Quantitative assessment of observed versus predicted responses to selection

Christophe Pélabon,1,2 Elena Albertsen,1,3 Arnaud Le Rouzic,4 Cyril Firmat,5 Geir H. Bolstad,6 W. Scott Armbruster,7,8 and Thomas F. Hansen9

1 Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway
2 E-mail: Christophe.pelabon@ntnu.no
3 Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway
4 Évolution, Génomes, Comportement, Écologie, Université Paris-Saclay CNRS-IRD, Gif sur Yvette, France
5 INRAE, Université de Toulouse, UMR AGIR, Castanet-Tolosan Cedex, France
6 Norwegian Institute for Nature Research (NINA), Trondheim, Norway
7 School of Biological Sciences, University of Portsmouth, Portsmouth, UK
8 Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska, USA
9 Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

Received April 18, 2018
Accepted May 26, 2021

Although artificial-selection experiments seem well suited to testing our ability to predict evolution, the correspondence between predicted and observed responses is often ambiguous due to the lack of uncertainty estimates. We present equations for assessing prediction error in direct and indirect responses to selection that integrate uncertainty in genetic parameters used for prediction and sampling effects during selection. Using these, we analyzed a selection experiment on floral traits replicated in two taxa of the Dalechampia scandens (Euphorbiaceae) species complex for which G-matrices were obtained from a diallel breeding design. After four episodes of bidirectional selection, direct and indirect responses remained within wide prediction intervals, but appeared different from the predictions. Combined analyses with structural-equation models confirmed that responses were asymmetrical and lower than predicted in both species. We show that genetic drift is likely to be a dominant source of uncertainty in typically-dimensioned selection experiments in plants and a major obstacle to predicting short-term evolutionary trajectories.

**KEY WORDS:** breeder's equation, evolvability, G-matrix, indirect selection, Lande equation, correlated traits, artificial selection, Dalechampia.
forecasting in specific cases (e.g., Hansen et al. 2019; Shaw 2019; and see Hendry 2017, chapter 3 for a review of the performance of the Lande equation in predicting short-term evolution in natural populations).

The breeder’s equation also has somewhat mixed performance in predicting the outcome of univariate selection experiments in the lab (Sheridan 1988; Eisen 2005; Walsh and Lynch 2018, chapter 18), and its success in predicting correlated responses in traits under indirect selection is generally considered to be poor (e.g., Bohren et al. 1966; Rutledge et al. 1973; Palmer and Dingle 1986; Gromko et al. 1991, 1995; Cortese et al. 2002; Roff 2007). This is a serious concern, because indirect selection may well be the major component of selection on most traits in natural populations. The main advance brought by the G-matrix was the ability to predict correlated responses, and if this cannot be done reliably, then inferences about genetic constraint based on $G$ must be poor.

Most assessments of evolutionary predictions have been qualitative, however, and when prediction uncertainties are presented, they are usually given as estimation errors in realized heritabilities based on very simple models of the underlying genetic architecture, and further reduced to statements about “significant” difference between predictions and observations (e.g., Hill and Caballero 1992; Walsh and Lynch 2018, chap. 18). For correlated responses, uncertainties have almost never been presented, and claims of poor prediction of correlated responses are based largely on qualitative assessments (e.g., Roff 2007).

There are three main sources of error in predicting evolutionary responses: (i) errors due to discrepancies between the model used for prediction and the actual evolutionary process; (ii) errors made in estimating the parameters in the chosen prediction model, and (iii) errors due to inherent stochasticity in the response.

Deterministic discrepancies from simple predictive models such as the Lande or breeder’s equation are unlikely to be substantial in the first few generations, but as selection extends over more generations, the response may deviate due to a number of issues related to details of genetic architecture, inbreeding and counteracting natural selection (Le Rouzic et al. 2011). In principle, more detailed models can be fitted to evolutionary time series to estimate parameters describing such effects (e.g., Le Rouzic et al. 2010, 2011; Walsh and Lynch 2018, chapter 19), but this has rarely been done.

Various methods have been suggested to assess the effects of sampling error in quantitative genetic parameters on prediction variance (Tai 1979; Knapp et al. 1989; McCulloch et al. 1996; Conner et al. 2011). Stinchcombe et al. (2014) used a Bayesian approach and combined the Price equation with the Lande equation to estimate uncertainties of the predicted response to a single generation of selection from the posterior distribution of the G-matrix and the selection gradients (see also Careau et al. 2015).

As for the last source of error, few studies have assessed the effects of genetic drift on the prediction uncertainty. Building on Prout (1962) and Hill (1971), Hill (1974) provided an estimate for the drift variance of selection lines, and Sorensen and Kennedy (1983, 1984) showed how pedigree information analyzed with mixed-effect models could be used to incorporate these effects into the estimation of realized genetic parameters (see also Walsh and Lynch 2018, chapter 19). Combined with a Bayesian approach, this method has been used to estimate genetic parameters when the pedigree is known, but to our knowledge, it has not yet been implemented to estimate prediction intervals of selection responses.

In this paper, we report the results of a selection experiment on floral traits replicated in two taxa of the Dalechampia scandens (Euphorbiaceae) species complex and use these to illustrate some difficulties in predicting multivariate selection responses from estimated G-matrices. We present a simple pedigree-free equation to calculate the expected variance in the discrepancy between predicted and observed responses under truncation selection that incorporates both stochasticity in the observed response and uncertainty in the predicted response. With this, we assess the relative importance of the different sources of error in short-term selection experiments. To assess the discrepancy between the evolutionary model chosen to make the predictions (i.e. the Lande equation) and the evolutionary process that produced the selection responses, we further analyze the temporal dynamics of responses with structural-equation models that assume different genetic architectures. Finally, by reviewing parameter uncertainties in breeding experiments and the design of artificial-selection studies, we show that the large prediction uncertainties found in our experiment are not unusual for artificial-selection studies on plants.

**Theory: Prediction Error in Artificial Selection**

Consider a focal trait, $z$, which can be under direct selection with selection gradient $\beta_z$, or under indirect selection due to its correlation with another trait, $y$, under selection with gradient $\beta_y$. The Lande equation (Lande 1979; Lande and Arnold 1983) predicts the mean of the focal trait, $z$, in generation $t$ as

$$
\mu_z = \mu_{z_{t-1}} + V_A \beta_z + G_{z y} \beta_y,
$$

where $V_A$ is the additive genetic variance in $z$, $G_{z y}$ is the additive genetic covariance between $z$ and $y$. Motivated by our selection experiment with *Dalechampia*, we focus on the situation in which only one of the two traits is under selection, but we give the
key equations with selection on both traits to allow general application. We assume that the G-matrix and the selection gradient stay constant and hence do not include time notation on these. Equation 1 gives the partial change in the mean additive genetic value due to selection. In absence of other effects on mean phenotype (e.g., biased transmission, migration from genetically differentiated populations, non-random mating, or non-additive gene action), this change in mean genetic value will be the expected change in mean phenotype, around which the realized mean will be distributed due to sampling effects. Additionally, uncertainties in the estimation of genetic and selection parameters may lead to errors in the prediction that need to be considered. To combine these uncertainties, we derive the variance of the deviation between the predicted, $\mu$, and observed, $\bar{z}$, trait means at a given generation as:

$$Var[\bar{z} - \mu] = Var[\bar{z}] + Var[\mu], \quad (2)$$

which assumes that the statistical error in the prediction is independent from stochasticity due to sampling in the observed response. If the selection gradient is known without error and the G-matrix stays constant throughout the experiment, the variance in the prediction after $t$ generations is:

$$Var[\mu_t] = (Var[V_A] \beta^2 + Var[G_{xy}] \beta^2 + 2Cov[V_A, G_{xy}] \beta \beta_t) t^2, \quad (3)$$

where $Var[V_A]$, $Var[G_{xy}]$, and $Cov[V_A, G_{xy}]$ are the sampling variances and covariance in the estimates of genetic parameters. With selection on one trait only, this expression reduces to $Var[\mu_t] = Var[V_A] \beta^2 t^2$ for the direct response and $Var[\mu_t] = Var[G_{xy}] \beta^2 t^2$ for the correlated response. If all potential parents are phenotyped and their fitness known, as in our study, uncertainty in the selection gradient is small and limited to the measurement error of the phenotype. Known changes in the selection gradient from generation to generation can be accommodated by replacing the expressions $\beta^2 t^2$ with $(\sum \beta(t))^2$.

The variance in the observed selection response due to sampling is more complex. First, note that there are two distinct sampling effects on the mean of a quantitative trait. The first is the sampling of alleles we call genetic drift, and the second is the “sampling” of environmental effects to form the phenotypes in the new generation. We model the latter as the sampling variance of a mean from a normal distribution, which is $V_e/N$, where $V_e$ is the environmental variance, and $N$ is the population size (i.e., number of offspring). In contrast to genetic drift, this component does not accumulate over time.

In appendix 1, we derive the following equation for the variance in the per generation changes of an additive trait due to genetic drift under truncation selection in which exactly $N_p$ parents are picked from a population of $N$ individuals to make exactly $2N/N_p$ offspring each:

$$Var[\Delta z] \approx V_A \left( \frac{1}{N_p} - \frac{1 + 3F}{2N (1 + F)} - \frac{Var[w]}{N} \right), \quad (4)$$

where $F$ is the average inbreeding coefficient of the population, and $Var[w]$ is the variance in relative fitness among the genotypes in the population. The derivation assumes two alleles per locus and infinitesimal effects so that changes in allele frequency due to selection can be ignored. Linkage disequilibrium, dominance, and epistasis are also ignored.

Two special cases can help illustrate this equation. First, if there is no selection and parents are picked at random, then the variance in relative fitness is zero, and if also $F = 0$ then

$$Var[\Delta z] \approx V_A \left( \frac{1}{N_p} - \frac{1}{2N} \right), \quad (5)$$

which can be used to estimate the effect of genetic drift in control lines. Because the variance in relative genotypic fitness is small for low heritabilities, this equation will also be a good approximation under selection if the heritability of the selected trait is less than about 30% (Fig. 1). Note also that assuming $N = \infty$ yields the standard equation for the drift variance (Lande 1976): $Var[\Delta z] = V_A/N_p$. Second, if there is truncation selection and the heritability in the population is unity, so that the genotypic value equals the phenotypic value, then the variance in relative
fitness of genotypes is \( \text{Var}[w] = (N - N_p)/N_p \), and if \( F = 0 \), equation 4 reduces to

\[
\text{Var}[\Delta z] \approx \frac{V_A}{2N^2}
\]

In this case, the sampling of parental genotypes is deterministic, and the only stochasticity comes from sampling alleles from parents during mating. If the heritability is not unity, then a given genotype may or may not be picked as a parent in different realizations due to its random environmental effect, and this will reduce the variance in relative fitness. Hence, the sampling variance of the mean will be bounded between equations 5 and 6 and move from equation 5 towards equation 6 as heritability of the selected trait increases (Fig. 1). In Appendix 2, we outline an approximation to the variance in relative genotypic fitness as a function of the heritability of the trait under selection that we used to make prediction intervals.

To summarize, the expected variance of the prediction error after \( t \) generations is

\[
\text{Var}[z - \mu_z] = \text{Var}[G_z] \beta_t^2 \text{Var}[w] t + \frac{V_A}{N} \left( \frac{1 + 3F}{2N(1 + F)} - \frac{\text{Var}[w]}{N} \right) t + \frac{V_e}{N}.
\]

\( G_z \) is the additive covariance between the focal trait, \( z \), and the trait under selection, \( y \) (which could be the focal trait itself). The variance in fitness is a function of the heritability of the trait under selection, which is not necessarily the focal trait. This prediction ignores the effects of selection and genetic drift on the G-matrix, which is assumed to stay constant throughout the experiment. Ideally, drift variance in the G-matrix should be added to the variance in the first term. The prediction also ignores the effects of sampling on the realized selection gradient, which will vary stochastically in a finite population, but this effect is not an error in the prediction from an observed selection gradient.

**Materials and Methods**

**STUDY SPECIES AND TRAIT MEASUREMENTS**

*Dalechampia scandens* (Euphorbiaceae) is a perennial Neotropical vine with flowers arranged in pseudanthial inflorescences (blossoms), each consisting of a male subinflorescence of 10 staminate flowers and a female subinflorescence of three pistillate flowers (Fig. 2). The male subinflorescence also contains a gland producing a triterpenoid resin as reward for pollinating bees that use resin in nest construction (Armbruster 1984, 1985, 1986, 1988, 1993), and the amount of resin offered to the pollinator depends on the gland size (Armbruster 1984; Pélabon et al. 2012). In interpopulation and interspecies comparisons, blossoms with larger resin glands tend to attract larger pollinators (Armbruster 1985, 1988). The blossom is subtended by two large involucral bracts, which are white or light green in the study species. Phenotypic selection studies on several *Dalechampia* species and populations have shown that pollinators choose blossoms based either on the size of the involucral bracts (the signal), or on the size of the gland (the reward), thus causing selection on these traits (Armbruster et al. 2005; Bolstad et al. 2010; Pérez-Barrales et al. 2013; Albertsen et al. 2021). Additionally, the *Dalechampia* blossom is an integrated structure in which involucral-bract size is phenotypically and genetically correlated with resin-gland size (Armbruster 1991; Hansen et al. 2003b; Armbruster et al. 2004; Pélabon et al. 2012). In this study, we performed artificial selection on the size of the resin-producing gland and recorded the
direct response in gland size and the correlated response in bract size.

Gland size was measured as the area of the resin-secreting surface (Gland Area, GA), and bract size was measured as the area of the upper bract (Upper Bract Area: UBA, see Fig. 2 for measurements details). Blossoms go through a series of ontogenetic stages during which they increase in size (Armbruster 1991; Opedal et al. 2016b). To reduce ontogenetic variation in size, we measured the blossoms on the first day of the bisexual phase, that is, when the first one-to-three male flowers were open. Measurements of the blossoms in the two diallels, the starting generation (F₀) and the first three episodes of selection were performed by one observer (CP) using 5× stereo magnifying lenses (Optivisor) and digital calipers with 0.01 mm precision. Measurements of the last generation were done by one observer (EA) using the same measuring devices. There was no evidence of systematic difference between observers in measurements on a common subset of blossoms, and because all selected lines were measured by a single observer at each generation, we do not expect the different observers to have affected the outcome of the study. During the diallels and the artificial-selection experiment, plants were placed randomly on four tables in a single room in the greenhouse of the Department of Biology, NTNU (Trondheim, Norway) and moved regularly during the measurement period to reduce positional effects.

Interpopulation crosses (Pélabon et al. 2004a, 2005) and molecular studies (Falahati-Anbaran et al. 2013, 2017) have shown that D. scandens is a complex of two or more distinct, yet undescribed species. In this study, diallels and artificial-selection experiments were performed on two populations from distinct species. Individuals from the first species are descended from seeds collected near Tulum, Mexico (20°13’ N; 87°26’ W), and individuals from the second species are descended from seeds collected near Tovar, Venezuela (8°21’ N; 71°46’ W).

**DIALLEL EXPERIMENT**

We estimated the G-matrix for several blossom traits in the two species with two diallels completed in 1999 to 2000 and in 2005 to 2006 for Tulum and Tovar, respectively. Methods and results of these diallels are presented in Bolstad et al. (2014). Briefly, seeds collected from different blossoms in the field were sown in the greenhouse and, upon maturity, plants were crossed in a partial diallel design with twelve and nine 5 × 5 blocks for Tulum and Tovar, respectively. Self and reciprocal crosses were included. Two seeds per cross were sown to produce the offspring generation (Tulum: n = 523 individuals; Tovar: n = 419 individuals) and two blossoms per individual were measured.

**SELECTION EXPERIMENT**

We conducted four episodes of selection on each species. Due to space limitation in the greenhouse, we alternated by generation the species being grown and measured, but the up- and down-selected lines as well as the control line from a given species were always grown simultaneously in the same room in the greenhouse with plants from the different lines placed randomly on the tables. To form the starting populations (F₀), we performed a stratified sampling of 100 individuals among the diallel blocks and families to have populations as similar as possible to the ones from which G-matrices were estimated. We did not include the diagonal of the diallel (i.e., selfed offspring) in this sampling. For Tovar, we sampled plants directly from the diallel experiment. For Tulum, whose diallel was completed first, all plants were selfed at the end of the diallel experiment and preserved as seeds until the start of the selection experiment. We sampled among these seeds to form the F₀ in Tulum, grew them, and measured their blossoms at maturation. Hence, the first episode of selection on the Tulum species was performed on selfed individuals.

This episode of selfing in the Tulum line may affect the response to selection by altering the trait mean due to dominance and by inflating additive variance by a factor of 1.5 for the first generation of selection (i.e., by 1+F, where F is the inbreeding coefficient, Lynch and Walsh 1998; Shaw et al. 1998). We thus multiplied G by 1.5 in our prediction for the first episode of selection in Tulum. We did not correct for dominance effects on the mean, however, because previous experiments with this species have shown little evidence of either dominance variance (Hansen et al. 2003a) or inbreeding effects (Pélabon et al. 2004b; Opedal et al. 2015).

We performed direct selection to increase or decrease the area of the resin-producing gland. We started the experiment by measuring gland and bract area on four blossoms per individual in the 100 individuals forming the F₀ and chose the 16 individuals with the smallest or largest mean gland area to produce the first generation of the down- and up-selected lines, respectively. Within each line, 64 new families were produced by pollinating each of the 16 individuals with pollen from four other individuals among the 16 selected. Each individual thus contributed equally to the next generation, four times as sire and four times as dam. Reciprocal crosses were avoided so that none of the 64 families shared more than one parent. Details of the crossing method and seed collection are presented in Pélabon et al. (2015). We kept track of the pedigree and never crossed individuals with a coefficient of relatedness higher than 0.10 to reduce inbreeding.

We sowed two or three seeds from each of the 64 families and kept one individual per family to form the F₁ generation. We then measured three blossoms per individual and selected the 16 (25%) most extreme individuals to produce the next generation. Selection gradients were calculated at each generation as the selection differential (mean of the selected individuals minus the mean of the population before selection) divided by the phenotypic variance of the line in that generation. From the F₁
generation and onwards, we maintained a population size of 64 individuals, and selected 16 of them. In practice, we kept the 20 most extreme individuals at each generation to replace individuals with poor blossom production to assure a total of 16 reproducing individuals. The number of individuals measured in each selected line at each generation varied slightly due to occasional failures to flower (Supporting Information S1 and S2).

For each species, we generated a control line from random crosses among individuals of the F₀ to assess phenotypic changes due to uncontrolled variation in the greenhouse environment. At each generation, several individuals from these control lines were grown simultaneously with the selected lines and randomly crossed while avoiding selfing to provide seeds for the next generation. For logistical reasons, the size of these control lines varied each generation (Appendix S1 and S2) and we were unable to measure them at the third generation (F₃).

The phenotypic values observed in the last generation (F₃) were unusual, particularly for bract size in the Tulum population (Appendix S3). This was most likely due to unusual conditions in the greenhouse. We thus regrew the last generation from seeds from the same crosses and measured it anew for both species. This second set of measurements provided qualitatively similar results as the first set regarding the differences between the up- and down-selected lines, but the phenotypic values were closer to the expected ones (see results). Therefore, we used this second set of measurements for the last generation in the analyses that follow.

**STATISTICAL ANALYSES**

**Genetic parameters from the diallel experiment**

The analyses of the two diallels are presented in Bolstad et al. (2014). For each species, we estimated the additive genetic variances and covariance for gland area and bract area together with their credible intervals. Using the R package MCMCglmm (Hadfield 2010) we fitted the following model:

\[
\begin{align*}
z_{ijk} &= \mu_i + a_{ij} + b_{ij} + d_{ik} + s_{ijk} + q_{ijk},
\end{align*}
\]

where \(z\) is the trait value, \(\mu\) is the trait mean, \(a\) is the breeding value, \(b\) is the non-genetic plant effect, \(d\) is the measurement date, \(s\) is the number of male flowers open when the blossom was measured (1, 2, or 3), and \(q\) is the within-plant residual effect. The subscripts \(i\), \(j\), and \(k\) represent the trait, plant, and blossom, respectively. We accounted for temporal variation in the greenhouse environment by including measurement date \(d\) as a random factor. The random effects are assumed to be distributed as \(a \sim N(0, G \otimes A)\), \(b \sim N(0, B \otimes I)\), \(d \sim N(0, F \otimes I)\), and \(q \sim N(0, E \otimes I)\), where \(A\) is the relatedness matrix, \(I\) is the identity matrix and \(\otimes\) is the Kronecker product. The model estimates the additive genetic variance matrix \(G\), the among-plant environmental variance ma-

**Comparing observed and predicted response to selection**

We compared the observed responses to selection with the predictions from the multivariate Lande equation (Eq. 1) with genetic variances and covariances estimated from the two diallels. We constructed 95% prediction intervals from the variance in equation 7 by assuming that \(\bar{z}_i - \mu_i\) is normally distributed with mean zero. This is an approximation because the estimation errors of the elements in \(G\) are not normally distributed. We assumed no inbreeding (i.e., \(F = 0\)) in Tovar, and in Tulum, we multiplied \(G\) by 1.5 in the F₀ to account for the episode of selfing between

---

C. PÉLABON ET AL.
the estimation of $G$ and the start of the selection. Because crosses among selected parents avoided crosses among relatives, we further assumed no inbreeding from the $F_1$ and onward.

To control for variation in the greenhouse environment across generations, we centered the response to selection on the mean of the up- and down-selected lines or we corrected the responses for changes in the control line. In the former case, the responses of the up- and down-selected lines are forced to be symmetrical. This approach also assumes that environmental effects are identical in the up- and down-selected lines. Correcting for changes in the control line avoids these assumptions but yields less precise predictions due to the estimation error in the control mean.

**Analyzing the temporal dynamics of the responses to selection**

In a second set of analyses, we used the R package SRA (Selection Response Analysis; Le Rouzic et al. 2011) to estimate the realized genetic parameters from the observed response to selection. The SRA package fits deterministic population-genetics models with different genetic architectures (e.g., epistasis, linkage disequilibrium, or finite number of loci) to time series of selection responses (Le Rouzic et al. 2010, 2011, Le Rouzic 2014). These analyses therefore assess the influence of discrepancies between the model chosen to make the predictions (Lande equation) and the evolutionary processes that generated the selection responses. The code was modified to include two traits, one selected and one correlated, and to estimate the genetic covariance between them (Supporting Information S6). The model estimates the realized additive genetic variance of the selected trait, $V_A$, and covariance, $G_{zy}$, with the correlated trait, as well as the environmental variances $E_x$ and $E_y$. It is not possible to estimate the realized additive genetic variance for the correlated trait because direct selection on this trait is assumed to be zero. The full dataset needed to fit the time-series model included for each generation, the sample size, the phenotypic means for both traits before and after selection, and their associated phenotypic variances (Supporting Information S1 and S2).

Although often assumed to be constant over a few episodes of selection, $G$ may change due to allele-frequency changes, directional epistasis, and changes in linkage disequilibrium (the Bulmer effect; Bulmer 1971). We compared models fitted with or without a Bulmer effect, but due to the limited number of generations, we could not fit more complex models including epistasis or major-effect loci. We also tested for the occurrence of asymmetry in the response to selection (e.g., Bohren et al. 1966; Frankham 1990; Bell 2008, Walsh and Lynch 2018 chapter 18) by allowing variances to differ in the two selected directions with and without correcting for changes in the control line.

At each generation, we calculated the phenotypic correlation and the slope of the regression of log(UBA) on log(GA) using mixed-effect models with plant identity as a random effect. All statistical analyses were performed with R 4.0.2 (R core team, 2020).

**Results**

**GENETIC VARIATION AND EVOLVABILITY**

The G-matrices estimated from the two diallels are presented in Table 1. Because we conducted analyses on natural-log-transformed data, the additive genetic variances can be interpreted as mean-scaled evolvabilities sensu Hansen et al. (2003a, 2011). The unconditional evolvabilities of gland and bract area were 1.05% and 1.41% in Tovar and 0.73% and 0.84% in Tulum. These are moderately high evolvabilities for morphological traits (Hansen and Pélabon 2021). The genetic covariances between the two traits were 0.57 in Tovar and 0.49 in Tulum, which should generate robust correlated responses to selection. The correlations are not strong enough to constitute a major constraint on evolution, however, because conditioning traits on each other, sensu Hansen et al. (2003b), would only reduce their evolvabilities by 22% in Tovar and 40% in Tulum.

**DIRECT RESPONSE TO SELECTION**

Gland area responded to selection in both species, but the response diminished after the second generation (Fig. 3 and 4; Supporting Information S3), and despite a good fit in the first generation, responses after four episodes of selection were nearly half the prediction for the Tovar lines and 30% lower for the Tulum lines. Still, observed responses remained within the 95% prediction intervals, although barely so for Tovar. Thus, if we neglect the temporal dynamics of the response, the discrepancies could be explained by a combination of sampling stochasticity in the response and uncertainty in the prediction due to estimation error in the additive genetic variance (Table 2). For the mean-centered responses (Fig. 3), 60% to 70% of the prediction error variance was due to sampling effects in the first generation, but by the last generation this has shifted to almost 80% being due to estimation error in the additive variance. For the control-corrected responses (Fig. 4), the contribution of the sampling effects was larger and remained above 50% of the total error variance even after four episodes of selection (Table 2).

The realized evolvabilities estimated from the selection response analysis were smaller than the evolvabilities estimated from the diallels (Table 3). Including a Bulmer effect improved the fit for the Tovar data, but had little impact on the estimated evolvabilities. When corrected for changes in the control lines, the direct responses in gland area were asymmetrical, with a larger decrease for both species (Table 3).
Table 1. Trait means (SD), genetic (G), environmental (E), and phenotypic (P) variance matrices for the Tovar and Tulum populations of Dalechampia scandens estimated from the diallel experiments with the full data set (including selfed).

|         | Gland area (GA) | Upper bract area (UBA) |
|---------|----------------|------------------------|
| **Tovar** |                |                        |
| Mean (SD) | 17.56 (3.30) mm² | 385.0 (75.7) mm² |
| **G** | GA 1.05 ± 0.20 (0.69; 1.44) | 0.57 ± 0.15 (0.30; 0.87) |
|         | UBA 0.47 ± 0.08 (0.30; 0.62) | 1.41 ± 0.20 (1.07; 1.80) |
| **E** | GA 2.27 ± 0.16 (1.93; 2.55) | 1.14 ± 0.11 (0.95; 1.35) |
|         | UBA 0.57 ± 0.03 (0.51; 0.63) | 1.76 ± 0.12 (1.56; 2.02) |
| **P** | GA 3.32 ± 0.19 (2.98; 3.68) | 1.72 ± 0.15 (1.44; 1.99) |
|         | UBA 0.53 ± 0.03 (0.47; 0.58) | 3.17 ± 0.19 (2.84; 3.60) |
| **Tulum** |                |                        |
| Mean (SD) | 20.20 (5.26) mm² | 407.0 (81.9) mm² |
| **G** | GA 0.73 ± 0.22 (0.37; 1.16) | 0.49 ± 0.13 (0.23; 0.75) |
|         | UBA 0.63 ± 0.11 (0.43; 0.82) | 0.84 ± 0.14 (0.60; 1.13) |
| **E** | GA 5.63 ± 0.30 (5.03; 6.20) | 1.60 ± 0.15 (1.29; 1.87) |
|         | UBA 0.45 ± 0.03 (0.38; 0.50) | 2.27 ± 0.13 (2.06; 2.53) |
| **P** | GA 6.36 ± 0.28 (5.82; 6.91) | 2.09 ± 0.16 (1.78; 2.37) |
|         | UBA 0.47 ± 0.03 (0.41; 0.51) | 3.11 ± 0.15 (2.81; 3.41) |

For each matrix, variances are reported on the diagonal, covariances above, and correlations below the diagonal along with their standard error and credible intervals (genetic; calculated with HPDinterval.mcmc with default probability = 0.95) or confidence intervals (phenotypic) between parentheses. Means and SDs are in mm², variances and covariances are in log(mm²) × 100. The E variance matrices are sums of two components: the residual of the diallel model and the individual error level. Genetic correlations are estimated from the posterior distribution of the MCMCglmm models estimating genetic variances and covariances between GA and UBA. Sample size are 820 for Tovar and 1046 for Tulum.

**CORRELATED RESPONSE TO SELECTION**

The correlated responses of bract area differed between the two species. In Tovar, the correlated responses paralleled the direct responses by matching the prediction in the first generation before diminishing in the following generations. In Tulum, the correlated responses were smaller than predicted already after the first episode of selection and were practically absent in the following generations (Fig. 3 and 4). Nevertheless, even the weak response in Tulum remained within the prediction intervals, which were large relative to the predicted response. This was particularly striking for the control-corrected responses for which the error intervals always included zero response (Fig. 4). Accordingly, realized additive genetic covariances estimated from the selection-response analysis were smaller than the covariances estimated from the diallels, especially in Tulum (Table 3).

Neither the among-individual phenotypic correlations nor the slopes of the regression of bract area on gland area estimated within each line at each generation changed markedly during the experiment (Table 4).

**Discussion**

Discrepancies between observed and predicted responses to selection have been attributed either to imprecise estimation of genetic parameters (Sheridan 1988; Eisen 2005; Roff 2007) or to changes in G during selection (Hill and Caballero 1992; Roff 2007). Few studies have considered the impact of genetic drift on the responses to selection, however. This is not because the importance of drift has gone unrecognized (e.g. Falconer 1973; Nicholas 1980; Walsh & Lynch 2018), or due to a lack of methods for assessing its effects (Hill 1974; Sorensen & Kennedy 1983, 1984), but it may be due to the relative inaccessibility of these methods and the lack of a common framework to account for the different sources of uncertainty. We have presented a simple equation for the error variance due to the combined effects of genetic drift and uncertainty in genetic parameters. Our equation is still limited to truncation selection and does not incorporate changes in genetic parameters due to selection or drift, but it can still be used for a priori pedigree-independent assessment of uncertainty in predicting short-term responses to selection. When applied to our own selection experiment, this method showed how sampling effects may dominate the uncertainty during early generations, making it difficult to predict selection responses over few generations, especially for correlated traits.

Our treatment of genetic drift differs in many aspects from that of Hill (1971, 1974). First, we have based our sampling on alleles while Hill sampled breeding values. Hill further assumed a normal distribution of breeding values and environmental effects, while we did not make distributional assumptions for the breeding values. In contrast, we did assume two alleles.
Table 2. Contribution of the different sources of uncertainty to the prediction intervals for data centered on the mean of the Up- and Down-selected lines (Up-Down centered) or corrected for changes in the control line (Control corrected).

| Trait | Source of error | Up-down centered | Control corrected |
|-------|-----------------|------------------|-------------------|
|       | Trait | Gen. | G | D | E | Var\(^\#\) | %G | %D | %E | Resp. | 2SE | Rel. err. | Var\(^\$\) | %G | %D | %E | 2SE | Rel. err. |
| Tovar | GA 0 | 0.0 | 0.0 | 3.6 | 1.8 | 0.0 | 0.0 | 100.0 | 0.00 | 0.03 | NA | 7.1 | 0.0 | 0.0 | 100.0 | 0.05 | NA |
|       | 1 | 3.8 | 7.8 | 3.6 | 9.4 | 40.0 | 41.1 | 18.9 | 0.11 | 0.06 | 27.2 | 26.4 | 14.3 | 58.7 | 27.0 | 0.10 | 45.6 |
|       | 2 | 15.1 | 15.5 | 3.6 | 24.7 | 61.3 | 31.5 | 7.2 | 0.21 | 0.10 | 23.1 | 53.3 | 28.3 | 58.3 | 13.4 | 0.15 | 34.0 |
|       | 3 | 34.0 | 23.3 | 3.6 | 47.4 | 71.7 | 24.6 | 3.8 | 0.30 | 0.14 | 22.7 | 87.7 | 38.8 | 53.1 | 8.1 | 0.19 | 30.9 |
|       | 4 | 60.4 | 31.1 | 3.6 | 77.8 | 77.7 | 20.0 | 2.3 | 0.40 | 0.18 | 22.0 | 129.7 | 46.6 | 47.9 | 5.5 | 0.23 | 28.5 |
|       | UBA 0 | 0.0 | 0.0 | 2.8 | 1.4 | 0.0 | 0.0 | 100.0 | 0.00 | 0.02 | NA | 5.5 | 0.0 | 0.0 | 100.0 | 0.05 | NA |
|       | 1 | 2.2 | 10.6 | 2.8 | 8.9 | 24.8 | 59.7 | 15.5 | 0.06 | 0.06 | 48.0 | 28.9 | 7.6 | 73.3 | 19.1 | 0.11 | 86.6 |
|       | 2 | 8.8 | 21.2 | 2.8 | 20.8 | 42.4 | 51.0 | 6.6 | 0.12 | 0.09 | 38.5 | 56.7 | 15.5 | 74.8 | 9.7 | 0.15 | 63.7 |
|       | 3 | 19.8 | 31.8 | 2.8 | 37.0 | 53.4 | 42.9 | 3.7 | 0.17 | 0.12 | 36.5 | 88.8 | 22.3 | 71.5 | 6.2 | 0.19 | 56.6 |
|       | 4 | 35.2 | 42.4 | 2.8 | 57.7 | 60.9 | 36.7 | 2.4 | 0.22 | 0.15 | 34.5 | 125.4 | 28.1 | 67.6 | 4.4 | 0.22 | 50.9 |
| Tulum | GA 0 | 0.0 | 0.0 | 8.8 | 4.4 | 0.0 | 0.0 | 100.0 | 0.00 | 0.04 | NA | 17.6 | 0.0 | 0.0 | 100.0 | 0.08 | NA |
|       | 1 | 1.3 | 5.9 | 8.8 | 10.6 | 32.5 | 25.7 | 41.7 | 0.10 | 0.06 | 34.0 | 31.9 | 10.8 | 34.0 | 55.2 | 0.11 | 59.2 |
|       | 2 | 13.7 | 10.9 | 8.8 | 23.6 | 58.3 | 23.0 | 18.7 | 0.13 | 0.10 | 36.7 | 53.1 | 25.9 | 40.9 | 33.2 | 0.15 | 55.0 |
|       | 3 | 30.9 | 16.3 | 8.8 | 43.4 | 71.1 | 18.8 | 10.1 | 0.19 | 0.13 | 34.5 | 81.1 | 38.1 | 40.2 | 21.7 | 0.18 | 47.2 |
|       | 4 | 54.9 | 21.7 | 8.8 | 70.2 | 78.2 | 15.5 | 6.3 | 0.25 | 0.17 | 33.8 | 116.0 | 47.4 | 37.5 | 15.2 | 0.22 | 43.5 |
|       | UBA 0 | 0.0 | 0.0 | 3.6 | 1.8 | 0.0 | 0.0 | 100.0 | 0.00 | 0.03 | NA | 7.1 | 0.0 | 0.0 | 100.0 | 0.05 | NA |
|       | 1 | 1.3 | 6.2 | 3.6 | 6.2 | 20.6 | 50.5 | 28.9 | 0.06 | 0.05 | 39.0 | 20.9 | 6.1 | 59.7 | 34.2 | 0.09 | 71.6 |
|       | 2 | 5.1 | 12.5 | 3.6 | 13.1 | 38.8 | 47.5 | 13.6 | 0.09 | 0.07 | 40.9 | 37.1 | 13.7 | 67.1 | 19.2 | 0.12 | 68.9 |
|       | 3 | 11.4 | 18.7 | 3.6 | 22.6 | 50.7 | 41.4 | 7.9 | 0.13 | 0.10 | 37.3 | 55.9 | 20.5 | 66.8 | 12.8 | 0.15 | 58.6 |
|       | 4 | 20.3 | 24.9 | 3.6 | 34.6 | 58.8 | 36.0 | 5.2 | 0.17 | 0.12 | 35.6 | 77.3 | 26.3 | 64.5 | 9.2 | 0.18 | 53.2 |

\(^\#\) Var = V_A + 1/2 \times \text{Drift} + 1/2 \times \text{Env.}

\(^\$\) Var = V_A + 2 \times \text{Drift} + 2 \times \text{Env.}

Uncertainties in 100 \times (\log(mm^2))^2 are due to error in genetic parameters, G, genetic drift, D, and environmental variation, E (see Eq. 7 for the calculation of each contribution). The total error variance is reported in the Var columns and the relative contributions are reported in the columns %G, %D and %E, respectively. The width of the prediction intervals in Fig. 3 and 4 are calculated as 2SE = 2\sqrt{Var}/100. We also give the relative error (Rel.) in the predicted response (Resp.) as 100 \times SE/Resp.
Table 3. Estimated (diallel) and realized (artificial selection) genetic variance of gland area (GA) and covariance between gland area and upper bract area (UBA) in the two species of *D. scandens*.

| Species      | Data               | Method          | ΔAICc | Direction | \( G_{\log(GA)} \) (95% CI) | \( G_{\log(GA), \log(UBA)} \) (95% CI) |
|--------------|--------------------|-----------------|-------|-----------|-----------------------------|--------------------------------|
| Tovar        | Diallel            | MCMCglmm        | 359.9 | Up        | 0.49 (0.46; 0.52)           | 0.33 (0.30; 0.37)              |
|              | Symmetric          |                 |       |           |                             |                                |
|              | Symmetric + Bulmer |                 | 352.5 |           | 0.53 (0.50; 0.57)           | 0.33 (0.30; 0.37)              |
|              | Asymmetric         |                 | 1.4   | Up        | 0.26 (0.21; 0.32)           | 0.52 (0.46; 0.59)              |
|              | Down               |                 |       |           | 0.72 (0.66; 0.78)           | 0.17 (0.11; 0.24)              |
|              | Asymmetric + Bulmer|                 | 0     | Up        | 0.27 (0.21; 0.34)           | 0.52 (0.45; 0.59)              |
|              | Down               |                 |       |           | 0.80 (0.73; 0.87)           | 0.17 (0.11; 0.24)              |
|              | Symmetric          |                 | 239.8 |           | 0.49 (0.46; 0.52)           | 0.34 (0.31; 0.37)              |
|              | Symmetric + Bulmer |                 | 234.6 |           | 0.54 (0.50; 0.58)           | 0.34 (0.31; 0.37)              |
|              | Asymmetric         |                 | 1.6   | Up        | 0.40 (0.34; 0.46)           | 0.33 (0.27; 0.40)              |
|              | Down               |                 |       |           | 0.58 (0.53; 0.65)           | 0.35 (0.29; 0.41)              |
|              | Asymmetric + Bulmer|                 | 0     | Up        | 0.44 (0.38; 0.51)           | 0.34 (0.27; 0.40)              |
|              | Down               |                 |       |           | 0.63 (0.57; 0.70)           | 0.35 (0.29; 0.41)              |
| Tulum        | Diallel            | MCMCglmm        | 462.8 |           | 0.73 (0.37; 1.16)           | 0.49 (0.23; 0.75)              |
|              | Symmetric          |                 |       | Up        | 0.38 (0.34; 0.42)           | 0.14 (0.11; 0.17)              |
|              | Symmetric + Bulmer |                 | 462.8 |           | 0.39 (0.35; 0.43)           | 0.14 (0.11; 0.17)              |
|              | Asymmetric         |                 | 5.5   | Up        | 0 (0; Inf)                  | 0.17 (0.11; 0.19)              |
|              | Down               |                 |       |           | 0.84 (0.78; 0.9)            | 0.13 (0.07; 0.19)              |
|              | Asymmetric + Bulmer|                 | 0     | Up        | 0 (0; Inf)                  | 0.17 (0.11; 0.23)              |
|              | Down               |                 |       |           | 0.90 (0.83; 0.97)           | 0.13 (0.07; 0.19)              |
|              | Symmetric          |                 | 242.0 |           | 0.38 (0.35; 0.42)           | 0.15 (0.11; 0.18)              |
|              | Symmetric + Bulmer |                 | 243.0 |           | 0.40 (0.36; 0.44)           | 0.15 (0.11; 0.18)              |
|              | Asymmetric         |                 | 0     | Up        | 0.27 (0.22; 0.34)           | 0.16 (0.10; 0.22)              |
|              | Down               |                 |       |           | 0.51 (0.45; 0.58)           | 0.13 (0.07; 0.20)              |
|              | Asymmetric + Bulmer|                 | 1.6   | Up        | 0.28 (0.22; 0.35)           | 0.16 (0.10; 0.22)              |
|              | Down               |                 |       |           | 0.53 (0.46; 0.61)           | 0.13 (0.07; 0.20)              |

*The bivariate SRA models assume constant genetic covariance (see Appendix S6), therefore estimates of the genetic covariance are not affected by the Bulmer effect. All estimates are calculated from log-transformed data and \( \times 100 \). Estimates for the realized responses are calculated for the pooled up- and down-selected lines assuming symmetric response, with or without Bulmer effect, or assuming different variances in the up and down lines, using the raw data or the data corrected for the control (95% credible intervals for the estimated parameters and 95% confidence interval for the realized values in parentheses).*
Per locus, infinitesimal effects of alleles, and no variation among selected parents in number of offspring. Second, we sampled parents without replacement from a finite population, while Hill assumed sampling of the measured individuals from an infinite zygote pool. This generates a difference in the equations even in the absence of selection. The treatment of selection is also different. Hill (1971) and Prout (1962) considered the variance in breeding values conditionally on phenotypes and thus neglected a component due to variation in phenotypes among selected parents. Hill (1974) did consider this but used a different approximation from the one we used and suggested that this effect could be ignored. We have shown that the effect of selection on the drift variance is a non-ignorable function of the genetic variance in relative fitness generated by the selection scheme. Unfortunately, under truncation selection, this variance can be given analytically only in special cases, and we could only provide a crude approximation for the general case. We also reiterate that we have not included the effects of either drift or selection on the G-matrix.

In the additive infinitesimal model, genetic drift will generate a pattern equivalent to Brownian motions of the population mean (Lande 1976), and the variance of the mean among independent replicate lines should increase linearly with time (i.e., generations). Thus, if the mean selection response increases linearly with time, the relative prediction error due to drift would decrease with the square root of time. In contrast, the prediction

---

Figure 3. Observed and predicted response to selection with data centered on the generation mean of the up- and down-selected lines. For the two species of *D. scandens* the responses are shown for gland area (GA), which is the selected trait, and bract area (UBA), which is the correlated trait. Observed responses in trait means (±2SE) are given as the dotted lines. The predicted responses with their prediction intervals (±2SE) are represented by the black lines and shaded area. Control lines are reported in grey with their prediction intervals in light grey.
Figure 4. Observed and predicted response to selection with responses corrected for changes in the control lines. No results are given for the 3rd generation due to the absence of a control. See Figure 3 for definitions of symbols and shaded areas.

To estimate asymmetry and correct for environmental variation and inbreeding, it is customary to subtract the changes from a control line. This method has the unfortunate consequence of increasing the imprecision of the predictions because the uncertainty of the control line is then incorporated into the imprecision of the selected lines (Nicholas 1980, and compare Fig. 3 and 4). Additionally, the effect of genetic drift is often larger in control lines due to smaller sample sizes (e.g. Worley and Barrett 2000; Sarkissian and Harder 2001), and because the more deterministic choice of individuals in selected lines reduces sampling effects (compare equations 5 and 6). This problem may be mitigated by increasing the size of the control line or maintaining replicated control lines with the same effective population sizes as the selected lines. This second method allows assessing inbreeding...
Table 4. Regression slope on natural-log-transformed data and phenotypic correlation (r) between gland area (GA) and upper bract area (UBA) in the different lines at each generation.

| Generation | Up-selected lines | Down-selected lines | Control |
|------------|-------------------|---------------------|---------|
|            | Slope (±SE)       | r (95% CI)          | Slope (±SE) | r (95% CI) | Slope (±SE) |  r (95% CI) |
| Tovar      |                   |                     |         |         |             |           |
| 0          | 0.56 (±0.04)      | 0.50 (0.33; 0.63)   | 0.56 (±0.04) | 0.50 (0.33; 0.63) | 0.56 (±0.04) | 0.50 (0.33; 0.63) |
| 1          | 0.78 (±0.05)      | 0.43 (0.21; 0.61)   | 0.75 (±0.06) | 0.44 (0.21; 0.62) | 0.64 (±0.04) | 0.56 (0.43; 0.67) |
| 2          | 0.44 (±0.05)      | 0.59 (0.40; 0.73)   | 0.55 (±0.06) | 0.46 (0.25; 0.63) | 0.54 (±0.05) | 0.38 (0.19; 0.54) |
| 3          | 0.58 (±0.06)      | 0.40 (0.17; 0.59)   | 0.32 (±0.06) | 0.31 (0.07; 0.52) | NA           | NA         |
| 4          | 0.52 (±0.05)      | 0.50 (0.32; 0.65)   | 0.55 (±0.04) | 0.47 (0.28; 0.63) | 0.60 (±0.05) | 0.56 (0.40; 0.69) |
| Tulum      |                   |                     |         |         |             |           |
| 0          | 0.57 (±0.04)      | 0.57 (0.42; 0.69)   | 0.57 (±0.04) | 0.57 (0.42; 0.69) | 0.57 (±0.04) | 0.57 (0.42; 0.69) |
| 1          | 0.51 (±0.06)      | 0.46 (0.25; 0.63)   | 0.55 (±0.06) | 0.50 (0.30; 0.66) | 0.60 (±0.04) | 0.56 (0.43; 0.66) |
| 2          | 0.51 (±0.06)      | 0.42 (0.19; 0.60)   | 0.36 (±0.06) | 0.36 (0.12; 0.57) | 0.44 (±0.05) | 0.43 (0.19; 0.62) |
| 3          | 0.55 (±0.07)      | 0.52 (0.31; 0.68)   | 0.41 (±0.06) | 0.30 (0.05; 0.51) | NA           | NA         |
| 4          | 0.37 (±0.05)      | 0.35 (0.14; 0.53)   | 0.24 (±0.04) | 0.40 (0.20; 0.57) | 0.40 (±0.06) | 0.46 (0.18; 0.67) |
Relative error in evolvability (in %)

Number of estimates

0  2  0  4  0  6  0  8  0  1  0  0

>100%

Figure 5. Distribution of relative errors in evolvability for 519 estimates from 40 studies. The median is 36% excluding 38 estimates with a relative error larger than 100% (see Supplement S6 for details). The relative error in the evolvability of gland area obtained from the diallel experiments were 18% in Tovar and 29% in Tulum. The relative error in evolvability is calculated as 100 × the standard error in evolvability divided by the evolvability.

Finally, the lower than expected response to selection may have resulted from an overestimation of additive genetic variance in the breeding experiments. Two possible causes of overestimation are epistatic variation and inbreeding among the parents used in the diallels. Neither of these mechanisms can explain the decline in the response only after the first generation, however. Furthermore, the similarity in the G-matrices estimated with or without the contribution of selfed individuals does not support the inbreeding hypothesis in the Tovar population, which was the one with the largest reduction of the observed responses.

Artificial selection has been instrumental in the development of the evolutionary theory from Darwin and onward (Robertson 1966; Wright 1977, Hill and Caballero 1992; Bell 2008), and it still provides valuable insights on the evolvability of quantitative traits (e.g., Beldade et al. 2002; Carlbog et al. 2006; Le Rouzic et al. 2008; Pavlicev et al. 2010; Carter and Houle 2011; Hine et al. 2011; Bolstad et al. 2015; Sztepanacz and Blows 2017; Morgan et al. 2020). The lack of consideration of uncertainty in this type of experiment, however, has limited our ability to infer underlying genetic architecture from the discrepancies between observed and predicted responses. Despite the Lande equation being 40 years old, models that explicitly incorporate estimation of uncertainties in the prediction of evolutionary changes or allow analysis and interpretation of the selection-response dynamics in terms of genetic architecture are only starting to be developed (Le Rouzic et al. 2010, 2011; Stinchcombe et al. 2014). We argue that, with such models, a better and more systematic quantification of the imprecision associated with genetic parameters...
and their predictions will help us to better understand the limitations of the current approaches, and design experiments that will bring progress in understanding multivariate evolution. This should also help us in assessing predictability in eco-evolutionary dynamics.

**AUTHOR CONTRIBUTIONS**

C.P., T.F.H. and W.S.A. initiated the study. C.P. and E.A. collected the data. T.F.H. did the derivation on the prediction intervals in collaboration with A.L.R., G.H.B. and C.P. E.A., A.L.R., C.F., G.H.B. and C.P. performed the statistical analyses. C.P. did the literature search. C.P., E.A. and T.F.H. wrote the manuscript with contributions from all authors.

**ACKNOWLEDGMENTS**

We are grateful to Grete Rakvaag and Liv Antonsen for taking care of the plants during the selection experiment. We thank Michael B. Morrissey, three anonymous reviewers, and many people at the Centre for Biodiversity Dynamics (NTNU) for fruitful comments on earlier versions of this paper. Elena Albertsen was supported by the Research Council of Norway through its Centre of Excellence funding scheme, project no. 223257. We also thank the Norwegian Research Council (grant no. 196494 and 287214 to C.P. and 244139 to C.P. and A.L.R., and 275862 to G.H.B.) and the US National Science Foundation (grant nos. DEB-0444157 to T.F.H. and DEB-0444745 to W.S.A.) for financial support. C.P., T.F.H., A.L.R., and W.S.A. were hosted by the Center of Advanced Study (CAS) in Oslo during the writing of the paper.

**DATA ARCHIVING**

All the data and scripts that were used to conduct analyses and produce the figures are deposited in the Dryad Digital Repository: https://doi.org/10.5061/dryad.ns1rn8psz

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**Appendix 1: Drift variance under truncation selection**

In this appendix, we derive equation 4 in the main text for the sampling variance in the mean breeding value of a quantitative trait after one generation of selection. We assume that exactly \( N_p \) parents are picked from a population of \( N \) individuals and then mated deterministically such that each parent produces exactly \( 2N/N_p \) offspring.

Let the phenotype of an offspring, \( i \), be \( z_i = g_i + e_i \), where \( g_i \) is the breeding value and \( e_i \) an environmental effect with mean zero and variance \( \sigma^2_e \). Let \( \bar{g} = \frac{1}{N} \sum g_i \) be the mean breeding value among the \( N \) offspring. To compute the variance in \( \bar{g} \) due to sampling we first write

\[
\text{Var} \left[ \bar{g} \right] = E \left[ \text{Var} \left[ \bar{g} \mid P \right] \right] + \text{Var} \left[ E \left[ \bar{g} \mid P \right] \right],
\]

where \( P \) denotes conditioning on the \( N_p \) parents of the offspring. To simplify the computation, we now consider a single locus with two alleles, B and b, with frequencies \( p \) and \( q \), and genotypic effects \( 2a \), \( a \), and 0 for BB, Bb, and bb, respectively. Assuming additivity and linkage equilibrium, the total genetic sampling variance will be the sum of the contribution from each locus. For each locus we have

\[
E \left[ \bar{g} \mid P \right] = 2ap',
\]

\[
\text{Var} \left[ \bar{g} \mid P \right] = a^2H'/2N,
\]

where \( p' \) and \( H' \) are the allele frequency and the heterozygosity among the parents. The first equation follows because each parent contributes exactly \( 2N/N_p \) alleles to the offspring’s total of \( 2N \) alleles, and each of the \( 2Np' \) B-alleles contribute an effect \( a \). To derive the second equation, note that every homozygous parent always contributes the same allele type to the offspring and thus no sampling variance. A heterozygote parent will contribute a variance of \( a^2/4 \). There are \( H'N_p \) heterozygote parents, each being the parent of \( 2N/N_p \) alleles, so we get

\[
\text{Var} \left[ \bar{g} \right] = \left( \frac{a^2}{4} \right) \left( H'N_p \right) \left( 2N/N_p \right) /N^2 = a^2H'/2N.
\]

This is the variance due to random sampling of alleles from individual parents during mating. Putting the two equations together we get

\[
\text{Var} \left[ \bar{g} \right] = \left( \frac{a^2}{2N} \right) E \left[ H' \right] + 4a^2 \text{Var} \left[ p' \right].
\]

We now need to compute \( E[H'] \) and \( \text{Var}[p'] \) over samples of parents. We start with the case of no selection, as in our control lines. Here, parents are picked at random without replacement. The joint distribution of the numbers of BB homozygotes and Bb heterozygotes in the sample is multivariate hypergeometric with moments

\[
E \left[ N'_{BB} \right] = N_p p_{BB},
\]

\[
E \left[ N'_{Bb} \right] = N_p p_{Bb},
\]

\[
\text{Var} \left[ N'_{BB} \right] = N_p p_{BB} (1 - p_{BB}) (N - N_p) / (N - 1),
\]

\[
\text{Var} \left[ N'_{Bb} \right] = N_p p_{Bb} (1 - p_{Bb}) (N - N_p) / (N - 1),
\]

\[
\text{Cov} \left[ N'_{BB}, N'_{Bb} \right] = - N_p p_{BB} p_{Bb} (N - N_p) / (N - 1),
\]

where \( p_{BB} \) and \( p_{Bb} \) are the frequencies of the two genotypes in the population before sampling. Using these moments, and \( p_{BB} = p^2 + Fpq \) and \( p_{Bb} = 2pq(1 - F) \), where \( F \) is the coefficient of inbreeding in the population before sampling, we derive

\[
E \left[ p' \right] = p,
\]

\[
\text{Var} \left[ p' \right] = \frac{pq(1 + F)}{2N_p} \frac{N - N_p}{(N - 1)^2}.
\]
\[ E[H'] = 2pq(1 - F). \]

This yields
\[
\text{Var}[g] = \frac{pqa^2(1 - F)}{N} + \frac{2pqa^2(1 + F)(N - N_p)}{N_p(N - 1)}
\]
\[= 2pqa^2(1 + F) \left( \frac{1 - F}{2N(1 + F)} + \frac{N - N_p}{N_p(N - 1)} \right)\]
\[\approx V_A \left( \frac{1}{N_p} - \frac{1 + 3F}{2N(1 + F)} \right),\]

because the additive variance contributed by the locus is \(V_A = 2pqa^2(1+F).\) If the population we sample from is in Hardy-Weinberg equilibrium (i.e., \(F = 0\)), this reduces to
\[\text{Var}[g] \approx V_A(1/N_p - 1/2N).\]

To compute \(E[H']\) and \(\text{Var}[p']\) in the case of truncation selection, we need to make some approximations. The most important is that each locus has small effects such that we can ignore the effects of selection on the expected genotype frequencies. Hence, we assume \(E[N'_{BB}] = N_pP_{BB}\) and \(E[N'_{BB}] = N_pP_{BB}.\) We do, however, need to consider that individual parents differ in their probabilities of being selected. If we take the extreme case of a population in which all variance is additive genetic, then the probability of a given genotype being selected under truncation selection is either zero or one. This means that if we repeat the sampling, we will always pick exactly the same parents, and then \(\text{Var}[p'] = 0.\) If there are environmental sources of variation, then the probability of a given genotype being picked is uncertain and there will be sampling variance in \(p'.\) To quantify this, we need to compute the probability of picking a given genotype in the presence of environmental variance.

The number of BB individuals in the selected sample can be written as \(N'_{BB} = \sum y_i,\) where \(y_i\) is an indicator for whether individual \(i\) was included or not, and the sum is over all \(N_{BB}\) individuals that could be selected. The variance of this is
\[\text{Var}[N'_{BB}] = \text{Var}\left[\sum_{i=1}^{N_{BB}} y_i \right] = \sum_{i=1}^{N_{BB}} \text{Var}[y_i] + \sum_{i=1}^{N_{BB}} \sum_{j \neq i} \text{Cov}[y_i, y_j].\]

The \(y_i\) are each sampled without replacement, and in each of the \(N_p\) sampling events this happens with a probability of \(w_i/N,\) where \(w_i\) is the relative fitness of the individual genotype. From this we can derive
\[\text{Var}[y_i] = \frac{N_p}{N} w_i \left(1 - \frac{N_p}{N} w_i\right),\]
\[\text{Cov}[y_i, y_j]_{i \neq j} = -w_i w_j \frac{N_p}{N^2} \left(1 - \frac{(N_p - 1)}{2N} \left( \frac{w_i}{1 - w_i/N} + \frac{w_j}{1 - w_j/N} \right) \right).\]

Fitting this in yields
\[\text{Var}[N'_{BB}] = N_pP_{BB} \left(1 - P_{BB} - \frac{N_p - 1}{N} \right) \left(1 + \text{Var}[w]\right)\]
\[+ P_{BB} \left(\frac{N_p - 1}{N}\right) E \left[ \frac{w^2}{1 - w/N} \right]\]
\[= \left(\frac{N_p - 1}{N^2}\right) E \left[ \frac{w^3}{1 - w/N} \right],\]

where variance and expectation are over relative fitness in the population before selection, and we have used \(E[w] = 1.\) To obtain an approximation in terms of the variance of fitness, we use a second-order Taylor approximation of the expectations around the mean relative fitness:
\[E \left[ \frac{w^2}{1 - w/N} \right] \approx \frac{N}{N - 1} + \frac{N}{N - 1} \left(1 + (N - 1)^2 (2N - 1) \right) \text{Var}[w],\]
\[E \left[ \frac{w^3}{1 - w/N} \right] \approx \frac{N}{N - 1} + \left(\frac{N(3N - 2)}{(N - 1)^3}\right) \text{Var}[w].\]

Fitting in, collecting terms, and ignoring some terms of lower order in \(1/N\) yields
\[\text{Var}[N'_{BB}] \approx N_pP_{BB} \left(1 - P_{BB} \right) \left(\frac{N - N_p}{N - 1}\right) \left(1 - \left(\frac{N_p - 1}{N - N_p}\right) \text{Var}[w]\right).\]

Hence, increasing variance in relative fitness, stronger selection will reduce the sampling variance from the expectation under random sampling. Eventually, when selection becomes deterministic, the sampling variance becomes zero. The variance in relative fitness under deterministic truncation selection is \((N - N_p)/N.\) Using this in our equation, we find that the sampling variance is reduced with a factor \(1/N_p\) relative to random sampling. Provided the number of selected parents is not extremely small, this is close to zero, and thus shows that the approximations are good even far from random sampling. Using the same approach, we compute
\[\text{Var}[N'_{BB}] \approx N_pP_{BB} \left(1 - P_{BB} \right) \left(\frac{N - N_p}{N - 1}\right) \times \left(1 - \left(\frac{N_p - 1}{N - N_p}\right) \text{Var}[w]\right),\]
\[\text{Cov}[N'_{BB}, N'_{BB}] = -N_pP_{BB}P_{BB} \left(N - (N_p - 1) \right) E \left[ \frac{w^2}{1 - w/N} \right]\]
\[\approx -N_pP_{BB}P_{BB} \left(\frac{N - N_p}{N - 1}\right) \times \left(1 - \left(\frac{N_p - 1}{N - N_p}\right) \text{Var}[w]\right).\]

Using these we get
\[\text{Var}[p'] = \text{Var} \left[ \frac{N'_{BB} + \frac{1}{2}N'_{BB}}{N_p} \right]\]
\[\approx pq(1 + F) \left(\frac{N - N_p}{N_p(N - 1)}\right) \left(1 - \left(\frac{N_p - 1}{N - N_p}\right) \text{Var}[w]\right).\]
If we ignore the effects of selection on inbreeding and allele-frequency change, we can use this together with
\[ E[H'] = 2pqf(1-F) \]
to calculate
\[
\text{Var}[\bar{g}] = \frac{\sigma^2 E[H']}{2N} + 4\sigma^2 \text{Var}[p]
\]
\[
\approx \frac{pq^2 (1-F)}{N} + \frac{2pq^2 (1+F)(N-N_p)}{N_p(N-1)} \left( 1 - \left( \frac{N_p - 1}{N-N_p} \right) \text{Var}[w] \right)
\]
\[
= 2pq^2 (1+F) \times \left( \frac{1-F}{2N(1+F)} + \frac{N-N_p}{N_p(N-1)} \left( 1 - \left( \frac{N_p - 1}{N-N_p} \right) \text{Var}[w] \right) \right)
\]
\[
\approx V_A \left( \frac{1}{N_p} - \frac{1 + 3F}{2N(1+F)} - \frac{\text{Var}[w]}{N} \right),
\]
where we have ignored terms of lower order in \(1/N\) and \(1/N_p\). If the population we select from is in Hardy-Weinberg equilibrium, this reduces to
\[
\text{Var}[\bar{g}] \approx V_A \left( \frac{1}{N_p} - \frac{1 + 2\text{Var}[w]}{2N} \right).
\]

Appendix 2: An approximation for the genotypic variance in fitness

While the variance in relative fitness of phenotypes under truncation selection is exactly \((N - N_p)/N_p\), what we need in the equations is the variance of the relative fitness of genotypes. This is not easily expressed in terms of observable population parameters. For our purpose, to make prediction intervals, we use a crude approximation based on dividing the population into three groups, one with breeding values for the selected trait at least one environmental standard deviation above the selection cutoff, \(k\), one with breeding values at least one environmental standard deviation below the selection cutoff and one with breeding values in between. The two first groups we assign fitness of one and zero, respectively. For the intermediate group, we assign fitness equal to the mean fitness \(N_p/N\). We do this even if the fitness of a breeding value exactly at the cutoff is 1/2, because in our situation with \(N_p = N/4\) most of the probability mass of the group will be below the cutoff. Using the mean fitness will also make the variance in fitness converge correctly to zero when the heritability goes to zero. We assume that the breeding values and the environmental effects of the selected trait are normally distributed. If \(F(g)\) is the cumulative normal probability function for the breeding values, we can write
\[
F(k \pm \sqrt{\chi}) = \frac{1}{2} \left( 1 + \text{Erf} \left( \frac{k - \bar{g} \pm \sqrt{\chi}}{\sqrt{\chi}} \right) \right)
\]
\[
= \frac{1}{2} \left( 1 + \text{Erf} \left( \frac{k - \bar{g}}{\sqrt{\chi}} \pm \sqrt{1 - h^2} \right) \right),
\]
where \(\text{Erf}[x] = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt\) is the error function. The probability masses of the three categories are \(1 - F(k + \sqrt{\chi})\), \(F(k - \sqrt{\chi})\), and \(F(k + \sqrt{\chi}) - F(k - \sqrt{\chi})\), respectively, and using these we can write
\[
\text{Var}[w] \approx \frac{1}{E[W]} \left( 1 - F \left( k + \sqrt{\chi} \right) \right) + (E[W])^2 \left( F \left( k + \sqrt{\chi} \right) - F \left( k - \sqrt{\chi} \right) \right) - 1
\]
\[
= \frac{N^2}{2N^2_p} \left( 1 - \left( \frac{N_p}{N} - 1 \right) \text{Erf} \left( \frac{k - \bar{g}}{\sqrt{2\chi V_p}} \right) + \sqrt{1 - h^2} \right)
\]
\[
- \frac{N^2}{2N^2_p} \text{Erf} \left( \frac{k - \bar{g}}{\sqrt{2\chi V_p}} - \sqrt{1 - h^2} \right) - 1,
\]
where \(W\) is the absolute fitness. From the assumed normal distribution of the breeding values we have
\[
F(k) = \frac{N - N_p}{N},
\]
which yields
\[
k - \bar{g} = \sqrt{\chi} A^{-1} \left( \frac{N - N_p}{N} \right) = \sqrt{2\chi} A^{-1} \left[ 2 \left( \frac{N - N_p}{N} \right) - 1 \right],
\]
where \(\sqrt{2\chi} A^{-1}[x]\) is the probit function. Fitting this in we get
\[
\text{Var}[w] \approx \frac{N^2 - 2N^2_p}{2N^2_p} \left( \text{Erf} \left( \frac{k - \bar{g}}{\sqrt{2\chi V_p}} \right) + \sqrt{1 - h^2} \right)
\]
\[
- \frac{N^2 - 2N^2_p}{2N^2_p} \left( \text{Erf} \left( \frac{k - \bar{g}}{\sqrt{2\chi V_p}} - \sqrt{1 - h^2} \right) \right)
\]
\[
- \frac{1}{2} \left( \text{Erf} \left( \frac{k - \bar{g}}{\sqrt{2\chi V_p}} \right) + \sqrt{1 - h^2} \right),
\]
which can be computed from the heritability of the selected trait. This approximation gives the correct values of \((N - N_p)/N_p\) when \(h^2 = 1\) and zero when \(h^2 = 0\). We used this variance computed from the estimated heritability of the selected trait in equation 7 to make prediction intervals.

LITERATURE CITED
Albertsen, E., Ø. H. Opedal, G. H. Bolstad, R. Pérez-Barrales, T. F. Hansen, C. Pélabon and W. S. Armbruster. 2021. Using ecological context to in- terpret spatiotemporal variation in natural selection. Evolution, 75: 294–309.
Armbruster, W. S. 1984. The role of resin in angiosperm pollination - Ecological and chemical considerations. Am. J. Bot. 71: 1149–1160.
———. 1985. Patterns of character divergence and the evolution of reproductive ectotypes of Dalechampia scandens (Euphorbiaceae). Evolution 39: 733–752.
———. 1986. Reproductive interactions between sympatric Dalechampia species - Are natural assemblages random or organized. Ecology 67: 522–533.
———. 1988. Multilevel comparative-analysis of the morphology, function, and evolution of Dalechampia blossoms. Ecology 69: 1746–1761.
Dalechampia. September 2021: Blossoms: Honest signaling affects pollination success. 
Drosophila buzzatii blossoms. Phil. Trans. R. Soc. B: Composition of blossoms: Genetic analysis of pollinator preference. 

Falahati-Anbaran, M., H. K. Stenøien, C. Pélabon, G. H. Bolstad, R. Pérez-Barrales, E. J. Eisen, E. J. 2005. Testing quantitative genetic selection theory. In E.J. Eisen, J. K. Karoly, C. Stewart, V. A. Koelling, H. F. Sahli, and F. Carter, A. J. R., J. Hermisson and T. F. Hansen. 2005. The role of epistatic interactions in the response to selection and the evolution of evolvability. Theor. Pop. Biol. 68: 179–196.

Carlborg, Ö., L. Jacobsson, P. Åhgren, P. Siegel, and L. Andersson. 2006. Epistasis and the release of genetic variation during long-term selection. Nature genetics, 38: 418–420.

Carter, A. J. R., J. Hermisson and T. F. Hansen. 2005. The role of epistatic gene interactions in the response to selection and the evolution of evolvability. Theor. Pop. Biol. 68: 179–196.

Carter, A. J., and D. Houle. 2011. Artificial selection reveals heritable variation for developmental instability. Evolution, 65: 3558–3564.

Conner, J. K., K. Karoly, C. Stewart, V. A. Koelling, H. F. Sahli, and F. H. Shaw. 2011. Rapid independent trait evolution despite a strong pleiotropic genetic correlation. Am. Nat. 178: 429–441.

Cortese, M. D., F. M. Norry, R. P. and E. Hasson 2002. Direct and correlated responses to artificial selection on developmental time and wing length in Drosophila buzzatii. Evolution 56: 2541–2547.

Eisen, E. J. 2005. Testing quantitative genetic selection theory. In E.J. Eisen Ed.: Mouse in animal genetics and breeding research, Imperial College Press London, Pp 9–27.

Falahati-Anbaran, M., K. H. Stensien, G. H. Bolstad, T. F. Hansen, R. Pérez-Barrales, W. S. Armbruster, and C. Pélabon. 2017. Novel microsatellite markers for Dalechampia scandens (Euphorbiaceae) and closely related taxa: application to studying a species complex. Plant Spec. Biol. 32: 179–186.

Falahati-Anbaran, M., K. H. Stensien, C. Pélabon, G. H. Bolstad, R. Pérez-Barrales, T. F. Hansen and W. S. Armbruster. 2013. Development of Microsatellite Markers for the Neotropical Vine Dalechampia Scandens (Euphorbiaceae). Appl. Plant. Sci. 1200492.

Falconer, D. S. 1973. Replicated selection for body weight in mice. Genet. Res. Camb., 22: 291–321.

Frankham, R. 1990. Are responses to artificial selection for reproductive fitness characters consistently asymmetrical? Genet. Res. Camb. 56: 35–42.

Frankham, R., and K. R. Nurthen. 1981. Forging links between population and quantitative genetics. Theor. Appl. Genet. 59: 251–263.

Gromko, M. H. 1995. Unpredictability of correlated response to selection – pleiotropy and sampling interact. Evolution 49: 685–693.

Gromko, M. H., A. Briot, S. C. Jensen, and H. H. Fukui. 1991. Selection on copulation duration in Drosophila melanogaster – predictability of direct response versus unpredictability of correlated response. Evolution 45: 69–81.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. J. Stat. Softw. 33: 1–22.

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J. Evol. Biol. 21: 1201–1219.

Hansen, T. F., J. M. Alvarez-Castro, A. J. R. Carter, J. Hermisson, and G. P. Wagner. 2006. Evolution of genetic architecture under directional selection. Evolution 60: 1523–1536.

Hansen, T. F., W. S. Armbruster, M. L. Carlson, and C. Pélabon. 2003b. Evolvability and genetic constraint in Dalechampia blossoms: Genetic correlations and conditional evolvability. J. Exp. Zool. 296B: 23–39.

Hansen, T. F., C. Pélabon, W. S. Armbruster, and M. L. Carlson. 2003a. Evolvability and genetic constraint in Dalechampia blossoms: components of variance and measures of evolvability. J. Evol. Biol. 16: 754–766.

Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. Evol. Biol. 38: 258–277.

Hansen, T. F., and C. Pélabon. 2021. Evolvability: A quantitative-genetics perspective. Annu. Rev. Ecol. Syst. Evol. In press.

Hansen, T. F., T. M. Solvin and M. Pavlicev. 2019. Predicting evolutionary potential: A numerical test of evolvability measures. Evolution 73: 689–703.

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

Hendry, A. P. 2017. Eco-evolutionary dynamics. Princeton University Press.

Hill, W. G. 1971. Design and efficiency of selection experiments for estimating genetic parameters. Biometrics, 27: 293–311.

———. 1974. Variability of response to selection in genetic experiments. Biometrics, 30: 363–366.

Hill, W. G. and A. Caballero. 1992. Artificial selection experiments. Annu. Rev. Ecol. Syst. 23: 287–310.

Hine, E., K. McGuigan, and M. W. Blows. 2011. Natural selection stops the evolution of male attractiveness. Proc. Natl. Acad. Sci. 108: 3659–3664.

———. 2014. Evolutionary constraints in high dimensional trait Sets. Am. Nat. 184: 119–131.

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

Kelly, J. K. 2008. Testing the rare-alleles model of quantitative variation by artificial selection. Genetica 132: 187–198.

Knapp, S. J., W. C. Bridges Jr, and M. H. Yang. 1989. Nonparametric confidence interval estimators for heritability and expected selection response. Genetics 121: 891–898.

Kopp, M., and S. Matuszewski. 2014. Rapid evolution of quantitative traits: theoretical perspectives. Evol. Appl. 7: 169–191.

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

2234 | EVOLUTION SEPTEMBER 2021
Pélabon, C., T. F. Hansen, M. L. Carlson, and W. S. Armbruster. 2004b. Variational and genetic properties of developmental stability in Dalechampia scandens. Evolution 58: 504–514.

Pélabon, C., P. Thone, T. F. Hansen, and W. S. Armbruster. 2012. Signal honesty and cost of pollinator rewards in Dalechampia scandens (Euphorbiaceae). Ann. Bot. 109: 1331–1339.

Pérez-Barrales, R., G. H. Bolstad, C. Pélabon, T. F. Hansen, and W. S. Armbruster. 2013. Pollinators and seed predators generate conflicting selection on Dalechampia blossoms. Oikos 122: 1411–1428.

Prout, T. 1962. The error variance of the heritability estimate obtained from selection response. Biometrics, 18: 404–407.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rzeznick, D. N., J. Losos and J. Travis. 2019. From low to high gear: there has been a paradigm shift in our understanding of evolution. Ecol. Letters 22: 233–244.

Robertson, A. 1966. Artificial selection in plants and animals. Proc. R. Soc. B 164: 341–349.

Roff, D. A. 2007. A centennial celebration for quantitative genetics. Evolution 61: 1017–1032.

Rutledge, J. J., E. E. Eisen and J. E. Legates. 1973. Experimental evaluation of genetic correlation. Genetics 75: 709–726.

Sarkissian, T. S., and L. D. Harder. 2001. Direct and indirect responses to selection on pollen size in Brassica rapa L. J. Evol. Biol. 14: 456–468.

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

Shaw, R. G. 2019. From the past to the future: considering the value and limits of evolutionary prediction. Am. Nat. 193: 1–10.

Shaw, R. G., D. L. Byers, and F. H. Shaw. 1998. Genetic components of variation in Nemophila menziesii undergoing inbreeding: morphology and flowering time. Genetics 150: 1649–1661.

Sheridan, A. 1988. Agreement between estimated and realised genetic parameters. Anim. Breed. Abstr. 56: 877–889.

Sorensen, D. A., and B. W. Kennedy. 1983. The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. Theor. Appl. Genet. 66: 217–220.

———, 1984. Estimation of response to selection using least-squares and mixed model methodology. J. Anim. Sci. 58: 1097–1106.

Stinchcombe, J. R., A. K. Simonsen, and M. W. Blows. 2014. Estimating Uncertainty in Multivariate Responses to Selection. Evolution 68: 1188–1196.

Szpepanacz, J. L., and M. W. Blows. 2017. Artificial selection to increase the phenotypic variance in g max fish. Am. Nat. 190: 707–723.

Tai, G. C. C. 1979. An interval estimation of expected response to selection. Theor. Appl. Genet. 54: 273–275.

Walsh, B., and M. Lynch. 2018. Evolution and selection of quantitative traits. Oxford University press, London.

Worley, A. C., and S. C. Barrett. 2000. Evolution of floral display in Eichhornia paniculata (Pontederiaceae): direct and correlated responses to selection on flower size and number. Evolution 54: 1533–1545.

Wright, S. 1977. Evolution and the genetics of populations, III: experimental results and Varevolutionary deductions. University of Chicago press, Chicago.

Zhang, X. S., and W. G. Hill. 2005. Predictions of patterns of response to artificial selection in lines derived from natural populations. Genetics 169: 411–425.
Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplement S1: Descriptive statistics for the selection experiment in Tovar

Supplement S2: Descriptive statistics for the selection experiment in Tulum

Figure S3. Observed response to selection of gland area (GA), and correlated response of upper bract area (UBA), in two species of *D. scandens*

Table S4. Genetic (G), environmental (E), and phenotypic (P) variance matrices for the Tovar (small-glanded species) and Tulum (large-glanded species) populations of *Dalechampia scandens* estimated from the diallel experiments when selfed individuals (diagonal from the diallel) are excluded from the analysis

Figure S5. Visual representation of the changes $G$ when selfed individuals (diagonal of the diallel) are included (solid line) or not (dashed line) in the two species

Supporting Information