

A modelling framework for the analysis of artificial-selection time series

ARNAUD LE ROUZIC^{1,2*}, DAVID HOULE^{1,3} AND THOMAS F. HANSEN¹

¹Center for Ecology and Evolutionary Synthesis, Department of Biology, University of Oslo, Norway

²Laboratoire Évolution, Génome, Spéciation, CNRS UPR9034, Gif sur Yvette, France

³Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

(Received 7 April 2010; revised 23 August, 22 November and 20 December 2010; accepted 21 December 2010; first published online 8 April 2011)

Summary

Artificial-selection experiments constitute an important source of empirical information for breeders, geneticists and evolutionary biologists. Selected characters can generally be shifted far from their initial state, sometimes beyond what is usually considered as typical inter-specific divergence. A careful analysis of the data collected during such experiments may thus reveal the dynamical properties of the genetic architecture that underlies the trait under selection. Here, we propose a statistical framework describing the dynamics of selection-response time series. We highlight how both phenomenological models (which do not make assumptions on the nature of genetic phenomena) and mechanistic models (explaining the temporal trends in terms of e.g. mutations, epistasis or canalization) can be used to understand and interpret artificial-selection data. The practical use of the models and their implementation in a software package are demonstrated through the analysis of a selection experiment on the shape of the wing in *Drosophila melanogaster*.

1. Introduction

The genetic architecture of a character conditions its evolutionary properties (Hansen, 2006). However, being able to determine the number of genes and alleles involved in trait expression, their respective effects, their interactions and their frequencies in a specific population remains a challenge for evolutionary biologists.

A common approach consists in dissecting genetic architectures by identifying the genomic regions influencing phenotypic polymorphism (e.g. QTL mapping) and measuring both their individual effects and their interactions with other genes. Empirical work both in natural and artificial populations generally reports that the genetic architecture of most traits is polygenic. Often, few large factors and many small ones are identified, with a prevalence of non-additive interactions between them (Barton & Keightley, 2002; Carlborg & Haley, 2004; Mackay, 2004). However, the statistical power of gene mapping remains unsatisfactory in terms of genetic architecture description, since most of the genetic variation, attributed to small genetic

factors, typically remains unexplained (Gudbjartsson *et al.*, 2008; Manolio *et al.*, 2009). Except for very simple cases (e.g. Colosimo *et al.*, 2004), the contribution of unidentified factors remains essential to understand and quantify the evolutionary properties of a trait (Le Rouzic *et al.*, 2007). Arguably, most traits might be too complex for such a reductionist approach to quantify evolutionary relevant parameters with satisfactory precision.

Alternatively, the general properties of the genetic architecture may be evaluated through quantitative genetics models from phenotypic measurements in controlled populations. Classical designs include line-cross analyses or parent–offspring studies (Lynch & Walsh, 1998). Of particular interest are selective breeding experiments, which mimic the evolution of a character on a shorter timescale. Through a non-random selection of individuals that will be allowed to breed, it is possible to accumulate, generation after generation, an impressive amount of phenotypic change in a population (Falconer, 1992; Hill & Caballero, 1992).

It is well known that many phenomena like genetic drift (Weber & Diggins, 1990), epistasis (Carter *et al.*,

* Correspondence author: e-mail: lerouzic@legs.cnrs-gif.fr

2005), joint effect of natural selection (Zhang & Hill, 2002), pleiotropy and genetic correlations (Lande, 1979; Hansen & Houle, 2008) and mutations (Hill, 1982a; Zeng *et al.*, 1989), may affect selection response. Consequently, a careful statistical analysis of the time series, combining information about the dynamics of the mean and the phenotypic variance in the population, should be able to (i) sort out the different hypotheses and ground the interpretation of the data (e.g. through the rejection of some models) and (ii) quantify the value of parameters of interest. It is of particular interest to challenge the constant genetic-architecture models based on the ‘infinitesimal’ expectation (infinite number of loci with infinitesimal effect on the phenotype).

So far, few statistical tools are available to measure, survey and predict the changes in the genetic architecture of the trait under selection. From selection-response experiments, researchers generally estimate by regression the average rate at which the population responds to selection (often expressed as the realized heritability). A few other characteristics of interest, such as the selection limits (the extreme phenotypic values beyond which the population does not seem to respond to selection), are sometimes roughly estimated from direct observation of the selection-response time series (Eisen, 1972). Estimating non-constant genetic architecture features from the whole time series requires specific models (e.g. Sorensen *et al.*, 2001) or, more commonly, slicing the time series into parts on which a parameter of interest will be evaluated independently (e.g. Meyer & Hill, 1991; Martinez *et al.*, 2000; Holt *et al.*, 2005). As suggested in Le Rouzic *et al.* (2010), large Mendelian factors could also be detected from artificial-selection responses. However, this becomes difficult when the number of loci increases, and quantitative genetics models have to be considered when the trait is truly polygenic. An individual-based random-effect model (the ‘animal model’) is often used to analyse high-quality data sets reporting individual phenotypes and individual crossing schemes, i.e. full pedigrees (Gianola & Fernando, 1986; Hill & Caballero, 1992; see Walsh & Lynch, 1998). Such models, however, rely on strong hypotheses about the behaviour of the genetic architecture and the way characters are transmitted between parents and offspring. Well suited for the estimation of breeding values in complex pedigrees, they lack flexibility when applied to evolutionary issues, long time series (for which simple heredity assumptions may not hold) or mass-breeding selection experiments (in which the pedigree is not known).

Arguably, the dynamical properties of the selection responses (including changes in the phenotypic mean and the phenotypic variance) are often underexploited. In spite of the availability of many data sets

accumulated for more than a century, and regardless of the scientific interest of comparative studies to address pending questions about the evolutionary properties of genetic architectures, there are few general tools to compare different models fitted to a data set, or to analyse and compare the dynamics of independent experiments on a quantitative basis.

With this paper, we provide a comprehensive and simple statistical framework to analyse the temporal pattern of polygenic artificial-selection time series. First, we propose purely descriptive (phenomenological) models, designed to catch the trends in population means and variances without assuming specific features of the genetic architecture (i.e. including patterns unexpected by classical models). The models can be fit to data by maximum likelihood, and models of different complexity can be compared according to model selection procedures. We then show how these simple models can be modified to account explicitly for mechanisms of biological interest, such as epistasis, canalization or linkage disequilibrium. Models are implemented in a documented software package for R named selection response analysis (sra), and the approach is illustrated by analysing the example of an artificial-selection procedure modifying the shape of the wing in *Drosophila melanogaster*.

2. Models

(i) General framework

Our aim is to describe general time-series models that can be used to analyse artificial-selection data, with a high degree of flexibility related to the dynamics of the process and its underlying genetic mechanisms. The genetic architecture of the trait of interest is assumed to be polygenic, and both genetic and environmental effects are normally distributed (the importance of scale will be addressed below). More complex models involving one or two major loci can be derived (Pong-Wong *et al.*, 1999; Le Rouzic *et al.*, 2010), but will be only briefly considered in this manuscript.

The mean of the trait at generation t is μ_t , and its dynamics are predicted by the breeder’s equation, $\mu_{t+1} = \mu_t + h^2 (\mu_t^* - \mu_t)$, where μ_t^* is the phenotypic average of the artificially selected individuals allowed to reproduce, and h^2 the narrow-sense heritability, i.e. the ratio between additive (σ_A^2) and phenotypic (σ_P^2) variances. Here, we will use Lande’s reformulation of the breeder’s equation: $\mu_{t+1} = \mu_t + \sigma_A^2 \beta$, where β is the selection gradient, and corresponds to the slope of the regression of fitness to the phenotype (Lande, 1979; Lande & Arnold, 1983); the larger $|\beta|$, the stronger the selection. In the context of truncation selection, the realized selection gradient β can be calculated as $(\mu_t^* - \mu_t) / \sigma_P^2$.

Traditionally, the phenotypic variance is split into a genetic component σ_G^2 and an environmental variance σ_E^2 , such as $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$, assuming that the genetic-by-environment interactions can be neglected. The genetic variance itself is split into $\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2$, where σ_D^2 stands for the dominance variance and σ_I^2 for the sum of all epistatic variance components. According to the approximation of the Lande equation, only the additive part of the variance, σ_A^2 , matters for the selection response, and in most of this manuscript, we will consider that $\sigma_P^2 = \sigma_A^2 + \sigma_E^2$. This does not necessarily mean that the models we describe are purely additive, but rather that environmental variance includes the non-additive genetic variance components. In a few cases (e.g. directional epistasis), a part of the non-additive variance contributes to the general dynamics and will be considered explicitly.

(ii) *Phenomenological models*

Accounting for the dynamics of the variance components is achieved by modelling their change in the course of time. We will explore two families of models, corresponding to opposite approaches of statistical modelling. One possibility is to consider a set of models, based on accumulated knowledge in biology and population genetics, and to express the dynamics on the genetic architecture in terms of biologically meaningful parameters. Such models will be referred to as ‘mechanistic models’, and some simple examples will be detailed in the last part of this ‘Model’ section. The alternative possibility is to build flexible models designed to describe trends in the time series, without connecting these observations explicitly to underlying mechanisms. Such models, referred to as ‘phenomenological models’, are the focus of the current section.

Contrary to most situations encountered when analysing time series, neither the population mean nor the variance components are expected to be stationary. The models thus have to include the possibility that the parameters vary in a systematic way. The trends that will be considered for variance components include exponential and linear changes, as well as a combination of both. Such simple models are, however, unlikely to capture the whole complexity of the time series. Ideally, models growing in complexity should be straightforward to build, and extra parameters should provide significant flexibility to the new models. A satisfactory framework was derived by predicting the trend from the previous generations.

The two variance components (the additive genetic variance σ_A^2 and the environmental variance σ_E^2) are assumed to be algebraically independent. To a first approximation, their change can be exponential only,

and the general dynamics can be written as:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_{A_1}^2, \sigma_{E_1}^2, k_A, k_E), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \beta_t, \\ \sigma_{A_{t+1}}^2 &= k_A \sigma_{A_t}^2, \\ \sigma_{E_{t+1}}^2 &= k_E \sigma_{E_t}^2, \\ \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_{E_{t+1}}^2. \end{aligned} \tag{1}$$

This model involves five parameters (in vector Θ), three describing the state of the population at the first generation ($\mu_1, \sigma_{A_1}^2$ and $\sigma_{E_1}^2$), and two featuring the dynamics of the variances (k_A for the additive variance and k_E for the environmental variance).

The model may be expanded by considering linear terms in the variance dynamics:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_{A_1}^2, \sigma_{E_1}^2, k_{A_0}, k_{E_0}, k_{A_1}, k_{E_1}), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \beta_t, \\ \sigma_{A_{t+1}}^2 &= k_{A_0} + k_{A_1} \sigma_{A_t}^2, \\ \sigma_{E_{t+1}}^2 &= k_{E_0} + k_{E_1} \sigma_{E_t}^2, \\ \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_{E_{t+1}}^2. \end{aligned} \tag{2}$$

Variances should remain positive or null, so that the previous setting has to be understood as e.g. $\sigma_{A_{t+1}}^2 = k_{A_0} + k_{A_1} \sigma_{A_t}^2$ if $k_{A_0} + k_{A_1} \sigma_{A_t}^2 > 0$, and 0 otherwise.

Real time series are expected to show more complex patterns, and relatively good generality can be achieved by using a framework based on the previous generations. Because the time series are trended, the model is not equivalent to a classical auto-regressive setting; however, it relies on the same principles: by letting the variances depend on the generations $t-2, \dots, t-n$, n being as large as necessary, it is possible to add up new parameters. The general formulation becomes:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_{A_1}^2, \sigma_{E_1}^2, k_{A_0}, k_{A_1}, k_{A_2}, \dots, k_{A_n}, k_{E_0}, k_{E_1}, \\ &\quad k_{E_2}, \dots, k_{E_n}), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \beta_t, \\ \sigma_{A_{t+1}}^2 &= k_{A_0} + k_{A_1} \sigma_{A_t}^2 + k_{A_2} \sigma_{A_{t-1}}^2 + \dots + k_{A_n} \sigma_{A_{t-n+1}}^2, \\ \sigma_{E_{t+1}}^2 &= k_{E_0} + k_{E_1} \sigma_{E_t}^2 + k_{E_2} \sigma_{E_{t-1}}^2 + \dots + k_{E_n} \sigma_{E_{t-n+1}}^2, \\ \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_{E_{t+1}}^2. \end{aligned} \tag{3}$$

The parameters k_n are purely descriptive, and they are not intended to be interpreted in terms of meaningful biological processes. In particular, non-Markov lags ($n > 1$) should not be understood as a real impact of remote generations on the current dynamics of the variance components, but rather as hidden variables (including e.g. skew or kurtosis of the distribution of genetic effects, complex dominance or epistasis patterns, changes in frequency of large-effect alleles, etc) whose effects happens to be captured by this parameterization.

Allowing the k_n to take negative values could lead to non-convergent series, which might indicate over-parameterization, but could also identify potentially relevant cyclic patterns in the data set. If considered as a nuisance, this behaviour can be avoided by forcing $k_n > 0$. Models have been explored up to $n = 3$, with satisfactory results, as evidenced by the analysis of real data detailed later.

(iii) *Alternative scales*

In many cases, there is no particular reason to expect that the description of the genetic architecture from the scale on which the trait was measured is *a priori* meaningful. Models of similar complexity can thus be defined just by allowing the scale to change. Scale changes include common logarithm and power transformations, or more specific operations for traits expressed as frequencies or probabilities (e.g. logistic transformation).

In addition to general scale transformation, it is possible to consider scaling operations that are more specific to quantitative genetics. Most of the time, artificial-selection procedures target characters that are on ratio scales, such as weights or lengths. Consequently, phenotypic and genetic standard deviations should change along with the mean of the population, as suggested by Houle (1992). Here, we use the ‘mean-scaled evolvability’, defined as $I_A = \sigma_A^2 / \mu^2$, as a measurement of the adaptation potential of a population (see Hansen *et al.*, 2003; Hansen & Houle, 2008 for justification). The environmental variance can also be affected, such that $I_E = \sigma_E^2 / \mu^2$ is the quantity of interest. In this context, equation (1) can be expressed as:

$$\begin{aligned} \Theta &= (\mu_1, I_A, I_E, k_A, k_E), \\ \mu_{t+1} &= \mu_t + I_A \mu_t^2 \beta_t, \\ I_{A,t+1} &= k_A I_{A,t}, \\ I_{E,t+1} &= k_E I_{E,t}, \\ I_{P,t+1} &= I_{A,t+1} + I_{E,t+1}. \end{aligned} \tag{4}$$

Direct comparison of the parameters can be achieved by rewriting equation (4) in terms of σ_A^2 and σ_E^2 , given that $I_A = \sigma_A^2 / \mu_t^2$, e.g.: $\sigma_{A,t+1}^2 = \mu_{t+1}^2 \cdot k_A \sigma_{A,t}^2 / \mu_t^2$.

Alternatively, it is possible to interpret the evolutionary properties of the population based on the narrow-sense heritability h^2 , the ratio between additive and phenotypic variance. A model centred on heritability is

$$\begin{aligned} \Theta &= (\mu_1, h_1^2, \sigma_P^2, k_h, k_P), \\ \mu_{t+1} &= \mu_t + h_t^2 \beta_t \sigma_P^2, \\ h_{t+1}^2 &= k_h h_t^2, \\ \sigma_{P,t+1}^2 &= k_P \sigma_P^2. \end{aligned} \tag{5}$$

All the models described in this paper can be applied to any scale (e.g. log/power scale, evolvability or heritability scales). These combinations provide a large number of settings designed to catch major trends in the data, and to describe changes in the genetic architectures.

(iv) *Mechanistic models*

The quantitative-genetics literature focuses on mechanistic models, describing how selection response is expected to change assuming some specific genetic architectures. Model selection may rule out some of these, while leaving others can appear as plausible explanations for the trends seen in the phenomenological analysis. In this section, some of these models will be described as special cases of equation (2). This will allow us to estimate biologically relevant parameters directly.

(a) *Constant additive variance*

The simplest model assumes that the genetic architecture of the population stays constant. This model thus considers that $\sigma_{A_t}^2 = \sigma_{A_1}^2$ and $\sigma_{E_t}^2 = \sigma_{E_1}^2$ at any time t , so that the response to selection is linear (provided that selection does not vary). The ‘infinitesimal model’ of genetic architecture (Lande, 1975) is generally cited as a justification for the constant variance model. By assuming an infinite number of loci, each of them having an infinitely small effect on the phenotype, the mean of the population can change without affecting allele frequencies (Bulmer, 1980; Turelli, 1988). Neglecting migration, mutation, linkage disequilibrium and genetic drift, the additive variance can remain constant. The infinitesimal model is widely used in the literature, either explicitly or implicitly.

Only three parameters are necessary to estimate the expected dynamics, with

$$\begin{aligned} \Theta &= (\mu_1, \sigma_A^2, \sigma_E^2), \\ \mu_{t+1} &= \mu_t + \sigma_A^2 \cdot \beta_t, \\ \sigma_{P,t+1}^2 &= \sigma_A^2 + \sigma_E^2. \end{aligned} \tag{6}$$

This is equivalent to equation (2) assuming $k_{A_0} = 0$, $k_{A_1} = 1$, $k_{E_0} = 0$, and $k_{E_1} = 1$.

This model constitutes a basis on which more realistic models can be built. Such models may introduce the effects of finite population sizes (genetic drift), mutations or linkage disequilibrium. Although described separately for clarity, they can be combined in order to get a more realistic picture of the properties of a population under selection (Keightley & Hill, 1987).

(b) *Genetic drift*

The constant additive-variance prediction can be interpreted as the consequence of the assumption that

individual allele frequencies do not change much. However, unless the population is virtually infinite, genetic drift will drive alleles to irreversible fixation, with a rate that is inversely proportional to the population size. Genetic drift thus has two major impacts, a stochastic effect that can be accounted for when calculating the likelihood (as detailed later in this manuscript or in e.g. Le Rouzic *et al.*, 2010) and a deterministic effect on the additive variance, which is the topic of the current paragraph. Drift in finite populations is indeed expected to decrease σ_A^2 by a factor of $1/(2N_e)$ in each generation (Robertson, 1960; Lande, 1975), where N_e is the effective population size. In small populations and long time series, this deterministic effect is cumulative over generations and can be strong enough to affect the population dynamics. It can be accounted for simply by considering the same model as in equation (6), but letting the additive variance σ_A^2 decrease with time:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_{A_1}^2, \sigma_E^2), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \cdot \beta_t, \\ \sigma_{A_{t+1}}^2 &= \sigma_{A_t}^2 \cdot \left(1 - \frac{1}{2N_e}\right), \\ \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_E^2. \end{aligned} \tag{7}$$

(c) Mutations

From mutation-accumulation experiments, it is known that spontaneous mutations tend to generate a small but significant amount of selectable additive variation, ranging from 0.01 to 1% of σ_P^2 per generation, the mutational coefficient of variation being 0.1–4% of the trait mean (e.g. mutational evolvability $I_M = \sigma_M^2/\mu^2$ between 10^{-6} and 1.6×10^{-3}) (Lynch, 1988; Houle *et al.*, 1996; Keightley, 1998; Houle, 1998; Lynch *et al.*, 1999; Keightley, 2004). The depletion in additive variance is thus rarely total and irreversible. Mutations can be accounted for, considering that a fixed (additive) mutational variance σ_M^2 (a composite parameter featuring the appearance of new mutations and their change in frequency under selection) is generated in each generation:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_{A_1}^2, \sigma_E^2, \sigma_M^2), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \cdot \beta_t, \\ \sigma_{A_{t+1}}^2 &= \sigma_{A_t}^2 + \sigma_M^2, \\ \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_E^2. \end{aligned} \tag{8}$$

Similar models can be defined according to the alternative scales detailed before. For instance, on the evolvability scale, the quantity of interest would be $I_M = \sigma_M^2/\mu^2$, and $I_{A_{t+1}} = I_{A_t} + I_M$. Similarly, on the heritability scale, the logical parameter should be $h_m^2 = \sigma_M^2/\sigma_P^2$; however, $h_m^2 = \sigma_M^2/\sigma_E^2$ could also be used

to correspond to a quantity that is often estimated empirically (e.g. Lynch, 1988).

This model remains a rough approximation of what happens at mutation–selection–drift equilibrium. The actual probability for mutations to appear and to be fixed depend on the distribution of mutational effects, strength of selection and population size (Hill, 1982*b*; Bürger & Ewens, 1995; Otto & Whitlock, 1997; Hermisson & Pennings, 2005; Engen *et al.*, 2009).

(d) Linkage disequilibrium

In artificial selection, breeders are chosen according to some non-random selection rule, such that the group of breeders is a biased sample of the population. Most of the time, breeders tend to be more similar to each other than the rest of the population: only big or small phenotypes are bred (directional selection), or rarely individuals close to the mean of the population are chosen (stabilizing selection). Exceptionally, the bias is towards dissimilar individuals (disruptive selection). In any case, the phenotypic variance among breeders is different from the phenotypic variance in the population, and so is the additive variance. Such a direct or indirect selection on the variance will generate some linkage disequilibrium, which will affect the selection response. This effect of linkage disequilibrium on selection response is known as the ‘Bulmer effect’ (Bulmer, 1971).

A simple model consists in decomposing the additive variance: $\sigma_A^2 = \sigma_a^2 + d$; the ‘genic’ variance σ_a^2 corresponding to the additive variance if all loci were at linkage equilibrium, the influence of linkage disequilibrium being d . According to Bulmer (1971), $d_{t+1} = \frac{1}{2}(d_t + \delta_t \sigma_{A_t}^2 / \sigma_{P_t}^2)$, where $\delta_t = (s_t^{2*} - s_t^2) / s_t^2$ stands for the relative difference between the phenotypic variance of the breeders (s_t^{2*}) and the phenotypic variance of the rest of the population s_t^2 . If the variance among breeders is unknown, there is no general formula to compute this effect in all situations, but it can be deduced from the properties of the normal distribution for specific selection regimes.

Assuming that the genic variance is constant, the change in additive variance between two generations can thus be written as $\sigma_{A_{t+1}}^2 = \sigma_a^2 + d_{t+1}$, and since $d_t = \sigma_{A_t}^2 - \sigma_a^2$,

$$\sigma_{A_{t+1}}^2 = \frac{1}{2} \sigma_a^2 + \sigma_{A_t}^2 \cdot \frac{1}{2} \left(1 + \delta_t \frac{\sigma_{A_t}^2}{\sigma_{A_t}^2 + \sigma_{E_t}^2}\right). \tag{9}$$

If $\delta_t < 0$ (breeders are more similar to each other than in the whole population, which is expected for most selection regimes, including directional and stabilizing), the additive variance is reduced. If $\delta_t > 0$ (disruptive selection), the additive variance in the population is larger than in a non-selected population.

Combining linkage disequilibrium with e.g. genetic drift as described in equation (7), the following model can be derived:

$$\begin{aligned}
 \Theta &= (\mu_1, \sigma_{a_1}^2, \sigma_E^2, N_e, d_1), \\
 \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \cdot \beta_t, \\
 \sigma_{A_{t+1}}^2 &= \sigma_{a_{t+1}}^2 + d_{t+1}, \\
 \sigma_{a_{t+1}}^2 &= \sigma_{a_t}^2 \cdot \left(1 - \frac{1}{2N_e}\right), \\
 d_{t+1} &= \frac{1}{2} \left(1 - \frac{1}{N_e}\right) \left(d_t + \delta_t \frac{(\sigma_{A_t}^2)^2}{\sigma_{P_t}^2}\right), \\
 \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_E^2.
 \end{aligned}
 \tag{10}$$

A similar result was derived by Dempfle (1975) using a different approach. While changes in μ depend on selection on trait value, changes in d occur because of selection on variance. It is thus possible to predict changes in additive variance even if $\beta=0$ (for instance, if selection is stabilizing or disruptive). If there is no selection on variance ($\delta_t=0$) and loci are unlinked, the effect on linkage disequilibrium is halved each generation. If selection is constant, the disequilibrium rapidly stabilizes. This model considers that the initial linkage disequilibrium d_1 is unknown, but if the initial population is derived from several generations of random mating, it can be safely assumed that $d_1=0$.

(e) *Finite number of loci*

The infinitesimal model relies on the assumption that a large number of small-effect loci influence the trait. When this assumption is violated, the additive variance is expected to decrease faster than predicted in the models described above. Although precisely modelling the effect of individual loci is unrealistic, a general model was proposed by Chevalet (1994). This model introduces an additional parameter, n , corresponding to the number of loci in the genetic architecture, provided that all loci have the same effect. The previous model, including drift and linkage disequilibrium, can then be rewritten as:

$$\begin{aligned}
 \Theta &= (\mu_1, \sigma_{a_1}^2, \sigma_E^2, N_e, d_1, n), \\
 \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \cdot \beta_t, \\
 \sigma_{A_{t+1}}^2 &= \sigma_{a_{t+1}}^2 + d_{t+1}, \\
 \sigma_{a_{t+1}}^2 &= \sigma_{a_t}^2 \cdot \left(1 - \frac{1}{2N_e}\right) + \left(1 - \frac{1}{N_e}\right) \delta_t \frac{(\sigma_{A_t}^2)^2}{2n\sigma_{P_t}^2}, \\
 d_{t+1} &= \frac{1}{2} \left(1 - \frac{1}{N_e}\right) \left(d_t + \left(1 - \frac{1}{n}\right) \delta_t \frac{(\sigma_{A_t}^2)^2}{\sigma_{P_t}^2}\right), \\
 \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_E^2.
 \end{aligned}
 \tag{11}$$

If loci actually have different effects, n can be interpreted as n_e , an effective number of loci (Chevalet,

1994). As a first approximation, when estimating the number of loci, n_e can be considered as constant. However, this assumption is likely to be inaccurate for long time series, since n_e is expected to change along with the allele frequencies.

(f) *Directional epistasis*

Epistasis refers to the situation in which the effect of an allele substitution at a given locus depends on the genotype at other loci. Originally defined for discrete Mendelian genes, the concept has been extended to quantitative genetics, and corresponds to genetic effects that cannot be attributed to single-locus (i.e. additive and dominance) effects (Cockerham, 1954; Kempthorne, 1954; Phillips, 1998). Although often neglected in evolutionary models, epistasis (as well as dominance, the intra-locus equivalent to epistasis) has proved to be common for quantitative traits (Carlborg & Haley, 2004), and their impact on the selection response may be important (Carter et al., 2005; Hallander & Waldmann, 2007; Hansen et al., 2006; Le Rouzic et al., 2007; 2008).

When there is epistasis, the effect of artificial or natural selection can still be predicted by the Lande equation over a single generation (i.e. only additive variance matters), but changes in the genetic background due to selection modify genetic effects, so that the additive variance can change (upwards or downwards according to the pattern of interactions) in the course of time. In addition, the dynamics of non-additive genetic variance terms (interaction variance) may lead to changes in what is considered as σ_E^2 in the current setting.

One impact of epistasis on a selection-response time series can be summarized by a directional epistasis coefficient ε (Carter et al., 2005), featuring the average curvature of the genotype-phenotype map in the current population. If directionality is positive, epistasis tends to amplify the effects of alleles that have a positive effect on the trait, and the selection response accelerates when selecting for higher phenotypic values (decelerates when selecting for lower phenotypes). If $\varepsilon < 0$, allelic effects tend to reduce each other, and ‘up’ selection becomes less and less effective, while ‘down’ selection becomes easier. If $\varepsilon = 0$, the population will respond to selection in a similar way as if there was no epistasis.

If the genotype-phenotype map is multilinear (Hansen & Wagner, 2001), the interaction between two loci i and j is quantified by an epistatic parameter ${}^{ij}\varepsilon$. Even if ${}^{ij}\varepsilon$ is different between all pairs of loci (the variance of all pairwise ε can be noted σ_ε^2), the properties of the genotype-phenotype map can still be summarized by the composite parameter ε (Carter et al., 2005). Two predictions can be made: (i) the change in additive variance scales with ε and with the

strength of selection, and (ii) the epistatic variance σ_{AA}^2 scales with the square of the additive variance. More specifically:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_{A_1}^2, \sigma_E^2, \varepsilon, \sigma_\varepsilon^2), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \cdot \beta_t, \\ \sigma_{A_{t+1}}^2 &= \sigma_{A_t}^2 + 2\beta_t \varepsilon (\sigma_{A_t}^2)^2, \\ \sigma_{AA_{t+1}}^2 &\simeq (\varepsilon^2 + \sigma_\varepsilon^2) \cdot (\sigma_{A_{t+1}}^2)^2, \\ \sigma_{P_{t+1}}^2 &= \sigma_{A_{t+1}}^2 + \sigma_{AA_{t+1}}^2 + \sigma_E^2. \end{aligned} \tag{12}$$

In practice, σ_ε^2 has very little impact on the overall time series, and can be ignored. The multilinear model remains a local approximation, and is probably less realistic when studying large phenotypic distances. In particular, long-term selection (directional or stabilizing) may lead to changes in ε (Hermisson *et al.*, 2003; Hansen *et al.*, 2006; Álvarez-Castro *et al.*, 2009). Directional epistasis predicts an asymmetric response to selection, additive variance increasing when selecting in the direction of epistasis.

(g) Canalization

Directional epistasis is an operational tool to model and describe genetic architectures, and contrary to most architecture parameters (e.g. mutational variance, effective population size), it can be estimated from QTL data (Le Rouzic & Álvarez-Castro, 2008; Pavlicev *et al.*, 2010), and thus has the potential to help linking the results of reductionist genetic-architecture dissections and selection-response approaches. Yet, directional epistasis does not allow e.g. additive variance to decrease or increase in both selection directions, corresponding to what may be referred to as genetic canalization and decanalization.

A simple way to model genetic canalization is to consider that additive variance decreases or increases according to the distance δ_{μ_t} between the mean of the population μ_t and a specific phenotypic value θ (the ‘canalization optimum’) at which the additive variance is maximum (or minimum). The distance might be linear ($\delta_{\mu_t} = |\mu_t - \theta|$) or more elaborated (e.g. $\delta_{\mu_t} = (\mu_t - \theta)^2$). The model can be written as $\sigma_{A_{t+1}}^2 = \sigma_{A_t}^2 \cdot e^{k_g \delta_{\mu_{t+1}}}$, and a recursive form, more similar to the previous models, can be easily derived since $\sigma_{A_{t+1}}^2 = \sigma_{A_t}^2 \cdot e^{k_g(\delta_{\mu_{t+1}} - \delta_{\mu_t})}$. The strength of the genetic canalization is quantified by the parameter k_g , if $k_g > 0$, σ_A^2 increases with δ_{μ_t} (decanalization), while $k_g < 0$ models some genetic canalization.

Environmental variance can also be affected by canalization or decanalization. Genotype-by-environment ($G \times E$) interactions occur when genotypes react in a different way to the same environmental conditions. When selecting for extreme phenotypes, it is likely that the selected genotypes may differ in their ability to deal with stress,

temperature changes, food-quality variation, etc. Modelling environmental canalization can be done in a very similar way as for genetic canalization, by considering that environmental variance may increase when the population gets farther from a specific phenotypic value θ . This model thus assumes that σ_E^2 depends on a measure δ_{μ_t} , defined in the same way as for genetic canalization. In this model, k_c quantifies the strength of the canalization; if $k_c > 0$, σ_E^2 increases when the population moves away from its initial state (the genotypes are less stable and more sensitive to micro-environmental effects); on the contrary, $k_c < 0$ indicates a canalizing effect when the mean is shifted.

Considering both genetic and environmental canalization gives:

$$\begin{aligned} \Theta &= (\mu_1, \sigma_A^2, \sigma_{E_1}^2, k_g, k_c, \theta), \\ \mu_{t+1} &= \mu_t + \sigma_{A_t}^2 \cdot \beta_t, \\ \sigma_{A_{t+1}}^2 &= \sigma_{A_t}^2 \cdot e^{k_g \delta_{\mu_{t+1}}}, \\ \sigma_{E_{t+1}}^2 &= \sigma_{E_t}^2 \cdot e^{k_c \delta_{\mu_{t+1}}}, \\ \sigma_{P_{t+1}}^2 &= \sigma_A^2 + \sigma_{E_{t+1}}^2. \end{aligned} \tag{13}$$

To simplify the model, as a first approximation, θ can be considered as identical to the initial population mean ($\theta = \mu_1$).

(h) Effect of natural selection

Even in carefully controlled conditions in the lab, it is impossible to avoid the effects of natural selection. Indeed, selected traits are often correlated with fitness, either by influencing directly the survival or the fecundity of the organism, or by sharing some pleiotropic genes with fitness-related traits.

If natural selection tends to have a stabilizing effect, the efficiency of artificial selection will decrease when the mean of the population moves away from the optimal phenotype. As a consequence, selection response is expected to decrease as the mean changes, while a respectable amount of additive variance is still present in the population (e.g. Zhang & Hill, 2002). These ‘joint-effect’ models should be considered as alternative models to the framework proposed in equation (2), since the additive variance is not really reduced (e.g. when computing the phenotypic variance), but only partially available for artificial selection. A simple way to model this phenomenon is to consider a slightly modified version of Lande equation, such as:

$$\mu_{t+1} = \mu_t + \sigma_{A_t}^2 [\beta_t + s(\theta - \mu_t)], \tag{14}$$

in which the efficiency of artificial selection is decreased as μ_t gets further from the phenotypic optimum θ , s being the strength of natural selection (stabilizing when $s < 0$, the larger $|s|$, the stronger the

selection). If artificial selection is relaxed, $\beta_t = 0$, and the mean of the population will tend to stabilize at θ . Under a constant directional artificial-selection pressure, the population will eventually stabilize at a point where artificial and natural selection intensities are equal, but in opposite directions, so that no more phenotypic change occurs. The selection function presented in equation (14) is chosen for simplicity, and realistic alternatives (e.g. quadratic or exponential call-back terms) may be considered. More complex mechanistic models of joint-effect selection have been elaborated (Lande, 1983; Zhang & Hill, 2002; 2005) and may be used under different assumptions on e.g. distribution of allele frequencies.

Another alternative is to consider that the focal trait is only correlated with characters constrained by stabilizing selection, without itself having a direct impact on fitness. These correlations slow down the selection response, but do not lead to a halt: the trait can evolve, but not as freely as if all other traits were free to change; the ability to evolve when all correlated characters are kept constant has been termed the conditional evolvability of the trait (Hansen, 2003; Hansen et al., 2003a; Hansen & Houle, 2008). Deriving a precise mechanistic model would require some knowledge about the strength of stabilizing selection on the other traits as well as their correlation with the focus trait. Nevertheless, a simple model such as:

$$\mu_{t+1} = \mu_t + k_a \sigma_A^2 \beta_t, \tag{15}$$

with $0 < k_a < 1$, is able to describe the expected effect of the correlation of the selected trait with one or other characters that is not allowed to vary. The factor k_a can be interpreted as the part of the additive genetic variation in the trait of interest that is independent with traits under strong natural selection (Hansen et al., 2003a). If $k_a = 0$, no change in the trait is possible, in spite of some genetic variation, because all genes underlying the trait also affect other characters that are strongly constrained.

3. Methods

(i) Maximum-likelihood estimates

(a) General setting

The models described in the previous section provide a way to predict the response to selection from a given set of parameters. They can also be used to estimate the dynamic properties of the genetic architecture, given an experimental data set. The next paragraph details the way to compute the likelihood of a specific model given some phenotypic time series Y , which can be used to estimate the parameters Θ by maximizing the likelihood function $L(\Theta|Y)$.

A typical artificial-selection experiment is carried out on a population of size N , in which the phenotype of interest $Y = (y_1, \dots, y_i, \dots, y_N)$ (i.e. the trait on which selection is performed) is measured for each individual i . A subset Y^* of individuals is then chosen for breeding according to some selection rule. The process is then being repeated over T generations, and the full data set is $Y = (Y_1, \dots, Y_T)$.

It is common to report and archive only summary statistics of the population at each generation t : the average phenotype, \bar{y}_t , the population variance of the phenotype, $s_{y_t}^2$ and the mean phenotype of the breeders, \bar{y}_t^* . The selection gradient at generation t can be computed as $\beta_t = (\bar{y}_t^* - \bar{y}_t) / s_{y_t}^2$ (i.e. the selection differential divided by the phenotypic variance). For convenience, the full data set, Y , will be split into $Y = (\bar{y}, s_y^2)$, where $\bar{y} = (\bar{y}_1, \dots, \bar{y}_T)$, and $s_y^2 = (s_{y_1}^2, \dots, s_{y_T}^2)$. In addition, two vectors, $\beta = (\beta_1, \dots, \beta_T)$, and $N = (N_1, \dots, N_T)$, describe the experimental setting (selection gradient and population size for each generation).

Deterministic models for the dynamics of the mean and the variance in the population provide the theoretical expectations μ_t and $\sigma_{p_t}^2$ at generation t . However, even if the population is large and the underlying model is correct, the observed means and variances \bar{y}_t and $s_{y_t}^2$ will be affected by sampling error: each individual phenotypic measurement y_{it} is expected to follow a Gaussian distribution of mean μ_t and variance $\sigma_{p_t}^2$. The probability density for the whole population at this generation is thus:

$$P(Y_t | \mu_t, \sigma_{p_t}^2) = \prod_{i=1}^{N_t} \phi(y_{it} | \mu_t, \sigma_{p_t}^2), \tag{16}$$

where $\phi(x | \mu, \sigma^2)$ is the Gaussian probability density with mean μ and variance σ^2 .

The polygenic models assume that the phenotypes in a population follow a normal distribution, so that the entire distribution can be described by its mean and its variance. For the whole time series (T generations), the total probability is

$$P(\bar{y}, s_y^2 | \mu, \sigma_{p_y}^2; \beta, N) = \prod_{t=1}^T P(\bar{y}_t, s_{y_t}^2 | \mu_t, \sigma_{p_t}^2; N_t). \tag{17}$$

The models will predict the expected means and variances (μ_t and $\sigma_{p_t}^2$) for all generations, based on a set of parameters describing the genetic architecture (Θ) and the strength of selection for all generations (β). The full likelihood function can thus be written as:

$$L(\Theta | Y) = \prod_{t=1}^T P(\bar{y}_t, s_{y_t}^2 | \Theta; \beta, N). \tag{18}$$

The estimates of interest are the values $\hat{\Theta}$ maximizing this function.

(b) *Combining time series*

This framework is flexible, and can easily be adapted to particular experimental designs. A common feature in selection-response experiments is that several time series are often initiated from the same population, so that the data comprise several series of selection responses. Typically, at least two lines are selected, one for high values of the trait, the other one for low values; sometimes an unselected control line is used as a reference, and some selection treatments can be replicated. Considering all these series in a single genetic-architecture model would improve the precision and the meaning of the parameters, since differently selected lines will explore different parts of the genotype–phenotype map.

As discussed in Le Rouzic *et al.* (2010), if Y_1, \dots, Y_R are R phenotypic series initiated from a unique population, it is possible to combine the likelihoods such that:

$$L(\Theta; Y_1, \dots, Y_R) = \prod_{r=1}^R L(\Theta; Y_r) \tag{19}$$

$$= \prod_{r=1}^R \prod_{t=1}^{T_r} P(\bar{y}_{rt}, s_{y_{rt}}^2 | \Theta; \beta_r, N_r).$$

(c) *Random-effect models*

The likelihood function described above is based on the assumption that the sampling effect is the only source of stochasticity. In some cases, however, this assumption is unreasonable. Including additional sources of stochasticity can be achieved by considering nuisance parameters (e.g. generation-specific effects) that are not estimated individually, but rather considered as random effects. Two families of random effect models are briefly described below and detailed as supplementary material: (i) generation-specific macro-environmental effects and (ii) genetic drift.

The possibility to include generation-specific environmental effects is detailed in the Supplementary Methods section. The model considers that the influence of environment can be split into two components. The micro-environmental variance σ_E^2 , as described above, explains the non-genetic dispersion of the phenotypes among individuals of the same population, while the macro-environmental variance σ_{me}^2 catches generation-specific environmental shifts. Macro-environmental effects are considered as random effects, and the likelihood can be computed analytically under the assumption that effects are not correlated between generations.

Another source of randomness is genetic drift. Although small when considering consecutive generations, deviations due to genetic drift accumulates

over time and may impact the time series substantially, especially in small populations. The possibility to account for drift on the mean phenotype of the population was detailed in Le Rouzic *et al.* (2010), and is briefly described below.

The model is based on the assumption that the genetic mean μ_t and the additive variance σ_t^2 at generation t are random effects, the source of stochasticity being genetic drift. Then, $\mu_t = \tilde{\mu}_t + x_1$, where $\tilde{\mu}_t$ is the expected mean at generation t (e.g. $\tilde{\mu}_t = \mu_{t-1} + \sigma_{A_{t-1}}^2 \beta_{t-1}$ or any alternative model describing changes in the mean), and x_1 is drawn in a normal distribution of mean 0 and variance σ_A^2/N , the sampling variance of the genetic mean. In a similar way, $\sigma_{A_t}^2 = \tilde{\sigma}_{A_t}^2 x_2 / (N-1)$, where x_2 follows a χ^2 distribution with $N-1$ degrees of freedom, and $\tilde{\sigma}_{A_t}^2$ is the expected value of the variance at generation t , e.g. $\tilde{\sigma}_{A_t}^2 = \sigma_{A_{t-1}}^2 [1 - 1/(2N_e)]$ or any alternative model. Assuming that drift affects genetic mean and genetic variance independently, equation (16) can be rewritten as:

$$P(Y_t | \Theta; \mu_t, \mu_{t-1}, \sigma_{A_t}^2, \sigma_{A_{t-1}}^2) = P(\mu_t | \tilde{\mu}_t, \sigma_{A_{t-1}}^2) \tag{20}$$

$$\times P(\sigma_{A_t}^2 | \tilde{\sigma}_{A_t}^2, \sigma_{A_{t-1}}^2) \prod_{i=1}^{N_t} \phi(y_{it} | \mu_t, \sigma_{A_t}^2 + \sigma_E^2).$$

The vector of genetic means μ and the vector of additive variances σ_A^2 being considered as random effects, the likelihood function has to be integrated over them:

$$L(\Theta | Y) = \int \dots \int_{\mu_1, \dots, \mu_T, \sigma_{A_1}^2, \dots, \sigma_{A_T}^2} \left[\prod_{t=1}^T P(Y_t | \Theta; \mu_t, \mu_{t-1}, \sigma_{A_t}^2, \sigma_{A_{t-1}}^2; \beta_t, N_t) \right] \tag{21}$$

$$\times d\mu_1 \dots d\mu_T d\sigma_{A_1}^2 \dots d\sigma_{A_T}^2.$$

Contrary to macro-environmental variance (described in the Supplementary Results 1), genetic drift generates random deviations that are correlated between consecutive generations. The likelihood cannot be computed independently for each generation, and analytical development is complicated. Instead, likelihoods were calculated based on Laplace approximation, using the random-effect module of the software ADM-B-RE (Skaug & Fournier, 2006). Laplace approximation is exact when the posterior distribution of the random effects is normal, which is not exactly the case here (variances are χ^2 -distributed). For simplicity, when implementing the model, it was considered that $(\log(\sigma_{A_t}^2 / \sigma_{A_t}^2))$ was normally distributed with a mean of 0 and a variance of $2/(N_t - 1)$, which is a good approximation provided that $2/(N_t - 1) \ll 1$.

(ii) *Implementation*(a) *Model fitting*

Some of the models described in the previous section have been implemented as a fully documented free package for the statistical software R (R Development Core Team 2007). The package *sra* and its documentation can be downloaded on <http://cran.r-project.org/web/packages/sra/>. It provides a user-friendly wrapper for the optimum function of R, which performs the numerical maximization of the likelihood function.

In general, numerical convergence is not problematic when meaningful starting values are provided. Positive-only parameters (such as the variances or the population size) are log transformed before being fit in order to improve the efficiency of the convergence algorithm. The likelihood function of the highly nonlinear directional-epistasis model (equation (12)) was challenging for the gradient-based algorithms, and the package implements directional epistasis as two distinct models (positive and negative epistasis) that are discriminated based on their likelihood. Models were compared by calculating their Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC) (Akaike, 1973; Anderson *et al.*, 2000; Burnham & Anderson, 2002).

The most complex models, involving more than ten parameters, were challenging for simple hillclimbing algorithms. In particular, some phenomenological models (those in which more than one-generation lags were considered) often produced a complex likelihood function with several peaks. Maximum-likelihood estimates thus depend on the starting values, and the results reported in the next section are based on the best likelihood score out of C convergence attempts with randomized starting values, C varying from 1 to 30 depending on the complexity of the model. This procedure gave convincingly stable results, even if there is no certainty that the global maximum was reached. In any case, this does not raise interpretation issues, since different likelihood peaks generally correspond to similar patterns for the dynamics of the means and the variances. Conversely, convergence in mechanistic models was less problematic, and the few optimization errors that were encountered were easily solved by using the results of simpler models (e.g. the constant-variance model) as starting values for the more complex ones.

The likelihood function in the model including macro-environmental effects could be derived analytically, and convergence was achieved in R. The drift model is more complex and was implemented in the software AD Model Builder (<http://admb-foundation.org>), which offers a more sophisticated convergence algorithm (Skaug & Fournier, 2006) based on Laplace approximation of the likelihood

function. The corresponding code is available on request.

The framework proposed in the previous sections is flexible and can be easily extended. Possible improvements or alternatives include: (i) the definition of additional mechanistic models in order to test new biological hypotheses, and (ii) the adaptation of the statistical framework for philosophical or pragmatic reasons, e.g. by replacing the maximum-likelihood approach with a Bayesian framework (see e.g. O'Hara *et al.*, 2008).

(b) *Simulations*

Simulated data sets were generated to assess the validity of the models and the accuracy of the parameter estimates. Two procedures have been used to generate simulated data: (i) stochastic simulation of the model and (ii) individual-based simulations.

The stochastic simulation procedure is based on deterministic times series generated from specific models (equations (6)–(15)) with known parameters. Theoretical means and variances are then randomly modified independently for each generation, according to $\mu_{\text{sim}} = \mu_{\text{th}} + x_1 + x_2$ and $\sigma_{p_{\text{sim}}}^2 = \sigma_{p_{\text{th}}}^2 \cdot y / (N - 1)$, where x_1 is drawn from a normal distribution of mean 0 and variance $\sigma_{p_{\text{th}}}^2 / N$ (sampling variance of the mean), x_2 is drawn from a normal distribution of mean 0 and variance σ_{me}^2 (macro-environmental variation), and y is drawn in a χ^2 distribution with $N - 1$ degrees of freedom (sampling variance of the variance).

Individual-based simulations rely on the constant-variance model, considering a genetic value g_i and a phenotypic value p_i for each simulated individual i . At each generation, the N individuals are ranked according to their phenotype, and the N_{sel} extreme individuals constitute the parental population. A pair of parents is then randomly drawn in this parental population for each offspring j , the genotypic value of which is $g_j = (g_f + g_m) / 2 + x_j$, where g_f and g_m are the genotypic value of the father and the mother, respectively, and x_j is a random number drawn in a normal distribution of mean 0 and variance $\sigma_A^2 / 2$. The phenotypic value of individual j is drawn in a normal distribution of mean g_j and variance σ_E^2 .

4. Statistical properties

The accuracy of the likelihood equation as well as the software implementation of the models have been evaluated by simulation. Several properties of the statistical models have been explored: the ability for the model to converge on the expected estimates when the exact model is simulated, the power of model selection to discriminate among models depending on the quantity of the data, and the robustness of the

parameter estimates when additional sources of noise that are not explicitly accounted for in the likelihood function (macro-environmental variance and stochastic genetic drift) are considered.

(i) Accuracy

First, the quality of the estimates provided by the model was checked in the constant-variance case, supposing that model assumptions are true (i.e. no stochastic effects apart from sampling). In order to simulate a realistic situation, parameter values were chosen close to the estimates obtained from the example detailed in example section below: $\mu_0 = 0.04$, $\sigma_A^2 = 0.0001$, $\sigma_E^2 = 0.0004$. The Supplementary Results 1 section reports the bias and the precision of parameter estimates for various population sizes and various lengths of the time series. Estimates appeared to be satisfactorily unbiased. Even poor data sets (e.g. 20 individuals selected for five generations) provide meaningful estimates.

The flexibility of phenomenological models was also confirmed by fitting arbitrary variance dynamics (Supplementary Results 2). The power of model selection was assessed by simulating the selection response expected from two different genetic architectures ((i) constant variance and (ii) directional epistasis) and running various mechanistic models on these time series. AIC and BIC were calculated for each model, and the differences between models are reported (Supplementary Results 3). For large data sets, model selection is satisfactory (the correct model is chosen most of the time). From the simulation results, it appears that two factors are important regarding the accuracy of model selection: (i) the length of the selection experiment and (ii) the number of selected lines. When the time series is short (ten generations in this example), the simplest model (the constant-variance model) seems to be favoured, which is an expected result. One selected line alone does contain enough information to reject the constant-variance model when the real genetic architecture is complex (here, directional epistasis), but not enough to discriminate among the complex ones (AIC and BIC scores remain very close). This points out the importance of maximizing the number of selection lines in order to maximize the amount of information available. In contrast, the impact of population size appears to be smaller. Indeed, with two lines selected for 30 generations, the right model (directional epistasis) has the lowest AIC in 95% of the simulations even with a population size as small as $N = 20$.

(ii) Robustness

The simplest models (likelihood equation (18)) assume that the stochasticity of the time series can be

completely attributed to sampling effects. However, it is not clear whether the models based on infinite-size populations behave when fit to 'real' data when both sampling variance and other sources of stochasticity affect the measurements of the means and variances. The impact of two independent sources of stochasticity, the stochastic effect of genetic drift and macro-environmental variations, have thus been checked by simulation.

Some of the models presented above do account for some properties of genetic drift, such as the deterministic loss of genetic variation due to inbreeding in small populations (equation (7)). However, genetic drift will also generate some departure from the expected dynamics, which is accounted for by more complex random-effect models (equation (21)). Although the magnitude of this departure is expected to be smaller than e.g. sampling effects in a given generation, deviations due to genetic drift are heritable, and may then accumulate over generations according to a Markov process. Therefore, it is expected that genetic drift may affect parameter estimates. Supplementary Results 4 report the mean parameter estimates obtained when running the model on individual-based simulated data, with various population sizes. Fixed-effect estimates of additive and environmental variance are slightly biased: additive variance is underestimated in small populations, while environmental variance tends to be overestimated. Confidence intervals are also less reliable, being too narrow for σ_A^2 , and too wide for σ_E^2 . Nevertheless, apart from very small populations, drift-related effects on parameter estimates remains limited, and are unlikely to affect significantly the conclusions of the analysis. Supplementary Results 5 illustrate how considering random-effect models improves the estimated standard error, and Supplementary Results 6 focus on the impact of considering replicated selection experiments on parameter estimates.

Finally, we investigated the impact of macro-environmental effects by adding a random deviate to the mean phenotype each generation in the simulated data, drawn from a normal distribution with variance parameter σ_{me}^2 (Supplementary Results 7). It appears that macro-environmental variation affects the estimate of the environmental variance (and environmental variance related parameters, such as the k_c in equation (13)), which reflects the residual variance when fitting the model, but other parameters (additive variance, initial population mean) remain unaffected, even when $\sigma_{me}^2 > \sigma_E^2$. The statistical framework developed in this paper thus appears to be robust to macro-environmental effects. Computing the likelihood while accounting for macro-environmental effects efficiently removes the bias on σ_E^2 . However, the additional flexibility provided by macro-environmental effects may be misleading when fitting simple models

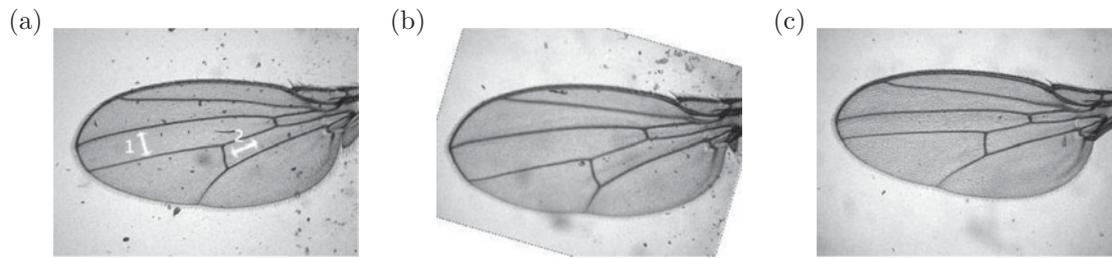


Fig. 1. *Drosophila melanogaster* wing pictures taken with the WINGMACHINE system. (a) Individual from the initial population; the two measures used to build the selection index are indicated: '1' is the the distance between veins III and IV and '2' is the relative position of the posterior cross-vein. (b) Individual wing from generation 29 of the 'up' selection line (selection for increasing the two indexes). (c) Individual wing from generation 29 of the 'down' line (selection for decreasing both indexes).

on complex data (Supplementary Results 8). To avoid interpretation issues, the example detailed below will thus be analysed without macro-environmental effects.

5. Example: artificial selection on wing shape in fruit flies

The properties of the statistical procedure will be illustrated by analyses of data from an artificial-selection experiment on morphology in *D. melanogaster*. From an initial stock, four populations of $N \simeq 100$ were selected in two directions (two 'up' and two 'down' lines). A picture of the wing of each individual fly was taken by the automatic system known as the WINGMACHINE (Houle *et al.*, 2003), and selection was performed on a composite, dimensionless index involving two relative distances (Fig. 1(a)), as described in Pélabon *et al.* (2006). Each generation, the most extreme 20 males and 20 females were bred randomly, and the selection procedure was carried on for 28 or 29 generations. At that point, the selected populations have accumulated easily recognizable wing-shape differences in the selected index (Fig. 1(b) and (c)), of a relative magnitude larger than what can be observed in the whole subgenus *Sophophora*.

The distribution of the selection index in males and females is virtually identical, and the sexes were merged for the analysis. All measured wings were considered except for three individuals from generation 20 of one of the 'up' line, which were clear outliers. A summarized data set reporting: (i) the phenotypic mean in the population, (ii) the phenotypic variance, (iii) the mean of the selected breeders and (iv) the population size before selection, was computed for model fitting.

The analysis was performed in three steps: (i) the best scale for the data was determined by comparing the fit of basic models on different scales, (ii) the properties of the time series and the potential explanatory power of simple models were assessed by fitting phenomenological models of increasing complexity, and (iii) mechanistic models were fit and the

corresponding biological parameters were estimated. For simplicity, the fixed-effect version of the model (infinite population size) was used in all cases.

The phenotypic measurement in this experiment is a complex variable. The selection index I is a combination of two traits (Pélabon *et al.*, 2006). I_1 is the average distance between veins III and IV normalized by the square root of the wing area. I_2 is the relative position of the posterior cross-vein, i.e. the average of two distance ratios. The total index is a weighted sum of these two traits, the weights being chosen in order to compensate for the fact that the phenotypic variance of I_1 was 2.6 times smaller than the variance of I_2 : $I = 2.6I_1 + I_2$. I is thus a dimensionless index; apart from the fact that it cannot be negative, its expected scaling properties are difficult to assess *a priori*. Table 1 reports the AIC of simple phenomenological models (exponential or linear changes in the variances) on four different scales (for the log scale, equation (16) was changed to a log-normal density function). For all models, the logarithmic scale outperformed other scales, and was thus retained for subsequent analysis. The main effect of the log scale here is to correct partially the asymmetry of selection responses, in a way that is very similar to the evolvability scale.

Fitting phenomenological models on the variance trends is a convenient way to introduce additional explanatory variables in the models without making specific assumptions about the underlying genetic mechanisms. Table 2 summarizes the gain in explanatory power when introducing an increasing number of variables, and the fit differences across models are illustrated in Supplementary Figures 1 to 3. This phenomenological analysis brings three main results: (i) The models fitting the data the most convincingly without over-parameterization have more than nine parameters. This illustrates the potential number of hidden variables in the system. Such variables might involve macro-environmental factors, measurement errors, genetic factors (segregation of large-effect alleles, complex epistatic patterns) or other biological mechanisms (e.g. maternal or epigenetic effects). (ii) The AIC obtained when fitting independently the

Table 1. AIC (expressed as the difference with the best model (AIC = 0)) for simple phenomenological models fitted on the whole data set (both ‘up’ and ‘down’ lines) on four different scales: original scale, logarithm scale, evolvability scale, and heritability scale. Log-transformed means and variances were computed from the original means and variances assuming log-normality on the original scale. The four different models correspond to subsets of equation (2): the constant variances model forces $k_{A_0} = 0$, $k_{A_1} = 1$, $k_{E_0} = 0$, $k_{E_1} = 1$. In the ‘exponential changes’ model, k_{A_1} and k_{E_1} can vary, in the linear change model, k_{A_0} and k_{E_0} can vary, and in the last model (linear and exponential change), all four parameters are active. The logarithmic scale appears to provide the best fit, whatever the model

	Original scale	Log scale	Evolvability scale	Heritability scale
Constant variances	4460.8	3855.7	3830.5	4460.8
Exp. changes	765.1	428.1	453.6	817.0
Lin. changes	1245.9	866.8	882.6	1146.7
Exp. and Lin.	240.6	0.0	37.2	300.2

four lines (44 parameters) is much better (by >1850 units) than the models considering all at the same time: any model predicting symmetric selection responses is thus necessarily approximate. Although less important, this very significant gap still exists when comparing two identically selected lines (the ‘up’ lines considered independently (22 parameters) perform 380 AIC units better than when pooled together (11 parameters), and the difference is 420 AIC units for the ‘down’ lines). (iii) The predicted trends systematically show unexpected patterns in variance dynamics: the additive variance trend is ‘V-shaped’, with an initial decrease and a late increase after 15–20 generations, which does not correspond to classical quantitative genetics expectations; possible explanations include, e.g. new mutations arising in both lines, a specific pattern of epistasis, or close linkage between two loci in repulsion (Hospital & Chevalet, 1996). In a similar way, environmental variance also increases at the end of the time series, especially in the ‘up’ lines, which also does not fit with common assumptions.

Fitting mechanistic models made it possible to estimate some parameters of direct biological interest. Table 3 reports maximum-likelihood estimates for six simple models. The initial additive variance is estimated to be between 8×10^{-5} and 1.4×10^{-4} depending on the model. The variance of the log-transformed data is approximately equal to the evolvability (as defined in equation (4)) on the original

scale ($\text{Var}(\log(x)) \simeq \text{Var}(x)/x^2 = I$), so that the evolvability of the selection index (on the original scale) is $I_{A_1} \simeq 0.01\%$: the index would evolve relatively slowly (0.01% change per generation) if selection on the wing index was as strong as selection on fitness (i.e. a change of 1% in the index represents 1% change in fitness). Consequently, this means that the genetic variation available on this wing-shape index was small, and that artificial-selection pressure had to be large in order to achieve a significant response (about 0.5%/generation). In other words, the evolvability of the wing shape is rather low, and rapid evolution is unlikely to occur unless the fitness function is particularly steep.

The estimate of the effective population size by the drift model ($\hat{N}_e = 9.36$, 95% confidence interval: 9.0–9.7) is compatible with the experimental procedure: 20 males and 20 females were mated, the ratio N_e/N being typically around 0.2 in short-term experiments, and 0.1 on a longer time scale (Frankham, 1995). The mutation model was forced to consider $N_e = 9.36$ (the estimate from the drift model alone) to provide more realistic estimates of the mutational variance: $\hat{\sigma}_M^2 = 2.0 \cdot 10^{-7}$ (confidence interval: 9.9×10^{-8} to 4.1×10^{-7}), which is less than 0.1% of the phenotypic variance. This estimate, low but not unrealistic (alternatively, mutational heritability $h_m^2 = \sigma_M^2/\sigma_E^2 \simeq 0.0007$), suggests that new mutations did not play an important part in the observed phenotypic change. The estimate of directional epistasis $\hat{\epsilon} = -0.37$ (CI 95%: -0.43 to -0.32) suggests small but significant negative directional epistasis (this number means that a mutation that increases the phenotype by 0.2 (about one phenotypic standard deviation) would on average reduce the effect of other genes by around 1%). Selection response tends to slow down over the course of the experiment, which might also be due to natural (stabilizing) selection. In this case, the joint-effect model predicts stabilizing selection ($\hat{s} = -264$ is equivalent to a pull-back force of $\simeq \beta/2$ when the population evolves by 0.1 phenotypic log unit). Important biological parameters can thus be estimated precisely through this procedure, some mechanisms can be ruled out while others are supported by model selection, and constitute interesting hypotheses to test with further experimental work.

The most convincing models are the ones in which environmental variance can evolve independently, as revealed by fitting six ‘canalization’ models (Table 4). The environmental variance increases in the course of selection, by about 17% when the mean changes by one (initial) phenotypic standard deviation (environmental decanalization: $\sigma_{E_t}^2 = \sigma_{E_0}^2 \exp(7.46 \cdot |\mu_t - 0.051|)$). Conversely, additive variance barely decreases when the population is selected away from its original state, by only 3% per phenotypic standard deviation change in the mean (genetic

Table 2. AIC score for phenomenological models of different complexity, expressed as the difference with the best model (AIC=0) for each data set. Seven data sets are considered, the two ‘up’ selected lines (independently and together, i.e. assuming an identical genetic architecture in both lines), the two ‘down’ selected lines (independently and together), and all four lines simultaneously. The first line corresponds to a model where the variance does not change, and is thus equivalent to the constant variance model of tables 1 and 3. The different models are described as in equation (2), i.e. the ‘lag 0’ model corresponds to a model in which only the constant parameters k_{A_0} and k_{E_0} are active, ‘lag 1’ to a model where both k_0 and k_1 parameters are estimated, etc. The simplest model (‘no change’) has three parameters, the most complex (‘lag 3’) has 11 parameters

	Up 1	Up 2	Both up	Down 1	Down 2	Both down	All
Cst	1307	755	1973	370	177	396	4191
lag 0	858	75	751	286	118	362	1203
lag 1	353	54	143	330	96	306	336
lag 2	10	23	7	151	22	10	70
lag 3	0	0	0	0	0	0	0

Table 3. Maximum-likelihood estimates for six quantitative genetics models (log-transformed data): constant variances (model (6)), genetic drift (model (7)), finite number of loci (model (11)), mutations (along with drift) (model (8)), directional epistasis (model (12)) and stabilizing natural selection on the focus trait (equation (14), with the optimum $\theta = \mu_1$). The mutation model and the finite number of locus model were fitted by fixing N_e to 9.36, the estimate from the drift model, because N_e could not be reliably estimated independently from mutational variance and the number of loci, respectively. Parameters that were fixed instead of being estimated are indicated by an asterisk. AIC values are compared with the constant variance model (reference model), the more negative the AIC, the better the fit. Variances are multiplied by 100 because this number can be directly interpreted in terms of percentage of evolvability (see text)

Model	$\hat{\mu}_1$	% $\hat{\sigma}_{A_1}^2$	% $\hat{\sigma}_{E_1}^2$	\hat{N}_e	\hat{n}_e	% $\hat{\sigma}_M^2$	$\hat{\epsilon}$	\hat{s}	ΔAIC
Const. var	0.042	0.008	0.037						0
Drift	0.042	0.014	0.028	9.36					-3065
Finite locus	0.042	0.014	0.028	9.36*	35.6				-3068
Mutation	0.042	0.014	0.028	9.36*		0.000020			-3074
Dir. epistasis	0.046	0.012	0.028	9.53			-0.37		-3188
Joint effect	0.043	0.013	0.029	7.14				-265	-3154

canalization: $\sigma_{A_t}^2 = \sigma_{A_1}^2 \exp(-1.55 \cdot |\mu_t - 0.051|)$, in addition to the $\approx 5\%$ decrease per generation due to inbreeding. The canalization optimum (the phenotype at which genetic variance is maximum and environmental variance is minimum) is $\hat{\theta} = 0.051$ in the largest model, close to the original mean of the population ($\hat{\mu}_1 = 0.043$): models including a free canalization optimum only slightly outperform those in which the optimum is the original mean. Figure 2 illustrates the fit of the worst (constant variances) and the best (environmental canalization with a free optimum) models. Although the canalization model convincingly describes the trends in genetic and environmental variances, the fit remains far from the best phenomenological model. In particular, phenomenological models predict a late raise in additive variance (Supplementary figures 1 to 3), which cannot be explained with simple quantitative genetics models.

Fitting both phenomenological and mechanistic models on the fly-wing selection data was thus

successful in quantifying the genetic architecture of this complex trait in an experimental population. By comparing the fit of classical quantitative-genetics models on different scales, we were able to identify the log scale as the most relevant for subsequent analysis. Phenomenological models revealed that the dynamics of the selection response is complex, and needs more parameters than are usually considered in quantitative genetics to be fully explained. There are strong indications that the change in both environmental and additive variances is non-monotonic, with a late increase after 15 generations in both selected lines. Mechanistic models showed that additive variance, if it were constant, would be around 8×10^{-5} in log units ($I_A = 0.008\%$ on the original scale). Nevertheless, more flexible models letting the variance decrease estimate a higher initial additive variance, around 1.5×10^{-4} ($I_A = 0.015\%$ on the original scale). This figure is rather low when scaled with the mean phenotype, and the trait is not very evolvable; selection

Table 4. Canalization and decanalization. Six genetic and environmental canalization models were fit independently and in combination, considering either a canalization optimum at the initial phenotypic mean (three first models) or independent (three last models). The effect of genetic drift is almost confounded with genetic canalization, and the effective population size was fixed in a similar way as explained in Table 3. The AIC score is displayed as the difference from the constant-variance model

Model	$\hat{\mu}_1$	$\% \hat{\sigma}_{A_1}^2$	$\% \hat{\sigma}_{E_1}^2$	\hat{N}_e	\hat{k}_g	\hat{k}_c	$\hat{\theta}$	ΔAIC
G. Canal	0.042	0.014	0.028	9.36*	-0.11			-3066
E. Canal	0.043	0.015	0.014	8.12		7.07		-3296
G.E. Canal	0.043	0.016	0.013	9.36*	-1.62	7.40		-3310
G. Canal. opt	0.046	0.014	0.027	9.36*	-0.85		-0.111	-3260
E. Canal. opt	0.044	0.015	0.017	8.05		6.21	0.065	-3348
G.E. Canal. opt	0.042	0.016	0.014	9.36*	-1.55	7.47	0.051	-3322

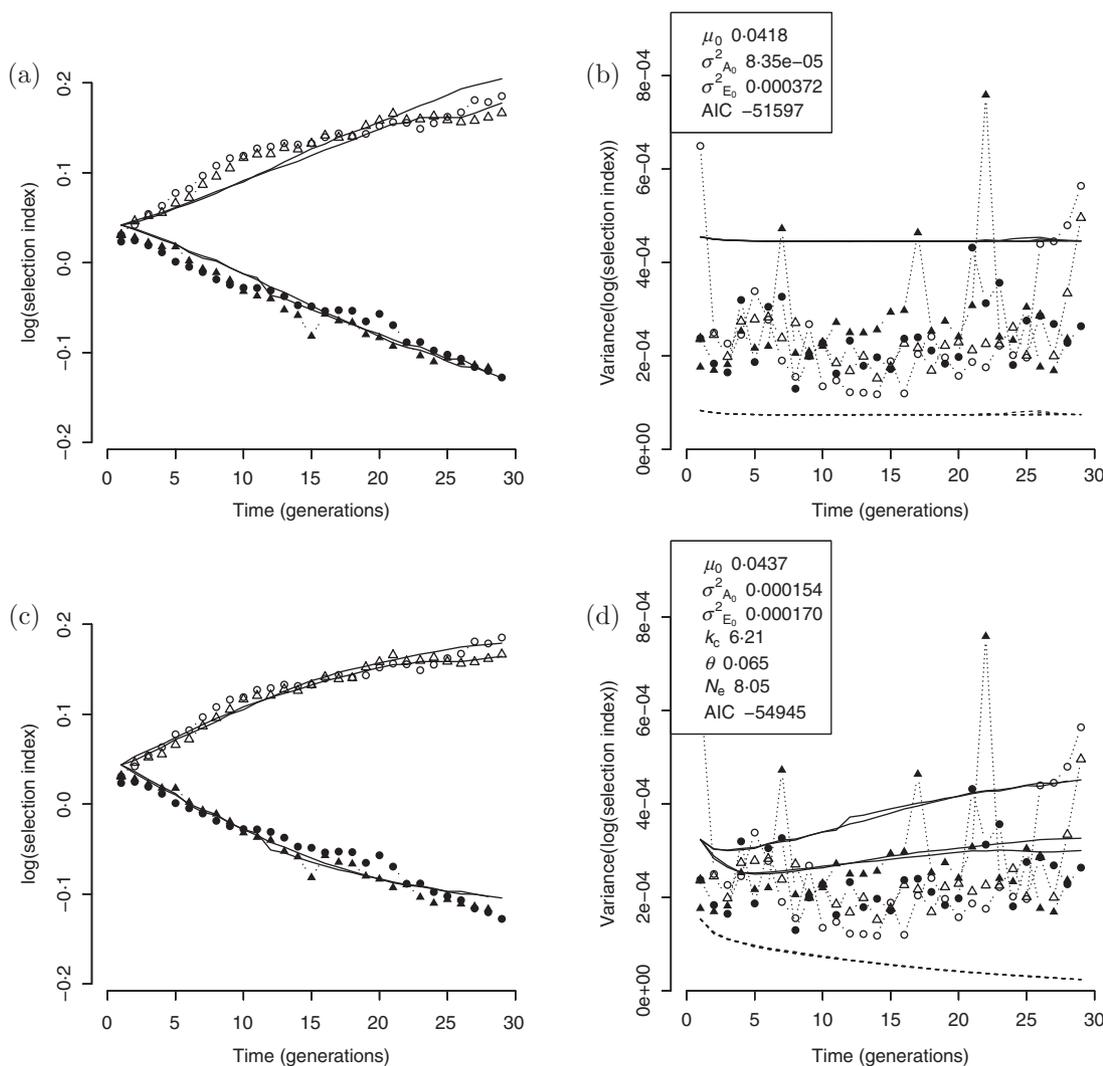


Fig. 2. Illustration of mechanistic model fitting on the *Drosophila* experiment (four lines). Symbols represent the data points (phenotypic mean on panels (a) and (c), phenotypic variance on panels (b) and (d)). Solid lines are model expectations for phenotypic means or variances, dashed lines are the expected additive variances. (a) and (b): Constant-variance model; (c) and (d): Environmental canalization with a free canalization optimum (one of the best mechanistic models). Predicted means and variances are not strictly identical across selected lines, because of slight differences in the realized selection gradients. In the decanalization model, the ‘up’ lines are closer to the canalization optimum, and so their predicted environmental variance is smaller than in the ‘down’ lines. The late raise in phenotypic variance in both ‘up’ lines can hardly be explained by genetic mechanisms, and affects the variance estimates for the whole time series, especially in the constant-variance model.

pressure had to be strong for obtaining the observed changes. Genetic drift can be partly responsible for the early decrease in variance, with an effective population size estimated between seven and ten depending on the model. Mutational variance could not be reliably estimated, but new mutations do not appear to play a major role in this experimental selection. Interestingly, the best model predicts environmental decanalization (environmental variance increases with deviance from an optimum), and changes in the genetic architecture seem to be more related to the evolutionary distance rather than the elapsed time.

6. Discussion

(i) *Scaling*

The models and the likelihood functions described above assume that phenotypes are normally distributed, and departure from normality may affect parameter estimates and model selection. Most models in quantitative genetics rely on the assumption that effects tend to combine additively, and that both genetic and environmental effects are normally distributed. There are solid reasons, both experimental and theoretical, to question this hypothesis. Most measured traits cannot be negative (such as weight, size, yield, fertility, etc), so that the normal distribution is at best an approximation of the real phenotypic distribution. Gaussian distributions seem to be empirically supported, although often outperformed by log-normal (e.g. Powers, 1950; Gingerich, 2000). Knowledge about molecular, cellular and physiological mechanisms does not suggest that genetic factors should combine additively, since the expression of complex phenotypes relies on multiple gene networks featuring multiple loops of regulation, and the overall process is expected to generate statistical interactions (Omholt *et al.*, 2000; Gjuvsland *et al.*, 2007).

According to the measurement-theory concepts (Hand, 2004; Hansen & Houle, 2008; Wagner, 2010; Houle *et al.*, 2011), many traits are on a ratio scale. They may result from the multiplication (rather than the addition) of multiple small factors, so that the resulting phenotypic distribution is not expected to be Gaussian. Since most models are not supposed to be applied to multiplicative traits, the easiest way to analyse them is to consider the logarithm of the measurements. Log transformation is common, but not universal.

Scaling issues are thus not only related to statistical problems, but they do affect biological conclusions. Directional epistasis or dominance on a linear scale might vanish on a different scale (Pavlicev *et al.*, 2010), or, on the contrary, apparent additivity might hide interesting physiological interactions. Since the literature is not particularly consistent in the use of the log

transformation on ratio-scale characters, general statements dealing with the symmetry or the linearity of selection responses are thus doubtful.

An ideal procedure would be to determine the scale on which the data are considered *a priori*, based on measurement theoretical considerations and the nature of the measured trait. A more pragmatic approach consists in estimating the best scale from the data itself. From the statistical framework previously defined, it is perfectly possible to fit the models on several scales. Model selection can be performed to compare them, and the procedure may constitute additional evidence towards the use of a particular scale.

(ii) *Evolvability and selection*

In the context of an artificial-selection experiment, in which the identity (and the phenotype) of the breeders is known, the between-generation expected change in the mean can be written $R = \sigma_A^2(\mu_t^* - \mu_t) / \sigma_P^2$, where R is the selection response $\mu_{t+1} - \mu_t$ and $\mu_t^* - \mu_t$ the selection differential (difference between selected individuals and the phenotypic mean of the population). We have chosen to write this equation as proposed by Lande & Arnold (1983), as $R = \sigma_A^2 \beta$, calculating the selection gradient $\beta = (\mu_t^* - \mu_t) / \sigma_P^2$.

Another common way to write this equation is the 'breeder's equation', $R = h^2(\mu_t^* - \mu_t)$, in which $h^2 = \sigma_A^2 / \sigma_P^2$ is the narrow-sense heritability. This form is often used in experimental genetics, as the selection differential is known from the selection procedure, h^2 describing the (supposedly constant) capacity for the population to respond to selection. However, it has been argued that this parameterization is confusing, or even misleading (Houle, 1992; Hansen *et al.*, 2003b; Wilson, 2008; Hansen & Houle, 2008). The additive variance σ_A^2 is both in the numerator and in the denominator of h^2 , and also influences the selection differential, so that changes in h^2 are harder to detect and interpret than changes in σ_A^2 . Understanding the dynamics of complex genetic architectures necessarily implies a clear separation between variation (σ_A^2) and selection (β), which is the case when using the formulation of Lande & Arnold (1983). This framework can easily be extended to multiple characters, replacing R and β by vectors, and σ_A^2 by the additive genetic variance-covariance \mathbf{G} matrix. Additionally, since β can also be defined as the regression coefficient of the fitness over the phenotype, the parameters of the equation remain meaningful when selection is not artificial, e.g. when there are fitness differences among breeders.

(iii) *Conclusion*

In this paper, we have proposed alternatives to the classical analysis of selection-response time series.

Rather than assuming a constant genetic architecture, we propose to fit phenomenological models designed to describe the changes in genetic and environmental variances (the dynamics of the phenotypic mean being constrained by the additive variance and the selection gradient). In the selection experiment on fly wings used as an example, this phenomenological approach pointed out an unexpected pattern (late increase in additive variance in both lines), which could have easily remained unnoticed otherwise. Temporal trends, which are direct observations from the data, can then be interpreted in the light of more or less elaborated mechanistic models, and parameters of biological interest can be estimated.

The estimates obtained with the deterministic framework remain satisfactory, even when sources of stochasticity that were not accounted for explicitly are simulated. Nevertheless, in extreme cases, imprecise confidence intervals might affect the accuracy of the results. It is then possible to account for these different sources of randomness through random-effect models. By considering that genetic means and genetic variances are themselves random variables with known distributions, depending on the effective population size and their status at the previous generation, models in which the cumulative effect of genetic drift can be implemented to estimate the genetic architecture parameters through a more realistic population genetics framework. The resulting statistical tools are, however, more complex. Because models are non-linear, random effects have to be integrated out through specialized algorithms, numerical convergence can be problematic and takes more time and more computational power. Additionally, random effects may have side effects raising interpretation issues. A detailed description of such models, their benefits and their limitations is provided in Le Rouzic *et al.* (2010) and in supplementary results. Whether or not random-effect models have to be preferred depends on multiple factors, including the experimental setting itself and the precision of the estimates that is desired.

Meticulous analysis of artificial-selection time series is likely to deepen our understanding of how selection affects the phenotypic changes in a population, of the impact of genetic architecture differences, and of the importance of measurement issues, including scaling. The two approaches (phenomenological and mechanistic) described in this paper are complementary, and may highlight how far common quantitative genetic models actually are from experimental results. Overall, the definition of a solid-statistical framework for selection time-series analysis enables the extraction of meaningful parameters from existing data sets, and highlights the benefits and the limitations of this experimental approach. Perhaps one of the most crucial points to investigate is the rate

at which the additive variance changes in a population and the consistency of this phenomenon. Indeed, the constancy (or at least the inertia) of the additive variance–covariance matrix is a key assumption when trying to bridge what has often been considered as a gap between intraspecific and interspecific evolution (Arnold *et al.*, 2001; Gingerich, 2001; McGuigan, 2006; Blows, 2007; Hohenlohe & Arnold, 2008; Arnold *et al.*, 2008). By empirically simulating large adaptive shifts, artificial-selection experiments provide valuable information to understand how genetic architectures behave when submitted to such directional selection pressures.

We thank H.J. Skaug, T. Reitan and T. Rouyer for helpful discussion, and A.J.R. Carter for helping assembling the data. The authors acknowledge W. G. Hill, B. Walsh, G. J. Geyer and two anonymous reviewers for their comments and suggestions on an earlier version of the manuscript. T.F.H. was supported by grant no 177857 from the Research Council of Norway and NSF grant DEB-0344417 to T.F.H. and D.H. A.L.R. was funded by the Marie Curie fellowships EIF-220558 and ERG-256507. D.H. was supported in this research by a visiting professorship at the Centre for Ecological and Evolutionary Synthesis at the University of Oslo, Norway.

7. Supplementary material

The online data can be found available at <http://journals.cambridge.org/GRH>

References

- Akaike, H. (1973). *Proceedings of the International Symposium on Information theory*, Information theory as an extension of the maximum likelihood principle, pp 267–281. Akademiai Kiado, Budapest.
- Álvarez-Castro, J. M., Kopp, M. & Hermisson, J. (2009). Effects of epistasis and the evolution of genetic architecture: exact results for a 2-locus model. *Theoretical Population Biology* **75**, 109–122.
- Anderson, D., Burnham, K. & Thompson, W. (2000). Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management* **64**, 912–923.
- Arnold, S. J., Burger, R., Hohenlohe, P. A., Ajie, B. C. & Jones, A. G. (2008). Understanding the evolution and stability of the g-matrix. *Evolution* **62**, 2451–2461.
- Arnold, S. J., Pfrender, M. E. & Jones, A. G. (2001). The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112–113**, 9–32.
- Barton, N. H. & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics* **3**, 11–21.
- Blows, M. W. (2007). A tale of two matrices: multivariate approaches in evolutionary biology. *Journal of Evolutionary Biology* **20**, 1–8.
- Bulmer, M. G. (1971). The effect of selection on genetic variability. *The American Naturalist* **105**, 201–211.
- Bulmer, M. G. (1980). *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.
- Bürger, R. & Ewens, W. J. (1995). Fixation probabilities of additive alleles in diploid populations. *Journal of Mathematical Biology* **33**, 557–575.

- Burnham, K. & Anderson, D. (2002). *Model Selection and Multi-Model Inference*. Springer-Verlag, New York, NY.
- Carlborg, Ö. & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* **5**, 618–625.
- Carter, A. J. R., Hermisson, J. & Hansen, T. F. (2005). The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology* **68**, 179–196.
- Chevalet, M. (1994). An approximate theory of selection assuming a finite number of quantitative trait loci. *Genetics Selection Evolution* **26**, 379–400.
- Cockerham, C. C. (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**, 859–882.
- Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D. & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biology* **2**, E109.
- Dempfle, L. (1975). A note on increasing the limit of selection through selection within families. *Genetics Research* **24**, 127–135.
- Eisen, E. J. (1972). Long-term selection response for 12-day litter weight in mice. *Genetics* **72**, 129–142.
- Engen, S., Lande, R. & Sæther, B. E. (2009). Fixation probability of beneficial mutations in a fluctuating population. *Genetics Research* **91**, 73–82.
- Falconer, D. S. (1992). Early selection experiments. *Annual Review of Genetics* **26**, 1–14.
- Frankham, R. (1995). Effective population-size/adult population-size ratios in wildlife – a review. *Genetical Research* **66**, 95–107.
- Gianola, D. & Fernando, R. (1986). Bayesian methods in animal breeding theory. *Journal of Animal Science* **63**, 217–244.
- Gingerich, P. D. (2000). Arithmetic or geometric normality of biological variation: an empirical test of theory. *Journal of Theoretical Biology* **204**, 201–221.
- Gingerich, P. D. (2001). Rates of evolution on the time scale of the evolutionary process. *Genetica* **112–113**, 127–144.
- Gjuvslund, A. B., Hayes, B. J., Omholt, S. W. & Carlborg, O. (2007). Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* **175**, 411–420.
- Gudbjartsson, D. F., Walters, G. B., Thorleifsson, G., Stefansson, H., Halldorsson, B. V., Zusmanovich, P., Sulem, P., Thorlacius, S., Gylfason, A., Steinberg, S., Helgadóttir, A., Ingason, A., Steinthorsdóttir, V., Olafsdóttir, E. J., Olafsdóttir, G. H., Jonsson, T., Borch-Johnsen, K., Hansen, T., Andersen, G., Jorgensen, T., Pedersen, O., Aben, K. K., Witjes, J. A., Swinkels, D. W., den Heijer, M., Franke, B., Verbeek, A. L., Becker, D. M., Yanek, L. R., Becker, L. C., Tryggvadóttir, L., Rafnar, T., Gulcher, J., Kiemeneý, L. A., Kong, A., Thorsteinsdóttir, U. & Stefansson, K. (2008). Many sequence variants affecting diversity of adult human height. *Nature Genetics* **40**, 609–615.
- Hallander, J. & Waldmann, P. (2007). The effect of non-additive genetic interactions on selection in multi-locus genetic models. *Heredity* **98**, 349–359.
- Hand, D. (2004). *Measurement Theory and Practice: The World Through Quantification*. Hodder Arnold, London.
- Hansen, T. F. (2003). Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Biosystems* **69**, 83–94.
- Hansen, T. F. (2006). The evolution of genetic architecture. *Annual Review of Ecology Evolution and Systematics* **37**, 123–157.
- Hansen, T. F., Alvarez-Castro, J. M., Carter, A. J. R., Hermisson, J. & Wagner, G. P. (2006). Evolution of genetic architecture under directional selection. *Evolution* **60**, 1523–1536.
- Hansen, T. F., Armbruster, W. S., Carlson, M. L. & Pélabon, C. (2003a). Evolvability and genetic constraint in *Dalechampia* blossoms: genetic correlations and conditional evolvability. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **296**, 23–39.
- Hansen, T. F. & Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* **21**, 1201–1219.
- Hansen, T. F., Pélabon, C., Armbruster, W. S. & Carlson, M. L. (2003b). Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. *Journal of Evolutionary Biology* **16**, 754–766.
- Hansen, T. F. & Wagner, G. P. (2001). Modeling genetic architecture: A multilinear theory of gene interaction. *Theoretical Population Biology* **59**, 61–86.
- Hermisson, J., Hansen, T. F. & Wagner, G. P. (2003). Epistasis in polygenic traits and the evolution of genetic architecture under stabilizing selection. *The American Naturalist* **161**, 708–734.
- Hermisson, J. & Pennings, P. (2005). Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335–2352.
- Hill, W. G. (1982a). Predictions of response to artificial selection from new mutations. *Genetics Research* **40**, 255–278.
- Hill, W. G. (1982b). Rates of change in quantitative traits from fixation of new mutations. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 142–145.
- Hill, W. G. & Caballero, A. (1992). Artificial selection experiments. *Annual Review of Ecology and Systematics* **23**, 287–310.
- Hohenlohe, P. A. & Arnold, S. J. (2008). MIPoD: a hypothesis-testing framework for microevolutionary inference from patterns of divergence. *The American Naturalist* **171**, 366–385.
- Holt, M., Meuwissen, T. & Vangen, O. (2005). Long-term responses, changes in genetic variances and inbreeding depression from 122 generations of selection on increased litter size in mice. *Journal of Animal Breeding and Genetics* **122**, 199–209.
- Hospital, F. & Chevalet, C. (1996). Interactions of selection, linkage and drift in the dynamics of polygenic characters. *Genetics Research* **67**, 77–87.
- Houle, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics* **130**, 195–204.
- Houle, D. (1998). How should we explain variation in the genetic variance of traits? *Genetica* **102–103**, 241–253.
- Houle, D., Mezey, J., Galpern, P. & Carter, A. (2003). Automated measurement of *Drosophila* wings. *BMC Evolutionary Biology* **3**, 25.
- Houle, D., Morikawa, B. & Lynch, M. (1996). Comparing mutational variabilities. *Genetics* **143**, 1467–1483.
- Houle, D., Pélabon, C., Wagner, G. & Hansen, T. F. (2011). Measurement and meaning in biology. *The Quarterly Review of Biology* **86**, 1–32.
- Keightley, P. D. (1998). Genetic basis of response to 50 generations of selection on body weight in inbred mice. *Genetics* **148**, 1931–1939.

- Keightley, P. D. (2004). Mutational variation and long-term selection response. *Plant Breeding Reviews* **24**, 227–247.
- Keightley, P. D. & Hill, W. G. (1987). Directional selection and variation in finite populations. *Genetics* **117**, 573–582.
- Kempthorne, O. (1954). The correlation between relatives in a random mating population. *Proceedings of the Royal Society of London, Series B* **143**, 102–113.
- Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genetics Research* **26**, 221–235.
- Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402–416.
- Lande, R. (1983). The response to selection on major and minor mutations affecting a metrical trait. *Heredity* **50**, 47–65.
- Lande, R. & Arnold, S. (1983). The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226.
- Le Rouzic, A. & Alvarez-Castro, J. M. (2008). Estimation of genetic effects and genotype-phenotype maps. *Evolutionary Bioinformatics* **4**, 225–235.
- Le Rouzic, A., Alvarez-Castro, J. M. & Carlborg, O. (2008). Dissection of the genetic architecture of body weight in chicken reveals the impact of epistasis on domestication traits. *Genetics* **179**, 1591–1599.
- Le Rouzic, A., Siegel, P. B. & Carlborg, O. (2007). Phenotypic evolution from genetic polymorphisms in a radial network architecture. *BMC Biology* **5**, 50.
- Le Rouzic, A., Skaug, H. J. & Hansen, T. F. (2010). Estimating genetic architectures from artificial-selection responses: A random-effect framework. *Theoretical Population Biology* **77**, 119–130.
- Lynch, M. (1988). The rate of polygenic mutation. *Genetics Research* **51**, 137–148.
- Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. & Willis, J. (1999). Spontaneous deleterious mutation. *Evolution* **53**, 645–663.
- Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc., Sunderland, MA.
- Mackay, T. F. (2004). The genetic architecture of quantitative traits: lessons from *Drosophila*. *Current Opinion in Genetics and Development* **14**, 253–257.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.
- Martinez, V., Bunger, L. & Hill, W. G. (2000). Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. *Genetics Selection Evolution* **32**, 3–21.
- McGuigan, K. (2006). Studying phenotypic evolution using multivariate quantitative genetics. *Molecular Ecology* **15**, 883–896.
- Meyer, K. & Hill, W. G. (1991). Mixed model analysis of a selection experiment for food intake in mice. *Genetics Research* **57**, 71–81.
- O'Hara, R. B., Cano, J. M., Ovaskainen, O., Teplitsky, C. & Alho, J. S. (2008). Bayesian approaches in evolutionary quantitative genetics. *Journal of Evolutionary Biology* **21**, 949–957.
- Omholt, S., Plahte, E., Oyehaug, L. & Xiang, K. (2000). Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. *Genetics* **155**, 969–980.
- Otto, S. P. & Whitlock, M. C. (1997). The probability of fixation in populations of changing size. *Genetics* **146**, 723–733.
- Pavlicev, M., Le Rouzic, A., Cheverud, J. M., Wagner, G. P. & Hansen, T. F. (2010). Directionality and scale of epistasis a murine intercross population. *Genetics* **185**, 1489–1505.
- Pélabon, C., Hansen, T. F., Carter, A. J. & Houle, D. (2006). Response of fluctuating and directional asymmetry to selection on wing shape in *Drosophila melanogaster*. *Journal of Evolutionary Biology* **19**, 764–776.
- Phillips, P. (1998). The language of gene interaction. *Genetics* **149**, 1167–1171.
- Pong-Wong, R., Haley, C. S. & Woolliams, J. A. (1999). Behaviour of the additive finite locus model. *Genetics Selection Evolution* **31**, 193–211.
- Powers, L. (1950). Determining scales and the use of transformations in studies on weight perlocule of tomato fruit. *Biometrics* **6**, 145–163.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Robertson, A. (1960). A theory of limits in artificial selection. *Proceedings of the Royal Society of London, B* **153**, 234–249.
- Skaug, H. J. & Fournier, D. (2006). Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics and Data Analysis* **56**, 699–709.
- Sorensen, D., Fernando, R. & Gianola, D. (2001). Inferring the trajectory of genetic variance in the course of artificial selection. *Genetics Research* **77**, 83–94.
- Turelli, M. (1988). Population genetics models for polygenic variation and evolution. In *Proceedings of the Second International Conference on Quantitative Genetics*, Raleigh, North Carolina (ed B. S. Weir *et al.*). Sinauer Associates, Sunderland, MA.
- Wagner, G. P. (2010). The measurement theory of fitness. *Evolution* **64**, 1358–1376.
- Weber, K. E. & Diggins, L. T. (1990). Increased selection response in larger populations. II. Selection for ethanol vapor resistance in *Drosophila melanogaster* at two population sizes. *Genetics* **125**, 585–597.
- Wilson, A. J. (2008). Why h^2 does not always equal V_A/V_P ? *Journal of Evolutionary Biology* **21**, 647–650.
- Zeng, Z. B., Tachida, H. & Cockerham, C. C. (1989). Effects of mutation on selection limits in finite populations with multiple alleles. *Genetics* **122**, 977–984.
- Zhang, X. S. & Hill, W. G. (2002). Joint effects of pleiotropic selection and stabilizing selection on the maintenance of quantitative genetic variation at mutation-selection balance. *Genetics* **162**, 459–471.
- Zhang, X. S. & Hill, W. G. (2005). Evolution of the environmental component of the phenotypic variance: stabilizing selection in changing environments and the cost of homogeneity. *Evolution* **59**, 1237–1244.