Phenotypic stasis with genetic divergence

François Mallard[1,*]

Luke Noble[1]

Thiago Guzella[1]

Bruno Afonso[1]

Charles F. Baer[2]

Henrique Teotónio[1,*]

1. Institut de Biologie de l'École Normale Supérieure, CNRS UMR 8197, Inserm U1024, PSL Research University, F-75005 Paris, France;

2. Department of Biology, University of Florida Genetics Institute, University of Florida, Gainsville, Florida 32611, U.S.A.;

∗ Corresponding authors: mallard@bio.ens.psl.eu, teotonio@bio.ens.psl.eu.

Keywords: phenotypic stasis, **G**-matrix, selection surface, genetic drift, locomotion behavior, transition rates, *Caenorhabditis elegans*, experimental evolution.

1 Abstract

Whether or not genetic divergence on the short-term of tens to hundreds of generations is compatible with phenotypic stasis remains a relatively unexplored problem. We evolved predominantly outcrossing, genetically diverse populations of the nematode *Caenorhabditis elegans* under a constant and homogeneous environment for 240 generations, and followed individual locomotion behavior. We find that although founders of lab populations show highly diverse locomotion behavior, during lab evolution the component traits of locomotion behavior – defined as the transition rates in activity and direction – did not show divergence from the ancestral population. In contrast, the genetic (co)variance structure of transition rates showed marked divergence from the ancestral state and differentiation among replicate populations during the final 100 generations and after most adaptation had been achieved. We observe that genetic differentiation is a transient pattern during the loss of genetic variance along phenotypic dimensions under drift during the last 100 generations of lab evolution. However, loss of genetic variances present in the founders may be due to directional selection. These results suggest that once adaptation has occurred, and on the short-term of tens of generations, stasis of locomotion behaviour is repeatable because of effective stabilizing selection at a large phenotypic scale, while the genetic structuring of component traits is contingent upon drift history at a local phenotypic scale.

2 Introduction

Stasis, the lack of directional change in the average values of a trait over time, is the most common phenotypic pattern observed over timespans reaching one million years (Arnold, 2014; Gingerich, 2019; Uyeda et al., 2011). Theory predicts phenotypic stasis when stabilizing selection, or when directional and other forms of selection cancel out over the period examined, acts upon standing genetic variation reflecting the phenotypic effects of mutational input (Charlesworth et al., 1982; Estes and Arnold, 2007; Hansen and Martins, 1996; Lande, 1986; Morrissey and Hadfield, 2012). When considering mutation-selection balance on the long-term (as scaled by the effective population sizes), theory has been successfully applied to explain, for example, fly wing evolution over a period of 40 million years (Houle et al., 2017), or nematode embryogenesis over 100 million years (Farhadifar et al., 2015). On the short-term of a few tens to hundreds of generations, however, many natural populations depend on standing genetic variation for adaptation or rescue from extinction, when mutation should be of little influence and founder effects, demographic stochasticity and genetic drift important (Chelo et al., 2013; Hill, 1982; Mallard et al., 2022b; Matuszewski et al., 2015).

On the short-term, before mutation-selection balance is reached, phenotypic stasis in extant natural populations is also commonly observed, often despite significant trait heritability and selection (Merilä et al., 2001; Pujol et al., 2018). Explanations for short-term phenotypic stasis have relied on showing that in many cases there were no changes in the breeding traits' values, that is, no genetic divergence, either because of selection on unmeasured traits that are genetically correlated with observed ones or because of correlated selection due to unknown environmental covariation between observed and unobserved traits with fitness e.g., (Czorlich et al., 2022; Kruuk et al., 2002), both instances of "indirect" selection. Short-term phenotypic stasis without genetic divergence has also been explained by phenotypic plasticity allowing the tracking of environmental fluctuations e.g., (Biquet et al., 2022; de Villemereuil et al., 2020). These studies indicate that phenotypic evolution cannot be understood when considering each trait independently of

others and that a multivariate description of selection and standing genetic variation is needed. Selection on multiple traits should be seen as a surface with potentially several orthogonal dimensions (Phillips and Arnold, 1989), each with particular gradients depicting selection strength and direction on each trait and between traits (Arnold et al., 2001; Lande and Arnold, 1983). Responses to selection in turn will depend on the size and shape of the **G**-matrix, the additive genetic variance-covariance matrix of multiple traits (Lande, 1979). For example, phenotypic dimensions with more genetic variation are expected facilitate adaptation, as selection will be more efficient (Lande, 1976, 1979; Schluter, 1996), even if indirect selection can confound predictions about phenotypic evolution (Mallard et al., 2022a; Morrissey and Bonnet, 2019; Stinchcombe et al., 2014).

The extent to which phenotypic stasis is compatible with the expected divergence of the **G** matrix in the short-term remains little unexplored cf. (Bohren et al., 1966; Gromko, 1995; Simões et al., 2019; Teotónio et al., 2004; Teotónio and Rose, 2000). Studies in natural populations cannot usually control environmental variation and estimates of **G** matrix dynamics are nearly impossible to obtain, while experiments employing truncation selection do not easily model the complexity of the selection surface. Under drift, and assuming an infinitesimal model of trait inheritance, the **G** matrix size (i.e., the total genetic variance) is reduced and diverges from ancestral states by a factor proportional to the effective population size (Lande, 1976; Lynch and Hill, 1986; Phillips et al., 2001). However, theory that includes the effects of finite population sizes, multivariate selection, and the pleiotropic effects of mutation remains out of reach for changes in genetic covariances between traits and thus **G** matrix shape (Barton and Turelli, 1987; Burger, 2000; Lande, 1980; Lynch and Walsh, 1998; Simons et al., 2018). We do expect, however, that once most adaptation has occurred divergence of the **G** matrix shape is caused by drift, and also know that different forms of selection might lead to further genetic divergence in the relatively local phenotypic space occupied after adaptation (Doroszuk et al., 2008; Haller and Hendry, 2014). Whether or not genetic divergence will also lead to phenotypic divergence should then depend on the distribution of pleiotropic effects of quantitative trait loci (QTL) alleles, and

linkage disequilibrium between them, created by past selection and drift, and ultimately on the developmental and physiological mapping of genetic onto phenotypic variation (Chebib and Guillaume, 2017; Hansen and Wagner, 2001; Morrissey, 2015; Riska, 1989).

Here we analyse the evolution of locomotion behavior on the hermaphroditic nematode *Caenorhabditis elegans*, spanning 240 generations of lab evolution in a constant and homogeneous environment, thus maximizing the chances of imposing and detecting stabilizing selection. We could obtain an accurate characterization of the fitness effects of component trait variation of locomotion behavior (transition rates between movement states and direction), by measuring essentially all individuals at the time of reproduction. We characterized the evolution of the broad-sense **G** matrix for hermaphrodite locomotion behavior, obtained by phenotyping inbred lines derived from the domesticated ancestral population at generation 140, and from three replicate populations during further 50 and 100 generations in the same environment. After domestication, selection gradients were estimated by regressing fertility onto transition rates. We seek to find if short-term evolution of the **G** matrix follows the directions of selection, or if there is loss of genetic variance just by drift, and to determine how genetic divergence is compatible with phenotypic stasis once most adaptation has been achieved.

3 Methods

3.1 Archiving

Data, R code scripts, and modeling results (including **G** matrix estimates) can be found in our github repository, and will be be archived in *Dryad.org* upon publication.

3.2 Laboratory culture

We analyzed the lab evolution of locomotion behavior during 273 generations (Figure 1A), the first 223 of which have been previously detailed (Noble et al., 2017; Teotónio et al., 2012; Theologidis et al., 2014). Briefly, 16 inbred founders were intercrossed in a 33 generation funnel to

obtain a single hybrid population (named A0), from which six population replicates (A[1-6]) were domesticated for 100-140 generations. Based on the evolution of several life-history traits such as hermaprodite self and outcross fertility, male mating ability or viability until reproduction we have previously shown that most adaptation to lab conditions had occurred by generation 100 (Carvalho et al., 2014a,b; Poullet et al., 2016; Teotónio et al., 2012; Theologidis et al., 2014). From population A6 at generation 140 (A6140) we derived 6 replicate populations and maintained them in the same environment for another 100 generations (CA[1-6]). Inbred lines were generated by selfing hermaphrodites from A6140 (for at least 10 generations), and from CA populations 1-3 at generation 50 and 100 (CA[1-3]50 and CA[1-3]100; Noble et al. (2019)). We refer to these last 100 generations as the focal stage. During the domestication and focal stages, populations were cultured at constant census sizes of $N = 10^4$ and expected average effective population sizes of $N_e = 10^3$ (Chelo et al., 2013; Chelo and Teotónio, 2013). Non-overlapping 4-day life-cycles were defined by extraction of embryos from plates, followed by seeding of starvation-synchronised L1 larvae to fresh food (Teotónio et al., 2012). Periodic storage of samples ($> 10^3$ individuals) were done by freezing (Stiernagle, 1999). Revival of ancestral and derived population samples allows us to control for transgenerational environmental effects under "common garden" phenotypic assays (Teotónio et al., 2017).

3.3 *Worm tracking assays*

3.3.1 *Sampling and design*

Population samples were thawed from frozen stocks on 9cm Petri dishes and grown until exhaustion of food (*Escherichia coli* HT115). This occurred 2-3 generations after thawing, after which individuals were washed from plates in M9 buffer. Adults were removed by centrifugation, and three plates per line were seeded with around 1000 larvae. Samples were maintained for one to two complete generations in the controlled environment of lab evolution. At the assay generation (generation 4-6 generations post-thaw), adults were phenotyped for locomotion behaviour

at their usual time of reproduction during lab evolution (72h post L1 stage seeding) in single 9cm plates. At the beginning of each assay we measured ambient temperature and humidity in the imaging room to control for their effects on locomotion.

Inbred lines from the experimental populations were phenotyped over 7 years in two different lab locations (Lisbon and Paris) by three different experimenters. In total, there were 197 independent thaws, each defining a statistical block containing 2-22 samples. 188 inbred lines from the A6140 population were phenotyped, with 52 CA150, 52 CA250, 51 CA350, 51 CA1100, 53 CA2100 and 68 from CA3100. Each line was phenotyped in at least two blocks (technical replicates). CA[1-3]50 and CA[1-3]100 lines were phenotyped within a year. A6140 lines were phenotyped over 2 consecutive years. We further phenotyped the outbred populations and the 16 founders. For these there were 9 independent thaws, of which 5 also contained founders. All founders and populations were phenotyped twice with the exception of A6140, included in six blocks.

In order to improve estimation of the selection surface in our lab evolution environment (see below), we also assayed locomotion bias in 56 inbred lines derived from populations evolved in a high-salt environment (GA[1,2,4]50) for which fertility data was available (Noble et al., 2017). These lines were phenotyped in the same blocks as the A6140 lines included in the gamma matrix analysis (Lisbon, single experimenter). Removing these lines from the analysis did not affect the mode of the posterior distribution estimates of our coefficients, and only led to the loss of statistical power reflected by wider credible intervals (analysis not shown).

3.3.2 Imaging

To measure locomotion behavior we imaged adults 72h post-L1 seeding using the Multi-Worm Tracker [MWT version 1.3.0; Swierczek et al. (2011)]. Movies were obtained with a Dalsa Falcon 4M30 CCD camera and National Instruments PCIe-1427 CameraLink card, imaging through a 0.13-0.16 mm cover glass placed in the plate lid, illuminated by a Schott A08926 backlight. Plates were imaged for approximately 20-25 minutes, with default MWT acquisition parameters.

Choreography was used to filter and extract the number and persistence of tracked objects, and assign movement states across consecutive frames as forward, still or backwards, assuming that the dominant direction of movement in each track is forward (Swierczek et al., 2011).

MWT detects and loses objects over time as individual worms enter and leave the field of view or collide with each other. Each individual track is a period of continuous observation for a single object (the mapping between individual worms and tracks is not 1:1). We ignored the first 5 minutes of recording, as worms are perturbed by plate handling. Each movie contains around 1000 tracks with a mean duration of about 1 minute. The MWT directly exports measurements at a frequency that can vary over time (depending on tracked object density and computer resource availability), so data were standardized by subsampling to a common frame rate of 4 Hz. Worm density, taken as the mean number of tracks recorded at each time point averaged over the total movie duration, was used as a covariate in the estimation of genetic variance-covariances below.

3.3.3 Differentiating males from hermaphrodites

A6140 and all CA populations are androdioecious, with hermaphrodites and males segregating at intermediate frequencies (Teotónio et al., 2012; Theologidis et al., 2014). We were able to reliably (97% accuracy) differentiate between the sexes based on behavioural and morphological traits extracted from MWT data.

We first evaluated a set of simple descriptions of individual size, shape and movement to find a subset of metrics that maximized the difference in preference for a two-component model between negative and positive controls: respectively, inbred founders and two monoecious (M) populations which contained no, or very few, males; and three dioecious (D) populations with approximately 50% males [M and D populations were derived from A6140, see Theologidis et al. (2014) and Guzella et al. (2018)]. Starting with worm area, length, width, curvature, velocity, acceleration, and movement run length as parent traits from Choreography output, derived descendant traits were defined by first splitting parents by individual movement state (forward, backward, still) and calculating the median and variance of the distribution for each track. Traits

with more than 1% missing data were excluded, and values were log-transformed where strongly non-normal (a difference in Shapiro-Wilk $-log_{10}(p) > 10$). Fixed effects of block and log plate density were removed by linear regression before model fitting. Two-component Gaussian mixture models were fit to tracks for each line/population (R package *mclust* Scrucca et al. (2016), VII spherical model with varying volume), orienting labels by area (assuming males are smaller than hermaphrodites). We sampled over sets of three traits (requiring three different parent trait classes, at least one related to size), and took the set maximizing the difference in median Integrated Complete-data Likelihood (ICL) between control groups (log area, log width, and velocity, all in the forward state). By this ranking, the 16 inbred founders and two monoecious populations fell within the lower 19 samples (of 77), while the three dioecious populations fell within the top 15 samples.

To build a more sensitive classifier robust to male variation beyond the range seen in control data, we then trained an extreme gradient boosting model using the full set of 30 derived traits on the top/bottom 20 samples ranked by ICL in the three-trait mixture model [R package *xgboost*, Chen and Guestrin (2016)]. Negative control samples were assumed to be 100% hermaphrodite, while tracks in positive controls were assigned based on mclust model prediction (excluding those with classification uncertainty in the top decile). Tracks were classified by logistic regression, weighting samples inversely by size, with the best cross-validated model achieving an area under the precision-recall curve of 99.75% and a test classification error of 3.1% ($max_depth = 4, eta = 0.3, subsample = 0.8, eval_metric = "error"$). Prediction probabilities were discretized at 0.5.

Males tend to move much faster than hermaphrodites (Lipton et al., 2004), and since individual collision leads to loss of tracking sex is strongly confounded with track length and number. To estimate male frequencies at the sample level, tracks were sampled at 1s slices every 30s over each movie in the interval 400-1200 seconds, and line/population estimates were obtained from a binomial generalized linear model (Venables and Ripley, 2002). Estimates appear to saturate at around 45%, due presumably to density-dependent aggregation of multiple males attempting to

copulate.

3.4 Locomotion behavior

3.4.1 Definition of transition rates

In a one-dimensional space individual locomotion behavior can be described by the transitions rates of activity and direction. We modelled the expected sex-specific transition rates between forward, still and backward movement states with a continuous time Markov process. We consider a system having $d = 3$ states with $P(t_1, t_2) \in \Re^{d,d}$, $t_2 > t_1$, denoting the transition probability matrix (Jackson, 2011; Kalbfleisch and Lawless, 1985):

$$p_{i,j}(t_1, t_2) = \mathbb{P}\left[s(t_2) = j \mid s(t_1) = i\right] \tag{1}$$

where $s(t) \in \mathcal{S}$, with $\mathcal{S} = \{\underline{s}till, \underline{f}orward, \underline{b}ackward\}$ being the movement state occupied in instant t. We consider a time-homogeneous process described by the transition rate matrix:

$$Q = \begin{bmatrix} -q_s & q_{s,f} & q_{s,b} \\ q_{f,s} & -q_f & q_{f,b} \\ q_{b,s} & q_{b,f} & -q_b \end{bmatrix} \tag{2}$$

where $q_{i,j} \geq 0 \, \forall i, j$, subject to the constraint:

$$q_i = \sum_{j \neq i} q_{i,j} \tag{3}$$

Hence, six of the nine possible transitions are independent. Let θ denote the parameters to be estimated, containing the off-diagonal elements from equation 2:

$$\theta = [q_{s,f}, q_{s,b}, q_{f,s}, q_{f,b}, q_{b,s}, q_{b,f}] \tag{4}$$

In this model, the time an object remains in a given state is on average $1/q_i$. Since the process is stationary, the probability of transition is a function of the time difference $\Delta t = t_2 - t_1$, such that $P(t_1, t_2) = P(\Delta t)$, and the elements of the $P(\Delta t)$ matrix:

$$p_{i,j}(\Delta t) = \mathbb{P}\left[S(\Delta t) = j \mid S(0) = i\right] \tag{5}$$

It then follows that:

$$P(\Delta t) = exp(\Delta t\, Q) \tag{6}$$

where exp(·) denotes the matrix exponential. The constraint in equation 3 ensures that:

$$\boldsymbol{P}(\infty) = \begin{bmatrix} f_s & f_f & f_b \\ f_s & f_f & f_b \\ f_s & f_f & f_b \end{bmatrix} \tag{7}$$

where f_i is the relative frequency of state i that no longer depends on the previous state (all three rows of the $\boldsymbol{P}(\infty)$ matrix converge). We find that the state frequencies from $\boldsymbol{P}(\boldsymbol{\infty})$ are a monotonic and mostly linear function of the observed frequencies of movement states (Figure S3), showing that violations of the Markov assumption of the model does not induce a large bias in the long-term predictions of our model.

3.4.2 *Estimation of transition rates*

To estimate transition rates, we have N objects (individual tracks) from each technical replicate (Petri plate), with the data on the k-th object denoted as:

$$\mathcal{D}_k = (x_{k,1}, x_{k,2}, \ldots, x_{k,n_k-1}) \tag{8}$$

$$x_{k,l} = (s_{k,l}, s_{k,l+1}, \Delta t_{k,l}),\ \Delta t_{k,l} = t_{k,l+1} - t_{k,l} > 0 \tag{9}$$

where $s_{k,l}$ is the state of the k-th object in the l-th time-point in which it was observed, and $t_{k,l}$ is the instant of time in which this observation was made. Then, given data $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_N\}$, the log-likelihood for the model for analysis is (Bladt and Sorensen, 2005; Kalbfleisch and Lawless, 1985):

$$\mathcal{L}(\theta \mid \mathcal{D}) = \sum_{k=1}^{N} \sum_{l=1}^{n_k-1} \ln(p_{i,j}(\Delta t)|_{i=s_{k,l}, j=s_{k,l+1}, \Delta t=\Delta t_{k,l}}) \tag{10}$$

where $p_{i,j}(\Delta t)$ was defined in equation 5, is calculated as a function of the parameters θ via equation 4. Therefore, the data on the N objects can be represented as the number of observations of $x = (i, j, \Delta t)$, which we denote as $\widetilde{n}_{i,j,\Delta t}$:

$$\widetilde{n}_{i,j,\Delta t} = \sum_{k=1}^{N} \sum_{l=1}^{n_k - 1} \mathbb{I}_{i,j,\Delta t}\left[s_{k,l}, s_{k,l+1}, \Delta t_{k,l}\right] \tag{11}$$

and where $\mathbb{I}_{i,j,\Delta t}[\cdot]$ is the indicator function:

$$\mathbb{I}_{i,j,\Delta t}\left[s_1, s_2, \delta t\right] = \begin{cases} 1, & \text{if } s_1 = i, s_2 = j \text{ and } \delta t = \Delta t \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

The input data can then be compressed by considering only the data:

$$\mathcal{Z} = \{z_1, z_2, \ldots, z_M\} \tag{13}$$

$$z_k = \left(\Delta t_k, \widetilde{N}_k\right),\ \Delta t_k \in \Re^+, \widetilde{N}_k \in \mathbb{N}_0^{d,d} \tag{14}$$

$$\widetilde{N}_k = \widetilde{n}_{i,j,\Delta t_k} \tag{15}$$

The log-likelihood to estimate transition rates can be finally rewritten as:

$$\mathcal{L}(\theta \mid \mathcal{Z}) = \sum_{k=1}^{m} \vec{1}_d^T \left(\widetilde{N}_k \odot \ln(P_k)\right) \vec{1}_d \tag{16}$$

where $\vec{1}_d$ is a d-dimensional vector of 1s, $\odot$ denotes the Hadamard product, and $\ln P_k$ is the matrix obtained by taking the logarithm of each value in matrix P_k.

These models were specified using RStan (Stan Development Team (2018), R version 3.3.2, RStan version 2.15.1), which performs Bayesian inference using a Hamiltonian Monte Carlo sampling to calculate the posterior probability of the parameters given the observed data. We used multi-log normal prior distributions with mean transition rate and a coefficient of variation: $\ln(q_{i,j}) \sim \mathcal{N}(\ln(2), 0.6)$.

Throughout, we denote non-self transition rates q_k the six off-diagonal elements of the $\mathbf{Q}$ matrix estimated by the above model.

3.4.3 *Male and inbreeding effects*

Using the transition rates measured in populations and inbred lines, we fit a series of linear mixed-effects models to test for phenotypic evolution in the outbred populations (see equation 19), for effects of male frequency on hermaphrodite transition rates in the outbred populations (equation 20), and for inbreeding effects in the inbred lines (equation 21). Given sparse temporal sampling, we make the conservative assumption of independence of observations within domestication and focal stages. For transition rate q_k:

$$\ln(q_k) = \alpha + \beta_{gen} G + \gamma_{anc} t + \delta_{anc} + \zeta_b + \epsilon \tag{17}$$

with α the trait mean, β_{gen} a fixed effect of generation number t, γ_{anc} and δ_{anc} random effects accounting for intercept and slope differences between the domestication and focal periods of lab evolution (both$\sim \mathcal{N}(0, \sigma^2)$), $\zeta \sim \mathcal{N}(0, \sigma^2)$ a random effect of block b and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ the residual error.

$$\ln(q_k) = \alpha + \beta F + \gamma_{popId} + \zeta_b + \epsilon \tag{18}$$

with β a fixed effect of male frequency F, $\gamma_{popId} \sim \mathcal{N}(0, \sigma^2)$ a random effect accounting for differences between populations, $\zeta \sim \mathcal{N}(0, \sigma^2)$ a random effect of block b, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ the residual error.

As we estimate the **G**-matrix from the line differences (see next section), it is likely that it does not reflect the true additive genetic (co)variance matrix (**G**-matrix) unless the mean trait values among lines are similar to the mean trait values of the outbred population from which the lines were derived (Kearsey and Pooni, 1996; Lynch and Walsh, 1998). Because the lines and the populations were phenotyped at different times, we included environmental covariates:

$$\ln(q_k) = \alpha + \beta + T * H * D + \gamma + \delta_{lineID} + \epsilon \tag{19}$$

where environmental covariates: temperature (**T**), relative humidity (**H**) and density (**D**) are fitted as fixed effects. β is a two-level categorical fixed effect (inbred lines or population). γ is a two-level categorical fixed effect accounting for differences between the years of phenotyping measurements of the A6140 lineages. $\delta \sim \mathcal{N}(0, \sigma^2)$ a random effect accounting for line identity within populations and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ the residual error.

Both male and inbreeding models were fit using the *lmer* function in R package *lme*4, and non-zero values of fixed effects were tested against null models without fixed effects with likelihood ratio tests. Marginal r^2 for the male frequencies were computed using the *r.squaredGLMM* function of the package MuMIn (Bartoń, 2020).

3.5 Transition rate genetics

3.5.1 *G-matrix estimation*

Genetic (co)variances of transition rates per population are estimated as half the between inbred line differences for lines separately derived from the evolving outbred populations. In the absence of selection during inbreeding and cancelling of directional non-additive gene action, this broad-sense **G**-matrix obtained from inbred lines is an adequate surrogate for the additive **G**-matrix of outbreeding populations (Kearsey and Pooni, 1996; Lynch and Walsh, 1998). We test these assumptions (see below).

G-matrices for the six non-self transition rates q_k were estimated from trait values for the inbred lines derived from focal populations. We estimated **G**-matrices separately for each of the seven populations (A6140, CA[1-3]50, CA[1-3]100). The 6 transition rates q_k were fitted as a multivariate response variable y in the model:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{T} * \boldsymbol{H} * \boldsymbol{D} + \boldsymbol{L} + \boldsymbol{B} + \boldsymbol{e}. \tag{20}$$

where the intercept ($\boldsymbol{\mu}$) and the environmental covariates: temperature (**T**), relative humidity (**H**) and density (**D**) were fitted as fixed effects. Environmental covariates were fitted individually

and with all possible interactions. Each covariate was standardized to a mean of 0 and standard deviation of 1. Block effects (**B**) and line identities (**L**) were modeled as random effects and **e** was the residual variance. We then estimated a matrix of genetic (co)variance as half the line covariance matrix (**L**). An additional two-level categorical effect was included when estimating the A6140 matrix that accounts for differences between the 2012 and 2013 phenotyping blocks.

Models were fit with the R package *MCMCglmm* (Hadfield, 2010). We constructed priors as the matrix of phenotypic variances for each trait. Model convergence was verified by visual inspection of the posterior distributions and by ensuring that the autocorrelation remained below 0.05. We used 100,000 burn-in iterations, a thinning interval of 2,000 and a total of 2,100,000 MCMC iterations.

*3.5.2 **G**-matrices under random sampling*

For each of our seven populations (A6140, CA[1-3]50, CA[1-3]100), we constructed 1000 randomised **G**-matrices to generate a null distribution against which to compare the observed estimates. We randomly shuffled both the inbred line and block identities and fit the equation 20. We then computed the posterior means of our 1,000 models to construct a null distribution.

We additionally generated 1000 matrices for the A6140 population using the same procedure on random subsets of 60 (of 188 total) inbred lines to determine the effects of sampling the same number of lines as those for CA[1-3]50 and CA[1-3]100 populations.

*3.5.3 **G**-matrix divergence and differentiation*

To compare the overall variance of the **G**-matrices during experimental evolution, we first computed the trace of the matrices. We then performed spectral analyses on the posterior ancestral **G**. This decomposition of describes the overall shape of the multivariate variance for each matrix, with the relative genetic variance between the six eigenvalues of each eigenvector indicating whether the matrix is elliptical (a few large eigenvalues) or round (homogeneous eigenvalues); the first eigenvector (defined as g_{max}) is the linear combination of traits where the genetic variance

is maximized.

We used spectral analysis to explore differences between the seven **G**-matrices, following (Aguirre et al., 2014; Hine et al., 2009). Genetic (co)variance tensors ($\mathbf{\Sigma}$) are fourth-order objects describing how phenotypic dimensions between transitions rates maximize differences between all the **G**-matrices. The genetic variation among multiple **G**-matrices can be described by $\mathbf{\Sigma}$ decomposition into orthogonal eigentensors (E_i, with i being the orthogonal dimensions), each associated with an eigenvalue quantifying its contribution to variation in $\mathbf{\Sigma}$ ($\boldsymbol{\alpha_i}$). In turn, eigentensors can be decomposed into eigenvectors (e_{ii}) each with associated eigenvalues (λ_i). Aguirre et al. (2014) implemented this approach in a Bayesian framework using *MCMCglmm*, and Morrissey and Bonnet (2019) made an important modification to account for sampling where the amount of variance in $\boldsymbol{\alpha_i}$ is compared to an expected distribution under sampling a finite number of lines.

3.6 Selection on transition rates

3.6.1 Selection surface

The log-transformed, covariate-adjusted fertility values (best linear unbiased estimates) for each inbred line were downloaded from Noble et al. (2017), exponentiated, and divided by the mean to obtain a relative fitness measure (w_l). See Results and Discussion for a justification of fertility as a fitness proxy.

Since we did not observe any directional change in locomotion behavior or component transition rates during lab evolution, and because the inbred lines were derived after domestication, most of adaptation to the lab environments has occurred and we do not expect linear (directional) selection to be significant. We estimated quadratic selection gradients using partial regression, following (Lande and Arnold, 1983), with the *MCMCglmm* R package:

$$w_l = \alpha + \sum_{k=1}^{6} \gamma_k z_{k,l}^2 + \sum_{k_1=1}^{5} \sum_{k_2=k_1+1}^{6} \gamma_{k_1,k_2} z_{k_1,l} z_{k_2,l} + \epsilon \tag{21}$$

with α being the mean relative fitness among all lines and γ the partial coefficients estimating quadratic selection on each transition rate k, or between pairs of transition rates k_1 and k_2. Environmental covariates (temperature, humidity, density) were defined and normalized as for the **G**-matrices estimation described above. Model residuals were normal and homocedastic (not shown).

We compared the results of this model (equation 21) with those of linear mixed effect models including as a random effect the additive genetic similarity matrix **A** between inbred lines, as defined in Noble et al. (2017) and Noble et al. (2019). We have also compared results from equation 21 to models including coefficients for linear selection on each transition rate. Under both circumstances parameter estimates are similar to those presented, albeit with changing credible intervals (not shown). Including other measured traits by the worm tracker, such as body size [a trait related to developmental time that is known to affect fertility in our populations (Theologidis et al., 2014)] similarly does not affect the qualitative conclusions we reach.

*3.6.2 **G**-matrix alignment with the selection surface*

We used canonical analysis (Phillips and Arnold, 1989) to visualize the selection surface as:

$$\Lambda = \boldsymbol{U}^T \gamma \boldsymbol{U} \tag{22}$$

with **U** being the matrix of eigenvectors of γ, and Λ the diagonal matrix of eigenvalues (denoted $\lambda_{[1-6]}$). **G** matrices were rotated to visualize them as:

$$\boldsymbol{G}' = \boldsymbol{U}^T \boldsymbol{G} \boldsymbol{U} \tag{23}$$

To sample a null distribution of the γ eigenvalues along the rotated dimensions, we fit the same model after permuting the relative fitness values of the lines. We then extracted the diagonal elements of these permuted γ after rotation using the estimated **U**.

To see the evolution of the **G** matrix in the selection surface, we calculated the Pearson product moment correlations between the eigentensor vectors explaining most of the genetic differences

between the 7 matrices (e_{11}, e_{12}) with the canonical selection dimensions (y_1-y_6). We estimated uncertainty in these values by sampling from the posterior distribution of γ 1000 times.

3.7 *Inference of effects*

Most of our analysis relies on Bayesian inference of genetic or phenotypic effects. As discussed in Walter et al. (2018), the "significance" of effects can be inferred when there is no overlap between the posterior null sampling distributions with the posterior empirical estimate of the expected values. Thus, we compare expected value estimates such as a mean or mode with the 95% credible intervals under random sampling of the expected value. However, when we compare with each other empirical posterior distributions, e.g. genetic (co)variance estimates or null distributions, we follow Austin and Hux (2002) and infer "significant" differences between them when their 80% credible intervals do not overlap (strictly 83%).

4 Results

4.1 *Laboratory culture*

Our lab evolution system is based on a hybrid population derived from 16 founder strains (Figure 1A). Replicate samples from the hybrid population were domesticated for 140 non-overlapping generations at census size N=10^4 to an environment in part characterized constant density, temperature and relative humidity, and little spatial structure during the life-cycle (see Methods). The dynamics of several life-history traits during domestication indicate that most adaptation to lab conditions occurred by generation 100 (Carvalho et al., 2014a,b; Poullet et al., 2016; Teotónio et al., 2012; Theologidis et al., 2014). From a single domesticated population we derived replicate populations and evolved them for another 100 generations in the same environmental conditions. Although we measured locomotion behavior throughout of lab evolution, we only follow the **G** matrix of its component traits during the last 100 generations, after adaptation, a stage that we call here the focal stage of lab evolution (Figure 1A).

C. elegans reproduces mostly by selfing in nature though there is considerable variance in male mating performance among the founders (Teotónio et al., 2006). By training a model on a suite of size- and locomotion-related metrics, we found that hermaphrodites could be clearly differentiated from males (see Methods), and estimated males frequencies were high during the entire experiment (Figure S1). Because *C. elegans* are androdioecious, and hermaphrodites cannot mate with each other, average expected selfing rates at a generation are 1 minus twice the male frequency at the previous generation (Teotónio et al., 2012), and we can conclude that outcrossing was the predominant reproduction mode during lab evolution. Previously, we showed that effective population sizes during domestication were of about N_e=10^3 (Chelo and Teotónio, 2013).

4.2 Evolution of locomotion behavior

We measured locomotion behavior at the time of reproduction for each outbred population and the inbred founders using worm video tracking (Swierczek et al., 2011). The output, after quality control and initial analysis, are individual worm tracks categorized at a given point in time by activity (moving, or not) and direction (forwards or backwards). We model a three-state memoryless (Markov) process with homogeneous spatial and temporal dynamics (see Methods, Figure S2). We view this as an obviously false but useful approximation of worm locomotion behavior under our conditions, which is only partially violated (worms tend to resume forward movement more often than expected; Figure S3). Component traits of locomotion behavior are the (sex-specific) six non-self transition rates between forward movement, backward movement, and immobility.

We find that while the founders of lab evolution show great diversity in locomotion behavior under lab conditions, evolved populations rapidly attained, and maintained, a stable level after hybridization for 240 generations. For example, considering the proportion of time individual worms are stationary (Figure 1B), we observe values of around 40% for hermaphrodites - much higher than most founders - while males are much more vagile (stationary around 10%). Neither hermaphrodite nor male transition rates showed directional change from the hybrid ancestral

state over the full 240 generation period (Table 1, Supplementary Figures S4 and S5). Differences between replicate populations can be explained by sampling error.

4.3 Broad-sense **G** matrix

To estimate **G** matrices, we used approximately 200 lines from the generation 140 domesticated population (A6140), and approximately 50 lines from each of three replicate populations derived from A6140 and sampled at generations 50 (CA[1-3]50) and 100 (CA[1-3]100) of the focal lab evolution. We use these broad-sense **G**-matrices as a surrogate for the narrow-sense (additive) **G**-matrices of the outbred populations (see Methods). These two kinds of matrices might not be identical because of selection during inbreeding or because of differential expression of non-additive genetic effects in inbred and outbred individuals. Such differences, if present, manifest as differences in the mean values of inbred and outbred samples as directional effects will statistically average out for polygenic traits (Kearsey and Pooni, 1996; Lynch and Walsh, 1998). We used the inbred lines and the focal A6140 ancestor to compare means for all transitions and we did not find any evidence of directional non-additive genetic effects (Table 2).

Our **G** matrices could also differ from the **G** matrices of outbred populations due to the absence of males in the inbred lines; which were abundant in the outbred populations. This is because males are known to disturb hermaphrodite locomotion behavior (Lipton et al., 2004). We tested for effects of male frequency on transition rates in outbred populations with univariate linear models and found that they were weak at best (Figure S6).

4.4 **G** matrix evolution

For the domesticated 140 population (A6140), ancestral to all CA populations during further 100 generations in the same environment after adaptation, there is significant genetic variance in all hermaphrodite transition rates, relative to a null distribution from permutations of line and technical replicate identity (Figure 2A). Likewise, the posterior distributions of most (12 of 15)

covariance estimates between transition rates do not overlap 0, and differ from the null distribution of posterior means. The A6140 **G**-matrix is structured in two main behavioral modules, with the transitions from still to forward or backward (i.e. leaving the still state) showing positive covariances with each other and negative covariances with other transition rates.

When looking at the evolved CA populations, we see that their **G**-matrices are reduced, particularly after 100 generations of evolution (Figure S7). The **G** matrices of CA populations after 100 generations do not differ from the permutation null. Reduced genetic (co)variance in generation 50 and 100 is particularly obvious when calculating the trace of the **G** matrices, although all populations contain more genetic variance than expected under sampling (Figure 2B). The loss of genetic (co)variances during focal evolution could be due to differences in statistical power or the result of continued lab evolution. Sub-sampling A6140 to the sampling sizes of CA[1-3]50 and CA[1-3]100 populations, while increasing the credible intervals did not affect the estimated modes, with many of them remaining different from the null (Figure S8).

Spectral decomposition of the A6140 **G** matrix further shows that, for the phenotypic dimension encompassing most genetic variation in this population (g_{max}), the projected variance of CA populations in this dimension is much reduced (Figure 2C). In this analysis we only find a difference between empirical variances with null expectations for A6140 and CA populations at generation 50, not CA populations at generation 100 .

4.5 Genetic divergence and differentiation

We tested for divergence of the **G**-matrices from the ancestral state, and for differentiation between derived replicate populations, using spectral analysis and comparing the seven matrices simultaneously (see Methods). This analysis identifies the phenotypic dimensions along which there are most differences between matrices. The first two eigentensors, $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$, explain more variation than the null expectation (Figure 3A), 54% and 21%, respectively.

G matrix coordinates in the space of $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$ (Figure 3B), show that the A6140 population drives most significant differences between all matrices, and thus encompasses most of the ge-

netic divergence. Along the first eigenvector of $\boldsymbol{E_1}$ (called e_{11}; Figure 3C) divergence is due to loss of genetic variance in the CA populations. We further find that e_{11} is highly collinear with the g_{max} of the A6140 population, the phenotypic dimension encompassing most ancestral genetic variation (not shown).

The second eigentensor $\boldsymbol{E_2}$ suggests differentiation between replicate populations at generation 50 (Figure 3C) and thus we further tested for differentiation between replicate populations during focal evolution by restricting the spectral analysis to only the three CA[1-3]50 **G**-matrices. We observe that a single eigentensor was different from the null expectations, explaining 53% of the differences between the matrices (Figure S9). The coordinates of these matrices in the space of the eigentensor indicate that CA150 and the remaining two populations contributed in opposite directions to the difference observed. Most of this difference is expressed along the first two eigenvectors (50% and 37%): CA[2-3]50 lost variance along the first eigenvector and CA150 along the second one.

4.6 Selection on locomotion behavior

In Noble et al. (2017) we reported the fertility of many of the inbred lines used to estimate the **G** matrices. This data encompasses hermaphrodite self-fecundity and progeny viability until early larvae, measured in an environment that closely mimicked that of lab evolution. With this data at hand we can estimate the selection surface of locomotion in our lab environment by applying equation 21, with relative fertility being partially-regressed onto the transition rates (see Methods).

We find that the 95% credible intervals for several coefficients for correlated selection between pairs of transition rates do not overlap zero: negative between still-forward (SF) and forward-still (FS) and positive between SB and FS, and FS and BS (Figure S10). To visualize the selection surface, we rotated the γ-matrix with canonical analysis (see Methods). The resulting selection surface suggests a saddle with three unstable equilibria in three canonical dimensions y_1-y_3, indicating disruptive selection, and three stable equilibria in three dimensions (y_4-y_6), indicating

stabilizing selection (Figure 4). We only find, however, evidence of weak and strong stabilizing selection on y_5 and y_6, respectively, because only these empirical estimates are unlikely under the null distribution.

4.7 *G matrix evolution in the selection surface*

Projection of the **G**-matrices onto the canonical selection dimensions shows that most genetic variance is concentrated in dimensions under neutrality (y_1-y_4), while the dimensions under stabilizing selection (y_5 and y_6) do not show much genetic variance (Figure 5). Loss of genetic variance along y_2-y_5 is clearly consistent with drift when assuming an infinitesimal model of trait inheritance (Barton et al., 2017) and effective population sizes of $N_e = 10^3$ (Chelo and Teotónio, 2013). For neutral y_1 and selected y_5 and y_6 dimensions it appears that replicate populations maintain more variation than expected under drift during the focal stage of lab evolution, and when compared with the founders before their hybridization.

To assess if **G**-matrix evolution aligned with the selection surface, we calculated the correlation between the directions of genetic divergence and differentiation (eigenvectors $e11$, $e12$; Figure 3) and the canonical selection dimensions (Figure 4). There is no alignment between the **G** matrix with the selection surface as no correlations were detected, except perhaps between y_3 and $e11$, and between y_5 and $e12$ (Figure S11).

5 Discussion

The evolution of *C. elegans* locomotion behavior during 240 generations in a fairly constant and homogeneous lab environment is characterized by stasis, following a genetically and phenotypically dynamic 33 generation period of hybridizing the founder strains. Most of the genetic variance along the several phenotypic dimensions under stabilizing selection or drift was lost, suggesting directional selection during hybridization of founders and domestication until generation 140. Despite phenotypic stasis, genetic divergence and differentiation continued during

further 50 to 100 generations of evolution in the same environment, a result that is sufficiently explained by drift. Because we previously showed that most adaptation had occurred by the start of the focal stage (Carvalho et al., 2014a,b; Poullet et al., 2016; Teotónio et al., 2012; Theologidis et al., 2014), our findings provide evidence for what Haller and Hendry (2014) termed "squashed" stabilizing selection: once populations adapt to their environment, most individuals will show close to optimal trait values and it becomes difficult to detect the effects of selection within the local phenotypic space occupied. Taken together, these findings suggest that directional selection outside a local phenotypic space after adaptation ensures phenotypic stasis, and that over a larger phenotypic space there was *effective* stabilizing selection.

The loss of genetic variance from the 16 founders to the domesticated population in the selection surface dimensions y_1, y_5 and y_6 is notable because it suggests that the rapid phenotypic evolution during intercrossing of founders to form the hybrid population was due to initially strong directional selection, which subsequently weakened during lab domestication. It can be argued that, with only 16 founders, we have little power to reject the hypothesis that there was no loss, and that the genetic (co)variances we found after domestication simply reflect natural standing genetic variation. At mutation-drift balance, the **G** matrix should reflect the patterns of mutational effects described by the **M** matrix, the equivalent measure of trait mutational variances, and covariances between them due to pleiotropy (Lande, 1979; Lynch and Hill, 1986). Elsewhere, we have estimated the **M** matrix in two of the founders of lab evolution, which show locomotion values divergent from those of lab evolution populations, by phenotyping a set of about 120 lines that accumulated mutations in a nearly-neutral fashion for 250 generations (Baer et al., 2005; Yeh et al., 2017). We found that the **M** matrices from these founders have similar sizes and are well aligned with each other, but not with the genetic **G** matrix of our A6140 domesticated population (Mallard et al., 2022b). Loss of genetic variances from the founders during hybridization and lab domestication was therefore at least partly due to directional selection. Future work should nonetheless try to understand if mutation-selection balance is responsible for the maintenance of genetic variation in locomotion behavior in nature by comparing **G** matrices

from natural populations, as they can be obtained from a large collection of wild isolates now available (Cook et al., 2017; Lee et al., 2021), with **M** matrices (Houle et al., 1996; Johnson and Barton, 2005).

Along the phenotypic dimensions of stabilizing selection (y_5 and y_6), more genetic variance might have been maintained than that expected under drift during the final 100 generations of lab evolution. If true, these findings are similar to those from many studies in nature and from artificially selected populations (Johnson and Barton, 2005; Turelli and Barton, 1994; Walsh and Blows, 2009): most quantitative traits have significant heritable genetic variation despite being under stabilizing selection. For example, when Sztepanacz and Blows (2017) applied six generations of disruptive truncation selection in *Drosophila serrata* to a multivariate dimension of maximal standing genetic variation (and low mutational variance) in male hydrocarbon traits, they observed a decrease of phenotypic variation, while disruptive selection in a dimension with little standing genetic variation (and high mutational variance, thus inferred to be under strong stabilizing selection in the base population), led to an increase in phenotypic variance. Assuming that there was little change in environmental variance during the experiment (Pelabon et al., 2010; Whitlock and Fowler, 1999), disruptive (stabilizing) selection is not only expected to increase (diminish) phenotypic variation but also to maintain (deplete) genetic variation relative to drift, though population genetic equilibria timescales differ between disruptive and stabilizing selection (Barton, 1990; Turelli, 1988; Vladar and Barton, 2014; Walsh and Lynch, 2018).

There are several explanations for potentially more than expected genetic variance with drift. Genetic variances in y_5 and y_6 are very low, and thus could be simply the result of measurement error and insufficient sampling. All three replicate populations showed similar posterior distributions, however, and no estimate overlaps zero. Another explanation is that we might have overestimated the strength selection or have biased estimates about the form of selection because environmental covariances with unmeasured traits could have caused correlated selection with transition rates (Blows and Brooks, 2003; Hunt et al., 2007; Mallard et al., 2022a). This seems an unlikely explanation. First, the strength of stabilizing selection on y_6 in particular is

of the same order of magnitude estimated in studies of natural populations (Johnson and Barton, 2005; Kingsolver et al., 2001). The regression equation 21 using standardized trait values returns an eigenvalue of -0.64 for y_6 when compared to the eigenvalue using unstandardized trait regression of about -10 shown in Figure 4. Second, stabilizing selection was estimated after domestication and adaptation to our lab environment and the regression approach employed is robust and unbiased to the inclusion of linear terms or other traits: when including directional selection coefficients or body size, a trait known to be genetically correlated with fertility (Noble et al., 2017; Theologidis et al., 2014), similar estimates are obtained (not shown).

A more serious concern is the role of unmeasured traits under potential selection that might be genetically correlated with the observed transition rates (Barton and Turelli, 1987; Lande, 1979; Mallard et al., 2022a; Shaw et al., 1995). Modelling additive genetic similarity among the inbred lines used in the regression (Noble et al., 2017, 2019) does not qualitatively change the inference about selection on transition rates (not shown), though we did not model the genetic architecture of locomotion behavior with multiple traits as the dependent variables. Several models have proposed that pleiotropic effects on unmeasured traits, which we did model, could explain maintenance of genetic variation in quantitative traits, though under weak selection and close to linkage equilibrium between QTL alleles in the long term of mutation-selection balance (Barton, 1990; Johnson and Porter, 2007; Simons et al., 2018; Zhang and Hill, 2005). While in our case transition rates between movement states should define the overall locomotion behavior of *C. elegans*, we cannot rule out that genetic covariation with unmeasured traits will better predict the evolution of genetic variances along the selected phenotypic dimensions.

Finally, potentially more genetic variance than expected by drift on y_5 and y_6 could also be explained by balancing selection generated by environmental or sex-specific effects (Johnson and Barton, 2005). In support of a role for balancing selection, we have previously shown that excess heterozygosity was maintained relative to that expected under drift and linked selection on deleterious recessives during the domestication stage of lab evolution (Chelo et al., 2019; Chelo and Teotónio, 2013). Further supporting balancing selection, QTL mapping of selection trait y_6,

using the sequence data of Noble et al. (2019) for a subset of the inbred lines used here, detects two QTL with high minor allele frequencies ($> 30\%$, unpublished analysis).

One of the major findings here is that of divergence and transient differentiation of the **G** matrix during the last focal 100 generations of lab evolution. The phenotypic dimensions of genetic divergence and differentiation among all populations were not obviously aligned with the phenotypic dimensions under selection, and most if not all of the genetic variance lost during this focal 100 generation period was expected with drift. Some of the phenotypic dimensions of genetic divergence were perhaps aligned with dimensions that were under weak stabilizing selection (y_5) or no selection (y_3), but there is no explanation for these observations, except as false positives, as drift is unknown to predictably change the shape of the **G** matrix (Phillips and McGuigan, 2006). In line with these observations are those of classic experiments in *D. melanogaster* by Fowler and colleagues where, after bottlenecking an outbred population, there was a reduction in the size of the **G** matrix for wing morphology in the derived bottlenecked populations, and size divergence among them, as expected under drift (Fowler and Whitlock, 1999; Phillips et al., 2001). Genetic divergence also occurred because the shape of the **G** matrix changed as derived populations showed different genetic covariances between traits. Unlike our finding of phenotypic stasis, however, variable drift history was consequential to the future phenotypic divergence of particular bottlenecked populations (Whitlock et al., 2002).

Most of our analyses and the underlying theoretical predictions are predicated on the assumption that the infinitesimal model of trait inheritance is a good approximation of the truth. However, that assumption may be violated, inasmuch as the genetic variances and covariances of locomotion behavior will not on the short-term of our hybridization and lab evolution be independent of allele frequency changes and linkage disequilibrium between smaller effect quantitative trait loci (QTL). QTL allele frequency independence is expected only in the long-term, when approaching strong recombination and weak selection, mutation and drift, steady-states (Barton, 1990; Barton et al., 2017; Vladar and Barton, 2014). Our findings pose the question of how genetic drift, together with effective stabilizing selection, generates variable allele frequency

changes at QTL so that pleiotropy or linkage disequilibrium between them eventually results in genetic covariances that diverge from the ancestral states and are not common among replicate populations. In our case, recombination during the focal stage should have remained much weaker than selection between 0.5-1 cM regions (Chelo and Teotónio, 2013; Noble et al., 2017, 2019); for total a total genome size of 300 cM. If after domestication several QTL alleles within these linked regions segregate at low frequency, it is possible that selection and drift was such that each replicate population during divergence fixed alleles with differently signed phenotypic effects that would not average out when comparing across populations (Bernstein et al., 2019; Cohan, 1984; Gromko, 1995). Inflation of the effects of drift is further expected because there is a correlation across generations between the traits' breeding values of successful parents and their offspring that results in a reduction in effective population sizes (Robertson, 1961; Santiago and Caballero, 1998). For intermediate frequency QTL alleles, linked selection could have dampened among replicate differentiation (Hill and Robertson, 1966; Zhang and Hill, 2005), but the strong polygenic sign epistasis observed previously in a subset of the same inbred lines used here for fitness-related traits could be at work in the opposite direction (Noble et al., 2017). Only investigation of allele and linkage disequilibrium frequency dynamics at exemplar QTL in our populations will shed light into these questions.

Short-term phenotypic stasis without genetic divergence in natural populations has been explained by indirect selection or phenotypic plasticity, despite heritability and direct selection on the traits that were followed. Our study suggests that phenotypic stasis can also occur with simultaneous genetic divergence, and we conclude that the adaptive landscape in our lab environment is best understood as a table-top mountain, where a saddled plateau with different optima are of little consequence to genetic or phenotypic divergence. Outside the plateau, directional selection explains phenotypic stasis and loss of genetic variation, within the plateau drift appears to be the main driver of evolution.

6 Acknowledgments

We thank A. Crist, J. Garcia, H. Gendrot, C. Goy, V. Pereira, F. Melo, and A. Silva for help with worm handling and data acquisition; R. Costa, R. Kerr, and N. Scwierczek for help with hardware and software implementation; N. Barton, C. Dillmann, A. Futschik, L. Kollar, S. McDaniel, P. Phillips, A. Le Rouzic, and A. Veber for discussion. We also thank the anonymous reviewers that greatly improved the presentation of this work. This work was supported by the European Research Council (ERC-St-243285) and the Agence Nationale pour la Recherche (ANR-14-ACHN-0032-01, ANR-17-CE02-0017-01) to HT, the National Institutes of Health (R01GM107227) to CB, and a Marie Curie fellowship (H2020-MSCA-IF-2017-798083) to LN.

7 Author contributions

Conceptualization FM, LN, CB, HT; hardware and software implementation BA, TG; data acquisition and analysis BA, FM, LN, TG; funding acquisition LN, CB, HT; project administration HT; resources CB, HT; writing, original draft FM, HT; writing, review and editing LN, CB; correspondence FM (mallard@bio.ens.psl.eu) and HT (teotonio@bio.ens.psl.eu).

8 Figures

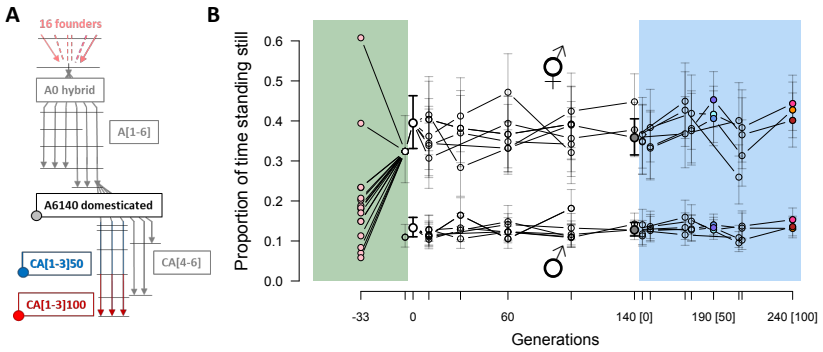

Figure 1: **A.** Experimental design. One hybrid population (A0) was created from the intercross of 16 inbred founders. Six replicate populations were then domesticated to a defined lab environments and after 140 generations one of these (A6140) was the ancestor to six other replicate populations maintained for an extra 100 generations under similar conditions (CA). Inbred lines were derived by selfing hermaphrodites (colored circles) from A6140 and three replicate CA populations at generation 50 and 100 (blue and red). Horizontal lines indicate outbred population samples that were phenotyped. **B.** Evolution of locomotion behavior. Stationary frequency in the founders (pink dots) and outbred populations during lab evolution. Colored overlays indicate three stages of lab evolution: hybridization, domestication and focal. Ticks are sampled time points, while colored points during the focal stage indicate populations from which inbred lines were derived. Point mean estimates are shown for 3-6 replicate populations at other generations, with 95% confidence intervals for each one of them. The evolution of the component traits of locomotion behavior in hermaphrodites and males, the transition rates between movement state and direction, can be found in Figures S4 and S5.

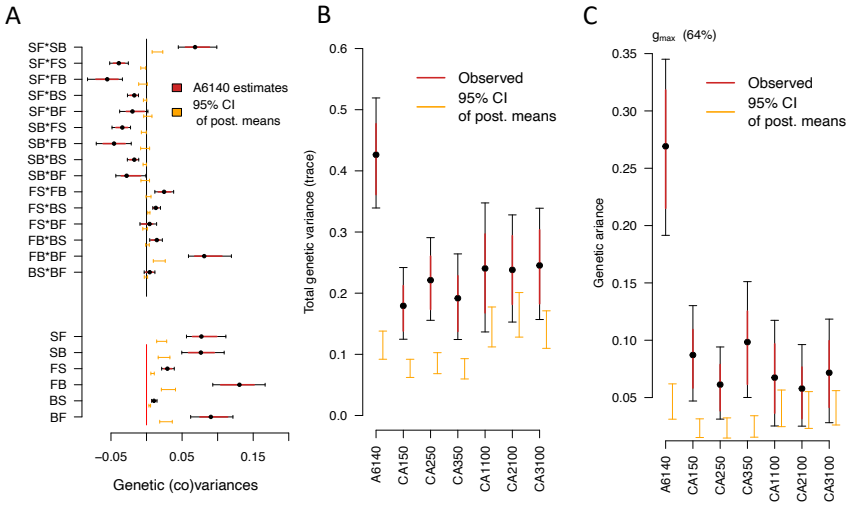

Figure 2: **G**-matrix evolution during the focal stage. **A.** A6140 **G**-matrix. Shown are the 15 genetic covariances between transitions rates (top) and six the genetic variances of transition rates (bottom), as bars and dots the 95% and 80% credible intervals (black and red) and mean of the posterior distribution, respectively. "S", "F", "B" stand for still, forward and backward movement states, with letter ordering indicating the direction of movement. **G** matrices of the CA populations can be found in Supplementary Figure S7. **B.** Total amount of genetic variance computed as the sum of the **G** diagonal elements (trace). All observed posterior means differ from the null 95% posterior means (orange). **C.** Genetic variance along the phenotypic dimension encompassing most genetic variation (g_{max}, red mean, 80% and 95% CI), when spectral decomposition is done on A6140 and CA[1-3] variances on this A6140 g_{max} dimension is calculated. A6140 g_{max} explains 64% of the total genetic variance.

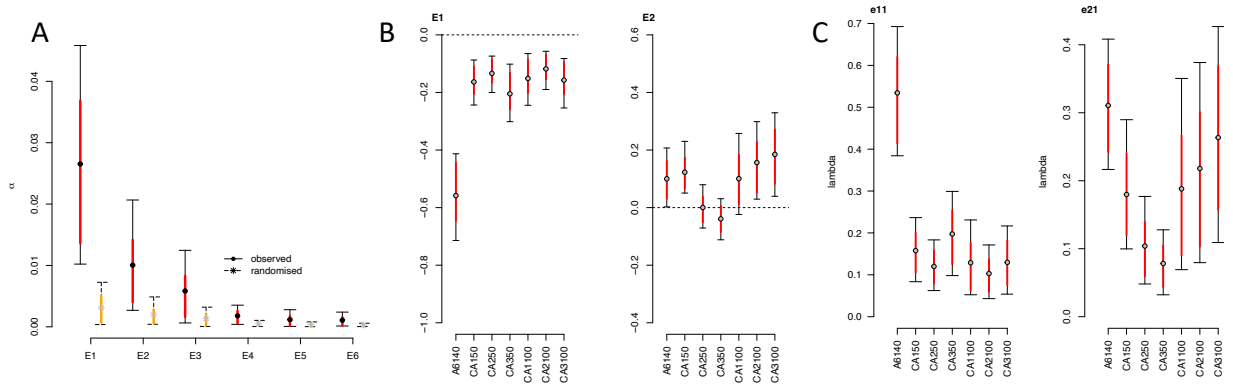

Figure 3: Genetic divergence and differentiation. **A.** Spectral decomposition of variation among **G**-matrices. The variance $\boldsymbol{\alpha_i}$ associated with the ith eigentensor $\boldsymbol{E_i}$ is compared to a null permutation model where variation among matrices is due to sampling (see Methods). Although several eigentensors are different from zero (bars, 95% credible interval) only the first two, $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$, are do not overlap the null (red and orange bars, 80% credible intervals). **B.** The coordinates of the **G**-matrices in the space of the first two eigentensors $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$, for each population. Absolute values of the coordinates in each eigentensor represent its contribution to the difference between matrices. **C.** Contribution of specific transition rate combinations to coordinated changes among **G**-matrices. Each panel shows the amount of genetic variance in the direction of the greatest variation among all **G**s for the first eigenvectors of $\boldsymbol{E_1}$ (e_{11}), $\boldsymbol{E_2}$ (e_{12}), for each population.

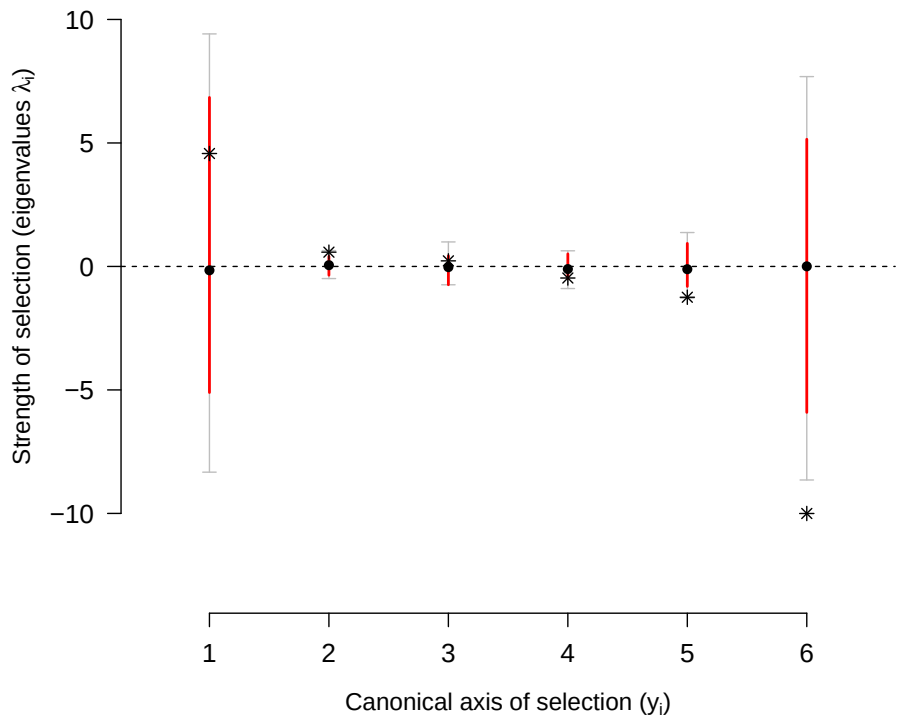

Figure 4: Selection surface of locomotion behavior. Canonical analysis of the γ-matrix shows positive phenotypic dimensions (y_1-y_3) of transition rate combinations under disruptive selection (as measured by the eigenvalue λ), and negative dimension (y_4-y_6) under stabilizing selection. Stars show the mode of the posterior empirical distribution (see Methods). These estimates are to be compared to the posterior distribution of null modes (dots and colored bars, the mean and 80% and 95% credible intervals). The γ-matrix before canonical rotation can be found in Figure S10.

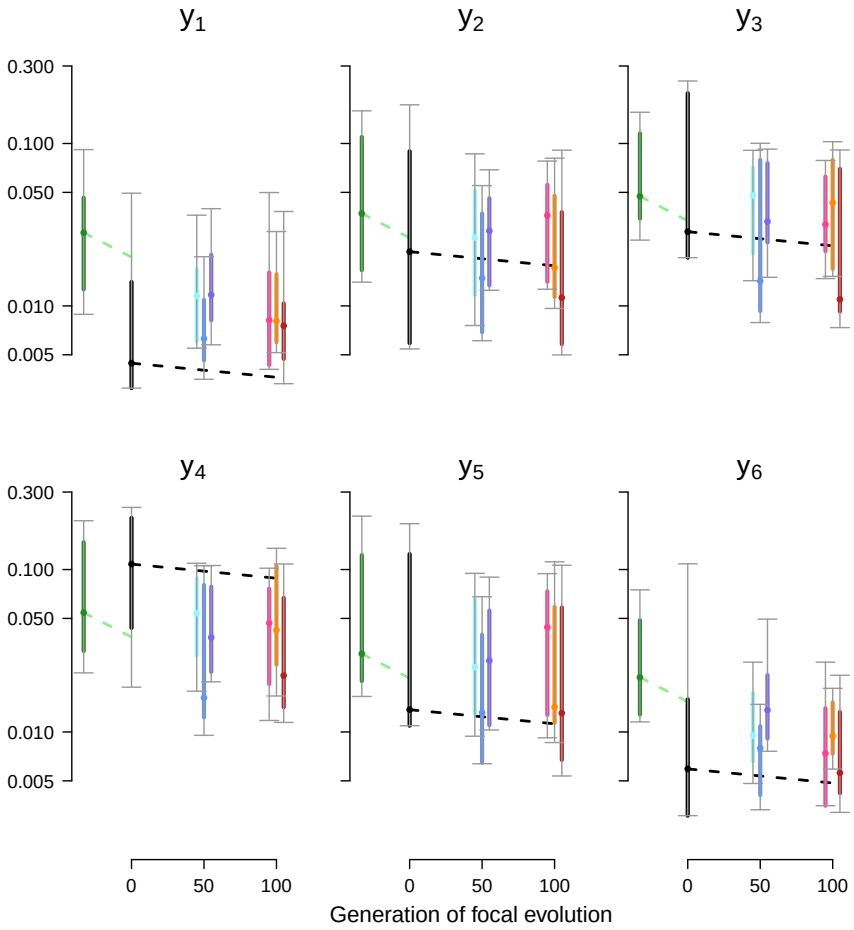

Figure 5: **G** matrix evolution in the selection surface. Loss of genetic variance along axes y_2-y_5, which contain most of the genetic variance in the evolved populations and are under very weak or no selection, is compatible with expectations from genetic drift under the assumption of infinitesimal trait inheritance (dashed lines, for $N_e = 10^3$). Along y_1, y_5 and y_6, genetic variance was much reduced relative to the founders of experimental evolution (green), but was subsequently maintained or increased. The genetic variance of each canonical axis y_i was obtained by rotation of the original **G**-matrices, with 95% (grey) and 80% (colored) credible intervals from sampling 400 matrices in the posterior distributions for each **G**-matrix. Dots show the median estimates.

9 Supplementary Figures

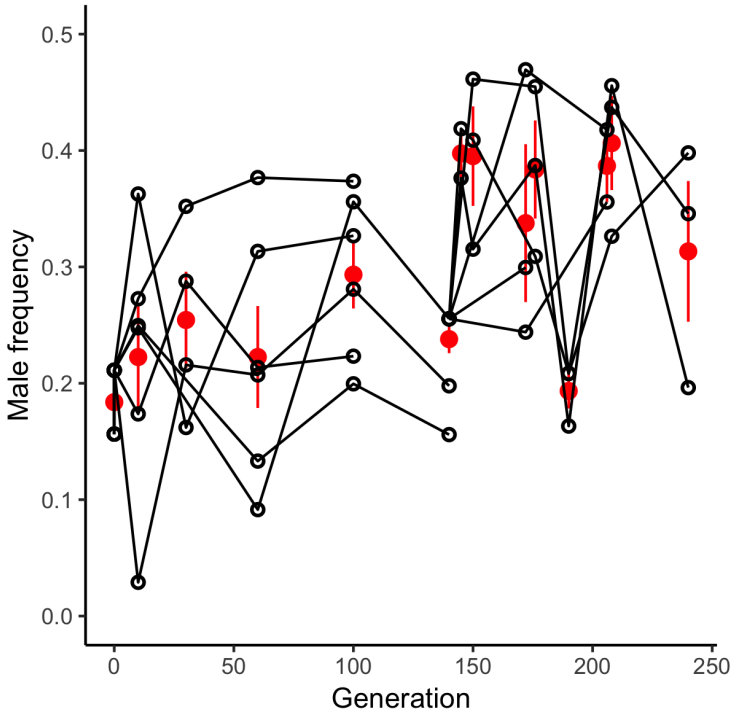

Figure S1: Male frequencies during lab evolution. Males and hermaphrodite tracks were differentiated with a 30-trait classifier based on moments of size, shape and velocity-related traits derived from Multi-Worm Tracker metrics, and frequencies were estimated from 1s slices across movies. Empty circles indicate the estimates for each replicate population (between 1 and 6 at each time point), red circles the mean among replicate populations ($\pm$ standard error). During the first 100 generations of domestication, the estimates are similar to those obtained by directly counting the number of males (Teotónio et al., 2012)

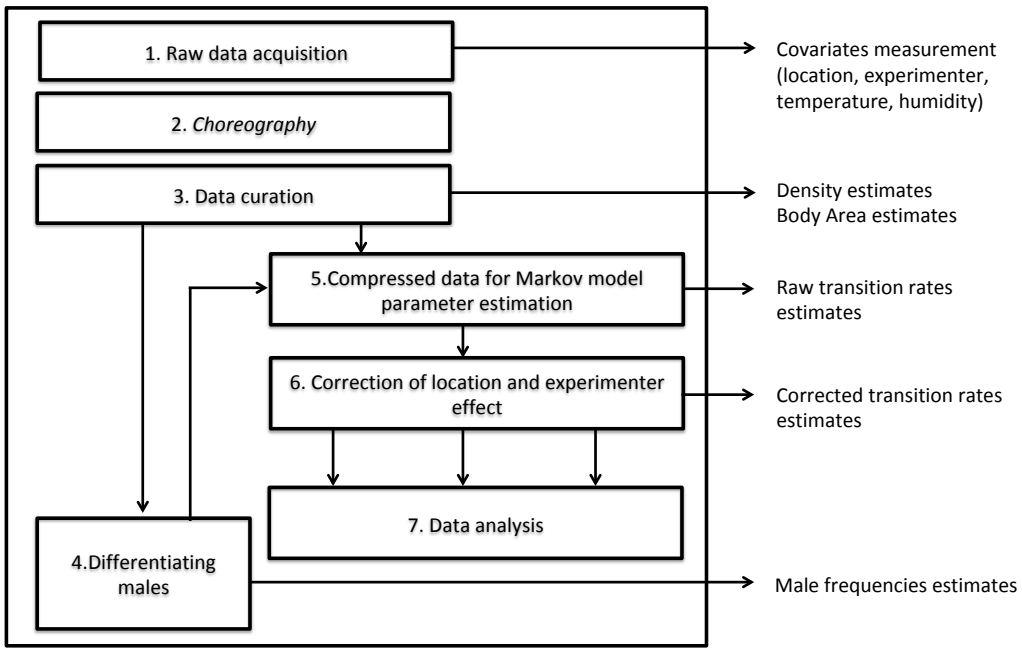

Figure S2: Schematic of data acquisition and analysis pipeline.

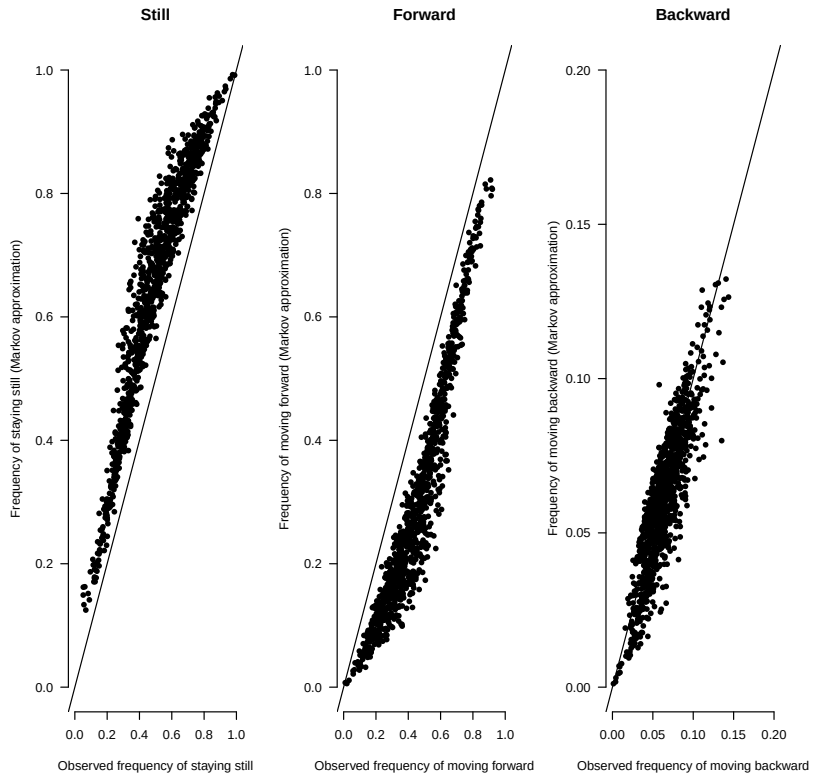

Figure S3: Correlation between the observed frequencies of each of the three movement states and the predicted values from the Markov model. There is a consistent bias in the long term predictions due to violation of the memoryless assumption of the model. Some moving worms tend to remain in this state longer than expected on the long term, that is, they can be briefly interrupted but are more likely to resume movement than predicted.

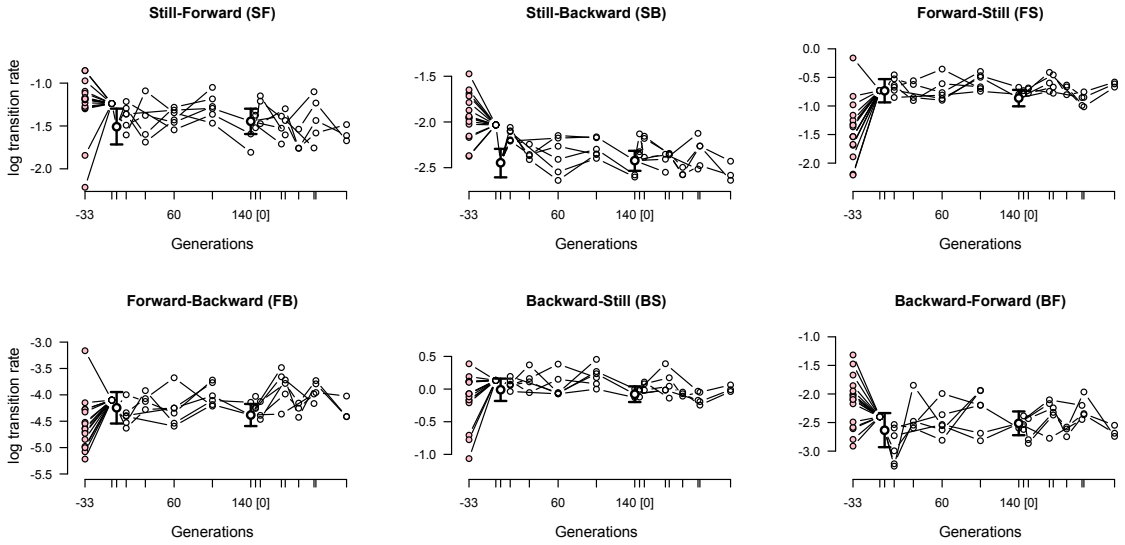

Figure S4: Evolution of mean hermaphrodite transition rates. Each panel shows the evolution of a transition rate in the founders (pink dots) and during experimental evolution (white dots). At the beginning of the domestication and focal stages there was one ancestral population, shown by empty circles with 95% credible intervals, while 3-6 replicate populations were measured at each sampled time point indicated by tick marks.

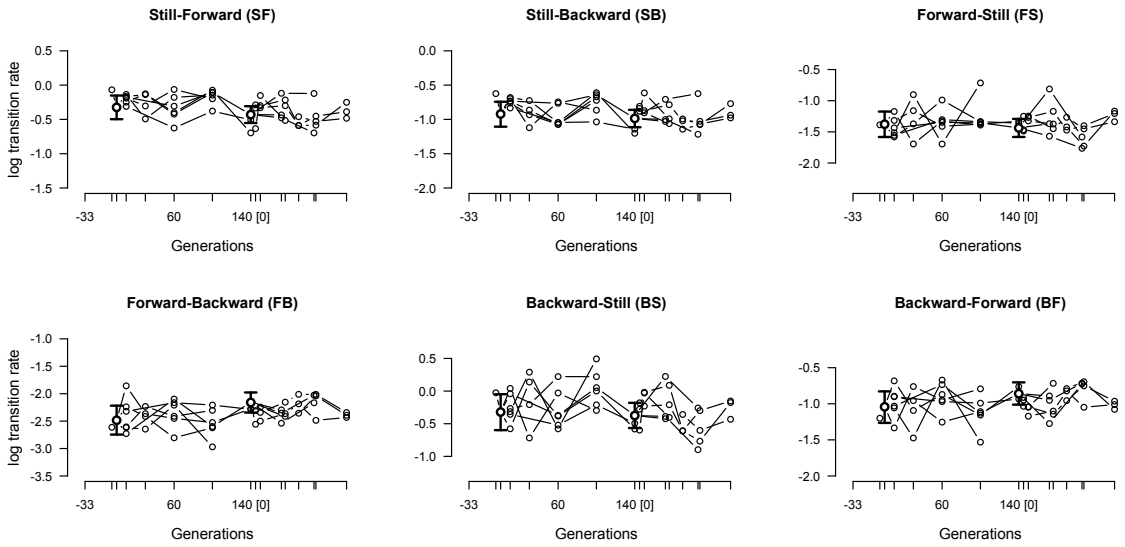

Figure S5: Evolution of mean male transition rates, as in Figure S4. Note that the founder inbred lines do not have any males.

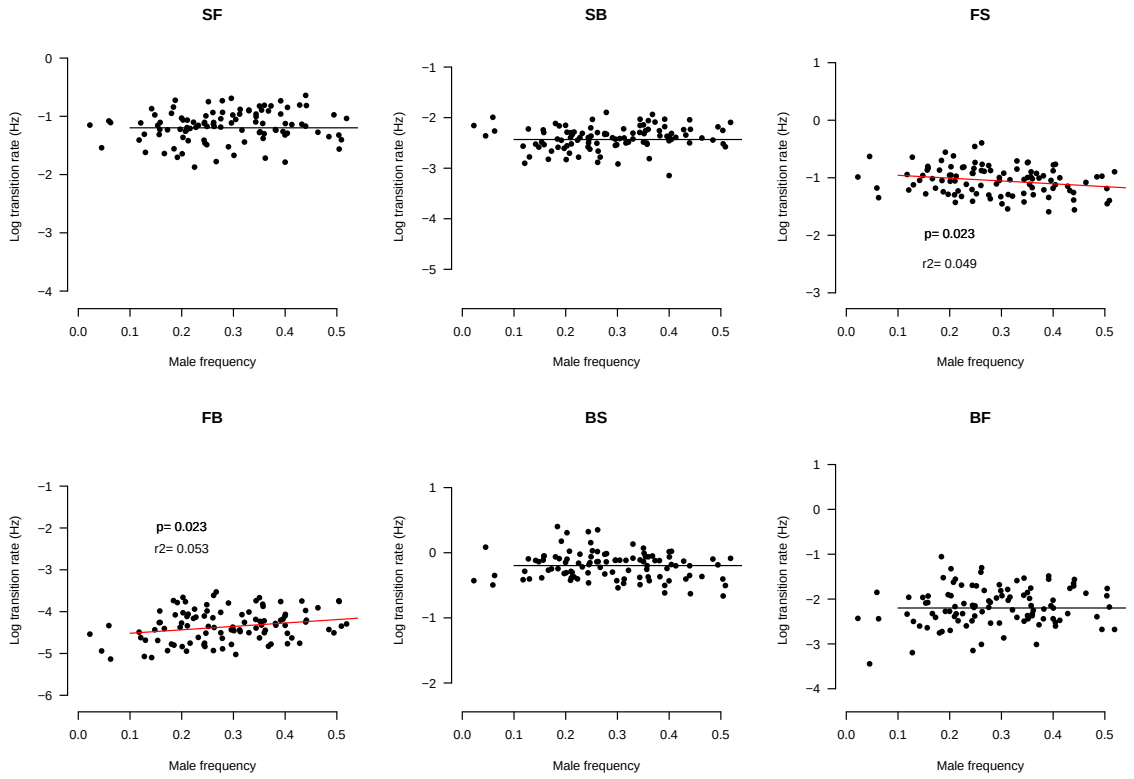

Figure S6: The effects of males on hermaphrodite transition rates in the outbred populations during lab evolution. Each point shows the relation between transition rates and male frequency for each replicate population at a given time point during lab evolution. Red (black) lines show significant (non-significant) linear effects of male frequency on transition rates. For all regression models the coefficient of determination is extremely low (r^2).

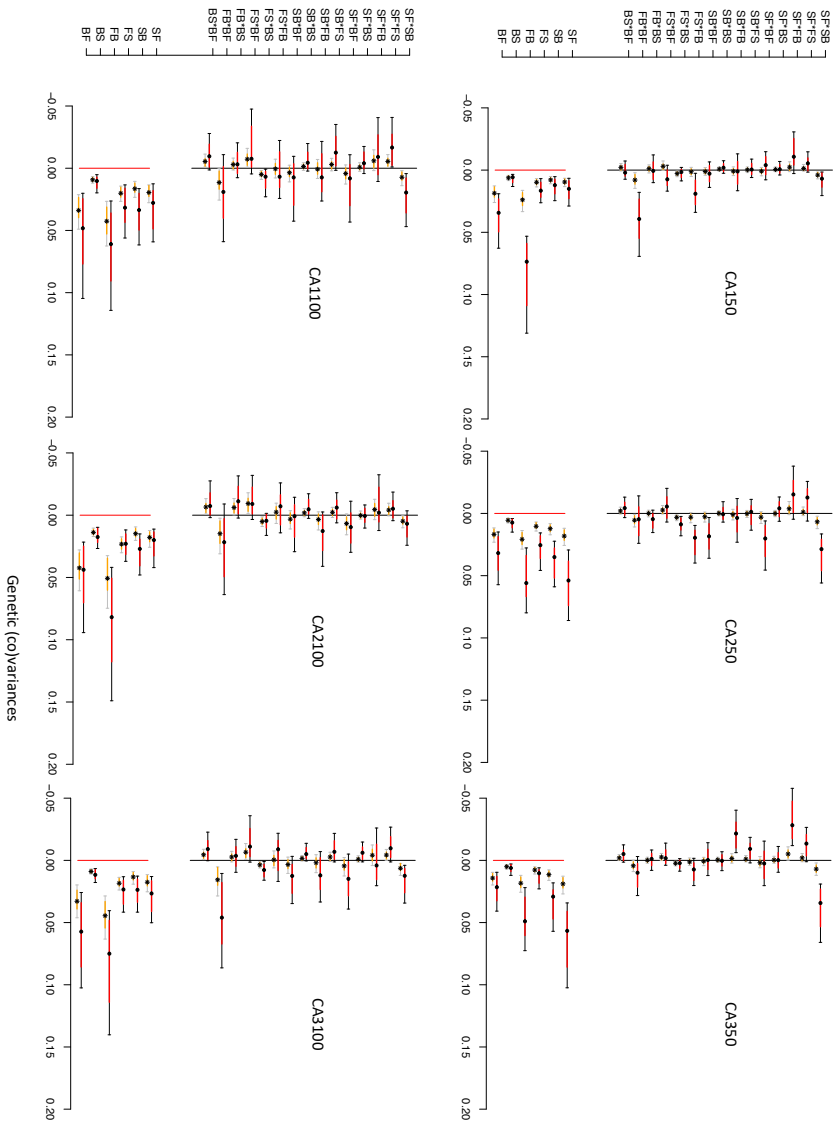

Figure S7: Median genetic covariance and variance estimates of the CA[1-3] populations at generation 50 and generation 100 (dots), as well as random expectations (stars). Intervals are shown with the 95% and 80% credible intervals (black and red, grey and orange). Many of the covariance estimates do not differ from zero (95% interval) and covariance and variance median estimates do not generally differ from a null distribution obtained from random permutations of the phenotypic values (black dots are within grey bars), particularly for generation 100 populations. This is explained by a loss of genetic (co)variances with continued lab evolution after domestication, and not sampling a limited number of lines in each population, see Figure S7).

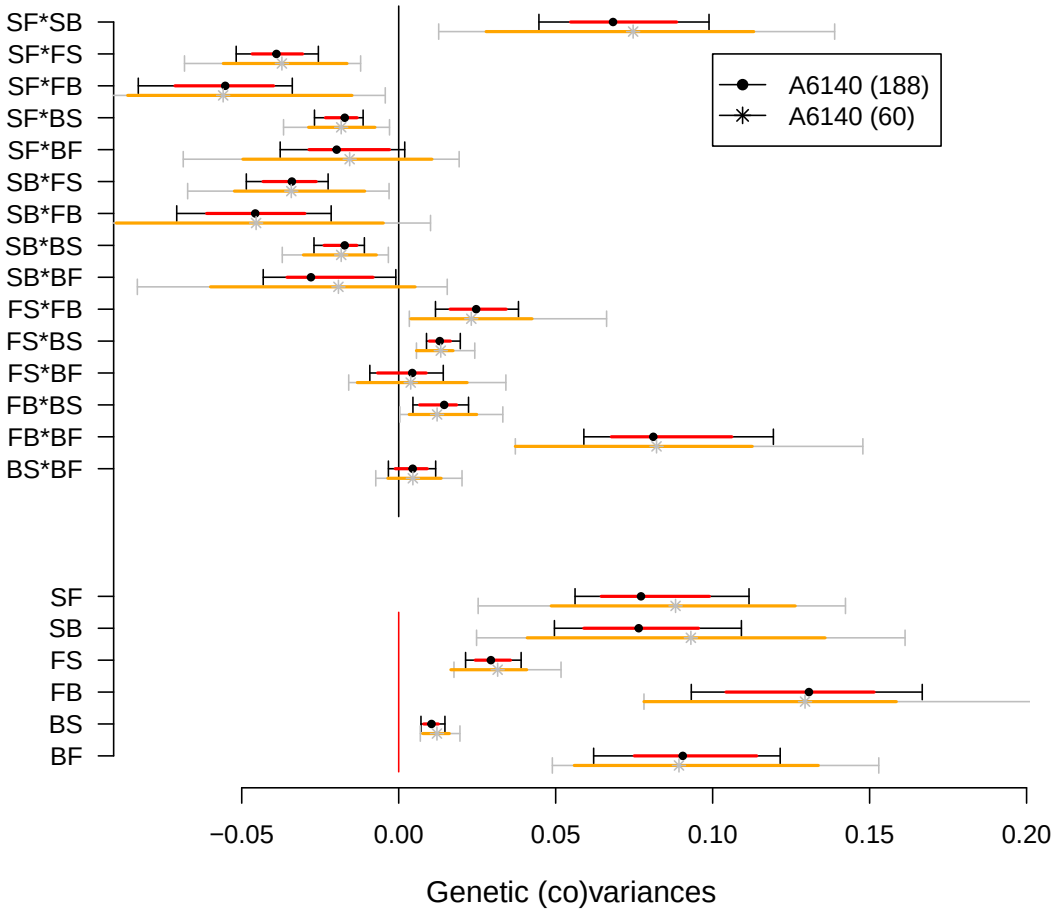

Figure S8: **G**-matrix estimates of the 140-generation domesticated A6140 population. Black and red show the estimated genetic (co)variances using all inbred lines as in main Figure 3. Grey and orange show genetic (co)variances after downsampling to 60 inbred lines, approximately the minimum number of lines phenotyped in the CA[1-3] populations. Median estimates are similar between data sets, though with larger intervals in the subsampled estimates.

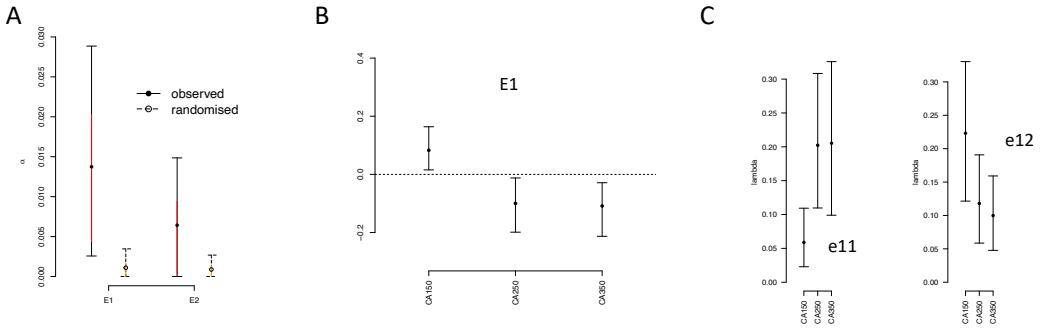

Figure S9: Genetic divergence and differentiation. **A.** Eigentensor decomposition of variation among **G**-matrices of the CA[1-3]50 populations. The variance $\boldsymbol{\alpha_i}$ associated with the ith eigentensor $\boldsymbol{E_i}$ is compared to a null permutation model where variation among matrices is due to sampling (see Methods). Here, only the first eigentensors is different from the null. **B.** The coordinates of the **G**-matrices in the space of the first eigentensor. The absolute values of the matrices coordinates in each eigentensor represent its contribution to the difference between matrices. Coordinates with opposing signs indicate that the matrices contribute in opposing directions. **C.** Contribution of specific trait combinations to coordinated changes among **G**-matrices. Each panel shows the amount of genetic variance in the direction of the greatest variation among **G** (eigenvector of $\boldsymbol{E_1}$ only).

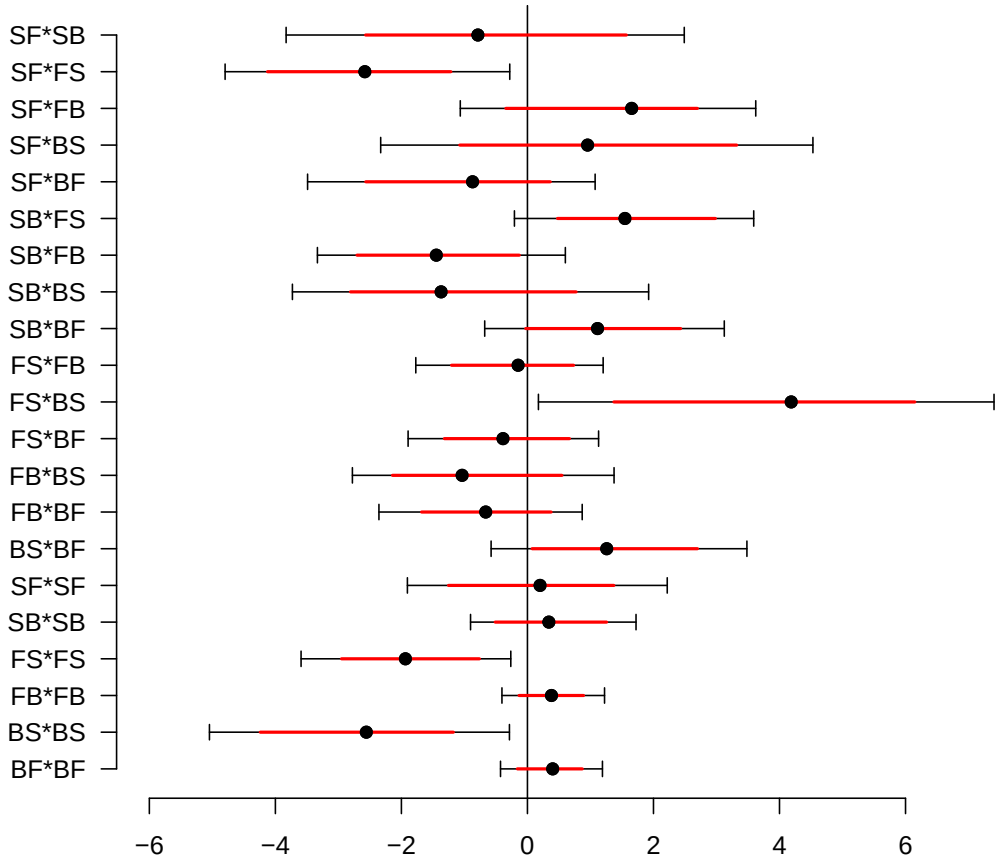

Figure S10: Quadratic selection coefficients. The partial regression coefficients of fertility on transition rates estimated by Bayesian inference. Each row shows the mode (dot), and 80% and 95% credible intervals (red bar and line bars, respectively) of the posterior distributions. The top 15 rows show coefficients of correlated selection between two transition rates, the bottom 6 rows show coefficients of stabilizing or disruptive selection on each transition rate.

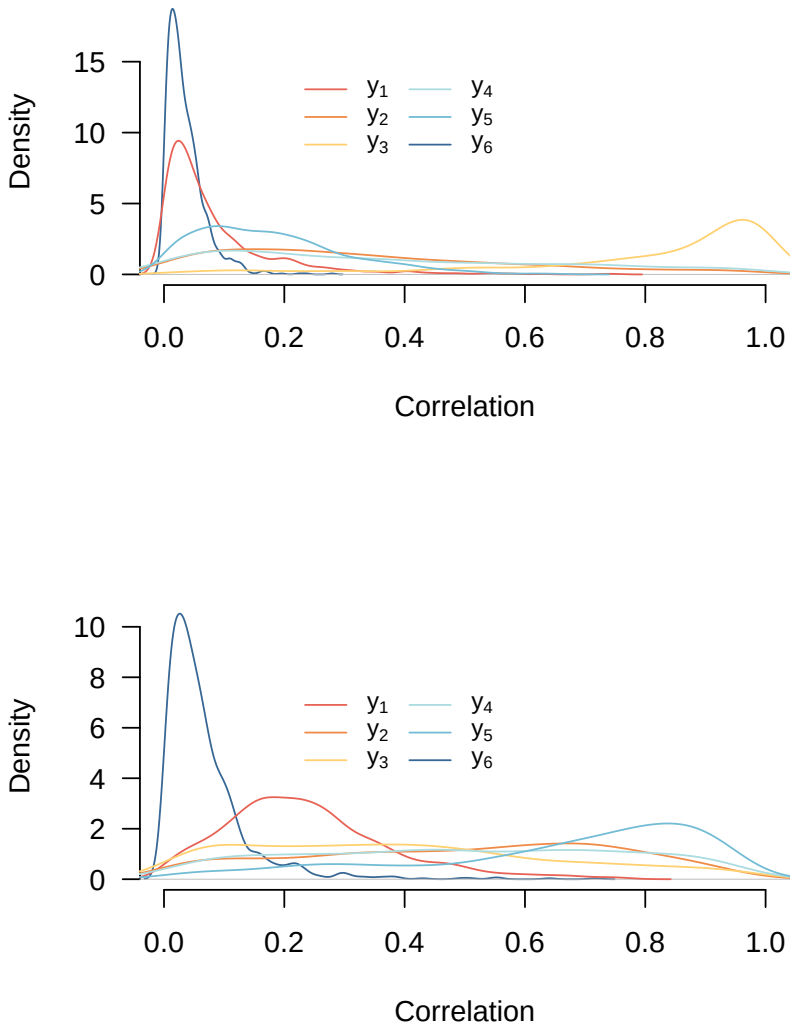

Figure S11: Alignment of **G**-matrices with the quadratic selection surface. Shown is the density distributions of Pearson product moment correlations between the first two eigenvectors of **E1** (top e_{11} and bottom e_{12}) for all 7 populations with the eigenvectors of the γ matrix (y_i). The density distributions are obtained from 1000 sampling in the posterior distribution of the γ matrix.

10 Tables

Transition rate	Sex	Chisq	P value	P adjusted
SF	H	0.542	0.462	0.799
SB	H	0.715	0.398	0.799
FS	H	0.389	0.533	0.799
FB	H	3.077	0.079	0.799
BS	H	0.037	0.847	0.924
BF	H	1.810	0.179	0.799
SF	M	0.414	0.520	0.799
SB	M	0.647	0.421	0.799
FS	M	0.542	0.462	0.799
FB	M	0.053	0.817	0.924
BS	M	0.001	0.980	0.980
BF	M	0.082	0.774	0.924

Table 1: Phenotypic stasis: Results of anova LRT χ^2_1 tests for directional changes in mean transition rates in hermaphrodites (H) and males (M), during the 240 generations of lab evolution. Corrected P values for multiple comparisons were obtained with the Benjamini-Hochberg method.

Transition rates	Chisq	P values	P adjusted
SF	5.445	0.020	0.118
SB	1.319	0.251	0.419
FS	0.107	0.743	0.860
FB	1.443	0.230	0.419
BS	1.170	0.279	0.419
BF	0.031	0.860	0.860

Table 2: Inbreeding effects: Results of anova LTR χ^2_1 testing for mean phenotypic differences between the mean of the inbred lines and the mean of the A6140 population from which they were derived. Corrected P values for multiple comparisons were obtained with the Benjamini-Hochberg method.

11 Literature Cited

Aguirre, J. D., Hine, E., McGuigan, K., and Blows, M. W. (2014). Comparing g: multivariate analysis of genetic variation in multiple populations. *Heredity (Edinb)*, 112(1):21–9.

Arnold, S., Pfender, M., and Jones, A. G. (2001). The adaptive landscape as a conceptual bidge between micro and macroevolution. *Genetica*, 112-113:9–32.

Arnold, S. J. (2014). Phenotypic evolution: the ongoing synthesis (american society of naturalists address). *Am Nat*, 183(6):729–46.

Austin, P. C. and Hux, J. E. (2002). A brief note on overlapping confidence intervals. *Journal of vascular surgery*, 36(1):194–195.

Baer, C. F., Shaw, F., Steding, C., Baumgartner, M., Hawkins, A., Houppert, A., Mason, N., Reed, M., Simonelic, K., Woodard, W., and Lynch, M. (2005). Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proc Natl Acad Sci U S A*, 102(16):5785–90.

Barton, N. H. (1990). Pleiotropic models of quantitative variation. *Genetics*, 124(3):773–782.

Barton, N. H., Etheridge, A. M., and Veber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theor Popul Biol*, 118:50–73.

Barton, N. H. and Turelli, M. (1987). Adaptive landscapes, genetic distance and the evolution of quantitative characters. *Genet Res*, 49(2):157–73.

Bartoń, K. (2020). *MuMIn: Multi-Model Inference*. R package version 1.43.17.

Bernstein, M. R., Zdraljevic, S., Andersen, E. C., and Rockman, M. V. (2019). Tightly linked antagonistic-effect loci underlie polygenic phenotypic variation in c. elegans. *Evol Lett*, 3(5):462–473.

Biquet, J., Bonamour, S., de Villemereuil, P., de Franceschi, C., and Teplitsky, C. (2022). Phenotypic plasticity drives phenological changes in a mediterranean blue tit population. *J Evol Biol*, 35(2):347–359.

Bladt, M. and Sorensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):395–410.

Blows, M. W. and Brooks, R. (2003). Measuring nonlinear selection. *Am Nat*, 162(6):815–20.

Bohren, B., Hill, W., and Robertson, A. (1966). Some observations on asymmetrical correlated responses to selection. *Genet Res*, 7:44.

Burger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. Wiley Series in Mathematical and Computational Biology. John Wiley Sons, Ltd., New York.

Carvalho, S., Chelo, I. M., Goy, C., and Teotónio, H. (2014a). The role of hermaphrodites in the experimental evolution of increased outcrossing rates in *Caenorhabditis elegans*. *BMC Evol Biol*, 14:116.

Carvalho, S., Phillips, P., and Teotónio, H. (2014b). Hermaphrodite life history and the maintenance of partial selfing in experimental populations of *Caenorhabditis elegans*. *BMC Evol Biol*, 14:117.

Charlesworth, B., Lande, R., and Slatkin, M. (1982). A neo-darwinian commentary on macroevolution. *Evolution*, 36(3):474–498.

Chebib, J. and Guillaume, F. (2017). What affects the predictability of evolutionary constraints using a g-matrix? the relative effects of modular pleiotropy and mutational correlation. *Evolution*, 71(10):2298–2312.

Chelo, I. M., Afonso, B., Carvalho, S., Theologidis, I., Goy, C., Pino-Querido, A., Proulx, S., and Teotónio, H. (2019). Partial selfing can reduce genetic loads while maintaining diversity during evolution. *G3 (Bethesda)*, 9:2811–2821.

Chelo, I. M., Nédli, J., Gordo, I., and Teotónio, H. (2013). An experimental test on the probability of extinction of new genetic variants. *Nature Communications*, 4:10.1038/ncomms3417.

Chelo, I. M. and Teotónio, H. (2013). The opportunity for balancing selection in experimental populations of *Caenorhabditis elegans*. *Evolution*, 67(1):142–56.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA. ACM.

Cohan, F. (1984). Can uniform selection retard random genetic divergence between isolated conspecific populations. *Evolution*, 38:495–504.

Cook, D. E., Zdraljevic, S., Roberts, J. P., and Andersen, E. C. (2017). Cendr, the Caenorhabditis elegans natural diversity resource. *Nucleic Acids Res*, 45(D1):D650–D657.

Czorlich, Y., Aykanat, T., Erkinaro, J., Orell, P., and Primmer, C. R. (2022). Rapid evolution in salmon life history induced by direct and indirect effects of fishing. *Science*, 376(6591):420–423.

de Villemereuil, P., Charmantier, A., Arlt, D., Bize, P., Brekke, P., Brouwer, L., Cockburn, A., Côté, S. D., Dobson, F. S., Evans, S. R., Festa-Bianchet, M., Gamelon, M., Hamel, S., Hegelbach, J., Jerstad, K., Kempenaers, B., Kruuk, L. E. B., Kumpula, J., Kvalnes, T., McAdam, A. G., McFarlane, S. E., Morrissey, M. B., Pärt, T., Pemberton, J. M., Qvarnström, A., Røstad, O. W., Schroeder, J., Senar, J. C., Sheldon, B. C., van de Pol, M., Visser, M. E., Wheelwright, N. T., Tufto, J., and Chevin, L.-M. (2020). Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. *Proceedings of the National Academy of Sciences*, 117(50):31969–31978.

Doroszuk, A., Wojewodzic, M. W., Gort, G., and Kammenga, J. E. (2008). Rapid divergence of genetic variance-covariance matrix within a natural population. *The American Naturalist*, 171(3):291–304.

Estes, S. and Arnold, S. (2007). Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist*, 169(2):227–244.

Farhadifar, R., Baer, C. F., Valfort, A. C., Andersen, E. C., Muller-Reichert, T., Delattre, M., and Needleman, D. J. (2015). Scaling, selection, and evolutionary dynamics of the mitotic spindle. *Curr Biol*, 16:732–740.

Fowler, K. and Whitlock, M. C. (1999). The distribution of phenotypic variance with inbreeding. *Evolution*, 53(4):1143–1156.

Gingerich, P. D. (2019). *Rates of evolution: a quantitative synthesis*. Cambridge University Press.

Gromko, M. H. (1995). Unpredictability of correlated response to selection: Pleiotropy and sampling interact. *Evolution*, 49(4):685–693.

Guzella, T., Dey, S., Chelo, I. M., Pino-Querido, A., Pereira, V., Proulx, S., and Teotónio, H. (2018). Slower environmental change hinders adaptation from standing genetic variation. *PLoS Genet*, 14:e1007731.

Hadfield, J. (2010). Mcmc methods for multi-response generalized linear mixed models: The mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.

Haller, B. C. and Hendry, A. P. (2014). Solving the paradox of stasis: squashed stabilizing selection and the limits of detection. *Evolution*, 68(2):483–500.

Hansen, T. and Martins, E. (1996). Translating between macroevolutionary process and microevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50:1404–1417.

Hansen, T. F. and Wagner, G. P. (2001). Modeling genetic architecture: a multilinear theory of gene interaction. *Theor Popul Biol*, 59(1):61–86.

Hill, W. G. (1982). Rates of change in quantitative traits from fixation of new mutations. *Proc Natl Acad Sci U S A*, 79(1):142–5.

Hill, W. G. and Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genet Res*, 8(3):269–94.

Hine, E., Chenoweth, S. F., Rundle, H. D., and Blows, M. W. (2009). Characterizing the evolution of genetic variance using genetic covariance tensors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1523):1567–1578.

Houle, D., Bolstad, G. H., van der Linde, K., and Hansen, T. F. (2017). Mutation predicts 40 million years of fly wing evolution. *Nature*, 548(7668):447–450.

Houle, D., Morikawa, B., and Lynch, M. (1996). Comparing mutational variabilities. *Genetics*, 143(3):1467–1483.

Hunt, J., Blows, M. W., Zajitschek, F., Jennions, M. D., and Brooks, R. (2007). Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (teleogryllus commodus). *Genetics*, 177(2):875–880.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8):1–29.

Johnson, N. A. and Porter, A. H. (2007). Evolution of branched regulatory genetic pathways: directional selection on pleiotropic loci accelerates developmental system drift. *Genetica*, 129(1):57–70.

Johnson, T. and Barton, N. (2005). Theoretical models of selection and mutation on quantitative traits. *Philos Trans R Soc Lond B Biol Sci*, 360(1459):1411–25.

Kalbfleisch, J. D. and Lawless, J. F. (1985). The Analysis of Panel Data Under a Markov Assumption. *Journal of the American Statistical Association*, 80(392):863.

Kearsey, M. and Pooni, H. (1996). *The Genetical Analysis of Quantitative Traits*. Chapman and Hall, New York.

Kingsolver, J., Hoekstra, H., Hoekstra, J., Berrigan, D., Vignieri, S., Hill, C., Hoang, A., Gibert, P., and Beer, P. (2001). The strength of phenotypic selection in natural populations. *Am Nat*, 157:245–261.

Kruuk, L., Slate, J., Pemberton, J., Brotherstone, S., Guiness, F., and Clutton-Brock, T. (2002). Antler size in red deer: heritability and selection but no evolution. *Evolution*, 56:1683–1685.

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30:314–334.

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

Lande, R. (1980). The genetic covariance between characters maintained by pleiotropic mutations. *Genetics*, 94(1):203–15.

Lande, R. (1986). The dynamics of peak shifts and the pattern of morphological evolution. *Paleobiology*, 12:343–354.

Lande, R. and Arnold, S. J. (1983). On the measurement of selection on correlated characters. *Evolution*, 37:1210–1226.

Lee, D., Zdraljevic, S., Stevens, L., Wang, Y., Tanny, R. E., Crombie, T. A., Cook, D. E., Webster, A. K., Chirakar, R., Baugh, L. R., et al. (2021). Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nature ecology & evolution*, 5(6):794–807.

Lipton, J., Kleemann, G., Ghosh, R., Lints, R., and Emmons, S. W. (2004). Mate searching in *Caenorhabditis elegans*: a genetic model for sex drive in a simple invertebrate. *J Neurosci*, 24(34):7427–34.

Lynch, M. and Hill, W. G. (1986). Phenotypic evolution by neutral mutation. *Evolution*, 40(5):915–935.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland.

Mallard, F., Afonso, and Teotónio, H. (2022a). Selection and the direction of phenotypic evolution. *bioRxiv*, page XXXX.

Mallard, F., Noble, L., Baer, C., and Teotónio, H. (2022b). Variation in mutational (co)variances. *bioRxiv*, page XXXX.

Matuszewski, S., Hermisson, J., and Kopp, M. (2015). Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics*, 200:1255–1274.

Merilä, J., Sheldon, B. C., and Kruuk, L. E. (2001). Explaining stasis: microevolutionary studies in natural populations. *Genetica*, 112-113:199–222.

Morrissey, M. B. (2015). Evolutionary quantitative genetics of nonlinear developmental systems. *Evolution*, 69(8):2050–66.

Morrissey, M. B. and Bonnet, T. (2019). Analogues of the fundamental and secondary theorems of selection, assuming a log-normal distribution of expected fitness. *J Hered*, 110(4):396–402.

Morrissey, M. B. and Hadfield, J. D. (2012). Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, 66(2):435–42.

Noble, L. M., Chelo, I., Guzella, T., Afonso, B., Riccardi, D. D., Ammerman, P., Dayarian, A., Carvalho, S., Crist, A., Pino-Querido, A., Shraiman, B., Rockman, M. V., and Teotónio, H. (2017). Polygenicity and epistasis underlie fitness-proximal traits in the *Caenorhabditis elegans* multiparental experimental evolution (cemee) panel. *Genetics*, 207(4):1663–1685.

Noble, L. M., Rockman, M. V., and Teotónio, H. (2019). Gene-level quantitative trait mapping in *C. elegans*. *G3 Genes— Genomes— Genetics*.

Pelabon, C., Hansen, T. F., Carter, A. J., and Houle, D. (2010). Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. *Evolution*, 64(7):1912–25.

Phillips, P. and McGuigan, K. (2006). *Evolution of Genetic Variance-Covariance Structure*. Oxford University Press, Oxford.

Phillips, P., Whitlock, M. C., and Fowler, G. R. (2001). Inbreeding changes the shape of the genetic covariance matrix in drosophila melanogaster. *Genetics*, 158:1137–1145.

Phillips, P. C. and Arnold, S. J. (1989). Visualizing multivariate selection. *Evolution*, 43(6):1209–1222.

Poullet, N., Vielle, A., Gimond, C., Carvalho, S., Teotónio, H., and Braendle, C. (2016). Complex heterochrony underlies the evolution of *Caenorhabditis elegans* hermaphrodite sex allocation. *Evolution*, 30:2357–2369.

Pujol, B., Blanchet, S., Charmantier, A., Danchin, E., Facon, B., Marrot, P., Roux, F., Scotti, I., Teplitsky, C., Thomson, C. E., and Winney, I. (2018). The missing response to selection in the wild. *Trends Ecol Evol*, 33(5):337–346.

Riska, B. (1989). Composite traits, selection response and evolution. *Evolution*, 43:1172–1191.

Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genetical Research*, 2(2):189–194.

Santiago, E. and Caballero, A. (1998). Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics*, 149(4):2105.

Schluter, D. (1996). Adaptive radiaation along genetic lines of least resistance. *Evolution*, 50(5):1766–1774.

Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1):205–233.

Shaw, F., Shaw, R. G., Wilkinson, G., and Turelli, M. (1995). Changes in genetic variances and covariances: G whiz! *Evolution*, 49:1260–1267.

Simons, Y. B., Bullaughey, K., Hudson, R. R., and Sella, G. (2018). A population genetic interpretation of gwas findings for human quantitative traits. *PLoS Biol*, 16(3):e2002985.

Simões, P., Fragata, I., Santos, J., Santos, M. A., Santos, M., Rose, M. R., and Matos, M. (2019). How phenotypic convergence arises in experimental evolution. *Evolution*, 73(9):1839–1849.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2.

Stiernagle, T. (1999). *Maintenance of C. elegans*. Oxford University Press, Oxford.

Stinchcombe, J. R., Simonsen, A. K., and Blows, M. W. (2014). Estimating uncertainty in multivariate responses to selection. *Evolution*, 68(4):1188–96.

Swierczek, N. A., Giles, A. C., Rankin, C. H., and Kerr, R. A. (2011). High-throughput behavioral analysis in *C. elegans*. *Nat Methods*, 8(7):592–8.

Sztepanacz, J. and Blows, M. W. (2017). Artificial selection to increase the phenotypic variance in gmax fails. *Am Nat*, 190:707–723.

Teotónio, H., Carvalho, S., Manoel, D., Roque, M., and Chelo, I. M. (2012). Evolution of outcrossing in experimental populations of *Caenorhabditis elegans*. *PLOS ONE*, 7(4):e35811.

Teotónio, H., Estes, S., Phillips, P., and Baer, C. (2017). Experimental evolution with caernohabditis nematodes. *Genetics*, 206(12):691–716.

Teotónio, H., Manoel, D., and Phillips, P. C. (2006). Genetic variation for outcrossing among *Caenorhabditis elegans* isolates. *Evolution*, 60(6):1300–1305.

Teotónio, H., Matos, M., and Rose, M. R. (2004). Quantitative genetics of functional characters in drosophila melanogaster populations subjected to laboratory selection. *J Genet*, 83:265–277.

Teotónio, H. and Rose, M. R. (2000). Variation in the reversibility of evolution. *Nature*, 408(6811):463–6.

Theologidis, I., Chelo, I. M., Goy, C., and Teotónio, H. (2014). Reproductive assurance drives transitions to self-fertilization in experimental *Caenorhabditis elegans*. *BMC Biology*, 12(1):93.

Turelli, M. (1988). Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution*, 42(6):1342–1347.

Turelli, M. and Barton, N. (1994). Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics*, 138:913–941.

Uyeda, J. C., Hansen, T. F., Arnold, S., and Pienaar, J. (2011). The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences*, 108:15908–15913.

Venables, W. and Ripley, B. (2002). *Modern Applied Statistics with S*. Statistics and Computing. Springer, New York.

Vladar, H. and Barton, N. (2014). Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics*, 197:749–767.

Walsh, B. and Blows, M. W. (2009). Abundant genetic variation + strong selection = multivariate genetic constraints: A geometric view of adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):41–59.

Walsh, B. and Lynch, M. (2018). *Evolution and Selection of Quantitative Traits*. Sinauer Associates, Inc., Sunderland.

Walter, G. M., Aguirre, J. D., Blows, M. W., and Ortiz-Barrientos, D. (2018). Evolution of genetic variance during adaptive radiation. *American Naturalist*, 191(4):E108–E128.

Whitlock, M. C. and Fowler, K. (1999). The changes in genetic and environmental variance with inbreeding in drosophila melanogaster. *Genetics*, 152(1):345–53.

Whitlock, M. C., Phillips, P. C., and Fowler, K. (2002). Persistence of changes in the genetic covariance matrix after a bottleneck. *Evolution*, 56(10):1968–75.

Yeh, S.-D., Saxena, A. S., Crombie, T. A., Feistel, D., Johnson, L. M., Lam, I., Lam, J., Saber, S., and Baer, C. F. (2017). The mutational decay of male-male and hermaphrodite-hermaphrodite competitive fitness in the androdioecious nematode C. elegans. *Heredity*, 120:1–12.

Zhang, X. S. and Hill, W. G. (2005). Predictions of patterns of response to artificial selection in lines derived from natural populations. *Genetics*, 169(1):411–25.