/evo.12200)
17. Blows MW, Hoffmann AA. 2005 A reassessment of genetic limits to evolutionary change. *Ecology* **86**, 1371–1384. (doi:10.1890/04-1209)
18. Swallow JG, Carter PA, Garland Jr T. 1998 Artificial selection for increased wheel-running behavior in house mice. *Behav. Genet*. **28**, 227–237. (doi:10.1023/A:1021479331779)
19. Falconer DS. 1952 The problem of environment and selection. *Am. Nat*. **86**, 293–298. (doi:10.1086/281736)
20. Kelly SA *et al.* 2010 Genetic architecture of voluntary exercise in an advanced intercross line of mice. *Physiol. Genomics* **42**, 190–200. (doi:10.1152/physiolgenomics.00028.2010)
21. Roff DA, Wilson AJ. 2014 Quantifying genotype-by-environment interactions in laboratory systems. In *Genotype-by-environment interactions and sexual selection* (eds J Hunt, D Hosken), pp. 101–136. Chichester, UK: John Wiley and Sons.
22. Hadfield JD. 2010 MCMC methods for multi-response generalised linear mixed models: the MCMCglmm R package. *J. Stat. Softw*. **33**, 1–22. (doi:10.18637/jss.v033.i02)
23. Sorensen DA, Kennedy BW. 1984 Estimation of genetic variances from unselected and selected populations. *J. Anim. Sci*. **59**, 1213–1223.
24. Fontanesi L, Schiavo G, Scotti E, Galimberti G, Calò DG, Samorè AB, Gallo M, Russo V, Buttazzoni L. 2015 A retrospective analysis of allele frequency changes of major genes during 20 years of selection in the Italian Large White pig breed. *J. Anim. Breed. Genet*. **132**, 239–246. (doi:10.1111/jbg.12127)
25. Meyer K, Hill WG. 1991 Mixed model analysis of a selection experiment for food intake in mice. *Genet. Res*. **57**, 71–81. (doi:10.1017/S0016672300029062)
26. Hansen TF, Houle D. 2008 Measuring and comparing evolvability and constraint in multivariate characters. *J. Evol. Biol*. **21**, 1201–1219. (doi:10.1111/j.1420-9101.2008.01573.x)
27. Hine E, Chenoweth SF, Rundle HD, Blows MW. 2009 Characterizing the evolution of genetic variance using genetic covariance tensors. *Phil. Trans. R. Soc. B* **364**, 1567–1578. (doi:10.1098/rstb.2008.0313)
28. Zuur A, Ieno EN, Walker N, Saveiliev AA, Smith GM. 2009 *Mixed effects models and extensions in ecology with R*. New York, NY: Springer.
29. Stinchcombe JR, Simonsen AK, Blows MW. 2014 Estimating uncertainty in multivariate responses to selection. *Evolution* **68**, 1188–1196. (doi:10.1111/evo.12321)
30. Gomes FR, Rezende EL, Malisch JL, Lee SK, Rivas DA, Kelly SA, Lytle C, Yaspelkis BBIII, Garland Jr T. 2009 Glycogen storage and muscle glucose transporters (GLUT-4) of mice selectively bred for high voluntary wheel running. *J. Exp. Biol*. **212**, 238–248. (doi:10.1242/jeb.025296)
31. Lande R. 1980 The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**, 203–215.
32. Garland Jr T, Morgan MT, Swallow JG, Rhodes JS, Girard I, Belter JG, Carter PA. 2002 Evolution of a small-muscle polymorphism in lines of house mice selected for high activity levels. *Evolution* **56**, 1267–1275. (doi:10.1111/j.0014-3820.2002.tb01437.x)
33. Careau V, Bininda-Emonds ORP, Ordonez G, Garland Jr T. 2012 Are voluntary wheel running and open-field behavior correlated in mice? Different answers from comparative and artificial selection approaches. *Behav. Genet*. **42**, 830–844. (doi:10.1007/s10519-012-9543-0)
34. Arnold SJ, Pfrender ME, Jones AG. 2001 The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112–113**, 9–32. (doi:10.1023/A:1013373907708)
35. Careau V, Garland Jr T. 2012 Performance, personality, and energetics: correlation, causation, and mechanism. *Physiol. Biochem. Zool*. **85**, 543–571. (doi:10.1086/666970)

Evolution of the additive genetic variance-covariance matrix under continuous directional selection on a complex behavioral phenotype

Vincent Careau[1], Matthew E. Wolak[2], Patrick A. Carter[3], Theodore Garland, Jr.[4]

[1]Canada Research Chair in Functional Ecology, Department of Biology, University of Ottawa, Ottawa, ON, Canada

[2]School of Biological Sciences, University of Aberdeen, Aberdeen, UK

[3]School of Biological Sciences, Washington State University, Pullman, WA, USA

[4]Department of Biology, University of California, Riverside, CA, USA

Author for correspondence:

Vincent Careau (vcareau@uottawa.ca)

Electronic Supplemental Material

Table of Contents

Table S1. Sample sizes. Number of individuals (and families) measured for their voluntary wheel running over a 6-days period in control (C) and high-runner (HR) lines of mice, for the entire dataset (generations 0 to 31) and different periods.

Generation	C	HR	Total
0 to 10			
Individuals	1,687	4,493	6,180
Families	443	441	884
11 to 20			
Individuals	1,569	3,740	5,309
Families	400	397	797
21 to 31			
Individuals	1,680	4,159	5,839
Families	432	432	864
0 to 31			
Individuals	4,936	12,392	17,328
Families	1,275	1,270	2,545

Table S2. Summary of the genetic covariance tensor applied to the 24 **G** matrices estimated for each of the 8 lines and 3 generation blocks. Shown are the eigenvectors (***e***; ordered by rows in terms of the absolute value) of the two leading eigentensors ($\mathbf{E}_1$ and $\mathbf{E}_2$), their eigenvalues and the percent of variation each eigentensor explains within their respective **E**, and loadings on each trait (wheel running on days 1 to 6). This table shows that the leading eigentensor ($\mathbf{E}_1$) captured a large part (75.6%) of the variance among **G** matrices and that the leading eigenvector $\boldsymbol{e}_{11}$ explained 99.6% of the variation captured by $\mathbf{E}_1$, with loadings ranging from -0.35 to -0.45 only. Applying the tensor analysis to the 24 **G** matrices estimated for each line and generation block yielded results that are consistent with the analysis in which control and selected lines are pooled (see table 1).

Eigentensors (**E**)	Eigenvalues of **E**	% of total variation explained	Eigenvectors (***e***) of **E**	Eigenvalues of ***e***	% of variation explained in **E**	Trait loadings					
						day 1	day 2	day 3	day 4	day 5	day 6
$\mathbf{E}_1$	0.2793	76.5%	$e_{1.1}$	-0.998	99.6%	-0.348	-0.401	-0.400	-0.435	-0.448	-0.410
			$e_{1.2}$	-0.046	0.2%	0.753	0.305	0.018	-0.147	-0.289	-0.485
			$e_{1.3}$	-0.034	0.1%	0.444	-0.286	-0.336	-0.536	0.279	0.493
			$e_{1.4}$	-0.021	0.0%	0.063	0.209	-0.829	0.505	-0.060	0.082
			$e_{1.5}$	-0.018	0.0%	-0.295	0.762	-0.067	-0.458	-0.202	0.278
			$e_{1.6}$	-0.013	0.0%	0.154	-0.200	0.187	0.195	-0.771	0.519
$\mathbf{E}_2$	0.0279	7.7%	$e_{2.1}$	0.711	50.5%	-0.788	-0.524	-0.263	-0.178	-0.067	-0.020
			$e_{2.2}$	-0.703	49.5%	0.289	-0.055	-0.292	-0.402	-0.578	-0.577
			$e_{2.3}$	-0.020	0.0%	0.267	-0.269	0.136	-0.779	0.439	0.194
			$e_{2.4}$	0.012	0.0%	-0.170	0.171	0.475	-0.272	-0.637	0.486
			$e_{2.5}$	-0.002	0.0%	0.413	-0.774	0.108	0.351	-0.228	0.209
			$e_{2.6}$	0.000	0.0%	0.160	0.150	-0.768	-0.043	-0.106	0.591

Table S3. Vector of realized selection gradients (β) applied to voluntary wheel running on each day of a 6-day period of wheel access. Posterior modes with their 95% highest posterior density (HPD) intervals obtained from a multiple regression of relative fitness (number of pups produced) were obtained using different blocks of generations. The target of selection was the average number of wheel revolutions on days 5&6. Boldface highlights selection gradients that were significantly different from zero (i.e., the 95% HPD did not overlap with zero). Note that some selection gradients on days 1 to 4 were negative and significantly different from zero.

generation			95% HPD	
block	trait	β	lower	upper
0-10	day 1	**-0.077**	**-0.152**	**-0.026**
0-10	day 2	**-0.092**	**-0.171**	**-0.017**
0-10	day 3	0.010	-0.068	0.094
0-10	day 4	-0.064	-0.154	0.008
0-10	day 5	**0.758**	**0.668**	**0.847**
0-10	day 6	**0.528**	**0.445**	**0.625**
11-20	day 1	-0.024	-0.111	0.038
11-20	day 2	0.003	-0.103	0.083
11-20	day 3	-0.018	-0.104	0.078
11-20	day 4	**-0.111**	**-0.230**	**-0.033**
11-20	day 5	**0.457**	**0.352**	**0.569**
11-20	day 6	**0.721**	**0.611**	**0.802**
21-31	day 1	**-0.072**	**-0.158**	**0.006**
21-31	day 2	0.065	-0.038	0.175
21-31	day 3	**-0.127**	**-0.239**	**-0.022**
21-31	day 4	-0.003	-0.136	0.096
21-31	day 5	**0.481**	**0.333**	**0.568**
21-31	day 6	**0.542**	**0.411**	**0.611**

Table S4. Predicted response to selection (**R**). Posterior modes and 95% highest posterior density (HPD) intervals for the six behavioural traits in control (C) and selected (HR; for "high runner") mice in different blocks of generations. To obtain the posterior distribution of **R** that incorporates the uncertainty in estimates of both **G** and β, we applied the multivariate breeder's equation to the i^{th} sample of the MCMC posterior distribution for **G** with the i^{th} sample of the MCMC posterior distribution for β (table S2). These calculations were made separately for each generation block (i.e., using the six **G**-matrices in figure S3 and the 3 β in table S3). This table demonstrates that the predicted response to selection was ~4-fold lower in HR than C mice, but that the **R** for days 5&6 in generations 21-31 was still significantly different from zero, even for HR mice.

generation		C			HR		
block	trait	**R**	lower	upper	**R**	lower	upper
0-10	day 1	0.13	0.06	0.21	0.04	0.01	0.06
0-10	day 2	0.15	0.08	0.24	0.04	0.02	0.08
0-10	day 3	0.15	0.08	0.25	0.04	0.01	0.07
0-10	day 4	0.16	0.09	0.27	0.03	0.02	0.07
0-10	day 5	0.20	0.14	0.32	0.06	0.03	0.08
0-10	day 6	0.24	0.14	0.33	0.06	0.04	0.10
11-20	day 1	0.17	0.08	0.25	0.02	0.00	0.05
11-20	day 2	0.13	0.08	0.24	0.03	0.01	0.05
11-20	day 3	0.14	0.06	0.23	0.02	0.00	0.05
11-20	day 4	0.18	0.10	0.28	0.04	0.02	0.08
11-20	day 5	0.19	0.13	0.32	0.06	0.04	0.09
11-20	day 6	0.18	0.10	0.26	0.06	0.04	0.08
21-31	day 1	0.14	0.08	0.22	0.03	0.00	0.06
21-31	day 2	0.18	0.11	0.27	0.03	0.01	0.06
21-31	day 3	0.17	0.11	0.27	0.03	0.01	0.06
21-31	day 4	0.19	0.11	0.28	0.04	0.01	0.07
21-31	day 5	0.24	0.15	0.34	0.05	0.03	0.09
21-31	day 6	0.24	0.15	0.33	0.05	0.03	0.08

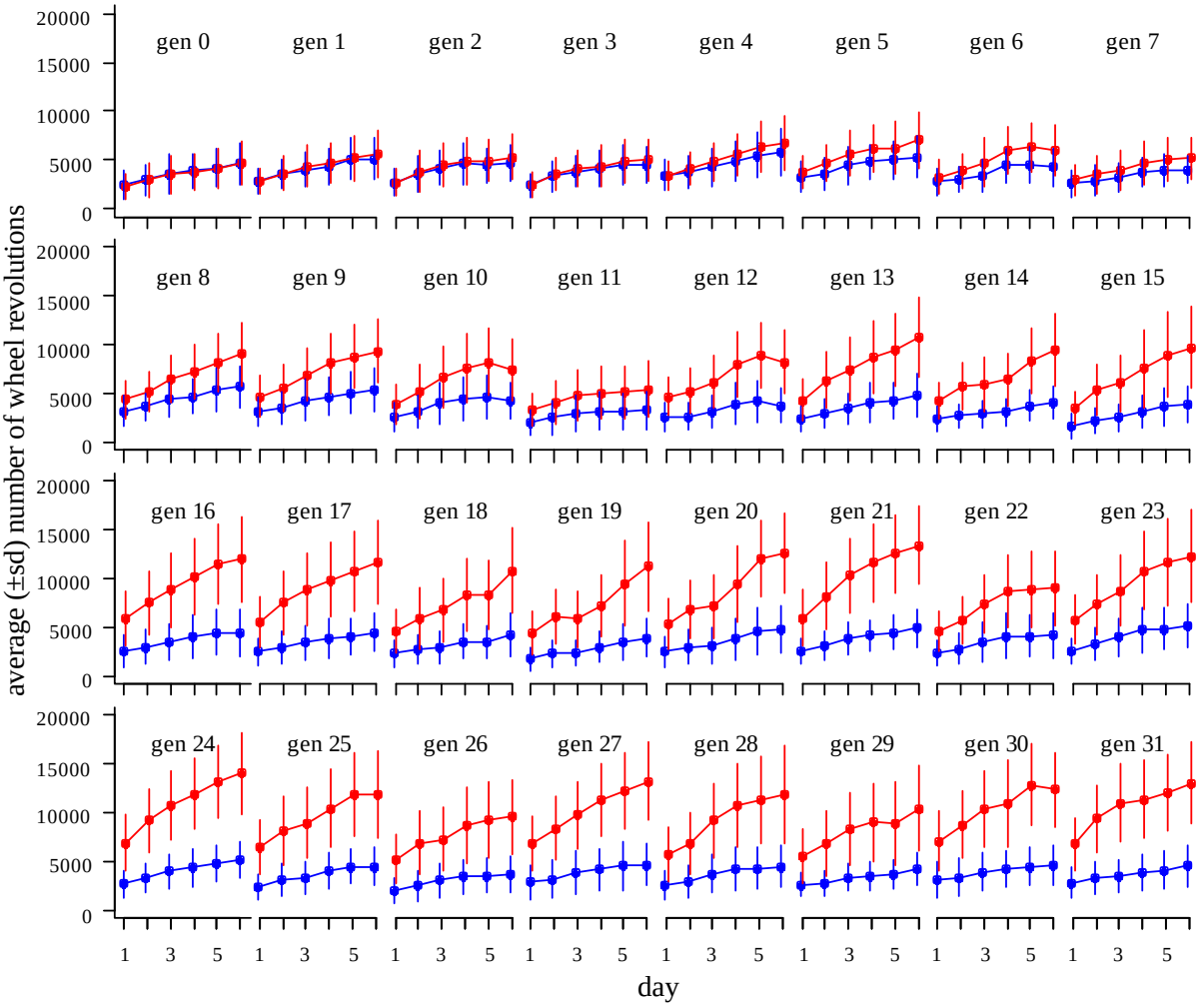

Figure S1. Daily and generational increases in wheel running in mice. Shown are the daily average number of wheel revolutions run (pooled means±sd [ignoring family structure] for four replicate lines in each selection group) over a 6-day period of wheel access in four replicate control (C; blue) and four replicate selected (HR, for “high runners”; red) mice over a 6-day period of wheel access, from generation 0 (before any selection was applied) to 31.

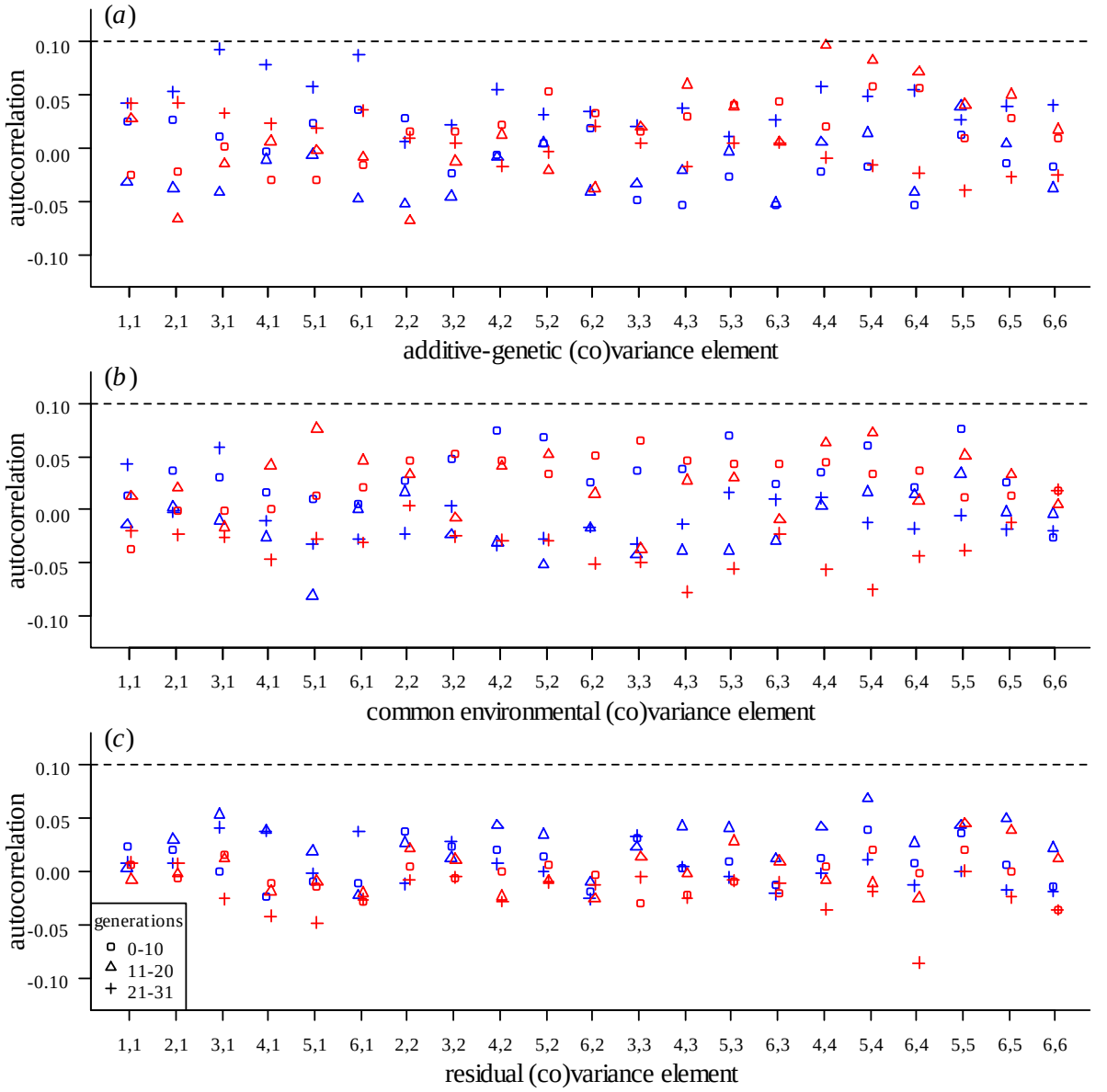

Figure S2. Convergence for MCMCglmm models. (*a*-*c*) Autocorrelation (i.e., at lag 1) of the 1,000 posterior samples for each of the estimated variance components in multivariate animal models run on pooled data for control (blue) and selected (red) mice for different generation blocks. No (co)variance component exceeded the nominal autocorrelation of 0.1. These diagnostic plots indicate that the MCMCglmm models converged properly.

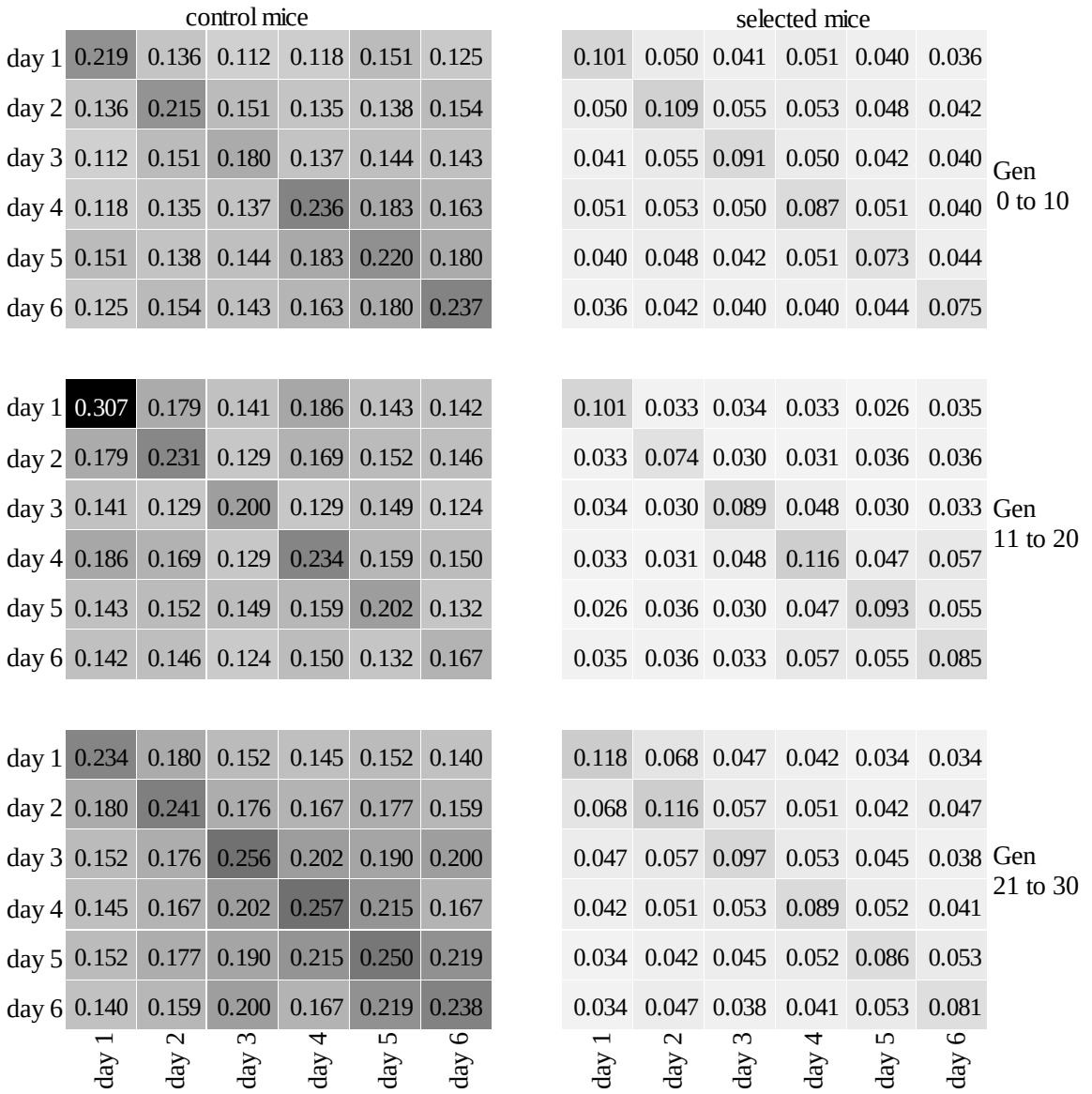

Figure S3. The estimated additive-genetic variance-covariance matrices (**G**) in control and selected mice, based on pooled analyses of all 4 replicate lines in each group). Shown are the posterior modes from Bayesian multivariate animal models fitted with MCMCglmm in R on different blocks of generations (0 to 10, 11 to 20, and 21 to 31). For each **G**, the main diagonal represents the additive-genetic variances (V_A) and the off-diagonals represent the additive-genetic covariances (COV_A) for pairs of traits (voluntary wheel running on each of 6 consecutive days). The cells of the matrix are shaded according to its value (darker cell = higher V_A or COV_A: all values are positive). (See figure S4*d-f* for the genetic correlations.)

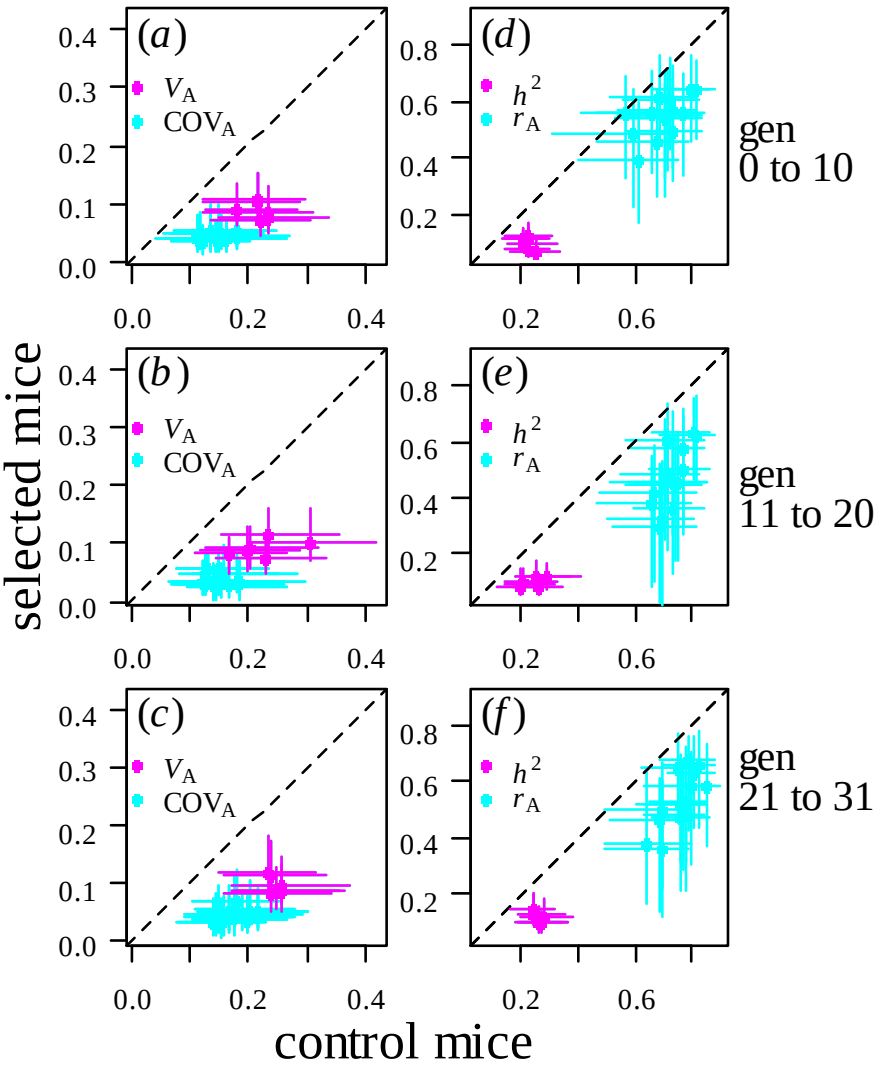

Figure S4. Elements of the additive-genetic variance-covariance matrices. (*a-c*) The posterior modes (dots) and 95% credible intervals (lines) of the highest posterior density from Bayesian multivariate animal models fitted with MCMCglmm in R on different blocks of generations (0 to 10, 11 to 20, and 21 to 31). Each point represents the same element of the matrix [additive-genetic variance (V_A) and co-variance (COV_A)] in selected mice (on the *y* axis) vs. control mice (on the *x* axis). The dashed line shows the 1:1 line. Note that compared to control, selected mice always have lower V_A for all 6 days of wheel running, and weaker positive COV_A among those traits. (*d-f*) Also shown are the corresponding narrow-sense heritability (h^2) and genetic correlations (r_A).

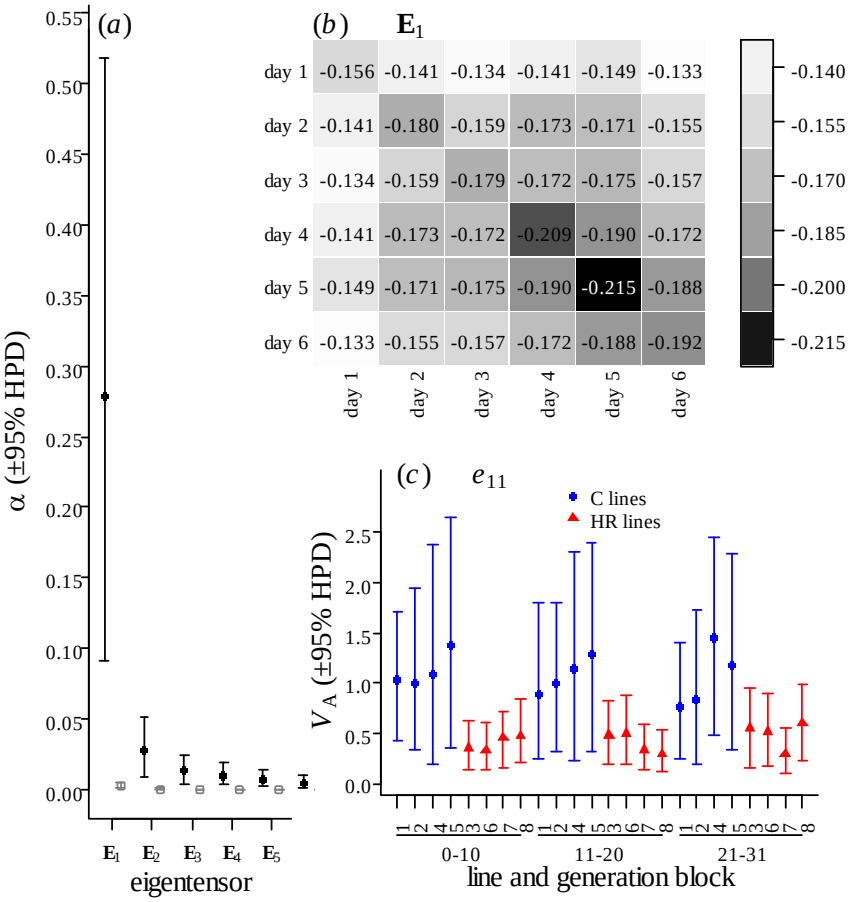

Figure S5. Changes in the additive-genetic variance-covariance matrix (**G**) induced by selection, presented at the level of replicate lines (control lines: 1, 2, 4, and 5; selected line: 3, 6, 7, and 8). (*a*) Variance [α; ±95% highest posterior density (HPD) intervals] accounted for by each eigentensor for the observed (black dots) and randomized (i.e., null hypothesis; grey dots) sets of **G**. (*b*) "Heat map" displaying the pattern of greatest variation among **G**s as captured by $\mathbf{E}_1$ (darker shading indicates greater variation among **G** matrices as measured by elements of $\mathbf{E}_1$, which reflect variance of the (co)variances among the 6 **G** matrices). Hence, variability among **G** matrices was distributed throughout the entire matrix, but slightly more intense for trait combinations involving days 4-6 compared to those earlier in the day sequence. (*c*) The additive-genetic variance (V_A) present in each of the four replicate control lines (C; blue) and four replicate selected lines (HR; red) along the direction of the first eigenvector of $\mathbf{E}_1$ (i.e., $\boldsymbol{e}_{11}$) across generation blocks. Altogether, this figure shows that the effect of selection on **G** was repeatable and that variance among **G** matrices was not caused solely by random genetic drift.

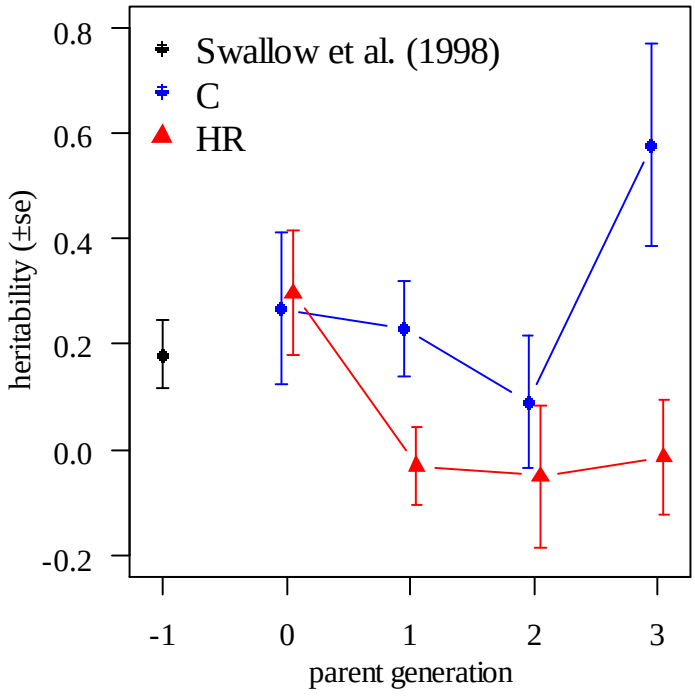

Figure S6. Heritability (±se) of voluntary wheel running on days 5&6 (the selected trait) estimated with offspring-on-midparent regressions in generation -1 and 0 by Swallow et al. (Behav. Genet. 28: 227-237) and in control (C; blue dots) and selected (HR; red triangles) mice using data from generation 0 to 1, 1 to 2, 2 to 3, and 3 to 4. For C and HR mice separately, we first calculated residual values in a model that included several fixed effects fit within generation [sex, age, inbreeding coefficient, line, and measurement block (batches 1-3 and rooms 1-2)]. The residuals of these models were then used in regressions weighted for litter size using an iterative process as described in Lynch and Walsh (1998, Genetics and Analysis of Quantitative traits, p541), using the "osw.R" function accompanying Careau *et al.* (2013, Evolution 67: 3102-3119). The weight (w_i) of the i^{th} family was calculated as $w_i=n_i/[n_i(t–B)+(1–t)]$, where n_i is the number of mice that were wheel-tested in that family, t the intraclass correlation coefficient (for that line at that generation), and B is the slope squared divided by two. This figure shows that heritability of wheel running was initially similar in C and HR mice, but rapidly declined in HR mice.

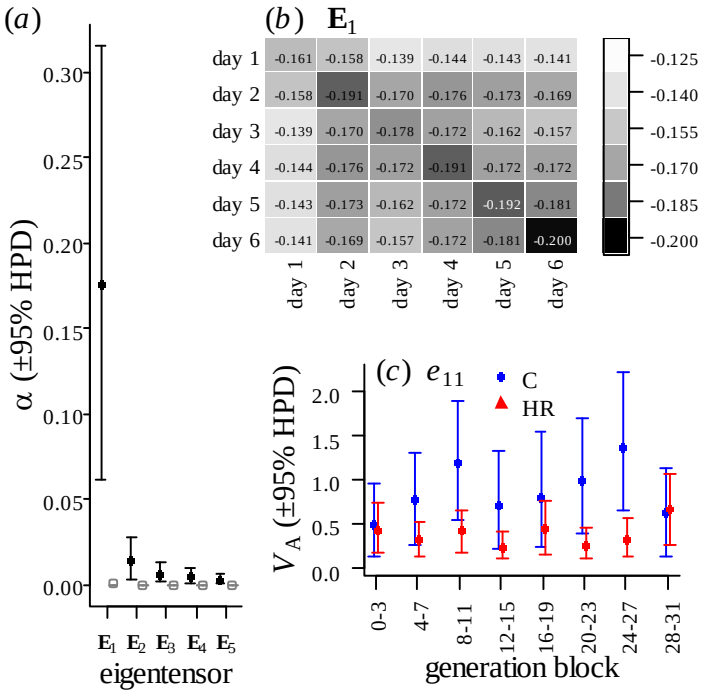

Figure S7. Changes in the additive-genetic variance-covariance matrix (**G**) induced by selection, presented as in Figure 2 (main text; see also figure S5 above), except that the genetic covariance tensor was applied to blocks of 4 generations in control (C) and selected (HR) mice. Altogether, this figure shows that V_A along $\boldsymbol{e}_{11}$ was similar in C and HR mice at the beginning of the experiment (generation 0-3), after which C and HR mice gradually diverged. Note that for this analysis, the sample size to estimate **G** was quite low (C mice: mean n = 617, range = 596-649; HR mice: mean n = 1,619, range = 1,405-1,631).

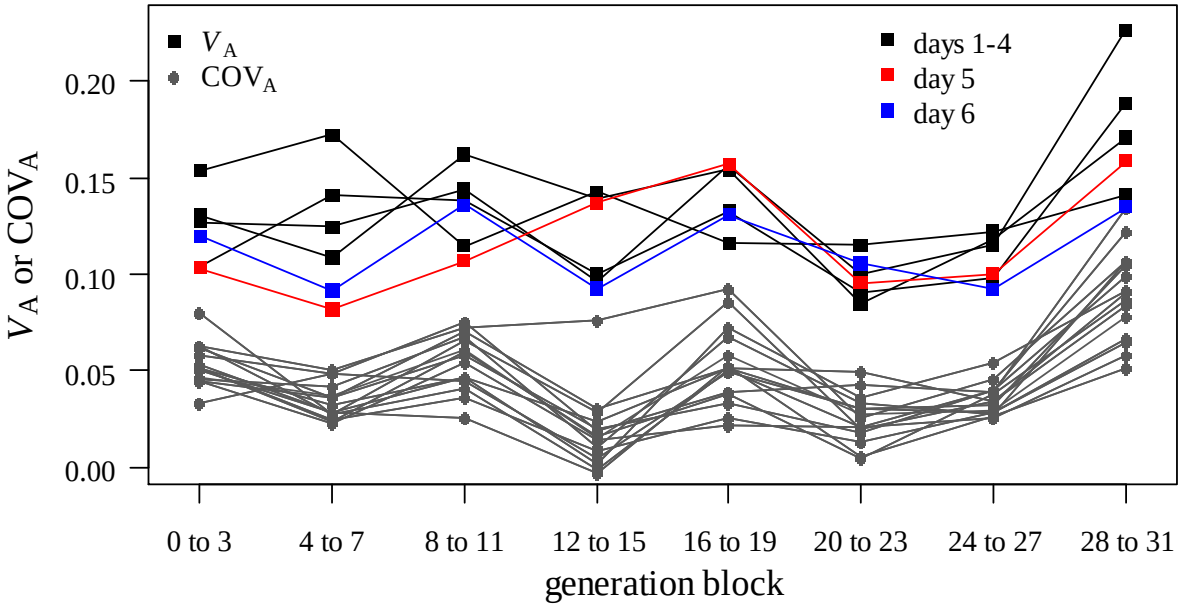

Figure S8. Absence of detectable change in the additive-genetic variance-covariance matrix (**G**) in selected mice, based on pooled analyses of all 4 lines. This shows each element of **G** [i.e., additive-genetic variance (V_A) in wheel running on days 1-4 (black squares), day 5 (red square), day 6 (blue square), and the additive-genetic covariances (COV_A) among all six traits (grey circles)] in mice selected for voluntary wheel running on days 5&6 of a 6-day exposure to wheels, as estimated in separate MCMCglmm models run on blocks of 4 generations. Note the absence of overall temporal change, thus indicating that the effect of selection on **G** occurred very rapidly (i.e., within the first three generations). Repeating these analyses, but breaking up the estimation of **G** into even smaller blocks (i.e., 2 or 3 generations), suffered from an obvious lack of statistical power to detect changes within the first 2-3 generations of selection.

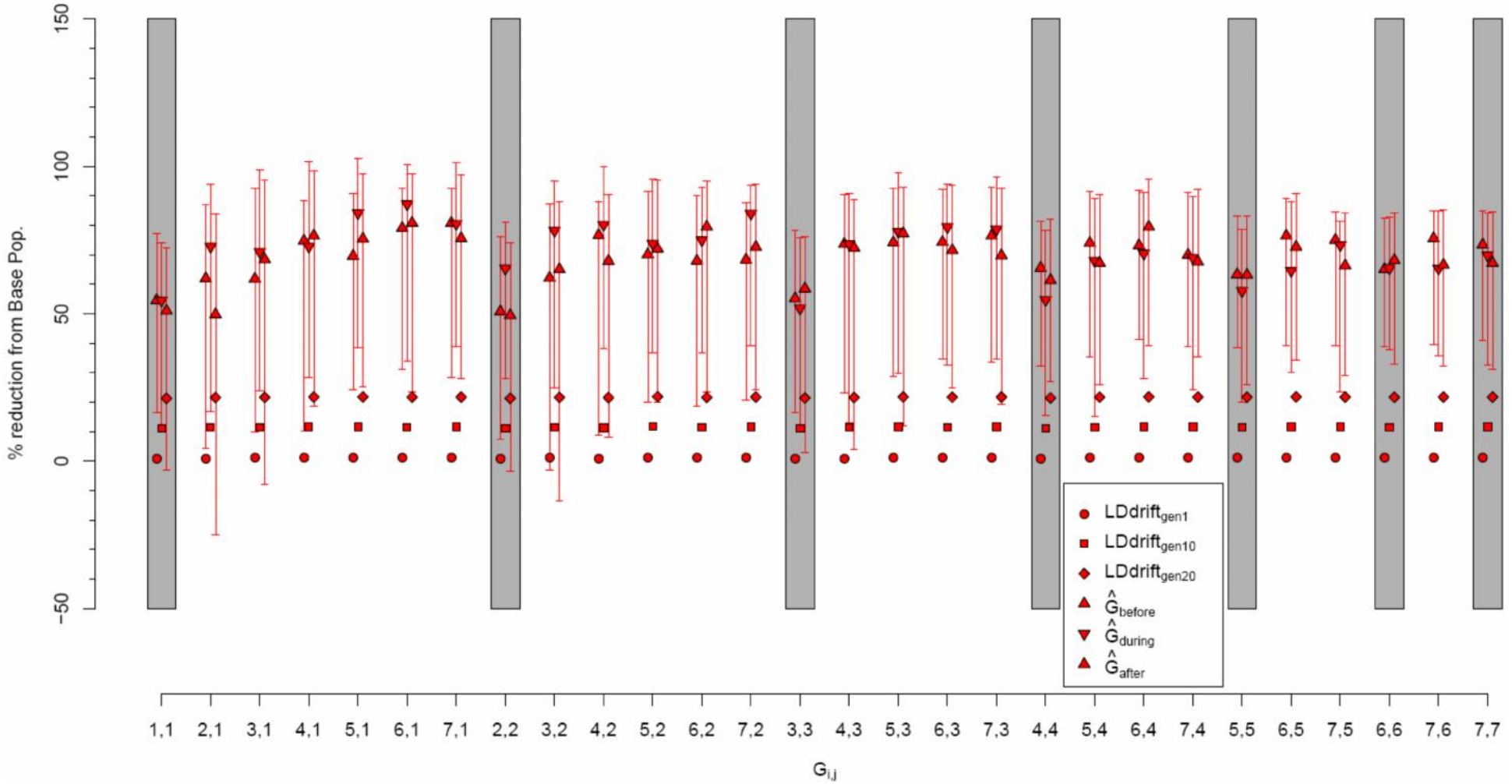

Figure S9: Observed and predicted percent reductions in each element of the additive-genetic variance-covariance matrix **G** for wheel running on days 1-6 in mice selected for voluntary wheel running on days 5&6 of a 6-day exposure to wheels. Values are the % reduction from the base population **G** matrix (estimated for the base stock of mice used to found the control and selected lines, using generations -2, -1, and 0) in each element of the lower triangle of the observed **G** from the MCMCglmm models in the before (up-pointing triangle that is the left-most point of each trio of triangles), during (down-pointing triangle), and after (up-pointing triangle that is the right-most point of each trio of triangles) generation blocks along with their 95% HPD intervals. The % reduction was calculated as [100 × (observed **G** – base **G**) / base **G**]. The expected **G** matrices were calculated based on within-family selection operating on an index trait (average of days 5&6) in the base population **G** matrix and changes occurring through linkage disequilibrium (LD; Bulmer, 1980, The mathematical theory of quantitative genetics. Oxford University Press, Oxford, UK) and drift (or inbreeding within each selection line). The expected **G** used in place of the observed **G** in the percent reduction equation were calculated for 1 (circles), 10 (squares), and 20 (diamonds) generations to match the observed **G** estimated in the three corresponding generation blocks. Grey, rectangular background shading indicates the diagonal elements of **G** (i.e., V_A) for reference. Overall, this figure demonstrates that the predicted changes in the (co)variance components due to selection generating LD are much lower than the observed changes in the selected mice. Details on the theory and calculations are available from the authors and various references (Falconer & MacKay, 1996, Introduction to Quantitative Genetics, 4th ed. Essex, U.K.: Longman; 1996; Shaw *et al*., 1995, Evolution 49: 1260-1267; Verrier *et al*., 1991, Livestock Production Science 29: 93-114; Villanueva and Kennedy 1990. Theoretical and Applied Genetics 80: 746-752).

Appendix S1: Establishment of experimental lines

Details regarding the establishment and maintenance of the control and selected lines can be found in previous publications (Swallow *et al.* 1998. Behav. Genet. 28: 227-237; Careau *et al.* 2013. Evolution 67: 3102-3119). Only the pertinent details are repeated here for clarity. A total of 112 male and 112 female (each specified to be from a different family) house mice (*Mus domesticus*) of the Hsd:ICR strain were purchased from Harlan Sprague Dawley, Indianapolis, IN (Building 202, Barrier A). The history and maintenance of the Hsd:ICR strain can be found in Swallow *et al.* (1998), Dohm *et al.* (2001. Genetics 159: 267-277), and Girard *et al.* (2002. Behav. Proces. 57: 37-50).

The 224 founder animals were designated generation -2. Males and females were chosen at random and placed into cages to form 112 pairs; the resulting offspring were designated generation -1. Generation -1 litters were randomly assigned to one of eight lines, and one male and one female were chosen randomly from each litter to be breeders for that line. Individuals within each line that had been chosen to be breeders were then paired randomly except that full-sibling mating was disallowed; 10 pairs were established for each line (plus 3 additional pairs as backups: see below). Lines were then randomly assigned into four non-selected control (C) lines and four selected “high-runner” (HR) lines. Their offspring were designated generation 0 and selection was first applied at this generation in HR lines. In each of the 4 HR lines, the male and female from each family that was ranked highest according to the selection criteria (see main text) were randomly paired with individuals from other families in that same line, avoiding full-sibling matings. In C lines, one male and one female were chosen at random from each family and randomly paired, avoiding full-sibling matings. For most generations, 13 pairs were established within each line where the first 10 litters weaned with at least two pups of each sex

were used to maintain that line. The three extra families per line were backups to ensure that each line was propagated with 10 families per generation.

Breeder mice were paired at approximately 10 weeks of age (except generation -2, which was paired at 7 weeks, and generations -1 and 0, which were paired at 8.5 weeks of age) by placing one male and one female together in a fresh cage. Males were removed and weighed 15-18 days after pairing; births began 19 days after pairing. From 19 to 33 days after pairing, pregnant females were checked daily between 1600 and 1800. On the day of parturition, the number of pups was recorded. At 21 days of age, offspring were weaned from the dam, weighed, toe clipped for individual identification, and housed in groups of four by sex. All offspring of HR families were kept, but for each C family only a random subsample of two males and two females was kept. At weaning, families were arbitrarily assigned to one of three wheel-running measurement batches such that each batch contained approximately 200 mice from nearly equal numbers of families from each line.

Probability calculations

Considering how lines were established (see above), it is very unlikely that they significantly differed in V_A at generation 0. Nevertheless, let's assume that 4 of the 8 lines had significantly lower V_A and calculate the probability that they all ended up in the selection group (see figure S5 above). We used two methods to calculate this probability, both of which yield the same answer ($P = 0.014$). The first method is based on simple probability calculation. The total number of ways to assign 4 of the 8 lines to each selection group is:

$$\binom{8}{4}$$

which is equal to 70. Assuming that 4 of the 8 lines have high V_A and 4 of 8 lines have low V_A, the number of ways of assigning all 4 low V_A lines to the HR selection group is

$$\binom{4}{4}$$

which is equal to 1, i.e., there is only one way in which all 4 low V_A lines can be assigned to the HR selection group. Thus, the probability of observing, by chance alone, 4 HR lines with low V_A is:

$$P = \frac{\binom{4}{4}}{\binom{8}{4}} = \frac{1}{70} = 0.014$$

The second method is based on simulation. The code below can be pasted directly into R, to show that the probability of observing, by chance alone, 4 HR lines with low V_A and 0 HR lines with high V_A is $P = 0.014$.

```
################      SIMULATION START      ####################
rm(list = ls())
set.seed(101)

# Do lines end up having high genetic variance: TRUE or FALSE
hiVar <- rep(c(TRUE, FALSE), each = 4)

calcFun <- function(...){
   # randomly select 4 out of 8 lines to be in the same group -
   # so the first 4 are one group
   draw <- hiVar[sample.int(8)][1:4]
   # Now, are those 4 lines ALL the high genetic variance lines
   # (value == TRUE)

   ## Tests the condition (one-tailed test) that:
   ## all C lines VA > HR line VA

   all(draw == TRUE)
 }

N <- 1000000
system.time(obs <- replicate(N, expr = calcFun()))
mean(obs)
################      SIMULATION END      ####################
```

Appendix S2: Genetic covariance tensor

To illustrate how the genetic covariance tensor (Hine *et al.*, 2009. Philos Trans R Soc Lond B 364:1567-1578) is applied within a Bayesian framework (Aguirre *et al*., 2014. Heredity 112:21-29), we simulated **G** matrices that differ in contrasting ways among populations. The R code below simulates 3 traits in two sets of 4 populations, with 4,600 individuals in each population. This code was used to create the datasets used to estimate **G** matrices shown in figure S8. In the first scenario (populations #1a, 1b, 1c, and 1d), the **G** matrices gradually differ among populations by a proportional amount (i.e., all elements of the matrix are changed by a constant value from population 1a through population 1d). In the second scenario (populations #2a, 2b, 2c, and 2d), **G** matrices gradually differ among populations but more for some pairs of traits than others (i.e., the top-left "corner" of the matrix is very similar among populations, but the lower-right "corner" grows more and more different among populations):

```
##############################################################################################
#######################  SIMULATIONS   START ##################################
##############################################################################################
rm(list = ls())
library(nadiv)
set.seed(100)

# number of traits:
t <- 3

#Because scaled (co)variances are easier to interpret/think about, we define
an 'H matrix' which contains heritabilities along the diagonal and
correlations on the off-diagonal. By first specifying a desired total
phenotypic variance for each trait, the additive genetic and environmental
covariances (G and E matrices, respectively) are easily calculated from the H
matrix.
#The first population H matrix for both scenario 1 and 2:
##heritability on the diagonal and genetic correlation on off-diagonal
H1a <- H2a <- matrix(c(0.6, 0.25, 0.25,
                       0.25, 0.6, 0.25,
                       0.25, 0.25, 0.6), t, t, byrow = TRUE)

#Now the other populations in scenario 1, creating proportional differences:
H1b <- 0.75 * H1a
H1c <- 0.50 * H1a
H1d <- 0.25 * H1a
```

```
#Now the other populations in scenario 2, creating gradual differences:
H2b <- matrix(c(0.6, 0.18, 0.14,
                0.18, 0.4, 0.2,
                0.14, 0.2, 0.2), t, t, byrow = TRUE)
H2c <- matrix(c(0.6, 0.16, 0.098,
                0.16, 0.3, 0.07,
                0.098, 0.07, 0.15), t, t, byrow = TRUE)
H2d <- matrix(c(0.6, 0.14, 0.06,
                0.14, 0.2, 0.04,
                0.06, 0.04, 0.05), t, t, byrow = TRUE)

# Store all of these in an array:
Harray <- array(data = c(H1a, H1b, H1c, H1d, H2a, H2b, H2c, H2d),
      dim = c(t, t, 8))

# Start with each trait and population having roughly the same phenotypic
variance of 1:
P <- diag(1, t, t)

# Fill in the diagonal elements of G for each population:
Garray <- sapply(seq(dim(Harray)[3]), FUN =
function(i){diag(diag(Harray[,,i]) * diag(P), t, t)}, simplify = "array")

# Fill in the diagonal elements of the environmental (co)variance matrix for
each population. Since P=G+E, then E=P-G:
Earray <- sapply(seq(dim(Harray)[3]), FUN = function(i){diag(diag(P) -
diag(Garray[,,i]), t, t)}, simplify = "array")

# Create and fill the diagonal for phenotypic (co)variance matrices of each
population:
Parray <- Garray + Earray

# Now, fill in the off-diagonal elements of the G, E, and P matrices
for(i in 1:dim(Harray)[3]){
  for(r in 1:dim(Harray)[1]){
     if(r != dim(Harray)[1]){
        for(c in (r+1):dim(Harray)[2]){
           Garray[r, c, i] <- Garray[c, r, i] <- Harray[r, c, i] *
sqrt(Garray[r, r, i] * Garray[c, c, i])
           Parray[r, c, i] <- Parray[c, r, i] <- Harray[c, r, i] *
sqrt(Parray[r, r, i] * Parray[c, c, i])
           Earray[r, c, i] <- Earray[c, r, i] <- Parray[r, c, i] - Garray[r,
c, i]
        }
     }
  }
}

# Create a pedigree
## HS pedigree with 100 sires, 5 dams per sire, and 8 offspring per dam =
4600 individuals (2 generations):
n <- 4600

# Create array to store all of the data:
Darray <- array(dim = c(n, 13, dim(Harray)[3]))
D <- lapply(seq(dim(Darray)[3]), FUN = function(i)
as.data.frame(Darray[,,i]))
```

```
# Create an Identity matrix for the population (so don't have to do it
repeatedly below):
I <- Diagonal(n, 1)

# For each population, create breeding values and residuals for each trait
according to the G and E covariance matrices, respectively:
for(j in 1:length(D)){
   D[[j]][, 1:4] <- simPedHS(s = 100, d = 5, n = 8, uniqueDname = TRUE,
   prefix = j)
   A <- makeA(D[[j]][ ,1:3])
         D[[j]][, 5:7] <- grfx(n, G = Garray[,,j], incidence = A, output =
      "matrix")   # Breeding values
   D[[j]][, 8:10] <- grfx(n, G = Earray[,,j], incidence = I, output =
   "matrix")       # residual deviations
   D[[j]][, 11:13] <- matrix(rnorm(3, 0, 0.1), nrow = n, ncol = 3, byrow =
   TRUE) + D[[j]][, 5:7] + D[[j]][, 8:10]     # phenotype of each trait with
   mean ~ 0
   names(D[[j]]) <- c("id", "dam", "sire", "sex", paste0("bvt", seq(t)),
   paste0("et", seq(t)), paste0("pt", seq(t)))
   # Add 2 columns to indicate the population and scenario
   D[[j]][, "pop"] <- if(j <= 4) j else j-4
   D[[j]][, "scen"] <- if(j <= 4) 1 else 2
}

# Organise into one data frame:
MasterD <- do.call(rbind, D)
save("Harray", "Garray", "Earray", "Parray", "D", "MasterD", file =
"TensorExampleSim.RData")

#load(file = "TensorExampleSim.RData")

# To show it worked
# We can also check and get the estimated covariance among breeding values in
a population and compare that to the G matrix (note, these won't be exactly
the same due to Monte Carlo error)
cov(D[[1]][, c("bvt1", "bvt2", "bvt3")])
Garray[,,1]

# Same for environmental deviations
cov(D[[1]][, c("et1", "et2", "et3")])
Earray[,,1]

##############################################################################
#########################  SIMULATIONS   END ################################
##############################################################################
####################  MCMCglmm MODELLING   START ############################
##############################################################################
library(MCMCglmm)
pops <- paste0(rep(seq(2), each = 4), rep(letters[1:4], 2))
# select the population to model [i is a number 1-8 (length of pops)]:
i <- 1

# Load the data and extract the 'popi' population data
load(file = "TensorExampleSim.RData")
assign("popi", eval(D[[i]]))
```

```
# Prepare for MCMCglmm
names(popi)[grepl("id", names(popi))] <- "animal"
PED.popi <- popi[, 1:3]

NITT <- 13000*100; THIN <- 10*100; BURNIN <- 3000*100
PRIOR <- list(R = list(V = diag(1/2,3), nu = 2.002),
              G = list(G1 = list(V = diag(1/2,3), nu = 2.002)))

M.popi <- MCMCglmm(cbind(pt1, pt2, pt3) ~ trait-1,
      random = ~ us(trait):animal,
      rcov = ~ us(trait):units,
      family = c("gaussian", "gaussian", "gaussian"),
      nitt = NITT, thin = THIN, burnin = BURNIN,
      pedigree = PED.popi,
      data = popi,
      prior = PRIOR)

# Now re-assign general 'popi' names to be specific for the particular
population chosen and then save
assign(paste0("pop", pops[i]), eval(popi))
assign(paste0("PED.pop", pops[i]), eval(PED.popi))
assign(paste0("M.pop", pops[i]), eval(M.popi))

save(list = c(paste0("pop", pops[i]), paste0("PED.pop", pops[i]),
paste0("M.pop", pops[i])), file = paste0("TensorExampleMod_pop", pops[i],
".RData"))

##############################################################################################
####################  MCMCglmm MODELLING   END ###############################
##############################################################################################
```

Looking at the simulated **G** matrices using heatmaps (figure S8 below), it is clear that the differences among populations #1a, 1b, 1c, and 1d are distributed throughout the matrix, whereas populations #2a, 2b, 2c, and 2d differ mostly in the lower-right corner of the matrix. This is because of the R code above, in which the diagonal elements are all equal for populations 1a, 1b, 1c, and 1d and the correlations between traits are also constant, with the only difference among populations are that the expected elements in H1b-d are 2/3, 1/2, and 1/3 the corresponding elements in H1a. For population #2a, 2b, 2c, and 2d, the R code above simulated substantial variation across the 4 **G** matrices in the lower-right corner, but not for elements in the top-left corner.

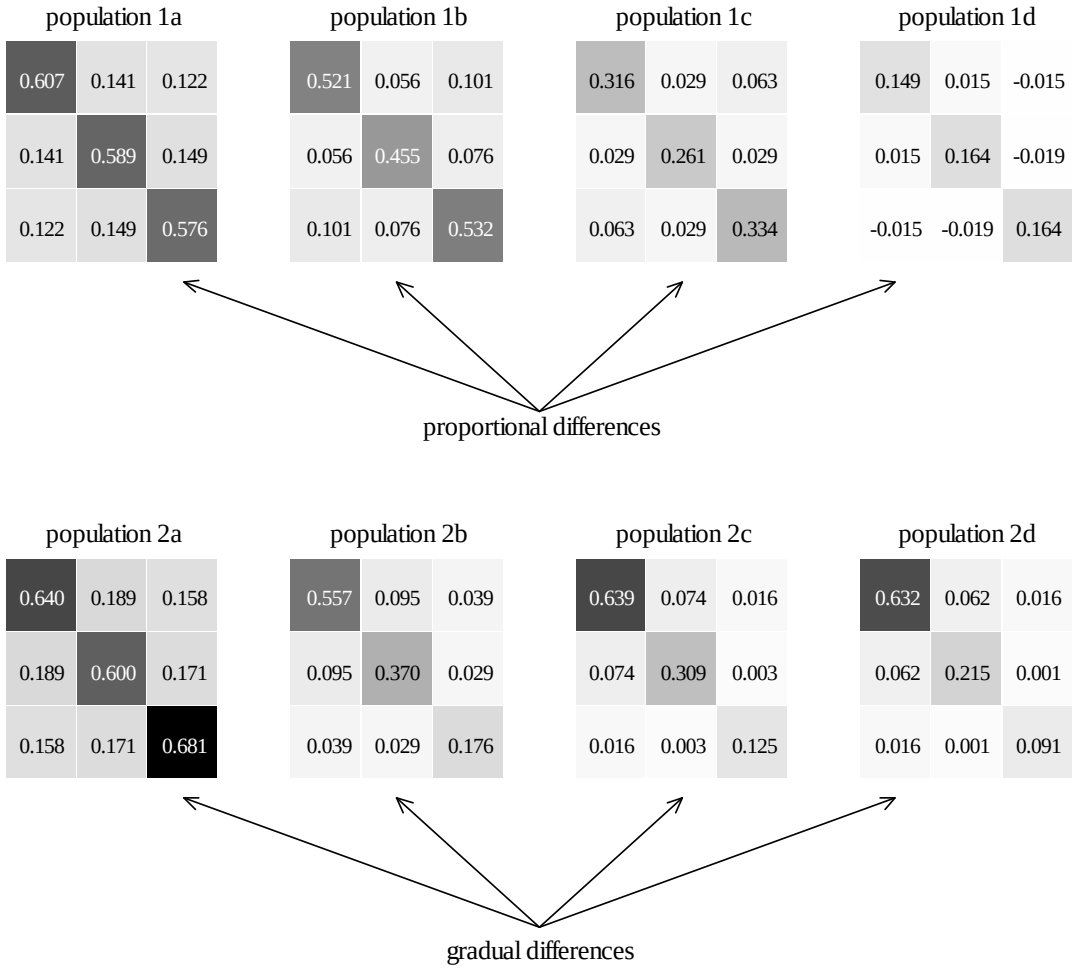

Figure S8. The estimated additive-genetic variance-covariance matrices (**G**) in the 8 simulated populations. Shown are the posterior modes from Bayesian multivariate animal models fitted with MCMCglmm (see example code above). For each **G**, the main diagonal represents the additive-genetic variance (V_A) and the off-diagonals represent the additive-genetic covariances (COV_A) for pairs of traits. The cells of the matrix are shaded according to its value (darker cell = higher V_A or COV_A), using the function "image()" in R.

We ran the code from Aguirre *et al*. (2014. Heredity 112:21-29) on these two sets of simulated **G** matrices. The covariance tensor results match what one would conclude from visually inspecting the patterns in heatmaps depicting the estimates of **G** (figure S8 above) or the expectations in the heritability matrices (R code above, H1a-d or H2a-d). For populations 1a, 1b, 1c, and 1d, the diagonal elements are all equal within each population and the correlations between traits are also constant. The only difference among populations are that the expected elements in H1b-d are 2/3, 1/2, and 1/3 the corresponding elements in H1a. The covariance tensor results reflect this, as the eigentensor $\mathbf{E}_1$ described 93.9% of the variation among the 4 **G** matrices (table S5*a*). The 95% highest posterior density (HPD) intervals of α for the nonzero eigenvalues of the genetic covariance tensor suggested that only $\mathbf{E}_1$ described significant variation among the 4 **G** matrices (figure S9*a*). As expected, $\mathbf{E}_1$ describes changes that were proportional because the variation across the 4 **G** matrices was substantial for all elements (figure S9*b*). The leading eigenvector $\boldsymbol{e}_{11}$ explained 72.9% of the variation captured by $\mathbf{E}_1$ (table S5*a*). Thus, most of the variation among the 4 **G** matrices is captured by a single combination of traits. The trait loadings of $\boldsymbol{e}_{11}$ are all roughly equal (table S5*a*); thus the 3 traits contribute roughly equally to the major axis of variation among **G** matrices. Calculating the V_A along this axis of variation shows that the changes captured in $\boldsymbol{e}_{11}$ were driven by the progressive decrease in V_A from population 1a (most V_A) to population 2d (least V_A) (figure S9*c*). This helps understanding results from the selected experiment, because it is obvious from figure 2*b* that $\mathbf{E}_1$ describes changes that were mostly proportional, that all 6 traits contributed to the major axis of variation among **G** matrices (table 1), and that selected mice harbour lower V_A along this direction ($\boldsymbol{e}_{11}$) than control mice (figure 2*c*).

Table S7. Summary of the genetic covariance tensor applied to (*a*) simulated populations #1a, 1b, 1c, and 1d, and (*b*) simulated populations 2a, 2b, 2c, and 2d. Shown are the eigenvectors (***e***; ordered by rows in terms of the absolute value) of each eigentensor (**E**), their eigenvalues, and the percent of variation they explain within their respective **E**, and loadings on each trait.

Eigentensors (**E**)	Eigenvalues of **E**	% of total variation explained	Eigenvectors (***e***) of **E**	Eigenvalues of ***e***	% of variation explained in **E**	Trait loadings		
						trait 1	trait 2	trait 3
(*a*) populations 1a, 1b, 1c, and 1d								
$\mathbf{E}_1$	0.1304	93.9%	$e_{1.1}$	-0.854	72.9%	0.579	0.574	0.579
			$e_{1.2}$	-0.415	17.2%	0.799	-0.541	-0.263
			$e_{1.3}$	-0.315	9.9%	-0.163	-0.615	0.772
$\mathbf{E}_2$	0.0035	2.5%	$e_{2.1}$	0.713	50.8%	0.461	0.887	-0.010
			$e_{2.2}$	-0.655	42.9%	0.590	-0.298	0.751
			$e_{2.3}$	-0.252	6.4%	-0.663	0.352	0.660
$\mathbf{E}_3$	0.0020	1.4%	$e_{3.1}$	0.729	53.1%	0.544	-0.678	-0.494
			$e_{3.2}$	-0.581	33.8%	0.812	0.278	0.513
			$e_{3.3}$	0.362	13.1%	-0.210	-0.681	0.702
$\mathbf{E}_4$	0.0013	0.9%	$e_{4.1}$	0.707	49.9%	0.971	0.123	-0.203
			$e_{4.2}$	-0.640	40.9%	-0.097	-0.572	-0.814
			$e_{4.3}$	0.302	9.1%	0.216	-0.811	0.544
$\mathbf{E}_5$	0.0009	0.3%	$e_{5.1}$	-0.835	69.8%	0.013	-0.640	0.768
			$e_{5.2}$	0.547	30.0%	0.850	-0.397	-0.346
			$e_{5.3}$	0.049	0.2%	-0.526	-0.658	-0.539
$\mathbf{E}_6$	0.0008	0.6%	$e_{6.1}$	0.724	52.5%	-0.563	-0.137	0.815
			$e_{6.2}$	-0.689	47.5%	-0.478	0.858	-0.186
			$e_{6.3}$	0.012	0.0%	0.674	0.494	0.549
(*b*) populations 2a, 2b, 2c, and 2d								
$\mathbf{E}_1$	0.1309	93.3%	$e_{1.1}$	-0.952	90.6%	-0.276	-0.466	-0.841
			$e_{1.2}$	-0.296	8.8%	-0.102	-0.856	0.508
			$e_{1.3}$	0.077	0.6%	0.956	-0.226	-0.188
$\mathbf{E}_2$	0.0041	2.9%	$e_{2.1}$	0.805	64.8%	0.959	-0.006	0.284
			$e_{2.2}$	-0.578	33.4%	0.031	0.996	-0.081
			$e_{2.3}$	0.134	1.8%	-0.282	0.086	0.955
$\mathbf{E}_3$	0.0022	1.6%	$e_{3.1}$	0.886	78.6%	0.741	0.646	-0.183
			$e_{3.2}$	0.355	12.6%	0.647	-0.760	-0.064
			$e_{3.3}$	-0.297	8.8%	0.181	0.072	0.981
$\mathbf{E}_4$	0.0015	1.1%	$e_{4.1}$	0.785	61.6%	0.700	-0.473	-0.535
			$e_{4.2}$	-0.463	21.4%	-0.694	-0.625	-0.356
			$e_{4.3}$	0.412	17.0%	0.167	-0.621	0.766
$\mathbf{E}_5$	0.0009	0.6%	$e_{5.1}$	-0.743	55.2%	0.649	-0.697	0.305
			$e_{5.2}$	0.663	43.9%	0.586	0.202	-0.785
			$e_{5.3}$	-0.093	0.9%	0.485	0.688	0.540
$\mathbf{E}_6$	0.0008	0.5%	$e_{6.1}$	-0.798	63.7%	-0.081	0.609	-0.789
			$e_{6.2}$	0.602	36.2%	0.337	-0.729	-0.596
			$e_{6.3}$	0.005	0.0%	0.938	0.314	0.146

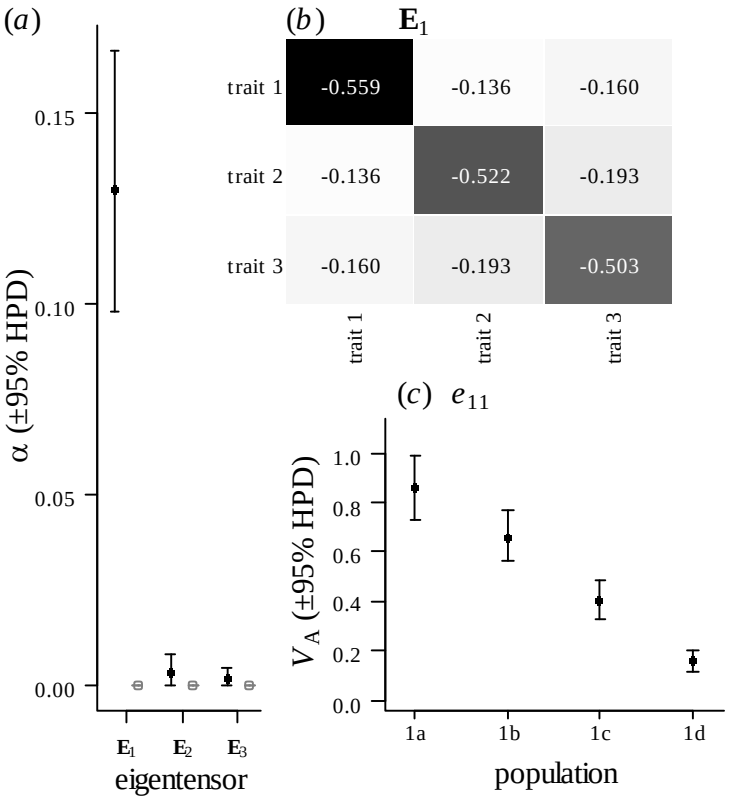

Figure S9. The genetic covariance tensor approach applied to simulated **G** matrices from fictive populations #1a, 1b, 1c, and 1d. (*a*) Variance [α; ±95% highest posterior density (HPD) intervals] accounted for by each eigentensor ($\mathbf{E}_n$) for the observed (black dots) and randomized (i.e., null hypothesis; grey dots) sets of **G**. Because the 95% HPD intervals of the observed vs. randomized sets of **G**s did not overlap for $\mathbf{E}_1$, this eigentensor described significantly more variation among the observed **G** than by chance. (*b*) "Heat map" displaying the pattern of greatest variation among **G** matrices as captured by $\mathbf{E}_1$ (darker shading indicates more variation among **G** matrices). As expected, variability among **G** matrices is distributed throughout the entire matrix. (*c*) Across populations, the additive-genetic variance (V_A) in the direction of the first eigenvector of $\mathbf{E}_1$ (i.e., $\boldsymbol{e}_{11}$).

We can also inspect the patterns from heatmaps depicting the estimates of **G** (figure S8 above) or the heritability matrices in the code for populations 2a, 2b, 2c, and 2d (see H2a-d in R code above). In this second scenario, the four **G** matrices gradually change from the top-left corner being fairly similar among populations to the bottom-right corner being very different among populations. In the covariance tensor analysis, the eigentensor $\mathbf{E}_1$ described 93.3% of the variation among the 4 **G** matrices (table S5*b*). The 95% HPD intervals of α for the nonzero eigenvalues of the genetic covariance tensor suggested that only $\mathbf{E}_1$ described significant variation among the 4 **G** matrices (figure S10*a*). As expected, $\mathbf{E}_1$ describes changes that were gradually distributed throughout the matrix because the variation across the 4 **G** matrices was substantial for elements in the lower-right corner, but not for elements in the top-left corner (figure S10*b*). The leading eigenvector $\boldsymbol{e}_{11}$ explained 90.6% of the variation captured by $\mathbf{E}_1$ (table S5*b*). The trait loadings of $\boldsymbol{e}_{11}$ (table S5*b*) for traits 1, 2, and 3 are -0.276, -0.466, and -0.841, suggesting that they respectively contribute slightly, moderately, and strongly to the major axis of variation among **G** matrices. Calculating the V_A along this axis of variation shows that the changes captured in $\boldsymbol{e}_{11}$ were driven by a gradual decrease in V_A from population 2a (most V_A) to population 2d (least V_A) (figure S10*c*). Again, this helps understanding results from the selected experiment, because in figure 2*b* we can see that $\mathbf{E}_1$ describes changes that were most intense in the lower-right corner of the **G** matrix (the area under most intense selection).

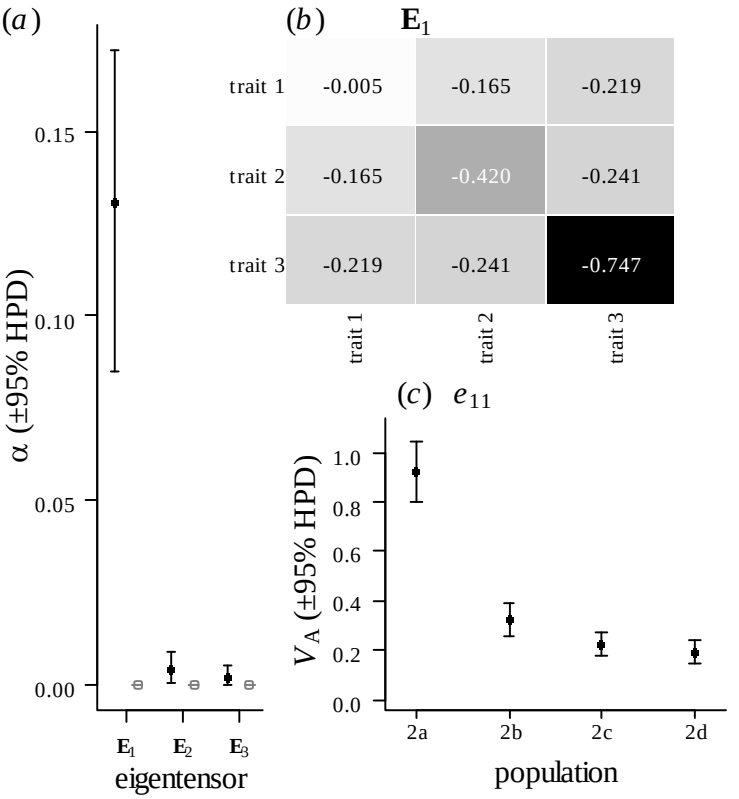

Figure S10. The genetic covariance tensor approach applied to simulated **G** matrices from fictive populations 2a, 2b, 2c, and 2d. (*a*) Variance [α; ±95% highest posterior density (HPD) intervals] accounted for by each eigentensor ($\mathbf{E}_n$) for the observed (black dots) and randomized (i.e., null hypothesis; grey dots) sets of **G**. Because the 95% HPD intervals of the observed vs. randomized sets of **G**s did not overlap for $\mathbf{E}_1$, this eigentensor described significantly more variation among the observed **G** than by chance. (*b*) "Heat map" displaying the pattern of greatest variation among **G** matrices as captured by $\mathbf{E}_1$ (darker shading indicates more variation among **G** matrices). As expected, variability among **G** matrices is more intense in the lower-right corner of the matrix. (*c*) Across populations, the additive-genetic variance (V_A) in the direction of the first eigenvector of $\mathbf{E}_1$ (i.e., $\boldsymbol{e}_{11}$).