Unidirectional response to bidirectional selection on body size

II Quantitative genetics

Arnaud Le Rouzic[1,*], Clémentine Renneville[2], Alexis Millot[3], Simon Agostini[3], David Carmignac[2], and Éric Édeline[2,4]

1. Évolution, Génomes, Comportement et Écologie, Université Paris-Saclay, CNRS, IRD, 91198, Gif-sur-Yvette, France.

2. Institut d'Écologie et des Sciences de l'Environnement de Paris (iEES-Paris), Sorbonne Université, Université Paris Diderot, UPEC, CNRS, INRAE, IRD, Paris, France.

3. Centre de Recherche en Écologie Expérimentale et Prédictive (CEREEP-Ecotron IleDe-France), École normale supérieure, PSL Research University, Saint-Pierre-lès-Nemours, France.

4. ESE Ecology and Ecosystem Health, INRAE, Agocampus Ouest, Rennes, France.

∗ Corresponding author; EGCE, Avenue de la Terrasse, Bâtiment 13, 91198 Gif-sur-Yvette, France. Tel: +33 (0)1 69 82 37 65, `lerouzic@egce.cnrs-gif.fr`

Running title: Quantitative genetics of selection response

Abstract

Anticipating the genetic and phenotypic changes induced by natural or artificial selection requires reliable estimates of trait evolvabilities (genetic variances and covariances). However, whether or not multivariate quantitative genetics models are able to predict precisely the evolution of traits of interest, especially fitness-related, life-history traits, remains an open empirical question. Here, we assessed to what extent the response to bivariate artificial selection on both body size and maturity in the medaka *Oryzias latipes*, a model fish species, fits the theoretical predictions. Three populations were selected for divergent body size while maintaining a constant selection pressure against late maturity. The observed evolutionary trends did not match the predictions from a bivariate quantitative genetics "animal" model. The most parsimonious model identified environmental, but not genetic, covariances between both traits, which cannot explain why body size did not evolve in the line selected for a smaller body length. We investigated alternative mechanisms (including genetic drift, inbreeding depression, natural selection, scaling or genetic asymmetry issues, and undetected genetic correlations) but could not attribute the deviation from theory to any single explanation. Overall, these results question the ability of multivariate quantitative models to provide valid and operational predictions of the evolutionary response to multivariate selection on complex traits.

Keywords Animal model, Artificial selection, Asymmetric response, Bayesian mixed models, Bivariate selection, Evolvability, G-matrix, Genetic constraints, Model selection.

1 Introduction

Quantitative genetics offer simple and practical models to understand the evolution of quantitative traits in populations (Falconer & MacKay, 1996; Lynch & Walsh, 1998). In practice, these models are used both to analyze past selection response (identifying the factors involved in phenotypic change), and to predict the potential response to selection in a population. The rate of phenotypic change per generation is estimated by multiplying a measurement of the standing genetic (heritable) variation by a measurement of the strength of selection. In the simplest univariate model (the "breeder's equation", Lush, 1937), the genetic variation can be quantified by the heritability h^2 (proportion of the phenotypic variance that is heritable between parents and offspring). This setting is convenient when a unique trait is under selection, such as in some selective breeding programs, but becomes rapidly limited when the selection pressure is more complex and targets multiple traits at once. Multivariate models generally propose a slightly different setting, and quantify evolvability through the "$\mathbf{G}$" matrix of additive genetic (co)variances across traits, and selection through a vector of selection gradients β (Lande, 1979; Lande & Arnold, 1983; Blows, 2007; McGuigan, 2006). This approach offers efficient tools to explore theoretically and estimate empirically the properties of multivariate evolution and genetic constraints in complex and integrated biological systems (Cheverud, 1984; Hansen & Houle, 2008; Houle et al., 2017).

Although the data from various organisms and different kinds of traits are heterogeneous and experimental results lack consistency, the general pattern that seems to emerge from artificial selection experiments considering several traits is that quantitative genetics may predict short-term direct responses (phenotypic change of a single selected trait) convincingly (Sheridan, 1988; Walsh & Lynch, 2018, p. 606), correlated responses (genetic change in a trait that is genetically correlated to a selected trait, without being the target of selection) at least qualitatively (Gromko, 1995), while the response to multivariate selection (in which the selection gradient affects several traits) may be complex and inconsistent in some cases (Roff, 1997, p. 188, Roff, 2007 for review). Whether this lack of prediction power roots

into experimental issues, unrealistic assumptions and flaws in the multivariate quantitative genetics theory (e.g. inappropriate abstraction of proximal physiological mechanisms, Davidowitz et al., 2016), or poor understanding of the nature and stability of genetic correlations (which could be extremely environmentally-labile, Gutteling et al., 2007), largely remains to be determined.

In a companion paper, Renneville et al. (Under Review) investigated the phenotypic consequences of artificial selection on the medaka fish (*Oryzias latipes*) for a broad set of morphological, physiological, and life history traits, among which two were under direct selection. Wild-caught fish were submitted to 6 generations of truncation selection on fish length at 75 days (and thus, somatic growth rate). The experimental procedure generated three populations; a Large line, in which only large fish were bred, a Small line, in which only small fish were bred, mimicking harvest-like selection regime, and a Control line, in which fish were bred independently from their size. As the experimental design discarded *de facto* immature fish from the breeding pool, all three lines were thus also affected by a selection pressure for early maturity, a trait that was phenotypically correlated with size. Selection was thus essentially bivariate, in divergent directions across lines for body size and in the same direction (but with different intensities) for maturity. As extensively described in Renneville et al. (Under Review), after 6 generations of artificial selection, all lines evolved, but phenotypic response did not follow the selection differentials. Fish body size evolved only in the Large line, but not in the Small line, which remained statistically indistinguishable from the Control line. Conversely, the frequency of mature fish did not increase in spite of a positive selection differential in all three lines. These results confirm that anticipating qualitatively and quantitatively the consequences of multivariate selection on fish morphology and physiology cannot be based on the fitness function, but also needs to account for a deeper understanding of the functional and evolutionary relationships between selected traits.

Here, we will investigate to what extent multivariate quantitative genetic models, which include explicit genetic covariance components, could explain and predict such a counter-

intuitive selection response. We estimated genetic and environmental (co)variance components with statistical mixed effect models accounting for the experimental pedigree structure on the > 5000 phenotyped fish, from which genetic trends could be estimated. The analysis indicated that the selection response on maturity was coherent given the direction of the selection gradients, but the asymmetric response of body size could not be explained by classical quantitative genetics. Overall, both the direction and the magnitude of the selection response remained inconsistent with multivariate predictions, even when accounting for possible environmental (non-genetic) trends.

2 Materials and Methods

2.1 Biological material and experimental procedure

The initial population was derived from 100 wild adult medaka (*Oryzias latipes*) individuals sampled in June 2011 in Kiyosu (Aichi Prefecture), Japan. In order to keep track of the pedigree, fish were kept as 15-individual full-sib "families" in 3L aquariums. After two generations of random mating, the F_1 mature individuals were split into three breeding groups (Large, Small, and Control), and artificial selection was further performed for 6 generations, up to generation F_7. Every generation, fish were artificially selected on body size at 75 days, conditional on sexual maturity (which, for practical reasons, was defined based on the presence of secondary sexual characters). Standard length (Sdl) was first measured at 60 days from the pictures of individual fish (length from snout to the base of the caudal fin), and 10 out of $\simeq$15 families were preselected in each line based on their average length. At about 75 days, pairs were formed by selecting the two largest (respectively, smallest and random) males and females in each family. Immature fish were discarded from the breeding pool. Selection on maturity was necessary to (i) ensure the synchronization of all three lines, (ii) avoid selecting individuals that would never reach the reproductive stage, and (iii) limit sex identification mistakes when making pairs, as sex determination in immature fish requires

molecular techniques. All fish (including non-selected families and non-selected individuals in selected families) were photographed, and their standard length (abbreviated as Sdl below) and maturity (Mat) status was individually recorded. Detailed experimental procedures are provided in Renneville et al. (Under Review).

The unavoidable increase in the inbreeding coefficient across generations was limited by a specific procedure. Every generation in all three lines, twenty theoretical pairs of fish (two males and two females from each of the 10 families pre-selected at 60 days) were determined by a computer resampling procedure (selection of the pairing pattern minimizing the median inbreeding coefficient), and this theoretical pairing pattern was followed as close as possible when fish were selected after 75 days. By generation F_7, assuming no inbreeding in the F_1 population, the mean inbreeding coefficients were $F = 0.11$ in the Large line, $F = 0.091$ in the Control line, and $F = 0.085$ in the Small line. As the inbreeding coefficient is expected to increase by a factor $1 - 1/2N_e$ every generation, inbreeding population size estimates were about $N_e \simeq 27$, $N_e \simeq 33$, and $N_e \simeq 35$ in Large, Control, and Small lines, respectively. This procedure thus made it possible to maintain an average effective population size around 30 in all three lines.

2.2 Data analysis

The dataset consists in the measurement of body size (in mm) at about 75 days, the sex, and the maturity status for each of the $n = 5285$ fish of the experiment. The father and the mother of each fish was recorded, except for generation F_0. All the data analysis was performed with R version 3.4.4 (R Core Team, 2018). Inbreeding and coancestry coefficients were calculated with the package kinship2 (Therneau & Sinnwell, 2015). An archive containing datasets and scripts to generate tables and figures is provided as a supplementary file.

Realized selection differentials for body size were calculated as the difference between the weighted average of the parental phenotype and the average offspring phenotype, weights being proportional to the number of surviving offspring. Averages were normalized by the

Control line mean body length in parental and offspring generations. The regression coefficient (linear model without intercept) of the cumulated response to selection (relative to the Control line) on the cumulated selection differential was considered to be an approximation for heritability h^2, the ratio between heritable (additive) genetic variance and phenotypic variance.

Selection gradients on standard length ($\beta_{S_{gl}}$) and on maturity ($\beta_{M_{gl}}$) were estimated independently every generation g for each selected line l as the coefficients of a multivariate linear regression $w_{igl}/\bar{w}_{gl} = \beta_{S_{gl}}\mathrm{Sdl}_{igl} + \beta_{M_{gl}}\mathrm{Mat}_{igl} + \varepsilon_{igl}$, where $w_{igl}/\bar{w}_{gl}$ is the relative fitness of individual i, Sdl_{igl} its body size, and Mat_{igl} its maturity status, encoded as 0 or 1. Both "artificial" and effective (or realized) gradients (Walsh & Lynch, 2018, p. 487) were reported; for artificial selection gradients, fitness was 1 or 0 depending on the breeding status, while for effective gradients fitness was approximated by the number of offspring recorded in the database (i.e. having survived for 75 days). The later accounts for potential differences in survival and fertility among artificially selected breeders, and reflects the real selection pressure including natural selection.

Genetic and environmental variance components were estimated from the selection response with a bivariate mixed-effect linear model framework ('animal' model) (Lynch & Walsh, 1998; Sorensen & Gianola, 2007; Thompson, 2008), which general setting was as follows. As we were considering two traits, the phenotype of an individual i ($1 \leq i \leq n$) is bivariate (y_{1_i}, y_{2_i}), each trait following the classical infinitesimal model in quantitative genetics, e.g. $y_{1_i} = \mu_1 + \alpha_{1_i} + e_{1_i}$, where μ_1 is the grand mean of trait 1 at the first generation (model intercept), α_{1_i} is the additive genetic (breeding) value of individual i for trait 1, and e_{1_i} is an environmental (non-heritable residual) deviation. The variance-covariance matrix of all $2n$ breeding values $[\alpha_{1_1}, \ldots, \alpha_{1_n}, \alpha_{2_1}, \ldots, \alpha_{2_n}]^T$ is the Kronecker product $\mathbf{G} \otimes \mathbf{A}$, where $\mathbf{G}$ is the 2×2 additive variance-covariance matrix between both traits (additive genetic variances on the diagonal, additive genetic covariance off-diagonal), and $\mathbf{A}$, a $n \times n$ square and symmetric matrix, is the genetic relationship matrix (which elements are twice the coefficient

of coancestry of each pair of individuals, calculated from the pedigree). In a similar way, the variance-covariance matrix of the $2n$ residuals is $\mathbf{E} \otimes \mathbf{I}_n$, where $\mathbf{E}$ is the 2×2 environmental variance-covariance matrix between both traits, and $\mathbf{I}_n$ is the identity matrix of size n.

This theoretical setting had to be slightly modified to fit our dataset. The second trait (maturity) is the stochastic realization of an underlying probability p_m of maturing before 75 days. Phenotypic values were thus considered to be on the probit scale (which fits with the assumption that maturity is a threshold character). In such a model, the mean and the variance are not independent, and the residual variance cannot be estimated. We also considered additional random effects: an aquarium effect (351 levels) to account for the fact that fish in the same aquarium shared a common environment (in addition of sharing the same parents), and a generation effect (8 levels) to account for inter-generational environmental variation. Both additional random effects were defined by 2×2 variance-covariance matrices with three independent parameters (variance for Sdl, variance for Mat, and covariance).

The model was fitted in a Bayesian framework with the package MCMCglmm (Hadfield, 2010). Markov chains were run for 10^5 iterations, with a burn-in of 10^4, and the state of the chain was stored every 100 iteration ($n_{\text{it}} = 900$ data points). In practice, 12 chains were run in parallel and analyzed together with the tools from the package coda (Plummer et al., 2006). Priors for the random effects were inverse-Wishart with two degrees of freedom ($\nu = 2$), and identity matrices $\Lambda = \mathbf{I}_2$ as variances. The prior for the residuals was the same, except that the variance of the binomial trait $V_E(\text{Mat})$ was fixed to 1. We used "parameter expanded priors" for random effects to improve convergence, with $\alpha_\mu = 0$ and $\alpha_V = 1000$. Such priors can be considered as informative compared to the improper $\nu = 0.002$ suggestion (de Villemereuil, 2012), but using informative priors was necessary to limit convergence and stationarity issues with the most complex models, while remaining denser around zero compared to the $\nu = 3$ possibility (uniform marginal distribution for correlations). The influence of the prior on the posterior distribution was assessed for the most parsimonious model, for which convergence issues were limited.

We performed model selection based on the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002), from a minimal model (only genetic and residual variances) to a full model (genetic, residual, aquarium, and generation variances, as well as genetic and residual covariances). DIC is only an approximate (but simple and operational) criterion for model selection in Bayesian models (Plummer, 2008; Hooten & Hobbs, 2015), its known caveats are probably not problematic for the current analysis (linear model and non-mixture distributions). We also assessed the goodness of fit of a model M by computing a posterior predictive p-value $p_M = \#(\mathrm{ss}_M < \mathrm{ss}_{\mathrm{data}})/n_{\mathrm{it}}$, which counts the frequency at which the residual sum of squares $\mathrm{ss} = \sum_i^n = [(\mathrm{pred}_i - \mathrm{obs}_i)/\mathrm{pred}_i]^2$ of a dataset simulated from the model posterior distribution is less than the residual sum of squares of the real data. The best balance between underfitting and overfitting is achieved when $p_M = 0.5$.

3 Results

3.1 Selection gradients

Selection gradients were constant and repeatable throughout the experiment (Figure 1 A). The mean realized gradient on Sdl in the Large line was 0.47 $\mathrm{mm}^{-1}\pm$ s.d. 0.08 (i.e. in average, being 1 mm larger increased relative fitness by 0.47), -0.32 $\mathrm{mm}^{-1} \pm 0.13$ in the Small line, and 0.05 $\mathrm{mm}^{-1} \pm 0.16$ in the Control line. Although the experimental procedure was identical in all three lines regarding maturity (only mature fish were kept for breeding), the fact that both traits were phenotypically correlated generated different selection gradients. Selection on maturity was rather large in the Small line (2.49 ± 0.78 expressed in inverse maturity probability, i.e. an increase in 10% in maturity probability raises the relative fitness by 25%), and more moderate in Control (0.86 ± 0.97) and Large (-0.71 ± 0.79) lines. The slightly negative gradient in the Large line is a consequence of the phenotypic correlation between size and maturity raw-scaled probability (large selected fish are enriched with individuals that should not have been mature if average sized). In sum, selection was bivariate and not

symmetric among lines; the gradients in the Small line were for smaller body size and high maturity, there was no gradient on body size in the Control line but a slight gradient for larger maturity, and the Large line was selected for a larger body size, with a small (or even slightly negative) gradient on maturity.

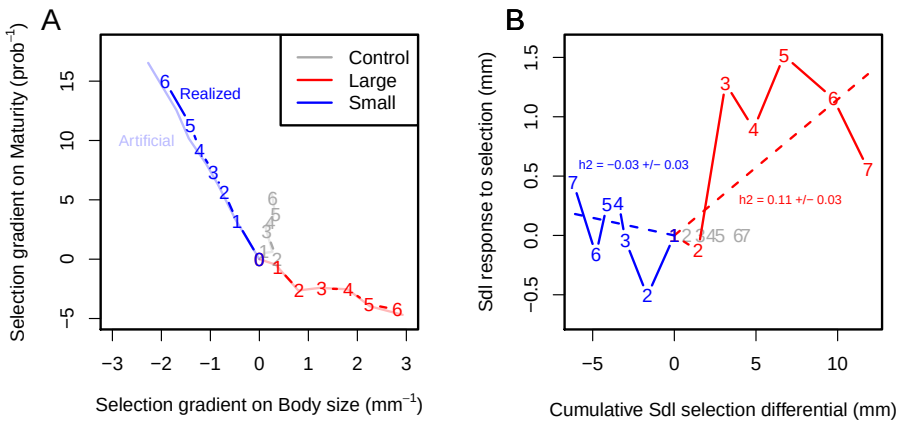


Figure 1: A: Bi-dimensional cumulative selection gradients in all three lines. Intended "Artificial" gradients (assuming no natural selection) are indicated in light colors, realized gradients are in dark. Generation numbers are indicated in the figure. B: Cumulative selection differentials vs. cumulative selection response for Body size in Large and Small lines, centered on the Control line. The regression coefficient (calculated independently for both lines) is an estimate of heritability.

3.2 Phenotypic response to selection

The phenotypic response to selection is presented in Figure 2 for fish length and maturity. For both traits, time series were characterized by substantial generation-specific effects. By generation F_6 (assuming that the size similarity in all F_7 fish is accidental), artificial selection has generated a $\simeq$ 1.2mm difference between the Large and the Control lines (about a 6% increase in length, equivalent to a $\simeq$ 18% increase in mass). Virtually all the phenotypic difference was built in two generations of selection, and there was no more phenotypic progress from generations F_3 to F_6. Surprisingly, there were no significant differences between the Control and the Small line, i.e. the Small line did not respond to selection on size. Realized

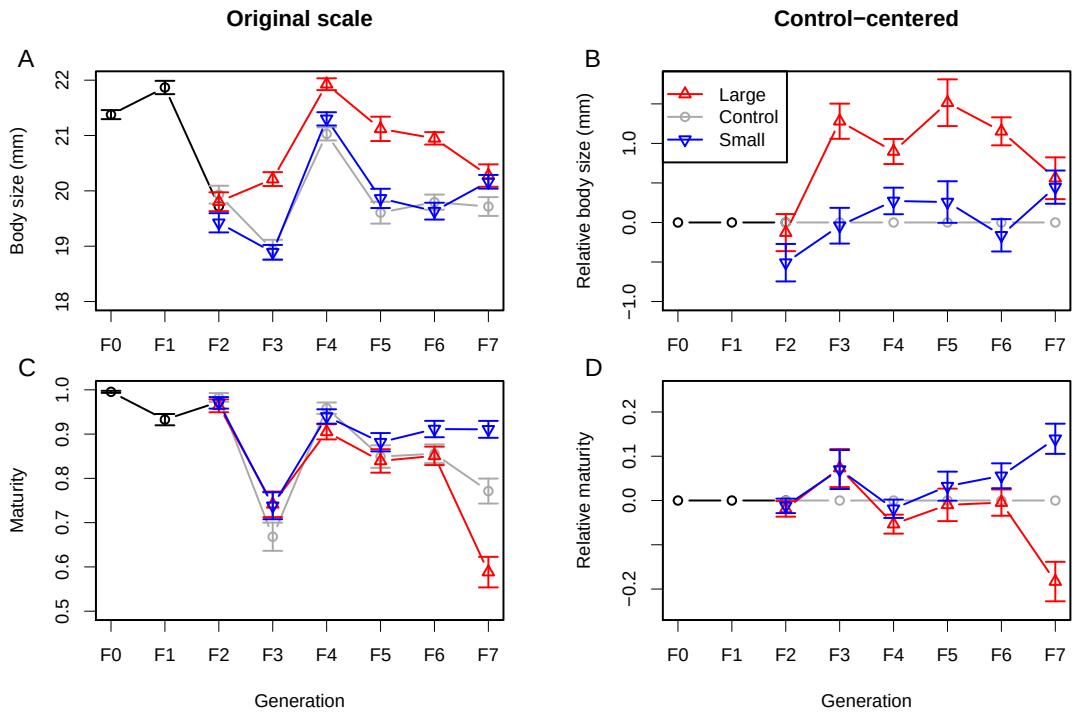


Figure 2: Response to selection for body length (top) and maturity (bottom) for all three lines (red: Large, blue: Small, gray: Control). Left: phenotypic means, Right: control-centered responses. Generations F_0 and F_1 (common to all lines) are drawn in black. Error bars stand for standard errors of the means.

heritabilities on body size was positive in the Large line ($h^2 = 0.11 \pm$ s.e. 0.03), and virtually zero in the Small line ($h^2 = -0.03 \pm 0.03$) (Figure 1 B).

The evolution in maturity was characterized by an irregular decrease, especially in the Control and Large lines. As for fish length, maturity was largely affected by generation-specific effects, especially in F_3, when maturity dropped from 95% to 75% in all three lines before increasing again to 90% in F_4. Overall, the general pattern for the bivariate selection response is featured by (i) a modest selection response in the Large line for body size, but not in the Small line, and (ii) a global evolutionary trend for maturity in the direction predicted by selection gradients relative to the Control line. The theoretical qualitative predictions are thus partially fulfilled for the Large line, but not for the Small line.

3.3 Genetic response to selection

A series of mixed-effect "animal" models of various complexities were defined, the simplest one consisting in genetic variances for both traits and a residual variance only for body size (no residual variance to be estimated for maturity due to the binomial nature of the trait), while the most complex one featured genetic, residual, aquarium, and generation variances, as well as genetic and residual covariances (Sup. Mat. 1). The most parsimonious model based on the Deviance Information Criterion (DIC) had six variance components: two genetic variances V_G (but no covariance), residual variance V_E for body length (by construction, the residual variance for maturity is fixed) and residual covariance C_E between both traits, and two generation effect variances V_F. The differences in both information criteria (DIC) and goodness of fit estimates (predictive posterior p-values) were substantial, and model selection was conclusive about the exclusion of genetic covariances between body length and maturity, as well as aquarium effects.

Note, however, that the variance components were not independent. For instance, aquarium effects and genetic effects were partly confounded, as fish sharing the same aquarium were full sibs. Including aquarium effects in the model thus decreased substantially the variance of genetic effects (Sup. Mat. 2). In a similar way, models including genetic covariances, which were discarded by the model selection procedure, tended to estimate large and significant genetic covariances (correlation about 0.7), as well as larger genetic variances (twice or even more). This is an indication that the exclusion of genetic covariance in the most parsimonious model was not due to a lack of statistical power, but rather to a poor model fit.

Table 1 reports the variance components from the best model as the median and support interval of the posterior distribution in a Bayesian framework. Model stationarity was unproblematic with the default (expanded) priors, but posteriors displayed a substantial amount of autocorrelation (Sup. Mat. 4). The additive genetic variance for body size was $V_G(\text{sdl}) = 1.6$ mm^2, which is more than estimated from realized heritabilities ($V_A = h^2 V_P \simeq 0.11 \times 5.3 = 0.6$

	2.5%	50%	97.5%
$V_G(\text{Sdl})$	1.32	1.58	1.85
$V_G(\text{Mat})$	0.11	0.21	0.37
$V_E(\text{Sdl})$	3.49	3.71	3.94
$C_E(\text{Sdl}, \text{Mat})$	1.78	1.85	1.91
$V_F(\text{Sdl})$	0.47	1.35	5.83
$V_F(\text{Mat})$	0.49	1.39	7.16

Table 1: Posterior median and 95% support interval from the MCMCglmm model fit. V and C stand for variance and covariances, respectively, and subscripts indicate different random effects: additive genetic effects G, residual effects E, and generation effects F. Sdl is assumed to be Gaussian, and Mat is binomial, on a probit scale (its residual variance $V_E(\text{Mat})$ is fixed to 1 instead of being estimated).

mm^2). The residual covariance between both traits was substantial, corresponding to a correlation $r \approx 0.88$ between residual body size and residual probit maturity probability. Most variance components happened not to be sensitive to the prior, except $V_G(\text{Mat})$ and $V_F(\text{Mat})$, which estimates were substantially lower when the prior was denser around 0 (Sup. Mat. 3), although all support intervals largely overlapped. This suggests that the data was weakly informative when estimating variances on maturity, and illustrates the difficulty in defining good priors for covariance matrices (Gelman et al., 2006; Alvarez et al., 2014).

The generation effect variance was relatively high, especially for Maturity. This generation effect not only captures inter-generational fluctuations, but it also corrects for a general trend in the time series for both traits. Figure 3 displays the genetic (average of breeding values posterior distributions) and non-genetic (from the generation effects) trends from the best model.

3.4 Selection response prediction

In addition to providing the theoretical framework to design statistical models for the estimation of variance components, quantitative genetics also aims at predicting the selection response from the genetic architecture of phenotypic traits. In practice, testing the predictive power of such models requires a specific protocol, as genetic variance components need

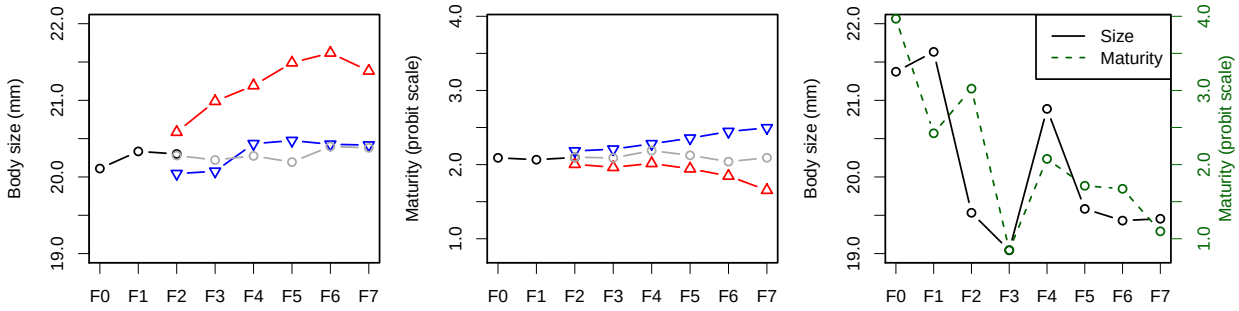


Figure 3: Estimated dynamics of genetic (left and center) and generation (right) effects. The figure is based on the average of the posterior distribution for each breeding value (genetic trend) and for each level of the generation effect. For genetic trends, breeding values were averaged conditional on the line (Large, Small, and Control lines).

to be estimated from the starting population (from e.g. an experimental cross design) and compared to the realized selection response. Although we do not have access to a direct measurement of additive variance components in the starting population here, we estimated the **G** matrix by fitting the animal model on the Control line individuals (including F_0 and F_1 generations), and compared the predicted selection response to the observed response from the selected lines (no overlap between both datasets). The best model for the Control-only data also excluded genetic covariances. Even when considering uncertainties due to the estimation procedure and genetic drift, there was no overlap between predicted and observed evolution for body size, while the prediction was convincing for maturity (Figure 4). Therefore, even if the selection response of body size appeared to occur in the expected direction in the Large line, the magnitude of the realized response was largely overestimated.

4 Discussion

4.1 Selection response

We performed a large-scale artificial selection experiment on medaka body size for 6 generations, keeping track of individual pedigrees. Three lines (Large, Small, and Control lines) were differentially selected for body length at 75 days of age, conditional on maturity. As

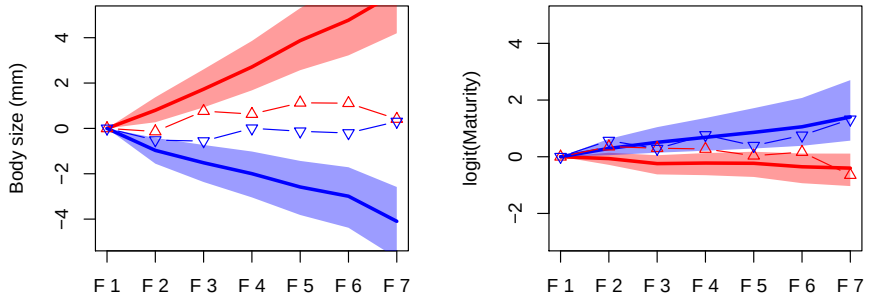


Figure 4: Predicted vs. realized selection responses. The **G** matrix was estimated by running the animal model on the control line, and the expected selection response was simulated by applying the Lande equation $\Delta z = \mathbf{G}\beta$ (Lande & Arnold, 1983) recursively over six generations, using the selection gradients β estimated in Figure 1. Plain lines: expected genetic responses (Large line in red, Small line in blue); symbols: actual phenotypic responses corrected for the generation effect estimated in Figure 3. Shaded areas represent the 95% support interval of the predicted selection response from 10,000 simulations (**G** matrices being sampled randomly from the Markov chain, and genetic drift being simulated by adding a random (Gaussian) cumulated deviation of variance V_G/N_e every generation, taking $N_e = 30$).

maturity and body size are phenotypically correlated in fish, this selection procedure generated a complex bi-dimensional selection pattern on two life history traits. After removal of non-heritable trends and noise with a random effect ('animal') model, the selection response pattern matched only partially the expected bidirectional response. Maturity evolved according to the predicted response (the Large line evolved late maturity compared to the Control line, while the Small line evolved early maturity). The phenotypic response on maturity was affected by a (supposedly non-genetic) downward trend, opposite to the selection gradient. For body size, Large and Control lines responded in the direction of selection gradients (larger body size and stasis, respectively), but, surprisingly, the Small did not evolve a smaller body length, and remained identical to the Control line throughout the experiment. Here again, even in the Large line, the magnitude of the empirical response was smaller than the theoretical prediction.

Environmental effects In spite of the tight control over environmental conditions (constant food, lightning, water quality and temperature), the data analysis highlighted a substantial amount of generation-specific effects that obscured the genetic selection response. Generations F_0, F_1, and F_4 appear to be substantially "better" (larger body size and higher maturity frequency) than e.g. F_3 and F_7. The fact that "good" generations were closer to the beginning of the experiment tends to generate an overall decreasing trend, which was difficult to interpret. A candidate explanation relies on an increase in inbreeding, which is unavoidable in such an experiment. However, the optimized pairing protocol limited the increase in inbreeding below 10% from generations F_0 to F_7 (Sup. Mat. 5). Furthermore, we found no correlation between phenotypic traits and inbreeding coefficient within generations. Thus, we deem it unlikely that inbreeding depression could cause the observed drop in maturity frequencies (and, to a lower extent, in body size). We also noticed "outlier" generations for various indicators; for instance, generation F_5 was featured by a neat increase in within-family variance in body size for all three lines (across-family variance and average phenotypes were not affected by this phenomenon, Sup. Mat. 6), and generation F_7 was characterized by an almost-complete loss of phenotypic differences in body size across lines, paralleled by a divergence for maturity. Again, it was not possible to relate these observations to particular events in the laboratory.

Evolutionary trends One of the most unexpected result of this experiment was the lack of response to selection on body size in the Small line, in spite of a substantial and consistent selection gradient. This lack of response was confirmed by the breeding value predictions from the animal model, and is necessarily associated with genetic or physiological factors that broke the assumptions of the infinitesimal model. The failure of the infinitesimal model was confirmed by the differences in variance components estimated from different generations or different selected lines (Sup. Mat. 7), which suggests that some model assumptions did not hold. However, genetic variances computed from the posteriors of the estimated breeding values did not show any strong trend (Sup. Mat. 8), excluding a drastic change in the genetic

architecture in the course of the experiment. Walsh & Lynch (2018, p. 611) proposed a list of 13 possible explanations for asymmetric selection responses, which we tried to address as thoroughly as possible (Table 2).

None of these explanations, taken individually, was particularly convincing. The potential for genetic drift to generate unexpected evolutionary patterns is substantial, and ruling out the influence of drift in laboratory experiments in notoriously difficult (Lynch, 1988). Contrary to common belief, replicating selection lines would not have helped to rule out genetic drift, as the sampling variance on the mean of two replicates of size $N_e/2$ is the same as on the mean of a single replicate of size N_e. Nevertheless, several independent lines of evidence tend to exclude genetic drift as a major explanatory factor of the observed response. First, the observed response does not lie within the theoretical support interval (Figure 4), showing that the response was significantly less than the theoretical expectation even when accounting for drift. Second, estimated additive genetic variances, as estimated by the animal model, are rather small compared to phenotypic variances (estimated heritabilities between 10 and 30% depending on the model). As the variance of the deviation due to drift is proportional to the genetic variance, low-heritability traits are expected to be rather insensitive to drift. Finally, the average breeding values estimated by the animal model (which accounts for drift) are quite stable across generations both for body size and maturity, which is compatible with the inbreeding effective population size $Ne \simeq 30$ estimated from the pedigree, and excludes drift as a major driver of the evolution of selected lines.

The fact that Large and Small lines were set in different competitive environments appears as an appealing explanation for the asymmetric response. Indeed, in order to keep track of the pedigree, fish were raised in the same tank as their full-sibs. If small or large body size was partly correlated to any competitive behavioral trait, the environment was varying during the experiment, as fish from the Large line were competing with better competitors every generation. This mechanism could have biased the selection response estimate in the Large line, and explain the lower-than-expected body-size response to selection. However, it

Design artefacts	
Drift	Difficult to exclude formally, but convergent evidence dismisses this possibility (further discussed in the text).
Scale effects	Could not explain the absence of selection response.
Different effective differentials	The selection procedure normalized the number of offspring irrespective to fertility differences. Effective selection gradients were calculated. No difference in hatching nor mortality rates across treatments. Natural selection against extreme phenotypes was detected (Renneville et al., Under Review), but both selected lines were affected symmetrically.
Undetected environmental trends	Less likely in controlled laboratory environments. An overall trend was detected, but no reason to expect different abiotic environmental conditions across lines. The competitive environment evolved (fish were raised with sibs), this possibility is further discussed in the text.
Effects of previous selection	Two generations of random mating were performed before the first generation of selection, which is expected to limit linkage disequilibrium in F_1.
Selection on correlated characters	The analysis accounts for the two traits that were artificially selected.
Nonlinear parent-offspring regression	
Major genes with dominance	Parent-offspring regression is rather linear (Sup. Mat. 9).
Genotype × Envir. interactions	The abiotic environment was identical across lines (randomized aquariums). Differences in competitive environments are expected to build up progressively, not to stop selection response from the first generation.
Departure from normality	Could not explain the absence of selection response.
Other sources	
Genetic asymmetries	Could not explain the absence of selection response in spite of the presence of additive genetic variance.
Inbreeding depression	Limited increase in inbreeding coefficient, and no within-generation correlation with selected traits (Sup. Mat. 5).
Maternal effects	Father-offspring and Mother-offspring regressions were very similar (Sup. Mat. 9).
Associative effects	Unlikely in controlled laboratory conditions.

Table 2: Why do quantitative genetics predictions fail? Thirteen possible explanations for asymmetric responses, as proposed by Walsh & Lynch (2018).

is less convincing that fish from the Small line could not become smaller because they were competing with worse competitors. Additional indirect evidence comes from an independent phenotyping experiment in which fish from the present selection experiment at later generations were raised in individual tanks without competition (Diaz-Pauli et al., 2019). In this phenotyping experiment, the genetic difference between the Large and Small lines was not larger than we found here, indicating that removing competition did not magnify the phenotypic effects of selection on medaka body size.

Finally, natural selection against a small body size could not be formally excluded as a cause of the asymmetric response, although it appears unlikely due to several consistent observations: (i) the Control line was affected by a general phenotypic (but not genetic) trend towards smaller body size, not larger; (ii) differences in fertility and mortality rates were limited by the experimental procedures, as the number of progeny per fish pair was normalized whenever possible; (iii) the difference between artificial selection gradients and realized gradients was reasonably small, (iv) natural selection on body length could be indirectly estimated (Renneville et al., Under Review), and appeared to affect both small and large fish.

Selection response on maturity also followed an unexpected pattern. First, it was affected by a general downward trend in all three lines, in spite of a positive gradient in Small and Control lines. As noticed in companion paper I (Renneville et al., Under Review), this trend vanished when maturity was corrected for body size, i.e. the intergeneration phenotypic trends affecting both selected traits (Figure 3) were consistent, and the decrease in maturity rate can be interpreted as a consequence of the downward environmental trend on body size, combined with a strong positive phenotypic correlation between both traits. Second, once corrected for this general trend, the genetic response of maturity followed the selection gradients qualitatively, but not quantitatively, as (i) the amplitude of the response was weak compared to the expectation from the estimates of additive variance, and (ii) the response was symmetric in spite of asymmetric selection gradients (Figure 4).

4.2 Strengths and weaknesses of the experimental design

Generalizing results obtained from model species in laboratory to wild species of interest is not straightforward, as differences in environment may condition trait means, trait variances, and genetic correlations (Gutteling et al., 2007; Postma, Visser, et al., 2007). The development of the 'animal' statistical model makes it possible to evaluate genetic components from observations in unmanipulated wild populations (Kruuk, 2004; Wilson et al., 2010). However, this approach remains particularly sensitive to e.g. gene-by-environment interactions, and the causal factors of observed trends may be difficult to identify formally (Postma & Charmantier, 2007; Walsh & Lynch, 2018). In contrast, controlled experiments (typically, complex breeding schemes) can only be carried out in laboratory conditions, and experimental approaches are often the only way to study key questions in population management, even when studying complex marine ecosystems (Suquet et al., 2005).

Artificial selection has long been proven to be an efficient way to simulate evolutionary processes in controlled conditions (Hill & Caballero, 1992; Conner, 2003). Here, we applied a classical truncation selection scheme, with substantial improvements compared to classical mass breeding experiments: (i) we kept track of the pedigree during the whole experiment, (ii) crosses were optimized to limit inbreeding, which kept the effective population size above $N_e = 27$ in all three lines, (iii) we recorded fecundity and mortality rates in all families, making it possible to evaluate the potential strength of natural selection, (iv) we raised a control line in the same conditions as selected lines, which helps distinguishing non-genetic and genetic trends, and (v) we selected explicitly on both body size and maturity, and considered both life history traits simultaneously in our analysis. The main drawback of this approach is an increased cost and human power involved, which necessarily limits the size of the experiment in terms of replicates (three lines in total) and duration (almost 3 years, for 8 generations and 6 episodes of selection, which may not be conclusive for low-evolvability traits). Nevertheless, in spite of such logistic limitations, our > 5000 fish pedigree displayed sufficient statistical power to (i) exclude models involving genetic covariances between body

size and maturity, (ii) reject the infinitesimal model predictions, and (iii) discard genetic drift as a major explanatory factor. In sum, the size of the experiment might be too limited to fully understand how life-history traits respond to complex multivariate selection, but is sufficient to conclude that this response does not follow quantitative genetics predictions.

4.3 Consistency with previous results

Due to their close relationship with fitness, life history traits are suspected to behave differently from other (morphological, physiological, behavioral) characters. Their heritability is lower (Price & Schluter, 1991; Roff, 1997) (probably due to a larger residual variance rather than a low genetic variance, Houle (1992)), and fitness-related traits have long been expected to display specific correlation patterns. In the wild, selection should indeed erode genetic variation for fitness, and thus only leave alleles with negatively correlated effects on positively-selected traits. Yet, such a negative genetic correlation may not always translate into a negative phenotypic correlation, as non-genetic residual positive covariances (typically, access to abundant food resources can improve all fitness-related components simultaneously) may dominate the phenotypic covariance structure (Lande, 1982; Reznick, 1985). Moreover, genetic polymorphism in genes involved in the efficiency of resource acquisition could also generate positive (genetic) covariances (Houle, 1991), hampering the derivation of strong theoretical predictions about the sign of genetic covariances among life history traits. Meta-analyses support the idea that correlations between life history traits range between -1 and 1 depending on traits and organisms, being slightly positive on average, but lower than between other kinds of traits (Roff, 1996). Our results featuring the absence of detectable genetic correlation between growth and maturity, associated with a strong residual correlation are thus not unexpected.

Less expected, however, was the lack of response to directional selection in the Small line. Fish artificially-selected for large or small size generally respond to selection in both directions (Diaz Pauli & Heino, 2014), symmetrically (as in the Atlantic silverside, Conover

& Munch, 2002, or in zebrafish, Amaral & Johnston, 2012) or slightly asymmetrically with a slower response in the Small line (in guppy, van Wijk et al., 2013). In the only experiment in which maturity was probably selected together with size (zebra fish, Uusi-Heikkilä et al., 2015), the response was complex and asymmetric (no size change and later age maturation in the Large line, smaller adult size and maturation at a smaller size — but not age — in the Small line).

4.4 Consequences on the response of life history traits to selection pressure

One of the most appealing applications of quantitative genetics outside of their original plant and animal breeding field is related to the prediction of the evolutionary consequences of human activity and/or environmental change on natural populations (Shaw, 2019). For instance, size-selective harvesting induces direct selection pressures on body size, and reduces life expectancy, which generates complex selection pressures on correlated life history traits (including growth rate, fertility, survival, and age at maturity) (Heino et al., 2015). Long-term evolutionary trends towards smaller body size, earlier maturity, and as a consequence, lower fecundity are frequent in highly-harvested species (Trippel, 1995; Law, 2000). It is therefore increasingly recognized that fisheries management programs should account for evolutionary change in life history traits (Kuparinen & Merilä, 2007; Fenberg & Roy, 2008; Laugen et al., 2014).

Accounting for evolutionary response management strategies for in wild populations generally relies on standard models in ecology and quantitative genetics, which assume that evolution can be reliably predicted when genetic trait variances and covariances are known (Diaz Pauli & Heino, 2014), which is generally not the case. Here, we show that such standard expectations may not be fulfilled, which questions the possibility to apply general recipes. First, the absence of genetic covariance between growth rate and maturity rules out the correlated-response hypothesis, and suggests that, at least in some species, maturity could

evolve independently from body size. In this context, observed evolution towards small body size and earlier maturity, which is widely observed in long-term exploited populations, has to be interpreted as the result of bivariate selection rather than the result of indirect selection due to genetic correlations. However, our results support the idea that bivariate selection response is hardly predictable even in a controlled environment, which questions the robustness of fishery management genetic models. Although we lack a clear explanation about why some heritable characters may not evolve when selected together, this phenomenon may decrease our confidence in the estimates of phenotypic trajectories for populations under anthropic pressure.

References

Alvarez, I., Niemi, J., & Simpson, M. (2014). Bayesian inference for a covariance matrix. *arXiv*, 1408.4050.

Amaral, I. P. & Johnston, I. A. (2012). Experimental selection for body size at age modifies early life-history traits and muscle gene expression in adult zebrafish. *Journal of Experimental Biology*, *215*(22), 3895–3904.

Blows, M. W. (2007). A tale of two matrices: Multivariate approaches in evolutionary biology. *Journal of Evolutionary Biology*, *20*(1), 1–8.

Cheverud, J. M. (1984). Quantitative genetics and developmental constraints on evolution by selection. *Journal of theoretical biology*, *110*(2), 155–171.

Conner, J. K. (2003). Artificial selection: A powerful tool for ecologists. *Ecology*, *84*(7), 1650–1660.

Conover, D. O. & Munch, S. B. (2002). Sustaining fisheries yields over evolutionary time scales. *Science*, *297*(5578), 94–96.

Davidowitz, G., Roff, D., & Nijhout, H. F. (2016). Synergism and antagonism of proximate mechanisms enable and constrain the response to simultaneous selection on body size

and development time: An empirical test using experimental evolution. *The American Naturalist*, *188*(5), 499–520.

de Villemereuil, P. (2012). *Estimation of a biological trait heritability using the animal model*. Retrieved from http://devillemereuil.legtux.org/wp-content/uploads/2012/12/tuto_en.pdf

Diaz Pauli, B. & Heino, M. (2014). What can selection experiments teach us about fisheries-induced evolution? *Biological journal of the Linnean Society*, *111*(3), 485–503.

Diaz-Pauli, B., Garric, S., Evangelista, C., Vøllestad, L. A., & Edeline, E. (2019). Selection for small body size favours contrasting sex-specific life histories, boldness and feeding in medaka, oryzias latipes. *BMC Evolutionary Biology*, *19*(1), 127.

Falconer, D. S. & MacKay, T. F. C. (1996). *Introduction to quantitative genetics* (Fourth). Pearson Education Ltd.

Fenberg, P. B. & Roy, K. (2008). Ecological and evolutionary consequences of size-selective harvesting: How much do we know? *Molecular ecology*, *17*(1), 209–220.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, *1*(3), 515–534.

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

Gutteling, E., Doroszuk, A., Riksen, J., Prokop, Z., Reszka, J., & Kammenga, J. (2007). Environmental influence on the genetic correlations between life-history traits in caenorhabditis elegans. *Heredity*, *98*(4), 206.

Hadfield, J. D. (2010). Mcmc methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22. Retrieved from http://www.jstatsoft.org/v33/i02/

Hansen, T. F. & Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, *21*(5), 1201–1219.

Heino, M., Diaz Pauli, B., & Dieckmann, U. (2015). Fisheries-induced evolution. *Annual review of ecology, evolution, and systematics*, *46*, 461–480.

Hill, W. G. & Caballero, A. (1992). Artificial selection experiments. *Annual Reviews of Ecology and Systematics*, *23*, 287–310.

Hooten, M. B. & Hobbs, N. T. (2015). A guide to bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28.

Houle, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics*, *130*, 195–204.

Houle, D. (1991). Genetic covariance of fitness correlates: What genetic correlations are made of and why it matters. *Evolution*, *45*(3), 630–648.

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

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1446), 873–890.

Kuparinen, A. & Merilä, J. (2007). Detecting and managing fisheries-induced evolution. *Trends in ecology & evolution*, *22*(12), 652–659.

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

Lande, R. & Arnold, S. J. (1983). The measument of selection on correlated characters. *Evolution*, *37*(6), 1210–1226.

Lande, R. (1982). A quantitative genetic theory of life history evolution. *Ecology*, *63*(3), 607–615.

Laugen, A. T., Engelhard, G. H., Whitlock, R., Arlinghaus, R., Dankel, D. J., Dunlop, E. S., ... Matsumura, S. et al. (2014). Evolutionary impact assessment: Accounting for evolutionary consequences of fishing in an ecosystem approach to fisheries management. *Fish and Fisheries*, *15*(1), 65–96.

Law, R. (2000). Fishing, selection, and phenotypic evolution. *ICES Journal of Marine Science*, *57*(3), 659–668.

Lush, J. L. (1937). Animal breeding plans. ames. Collegiate Press, Inc.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc.

Lynch, M. (1988). Design and analysis of experiments on random drift and inbreeding depression. *Genetics*, *120*(3), 791–807.

McGuigan, K. (2006). Studying phenotypic evolution using multivariate quantitative genetics. *Molecular Ecology*, *15*(4), 883–896.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, *9*(3), 523–539.

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, *6*(1), 7–11. Retrieved from https://journal.r-project.org/archive/

Postma, E., Visser, J., & Van Noordwijk, A. J. (2007). Strong artificial selection in the wild results in predicted small evolutionary change. *Journal of Evolutionary Biology*, *20*(5), 1823–1832.

Postma, E. & Charmantier, A. (2007). What 'animal models' can and cannot tell ornithologists about the genetics of wild populations. *Journal of Ornithology*, *148*(2), 633–642.

Price, T. & Schluter, D. (1991). On the low heritability of life-history traits. *Evolution*, *45*(4), 853–861.

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Renneville, C., Millot, A., Agostini, S., Carmignac, D., Maugars, G., Dufour, S., ... Edeline, E. (Under Review). Unidirectional response to bidirectional selection on body size. i. phenotypic and endocrine response. *Submitted*.

Reznick, D. (1985). Costs of reproduction: An evaluation of the empirical evidence. *Oikos*, 257–267.

Roff, D. A. (1997). *Evolutionary quantitative genetics*. New York [etc.]: Chapman & Hall.

Roff, D. A. (1996). The evolution of genetic correlations: An analysis of patterns. *Evolution*, *50*(4), 1392–1403.

Roff, D. A. (2007). A centennial celebration for quantitative genetics. *Evolution*, *61*(5), 1017–1032.

Shaw, R. G. (2019). From the past to the future: Considering the value and limits of evolutionary prediction. *The American Naturalist*, *193*(1), 1–10.

Sheridan, A. (1988). Agreement between estimated and realised genetic parameters. In *Anim. breed. abstr* (Vol. 56, *877*, p. 89).

Sorensen, D. & Gianola, D. (2007). *Likelihood, bayesian, and mcmc methods in quantitative genetics*. Springer Science & Business Media.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, *64*(4), 583–639.

Suquet, M., Rochet, M.-J., & Gaignon, J.-L. (2005). Experimental ecology: A key to understanding fish biology in the wild. *Aquatic Living Resources*, *18*(3), 251–259.

Therneau, T. M. & Sinnwell, J. (2015). *Kinship2: Pedigree functions*. R package version 1.6.4. Retrieved from https://CRAN.R-project.org/package=kinship2

Thompson, R. (2008). Estimation of quantitative genetic parameters. *Proc. Biol. Sci. 275*(1635), 679–686.

Trippel, E. A. (1995). Age at maturity as a stress indicator in fisheries. *Bioscience*, *45*(11), 759–771.

Uusi-Heikkilä, S., Whiteley, A. R., Kuparinen, A., Matsumura, S., Venturelli, P. A., Wolter, C., ... Killen, S. S. et al. (2015). The evolutionary legacy of size-selective harvesting extends from genes to populations. *Evolutionary applications*, *8*(6), 597–620.

van Wijk, S. J., Taylor, M. I., Creer, S., Dreyer, C., Rodrigues, F. M., Ramnarine, I. W., ... Carvalho, G. R. (2013). Experimental harvesting of fish populations drives genetically based shifts in body size and maturation. *Frontiers in Ecology and the Environment*, *11*(4), 181–187.

Walsh, B. & Lynch, M. (2018). *Evolution and selection of quantitative traits*. Oxford University Press.

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., ... Nussey, D. H. (2010). An ecologist's guide to the animal model. *Journal of Animal Ecology*, *79*(1), 13–26.

Supplementary material 1

Model selection

	V_G^{S}	V_G^{M}	C_G^{SM}	V_E^{S}	V_E^{M}	C_E^{SM}	V_A^{S}	V_A^{M}	V_F^{S}	V_F^{M}	Δ DIC	PPP
gr	✓	✓		✓	–						9806	0.00
gR	✓	✓		✓	–	✓					304	1.00
Gr	✓	✓	✓	✓	–						7740	0.00
GR	✓	✓	✓	✓	–	✓					1738	0.00
gRa	✓	✓		✓	–	✓	✓	✓			559	0.02
GRa	✓	✓	✓	✓	–	✓	✓	✓			1132	0.00
gRaf	✓	✓		✓	–	✓	✓	✓	✓	✓	252	0.17
GRaf	✓	✓	✓	✓	–	✓	✓	✓	✓	✓	797	0.00
gRf	✓	✓		✓	–	✓			✓	✓	0	0.32
GRf	✓	✓	✓	✓	–	✓			✓	✓	1431	0.00

General overview of the models considered in the model selection procedure. The table indicates which variance (V) and covariance (C) components were considered for each model (uniquely indexed by a code indicated in the first column). Superscripts S and M stand for Standard lenght and Maturity, respectively. Note that the residual variance for maturity, V_E^M, is fixed and thus never included in any model. Four variance components were considered, genetic (g), residual (r), aquarium (batch) (a), and generation (f). Capital letters stand for the full covariance structure (variances and covariances), while lower-case letters indicate the absence of covariance (diagonal covariance matrix). The deviation to the best model in Deviance Information Criterion (DIC) units is indicated, as well as an estimate of the Bayesian Posterior Predictive P-value (PPP) (probability for a dataset simulated from the model to be closer to the predicted values than the real data, $p_M = 0.5$ is expected when the model M can perfectly generate the data).

Supplementary material 2

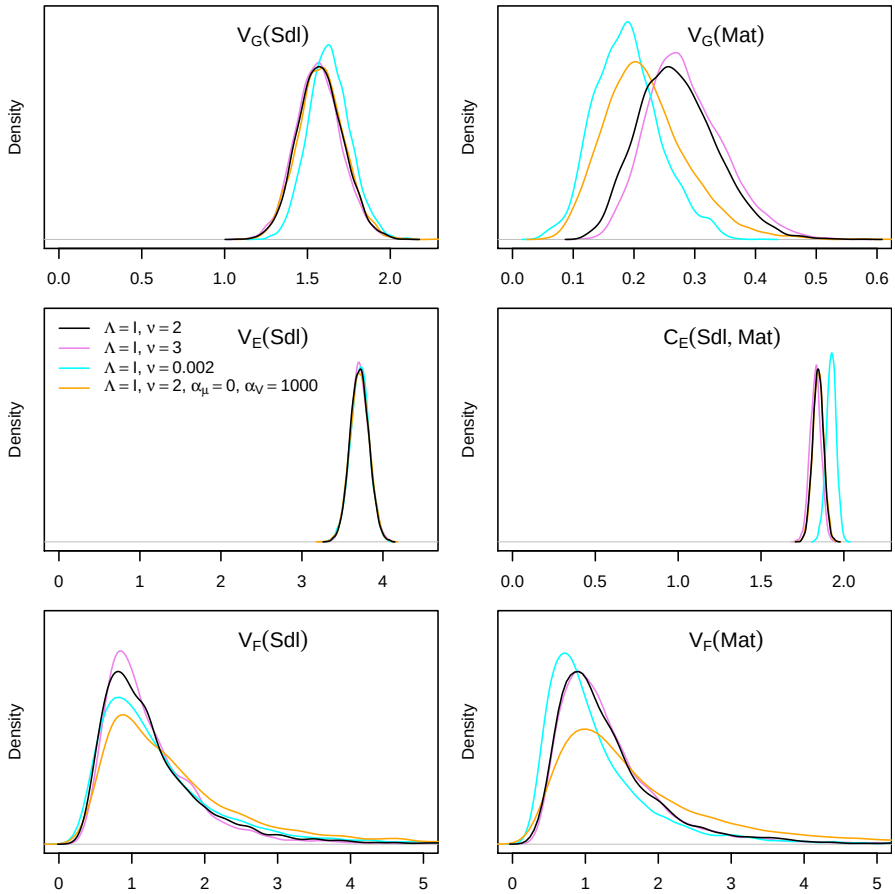
Posteriors for genetic variances and covariances

	2.5%	V_G(Sdl)	97.5%	2.5%	V_G(Mat)	97.5%	2.5%	C_G(Sdl,Mat)	97.5%
gr	2.29	2.64	3.03	4.99	9.86	30.98			
gR	1.61	1.88	2.16	0.52	0.73	1.02			
Gr	3.82	4.19	4.59	30.92	52.10	110.88	9.83	12.87	18.85
GR	2.49	2.85	3.24	2.18	3.23	4.81	1.58	2.09	2.74
gRa	0.10	0.28	0.52	0.03	0.18	0.39			
GRa	0.63	1.01	1.46	0.72	1.37	2.30	0.57	0.97	1.48
gRaf	0.10	0.25	0.46	0.01	0.08	0.20			
GRaf	0.67	1.14	1.66	0.41	0.98	2.10	0.41	0.88	1.55
gRf	1.32	1.58	1.85	0.11	0.21	0.37			
GRf	2.24	2.65	3.08	0.91	1.46	2.65	1.05	1.43	2.05

Median and 95% support interval of the posterior distribution for genetic variance components (additive variance for body size, additive variance for maturity, and additive genetic covariance when relevant) for all tested models. Models are labeled as described in Sup. Mat. 1. Variance components are not independent, genetic variances estimates decrease when the aquarium effect is included in the model (aquariums are shared by full-sibs, so that the variance between aquariums is expected to capture around half the genetic variance in the population), and the genetic covariance reaches high levels when residual covariances are excluded.

Supplementary material 3

Sensitivity to the prior



Posterior distribution of variance components from the best model under four prior Inverse-Wishart (ν, Λ) distributions. When the number of degrees of freedom ν is small ($\nu < 1$ in our two-dimensional case), the prior density is larger at the vicinity of zero, but using such an improper prior raised convergence and stationarity issues with models including more random effects. Expanded priors (represented in orange) implemented in MCMCglmm (marginally) improves the mixing of the chains, and were used as default priors for random effects.

Supplementary material 4

Model convergence

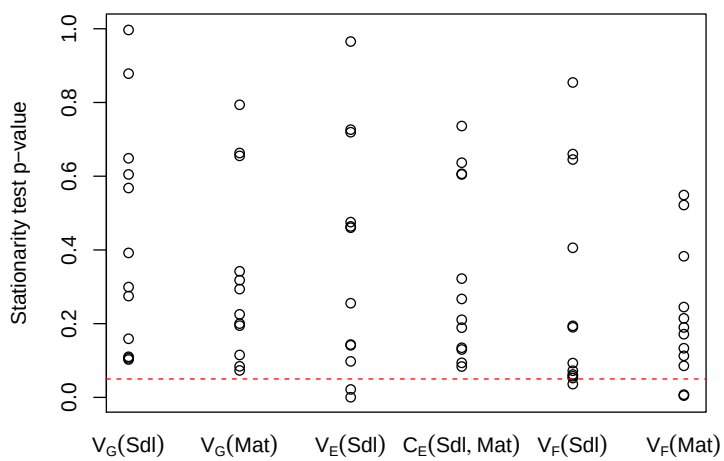
Autocorrelation and effective size

	Lag 1	Lag 5	Lag 10	Lag 50	Lag 100	Eff. Size
$V_G(\text{Sdl})$	0.93	0.82	0.74	0.40	0.21	2990.00
$V_G(\text{Mat})$	0.99	0.97	0.95	0.83	0.73	405.02
$V_E(\text{Sdl})$	0.94	0.77	0.64	0.25	0.14	3838.79
$C_E(\text{Sdl}, \text{Mat})$	0.95	0.80	0.69	0.33	0.19	2927.38
$V_F(\text{Sdl})$	0.28	0.15	0.15	0.15	0.14	958.43
$V_F(\text{Mat})$	0.19	0.08	0.08	0.07	0.06	2071.86

Autocorrelation for the random effect parameters was assessed with the `autocorr.diag` function from package `coda`. The effective size (sample size adjusted for autocorrelation) was evaluated with the function `effectiveSize` from the same package.

. Among all variance components, only the genetic variance for maturity displays a problematic autocorrelation.

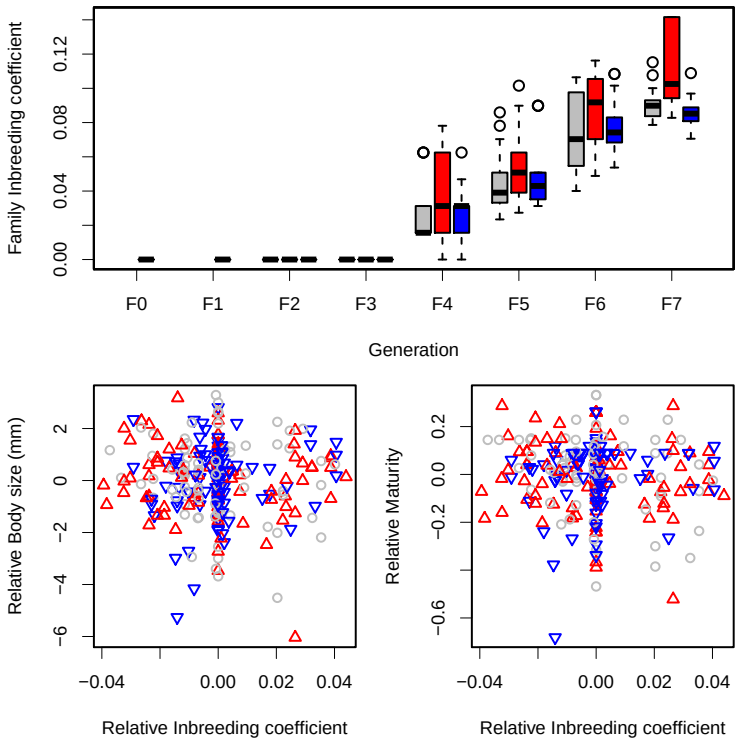
Stationarity



The vast majority of MCMC chains pass the Heidelberg stationarity test implemented in the `heidel.diag` function from the `coda` package (null hypothesis H_0: the chain is stationary at least over its last half, the dashed line illustrates the 5% threshold, 66 out of 72 chains — 92% — are above the threshold vs. 68 under H_0).

Supplementary material 5

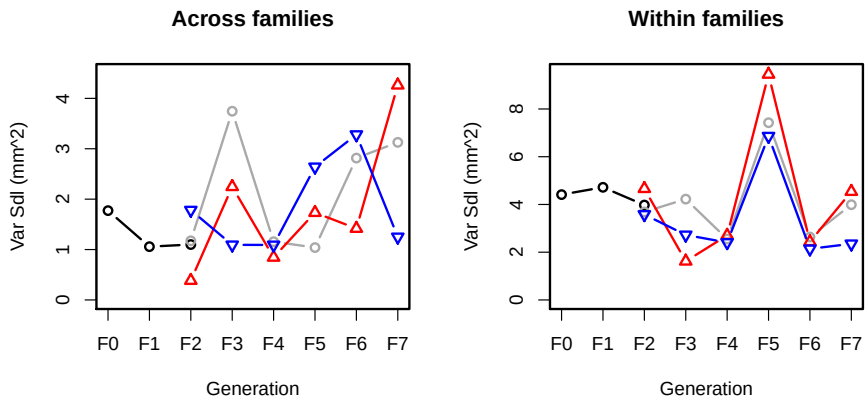
Inbreeding



Top: Distribution of the inbreeding coefficients across families in the course of the experiment (assuming no inbreeding in F_0). Bottom: relationship between the inbreeding coefficient of families (normalized by the average of the line each generation) to phenotypic traits (centered on the line and generation mean). None of these regressions are statistically significant.

Supplementary material 6

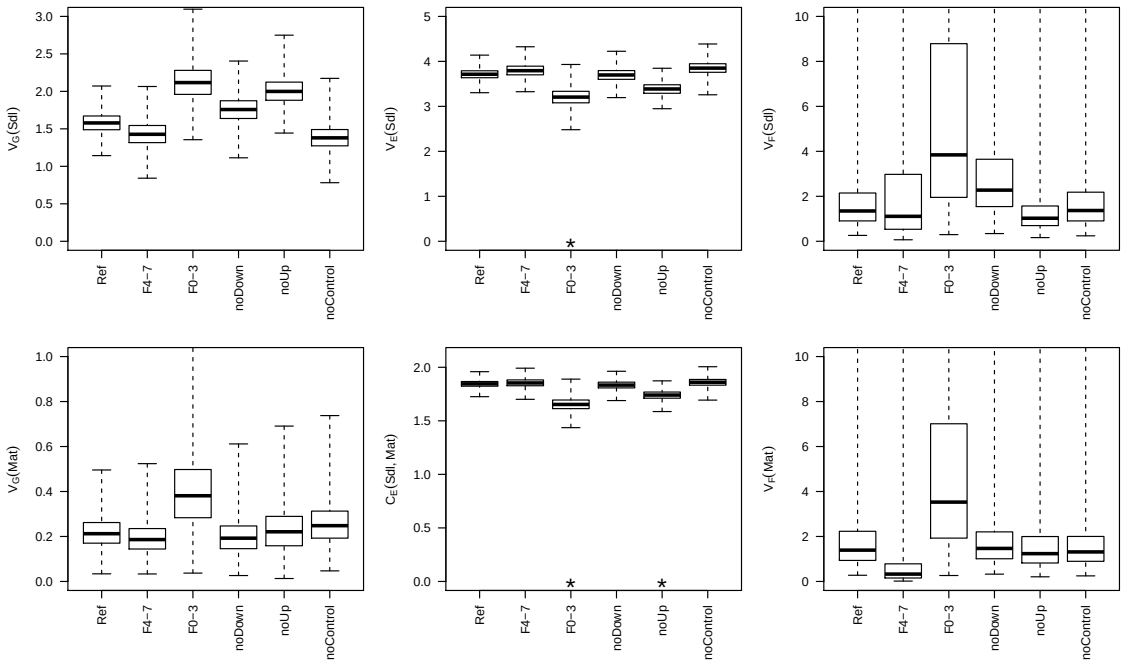
Phenotypic Variance



Dynamics of the phenotypic variance across (left) and within (right) aquariums during the selection experiment. The traditional decomposition of genetic variances states that $V_P = V_G + V_E$, where V_P, V_G and V_E stand for the phenotypic, genetic, and residual variances. The vast majority of quantitative genetics models assume a constant residual variance. In our dataset, the phenotypic variance for body size could further be split into two components; the variance across family means in each line at a given generation, and the variance within families. The variance across families (or aquariums) catches half the genetic variance, as well as the batch (aquarium) effect, which is part of the environment. It is rather constant through time, and almost identical in all three lines. The within-family variance catches the other part of the genetic variance (brother-sister genetic differences) as well as the residual variance (which includes microenvironmental, developmental, and measurement error). This is also very similar in all three lines, but it displays an unexplained peak (increase by more than 100%) in generation F_5.

Supplementary material 7

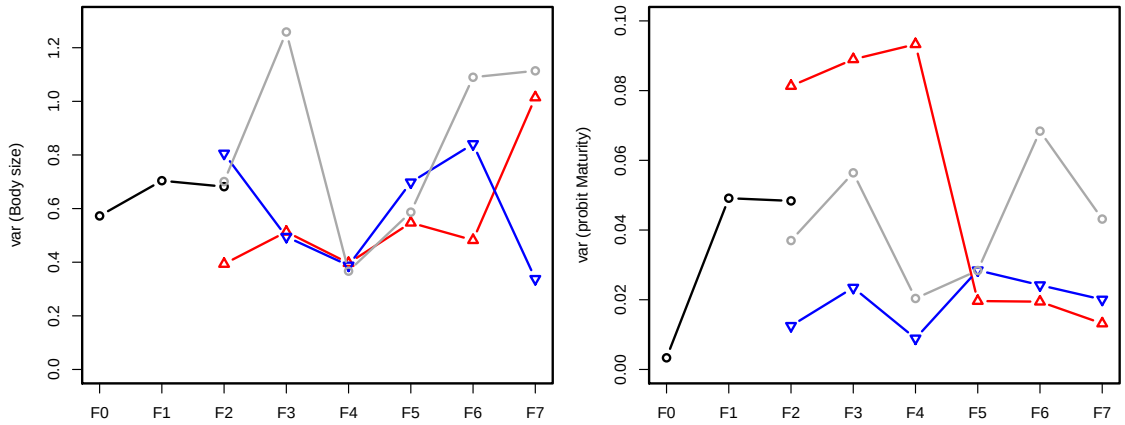
Model fitting on partial datasets



The animal model estimates variance components in the starting population (F_0) accounting for drift and selection in subsequent generations. As a consequence, if the assumptions of the infinitesimal model hold, fitting the model on partial datasets should not affect the estimates (while the posterior distribution is expected to be wider due to the decrease in information). We split the dataset according to (i) generations (fitting the model on generations F_0 to F_3, and from F_4 to F_7), and (ii) to the selected line (Large, Small, and Control lines), fitting the model excluding sequentially each line. In the figure, "Ref" stands for the posterior from the best model, asterisks indicate a significant ($\alpha < 5\%$) difference with the best model. The estimates for genetic variances increase for all sub-datasets, and residual variances and co-variances decrease accordingly. The most straightforward explanation is that the parameters estimated from the full dataset result from a compromise between early/late generations and selection lines, and that the goodness of fit of the model improves when fitted on partial data. Note that most posterior distributions largely overlap, suggesting that the estimated parameters remain meaningful.

Supplementary material 8

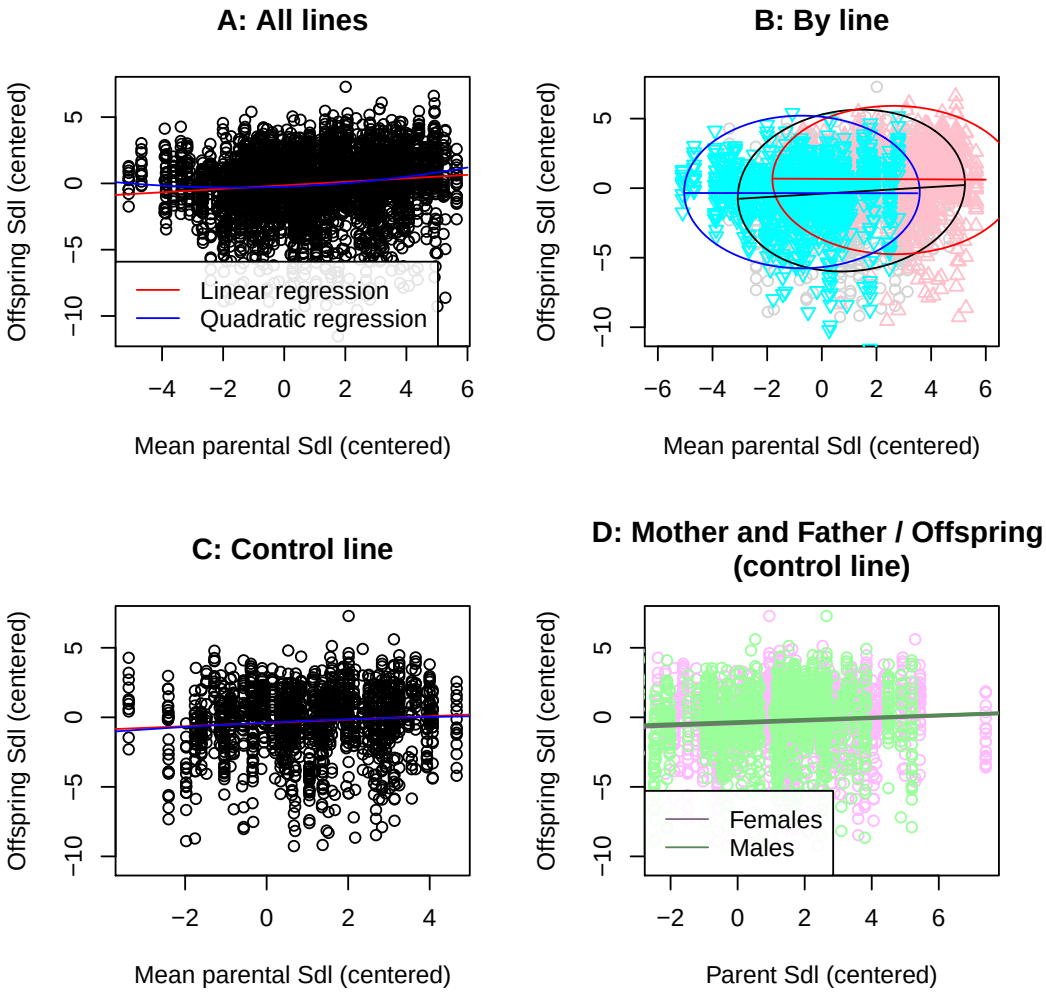
Genetic Variance



Estimated dynamics of the genetic variance, estimated as the variance of the mean posterior of individual breeding values for each selection line each generation. Posteriors were the same as the ones used to compute the dynamics of the mean genetic value for the best model (gRf), displayed in Figure 3.

Supplementary material 9

Parent-offspring regression



The mid-parent-offspring regression coefficient estimates trait heritability. In addition, the shape of the parent-offspring relationship is indicative of potential deviations from the infinitesimal model assumptions. In particular, a non-linear parent-offspring relationship may indicate strong dominance, epistasis, or genetic asymmetries, which could explain asymmetric responses to selection.

A. Taking all selected lines into account, normalizing by generation phenotypic averages to cancel out generation effects, the parent-offspring relationship appears to be slightly non-linear (significant quadratic component: $y = c + h^2x + k_2x$, with $h^2 = 0.083 \pm s.e.0.021$ being an estimate of heritability ($\Pr(h^2 = 0) = 6.57 \cdot 10^{-5}$), the quadratic term being also significant ($\Pr(k_2 = 0) = 5.81 \cdot 10^{-5}$).

B. However, considering each line separately, the pattern rather reflects different linear relationships in all three lines. The Large line response to selection shifts the offspring phenotype upwards, while the Small line lack of response sets the average offspring at the same level as the Control. Non-linearity in this case is the consequence, rather than the cause, of the asymmetric response.

C. When considering the Control line alone, which has the most statistical power because of the large variance in parental phenotypes, the quadratic term disappears, supporting the fact that the parent-offspring regression is linear ($h^2 \simeq 0.14, \Pr(h^2 = 0) = 0.00698, \Pr(k_2 = 0) = 0.62$)

D. Running mother-offspring and father-offspring regressions independently provide very similar results. Focusing on the control line sub-dataset, the mother-offspring regression leads to $h^2 = 0.089 \pm 0.031$ (s.e.), while the father-offspring regression results in $h^2 = 0.081 \pm 0.032$.