doi:10.1111/evo.13907

Is evolution predictable? Quantitative genetics under complex genotype-phenotype maps

Lisandro Milocco[1,2] and **Isaac Salazar-Ciudad**[1,3,4,5]

[1] *Institute of Biotechnology, University of Helsinki 00014, Helsinki, Finland*

[2] *E-mail: lisandro.milocco@helsinki.fi*

[3] *Centre de Recerca Matemàtica 08193, Barcelona, Spain*

[4] *Genomics, Bioinformatics and Evolution. Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona 08193, Barcelona, Spain*

[5] *E-mail: isaac.salazar@helsinki.fi*

Received May 16, 2019
Accepted November 27, 2019

A fundamental aim of post-genomic 21st century biology is to understand the genotype–phenotype map (GPM) or how specific genetic variation relates to specific phenotypic variation. Quantitative genetics approximates such maps using linear models, and has developed methods to predict the response to selection in a population. The other major field of research concerned with the GPM, developmental evolutionary biology, or evo-devo, has found the GPM to be highly nonlinear and complex. Here, we quantify how the predictions of quantitative genetics are affected by a complex, nonlinear map based on the development of a multicellular organ. We compared the predicted change in mean phenotype for a single generation using the multivariate breeder's equation, with the change observed from the model of development. We found that there are frequent disagreements between predicted and observed responses to selection due to the nonlinear nature of the genotype–phenotype map. Our results are a step toward integrating the fields studying the GPM.

KEY WORDS: evo-devo, genotype–phenotype map, G-matrix, mathematical modeling, quantitative genetics.

A fundamental aim of post-genomic 21st century biology is to understand how genomic variation relates to specific phenotypic variation. This relationship, called the genotype–phenotype map or GPM, is considered by many researchers to be critical factor for understanding phenotypic evolution (Houle et al. 2010). The GPM determines which phenotypic variation arises from which random genetic variation. Natural selection then acts on that realized phenotypic variation. Thus, natural selection and the GPM jointly influence how traits change over evolutionary time (Alberch 1982; Müller 2007; Polly 2008).

Quantitative genetics uses a statistical approach to describe the GPM and predict how the phenotype changes by natural or artificial selection. This approach has long made significant contributions to plant and animal breeding (Falconer and Mackay 1996; Roff 2007). A central equation of quantitative genetics, and to our understanding of the evolution of natural populations, is the breeder's equation: $R = h^2 s$ (Falconer and Mackay 1996). According to this equation, the change in the mean of a trait between two generations, R, is equal to the selection differential, s, multiplied by the heritability of the trait, h^2.

Natural selection can act on several traits at the same time, that is, an individual's fitness typically depends on multiple of its traits. In addition, there may be co-variation between traits. In this general case, it is more appropriate to use the multivariate extension of the breeder's equation:

$$\Delta \bar{z} = G P^{-1} s \qquad (1)$$

where $\Delta\bar{z}$ is the between-generation change of trait means, G is the matrix of additive genetic variances and covariances between traits, P is the matrix of phenotypic variances and covariances between traits, and s is the selection differential (Lande 1979; Lande and Arnold 1983). The multivariate breeder's equation can be used to infer past selection and predict future responses to selection (Lande 1979; Lande and Arnold 1983; Grant and Grant 1993; Morrissey et al. 2010). There is a large body of literature related to G and its evolutionary implications (Jones et al. 2004; Roff 2007; Arnold et al. 2008; Aguirre et al. 2013) and for some researchers the study of G is central to understand phenotypic evolution (McGuigan 2006; Roff 2007).

Quantitative genetics theory relies on a number of simplifying assumptions. The most important of these for our purposes is the assumption that only the additive contributions of genes to a phenotype play a significant role in evolution. Under this assumption, the variation in a trait is determined, to a large extent, by a large number of polymorphic loci. Alleles in each loci add a small effect to the trait and such effects are independent from each other. This is equivalent to assuming a linear GPM, that is, small genetic changes always lead to proportionally small phenotypic changes.

Developmental evolutionary biology or evo–devo (Alberch 1982; Müller 2007) is the other main field concerned with the GPM. Evo-devo views the GPM as highly nonlinear and complex (Alberch 1982; Newman and Müller 2000; Müller 2007; Gjuvsland et al. 2013). Most evo-devo studies, however, do not consider the population level. At that level, it has been suggested that the nonlinearities of the GPM average out because of the recombination of alleles occurring among individuals in sexual populations (Hansen 2008). According to these authors then, the nonlinearities of the GPM would not hinder the accuracy of quantitative genetics, at least in the short-term (Hansen 2008; Hill et al. 2008).

Some previous studies claim that the breeder's equation may be inaccurate if the GPM is not linear (Alberch 1982; Rice 2004). These studies, however, provide no direct information on how nonlinear real GPMs are. Studies in gene regulation, development, and physiology, on the other hand, provide direct functional information about real GPMs. These studies, however, are not usually framed in a population context that would allow the quantification of the accuracy of breeder's equation for these GPMs.

Our aim is to quantify how the predictions of the multivariate breeder's equation are affected when considering the complex and nonlinear GPMs found in the study of development. For that purpose, we combine a computational GPM model that is based on our current understanding of the development of a complex organ, the mammalian tooth (Salazar-Ciudad and Jernvall 2010; Salazar-Ciudad and Marín-Riera 2013), and a population genetics model with mutation, recombination, and selection (see Fig. 1). Through the combination of the development and population models, we simulate phenotypic evolution.

The tooth developmental model has been shown to be able to reproduce multivariate morphological variation at the population level (Salazar-Ciudad and Jernvall 2010) and has been used to model micro-evolution (Salazar-Ciudad and Marín-Riera 2013). The tooth developmental model starts from a small flat epithelium expressing some specific extracellular diffusible gene products. Over the development simulation time, these gene products regulate the expression of other gene products and cell behaviors such as cell division and adhesion. As a result, the epithelium starts to fold and change its morphology. These morphological changes affect the spatial distribution of the extracellular diffusible gene products and triggers further spatial changes in gene expression, cell behavior regulation and, ultimately, morphology. As a result a complex 3D morphology (i.e., a specific distribution of cells in space) is produced.

The developmental model includes a set of "developmental parameters." These specify how strong or weak the interactions between the gene products included in the model are, their regulation of cell behaviors and the strength of bio-mechanical interactions between cells during the simulation of a specific morphology. There are 21 of these developmental parameters and they specify, for example, the diffusion rates of the extracellular signaling molecules involved in the model, how strongly they regulate their own synthesis by the cells that receive them or the adhesion strength between cells (see Materials and Methods section and Salazar-Ciudad and Jernvall 2010 for details). The values of these parameters are genetically determined and genetic variation has an effect on phenotypic variation because it affects these values. However, there is no simple correspondence between these developmental parameters and genes. The diffusion rate of an extracellular gene product, for example, depends, among other things, on its size, shape, and hydration. Shape or 3D structure is determined by the folding of the primary sequence of a gene product. Such folding process is quite complex and is well-known to entail a complex GPM on its own (Stadler et al. 2001; Greenbury et al. 2014; Rodrigues et al. 2016).

Most mathematical models of development are based on similar interactions between gene products and cell behaviors (Oster and Alberch 1982; Raspopovic et al. 2014; Osterfield et al. 2017; Hirashima et al. 2017; Glen et al. 2019). These models give rise to different morphologies because they start from different initial conditions and include different networks of interactions between genes. Based on these models and on what is known about organ development, it has been suggested that the dynamics of interaction between gene products, cell behaviors, and morphology is, overall, of a similar complexity in most organs and body parts and that, thus, the overall complexity and properties of their GPM should be similar (Alberch 1982; Salazar-Ciudad

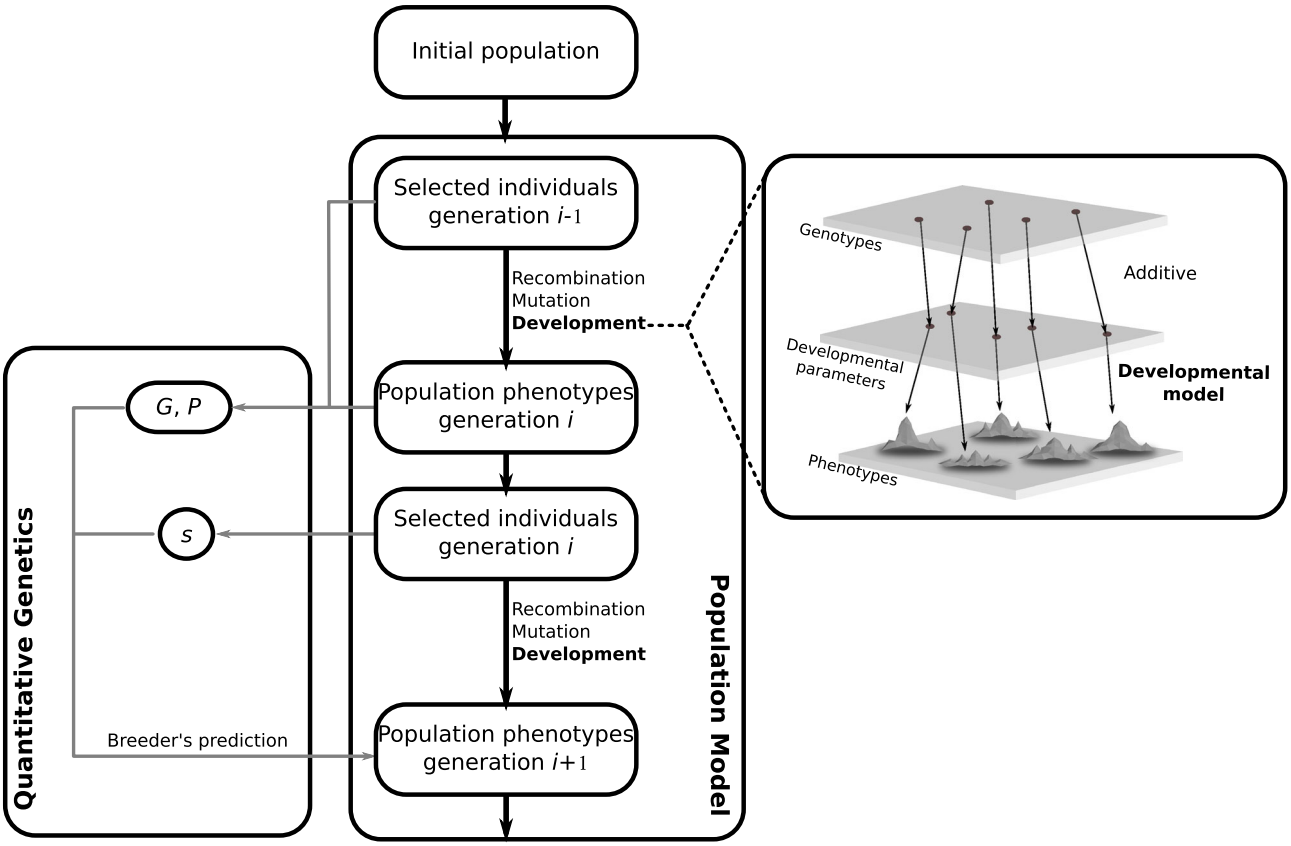

Figure 1. Our model combines a development model, leading from genotype to phenotype, and a population model, leading from phenotype to fitness and from fitness to next generation. The developmental model is depicted in the right. The developmental parameters of each individual are determined additively from a set of loci in each individual's genotype. Then the developmental model is run for the developmental parameter set of values in each individual. As a result a phenotype, a 3D morphology, is produced for each individual. For each individual phenotype a set of quantitative traits are measured. The distance between those and an optimal phenotype determine the fitness of each individual (middle panel in the figure). The 50% of males and females with highest fitness are selected as parents. Random mating then occurs among the selected individuals and mutation is applied to produce their offspring genotypes for the next generation. By iterating these processes of development, selection, recombination, and mutation in each generation, phenotypic evolution is simulated. In addition, shown in the right panel, the *G* and *P* matrices and the *s* vector are estimated for each generation (generation *i* in the figure). Estimations are done with data from a parallel half-sib breeding design in each generation. From these matrices, using the multivariate breeder's equations, we estimate the trait mean values for the next generation (generation *i*+1) and compare them with those observed from the evolutionary-developmental model.

et al. 2003; Salazar-Ciudad 2010; Urdy 2012). Thus, although our developmental model is based on tooth development, we should expect its GPM to be representative of a large class of phenotypes. This is specially the case for morphologies that form, like teeth, by the folding of epithelia. This applies to a large number of organs and body parts, including limbs, heads, kidneys, genitalia, lungs, insect wings, and early brain (Gilbert and Barresi 2016).

The population model includes a set of individuals. Each individual has a genotype, a set of values of the developmental parameters, a phenotype, and a fitness. The values of the developmental parameters of an individual are determined by its genotype. Each developmental parameter is determined additively by many loci. The phenotype of each individual is determined by running the developmental model on the values of its developmental parameters (see Fig. 1). Each individual phenotype is a 3D morphology (see Fig. 2A) on which five traits are measured: the position of specific morphological landmarks.

Individual fitness is calculated based on the distance between each trait in each individual's phenotype and these same traits in an optimal morphology (see Fig. 1;Fig. S1). Mutation is applied to individuals in each generation. The processes of mutation, development and selection are iterated over generations to simulate evolution. A different optimal morphology is used in each of these evolutionary simulations. Overall, thus, our model includes a genetic space, a developmental parameter space and a morphological trait space (see Fig. 1). Each individual occupies a point

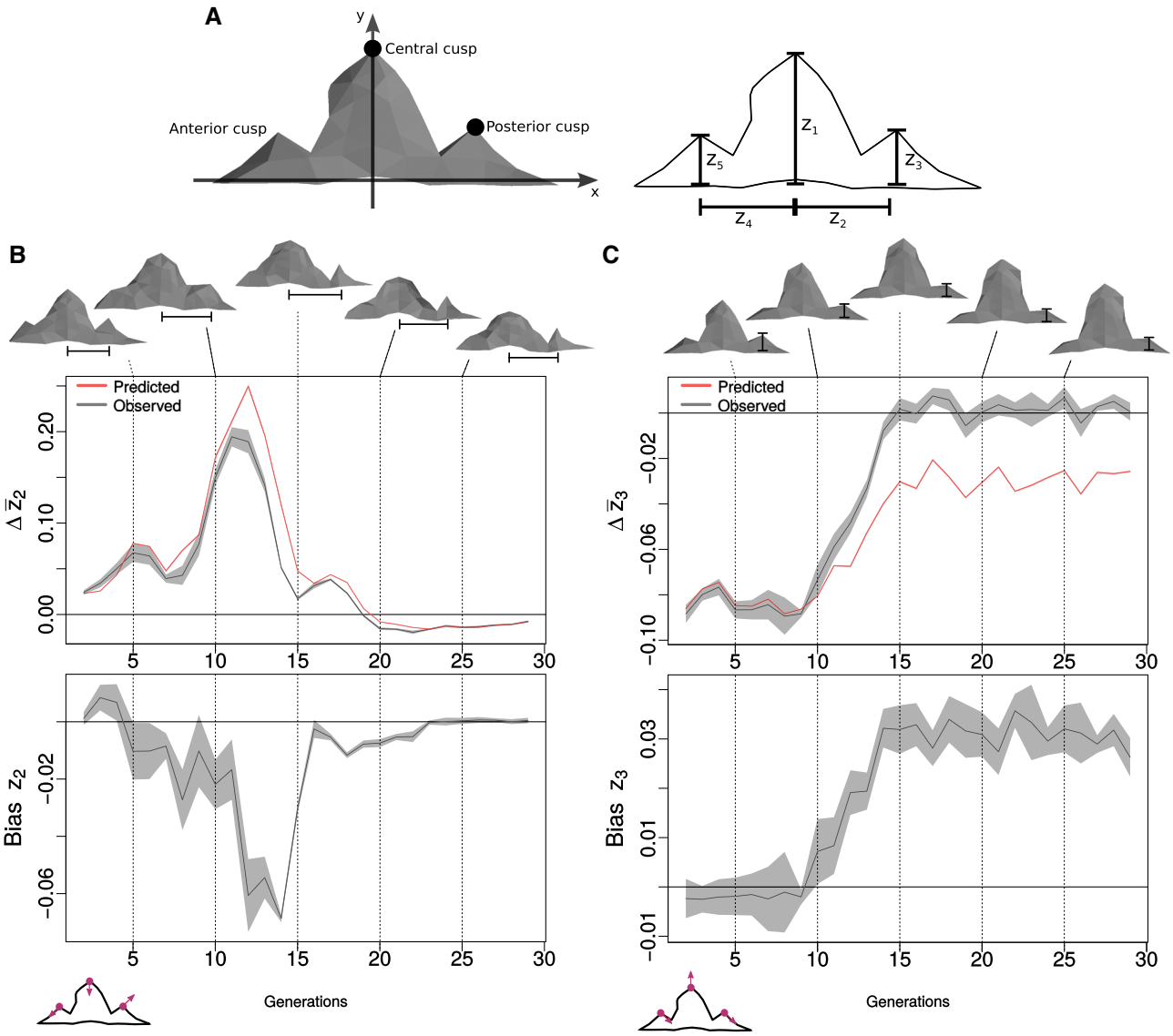

Figure 2. Observed and predicted change can differ in a systematic way over generations. Panel (A) shows, on the right, an example 3D morphology arising from the developmental model. Marked are the three cusps of the tooth, which are the landmarks. The outline on the right shows the five traits measured in each tooth, where z_i is the *i*-th trait. Panel (B) shows, on top, the observed (gray line) and predicted change (red line) for trait 2 in evolutionary simulation 18, and on the bottom the prediction bias in that simulation. Panel (C) shows the same but for trait 3 in evolutionary simulation 5. As explained in the main text, prediction bias is the difference between the predicted and observed change for each generation. Predicted change was calculated using the multivariate breeder's equation and estimates of *G*, *P*, and *s* in each generation. Observed change is shown with a 99% confidence interval (gray area) generated from repeating each generation 10 times to minimize the stochastic effect of drift, mutation, and recombination (the gray line is the mean of the repetitions, the gray area extends 3 SEM, see Materials and Methods section). Predicted change is shown with a 99% confidence interval (red area) calculated using the REML-MVN resampling method (the red line is the mean of the resamples, the red area is 3 SEM; notice that the interval is small compared to that of the observed change, due to the large sample size used to estimate each variance component). Prediction biases in the lower plots include 99% confidence intervals as well. Observed teeth with trait values closest to the population mean at five time points are included above the plots, with a black segment indicating the traits plotted. Direction of selection for each landmark is shown in the insets below each plot, with purple arrows pointing in the direction of selection. Evolution for the remaining traits for both simulations is shown in Figure S2. In the simulation shown in panel B, bias is big until generation 23. In the simulation shown in panel (C), bias is large from generation 10. There, the population reaches a region where it cannot further evolve toward the optimum, since genetic changes lead to teeth lacking lateral cusps.

in the genetic space (a genotype), a point in the developmental parameter space, and a point in the trait space (a phenotype). A population can then be seen as occupying a cloud in each of these spaces. In this article, as described above, the mapping between the developmental parameter space and the phenotypic space is determined by the developmental model while the mapping between the genotypic space and the developmental parameter space is additive. Notice then that any nonlinearity in the GPM is due to the developmental model alone. Because of that, one can talk about a region of the GPM being nonlinear when the population is in a region of the developmental parameter space where small parameter changes lead to relatively large phenotypic changes.

As we have the genotypes, traits, and fitness of each individual, we can estimate G, P, and s in each generation. From that (following equation [1]) we estimate, in each generation of each evolutionary simulation, the expected response to selection, that is, how much does each trait change. We then measure the difference between the response to selection expected from the multivariate breeder's equation and the observed response to selection in each generation of the evolutionary simulations (see Fig. 1).

Material and Methods

Evolution is modeled using an individual-based algorithm similar to that in Salazar-Ciudad and Marín-Riera (2013). The complete model is composed of a developmental model and a population model. The developmental model is used to generate each individual's phenotype from its genotype. The population model is used to determine the genotypes in each generation based on genotypes from the previous generation through selection, mutation, and recombination. Each model has a set of parameters that we refer to as developmental and population parameters, respectively. There are 21 developmental parameters and four population parameters. For each set of population parameters, we ran 32 simulations of 30 generations each using different optima to explore the behavior of the system. We refer to each of these simulations with a distinct optima and set of population parameters as an "evolutionary simulation." Both the population and developmental model were written in Fortran 90. The data from the evolutionary simulations was analyzed and visualized using R. In addition, in each generation, the matrix of additive genetic variances and covariances between traits, G, the matrix of phenotypic variances and covariances between traits, P, and the selection differentials, s, were estimated. From that a predicted response to selection was calculated from the multivariate breeder's equation and compared with the one observed in the evolutionary simulation.

DEVELOPMENTAL MODEL

The developmental model used is a computational model of tooth development (Salazar-Ciudad and Jernvall 2010). This model provides an example of a genotype–phenotype map for the morphology of a complex organ. The tooth developmental model is a mathematical representation of the current understanding of the basic gene network behind tooth development, and includes the basic cell behaviors (cell division and cell adhesion), cell mechanical interactions, and their regulation by gene products known to be involved in the process. The model is mechanistic in the sense that from this hypothesis and some very simple initial conditions—that is, a flat epithelium representing the initiation of tooth development—the model reproduces how the morphology and patterns of gene expression in 3D change during development until an adult tooth morphology is reached. The epithelium grows by cell division and folds owing to forces arising from cells. Cell division occurs after cell size reaches a threshold. Some gene products diffuse between cells: an activator autoregulates itself and, after a threshold, causes the differentiation of enamel knots (which will result in cusps) and induces the production of an inhibitor and secondary signals. These gene products interact with the cells and determine their mechanical behavior.

The dynamics of the developmental model are determined by the value of a set of 21 developmental parameters, which are genetically determined. These developmental parameters specify the magnitudes of several biological and physical properties of the cells, tissues, and molecules involved in the process such as proliferation rate, molecular diffusion rates, and interactions between gene products. The developmental parameters are listed in Table S1. The model's output is the three-dimensional position of tooth cells: the tooth morphology. For a more detailed description of the model we refer the reader to the original publication introducing it (Salazar-Ciudad and Jernvall 2010).

The tooth developmental model is suitable for our question because it is able to produce realistic population-level phenotypic variation (Salazar-Ciudad and Jernvall 2010). Furthermore, it is based on developmental biology and therefore includes epigenetic biophysical factors, such as extracellular cell signaling and diffusion in space and mechanical interactions between cells, which have been extensively proposed to lead to complex genotype–phenotype maps (Forgacs and Newman 2005).

For our study, we define measurable morphological traits in the tooth shape and track them through the generations of the population. We define the phenotypic traits as the coordinates of landmarks in the tooth morphology. We locate landmarks in the three tallest cusps in each tooth. In tooth development, cusp height correlates negatively with the time they appear in an individual's development (Jernvall et al. 1994). Thus, the three tallest cusps are the first three cusps to be formed during development. Because of that, we position the teeth on the (x,y) plane so that the x position of the central cusp is 0. The $y = 0$ axis of each tooth is at its base (calculated as the average of the positions of its margins), just as in previous studies using the same model (Salazar-Ciudad and Marin-Riera 2013). Since all teeth arise from the same

Table 1. Population parameters. Values for the population parameters. We defined a core population parameter set and studied deviations for each parameter at a time. Values where defined based on Jones et al. (2014). Note that the number of loci is per developmental parameter (i.e. total number of loci in the core population parameter set is five loci times 21 developmental parameters, 105 loci). The total number of loci affecting each trait is therefore always within the previously suggested ranges (50–100 in Goddard 2001; 2–100 in Falconer Mackay 1996).

Parameter	Low value	Core value	High value
Mutational effect size	0.05	0.1	0.2
Frequency of mutations	0.0005	0.001	0.002
Number of loci	2	5	10
Population size	100	600	1000

deterministic mathematical model, they are all in the same arbitrary space units and they are directly comparable without re-scaling or re-orientation. Note that the tooth model is deterministic and, thus, teeth develop in the same position (there is no noise displacing or rotating them that should be corrected by procrustes analysis). Figure 2A shows the location of the three landmarks in an example morphology. Trait 1 is the *y*-coordinate of the landmark located in the central cusp. Trait 2 and trait 3 are the *x*- and *y*-coordinates of the landmark located in the posterior cusp, respectively. Similarly, trait 4 and trait 5 are the *x*- and *y*- coordinates of the landmark located in the anterior cusp, respectively. Traits 1, 3, and 5 are therefore the heights of the three cusps.

POPULATION MODEL

The developmental model was embedded in a population model. Table 1 shows the four population parameters (population size, mutation rate, mutational effect size, and number of loci.) We defined a core set of population parameters and studied deviations from those core values. For each parameter set, we ran 32 evolutionary simulations each with a different optimum (see section "Mapping between phenotypes and fitness" below). All populations were composed of equal number of males and females, with no sexual dimorphism. The sexes were introduced primarily for consistency during the quantitative genetics analysis (i.e., sire and dam distinction, see "Quantitative genetics model" section).

The population model considers four steps per generation: (1) a mapping between genotypes and developmental parameters; (2) a mapping between developmental parameters and phenotypes (the developmental model); (3) a mapping between phenotypes and fitness; (4) reproduction, with recombination and mutation on the genotypes. By iterating steps (1) to (4) in each generation, we simulated how the genotypes and phenotypes of the population change over generations (Fig. 1).

(1) Mapping between genotypes and developmental parameters: The input of the developmental model is the values of the 21 developmental parameters. Each of these parameters corresponds to a developmentally relevant interaction, but not necessarily to a single gene (see Salazar-Ciudad and Jernvall 2010). In the model, for a given individual, several genes contribute to the value of each developmental parameter. Each individual's genotype is diploid with a fixed number of loci. Each allele in a loci has a specific quantitative value. These genetic values will determine the developmental parameters that will go into the tooth developmental model, and constitute the heritable information. The developmental parameter values are determined as the sum of the contributions of a fixed number of these genetic loci (similar to Rolian 2015). The population parameter number of loci (see Table 1) determines how many haploid loci additively make up a single developmental parameter. All developmental parameters are thus genetically determined. We purposely use this very simple yet unrealistic mapping because our focus is on how the part of the developmental dynamics we understand the better (i.e., the developmental model) affects phenotypic variation and evolution, without confounding effects. There is no pleiotropy since in the mapping from genes to parameters each loci contributes to the value of one and only one developmental parameter. Pleiotropy from genes to phenotype arises from the developmental dynamics which map parameters to phenotypes.

(2) Mapping between developmental parameters and phenotypes: This is accomplished by running the developmental model, for each combination of developmental parameters corresponding to each individual. As explained in the "Developmental model" section, phenotypes are measured as the coordinates of three cusps in the developed tooth (see Fig. 2A). All developmental parameters contribute to evolutionary change and some parameters have larger phenotypic effects than others (see Fig. S3).

(3) Mapping between phenotypes and fitness: Different regions of the GPM will have different properties and may lead to different behaviors. This is why we had to explore the GPM evenly in all directions. For this, we performed evolutionary simulations selecting in different directions (i.e., toward different optimal morphologies in this trait space). Each evolutionary simulation had a different optimal morphology. Each of the five morphological traits was chosen to increase or decrease in respect to an initial reference morphology (see the Initial population section). This leads to a total of $2^5 = 32$ trait combinations. The optimum value for each trait was set at an equal distance, ± 3 units from the mean value of that trait in the starting population (see Fig. S1). Optimal morphologies are therefore evenly distributed and equidistant around the initial morphology. Each of the optimal morphologies determines, thus, a distinct direction of selection. These directions, as defined in the trait space, result in different and non-colinear $\beta = P^{-1} s$ at the start of each evolutionary

simulation (see Fig. S12). We also ran simulations starting from different initial morphologies, to discard that our results are particular to a certain region of parameter space.

Selection was performed in each generation by selecting the 50% of males and females with least distance to the optimum morphology. The top males and females were then mated randomly. We do not allow siblings to mate to reduce inbreeding. If over evolution individuals with only one or two cusps arise (i.e., some of the five traits we study are missing), they are assigned 0 fitness and excluded from the calculation of G, P, and s.

(4) Reproduction, recombination, and mutation: Once the parents are selected, they produce one gamete each by randomly selecting one of the two alleles for each loci with equal probability. The gametes of the parents then fuse to form the diploid genome of the offspring. Each parental couple generates two males and two females for the next generations. This keeps the population size constant and results in all selected parents having exactly the same fitness.

In the gamete formation process, there is a per-loci probability of mutation, which is one of the population parameters called mutation rate (see Table 1). Mutation is implemented by adding a random number to the mutating loci, drawn from a normal distribution with a zero mean and a standard deviation equal to another of the population parameters, the "mutational effect size" (see Table 1).

(5) Initial population: The initial population is composed of exactly the same individuals in all evolutionary simulations with the same set of population parameters. The number of individuals is specified by the population parameter "population size" (see Table 1). To build the initial population, each locus starts with 10 equally frequent alleles with values drawn from a normal distribution. The mean of the distribution was such that when summing over all loci one obtains the reference developmental parameter values, that is, the developmental parameter values that reproduce the morphology of the postcanine tooth of the ringed seal (*Phoca hispida ladogensis*, see Salazar-Ciudad and Jernvall 2010). The variance of the distribution of genetic values was set equal to the population parameter mutational effect size. These genetic values were then used in simulations under stabilizing selection for 400 generations. Stabilizing selection was performed by using the mean traits of the population at the first generation as optimal morphology. In each generation, the fitness of each individual in the population was calculated as a function of the distance of its trait values and the optimal trait values as, $w(z) = \exp(-\alpha \sum_{i=1}^{n} (z_i - o_i)^2)$, where z_i is the i-th element of the vector of trait values for the individual, o_i is the i-th element of the vector of optimum trait values and is the fitness function. The steepness of the selection gradient is determined by the parameter α, which was set as 0.05 (Jones et al. 2004).

The parents of each individual in a generation were chosen at random from the individuals of the previous generation. For each individual in the previous generation, the probability of being chosen as a parent was equal to its fitness divided by the sum of the fitness of all individuals in the population in that generation. The resulting population after the 400 generations of stabilizing selection was used in the actual evolutionary simulations as the initial condition.

ENVIRONMENTAL EFFECTS

To simplify the study of the impact of the GPM on the accuracy of the breeder's equation, we did not include environmental effects for most evolutionary simulations. How the omission of environmental effects affects the system was studied by performing evolutionary simulations with the core population parameter set including environmental effects. In some simulations environmental effects were introduced at the level of the developmental parameters while in others they were introduced at the level of traits themselves. For one set of simulations, environmental noise was added to each of the loci of the genotype before being added to determine the developmental parameters. The noise added was 10^{-3} times the mutational effect size. For another set of simulations (one per optimal morphology), we introduced environmental effects at the level of the phenotype. After running the developmental model for each individual and obtaining an adult phenotype, a small environmental random effect was added to each trait. These effects were normally distributed, with mean 0 and standard deviation equal to 10^{-2} times the value of the trait.

REPETITIONS

Our objective is to bracket the ability of classical quantitative genetics to predict evolutionary change using a realistic GPM. Sources of prediction error aside from the characteristics of the GPM therefore must be carefully controlled. The two other main sources of error are (1) stochastic error due to drift; (2) errors in the estimation of the quantitative genetic parameters, mainly of the G-matrix. The latter will be discussed in the Quantitative genetics model section.

Because we are working with a finite population with a finite genome, there will be stochasticity in the change of the mean of the traits for the population due to sampling of individuals and alleles. This occurs in the couple formation during random mating and the random segregation of alleles in gamete formation. Stochasticity is also introduced through mutations. The randomness in the above processes will result in deviations from the predicted change using the breeder's equation that are not due to the nonlinearities of the GPM. To account for these types of error, we perform repetition simulations. After running the full evolutionary simulation for each of the optima for all 30 generations, we take the individuals of generation i and repeat the

processes of random mating, mutations, and allele segregation 10 times. This results in 10 "parallel" generations i+1. For each of these repeated, parallel generations i+1, we can calculate a change in mean value for the traits with respect to their mean values in generation i. These changes will only differ from each other due to the randomness of the processes described above. The selected parents from generation i are the same in all repetitions, but the couple formation, the mutations, and the recombination differ. What we call the "observed change" is the change in trait means in the simulated population between two consecutive generations, averaged over all repetitions. We also use the information of the spread of the repeated changes. Figure S4 shows that this method completely eliminates prediction errors due to stochasticity when using a linear GPM.

QUANTITATIVE-GENETIC MODEL

The observed change is to be compared with the predicted change using the multivariate breeder's equation (1). It is known that the parameters of the breeder's equation can change in time, although there is debate regarding how fast this occurs (Roff 2000; Doroszuk et al. 2008; Aguirre et al. 2013; Penna et al. 2017). To avoid errors arising from outdated estimations, G, P, and s were calculated for each generation.

We calculated variance components by building a half-sib breeding design on the evolutionary model. To estimate the variance components in generation i, we randomly took half of the males and half of the females from generation i and formed couples that produced four offspring each. This process was repeated 10 times, each time mating individuals randomly. Note that the parental population is always the same (i.e., the individuals in generation i). This design provides 11 times the population size of phenotypic data composed of parents, full-sibs, and half-sibs, to estimate each G and P (i.e., 6600 individuals for the core population parameter set).

Restricted maximum likelihood (REML) estimates of additive genetic variance were obtained using the software WOMBAT (Meyer 2007). The animal model used was the simplest possible, $y_i = \mu + a_i + e_i$, where y_i is the phenotype of individual i, μ is the population mean, a_i is the additive genetic merit of individual i, and e_i is a random residual error. When fitting an animal model using REML, initial estimates of variance components have to be provided. For each evolutionary simulation, initial estimates for additive genetic variance and residual variance in generation 1 were set to half the total phenotypic variance. For subsequent generations, the initial estimates were the estimates from the converged REML fitting carried out for the previous generation. In all cases, trait values were pre-multiplied by 100 to help with convergence.

Sampling variation in the estimation of G was accounted for using the REML-MVN method (Houle and Meyer 2015) implemented in WOMBAT. The method allows to approximate the uncertainty in evolutionary parameters estimated using animal models by resampling G-matrices from the distribution of its maximum-likelihood estimate. For each generation, we resampled 100 G- and P-matrices from this distribution and calculated 100 predicted multivariate changes using equation (1). What we call the "predicted change" is the mean change in trait means using the resampled variance components. We also use the information of the spread of the changes predicted using resampled variance components as a measurement of uncertainty.

MEASURING THE PREDICTION ERROR

As explained in the two preceding sections, the observed change for trait k is the average of the changes in the mean of trait k over the 10 repetitions. We symbolize it as $\Delta\bar{z}_k^o$. The standard deviation of the changes in the mean of trait k over the 10 repetitions is symbolized as σ_k^o. The predicted change for trait k is the average of the changes in the mean of trait k, obtained using the breeder's equation with 100 resamples of the G- and P-matrices from their distributions using the REML-MVN method. We symbolize it $\Delta\bar{z}_k^p$. The difference between the observed change for trait k and the predicted change for trait k, is what we call "prediction bias for trait k." We call it bias because it is not due to stochastic effects, which are removed by the averaging over repetitions (see section "Repetitions").

We calculate the relative bias as the absolute value of the bias divided by the average of the absolute values of the predicted and the observed changes, $2|\Delta\bar{z}_k^o - \Delta\bar{z}_k^p|/(\Delta\bar{z}_k^o + \Delta\bar{z}_k^p)$. This is a measure of how large the bias is, relative to the actual changes. We also measure the effect size of the bias, as the bias, divided by the standard deviation of the repeated observed changes $|\Delta\bar{z}_k^o - \Delta\bar{z}_k^p|/\sigma_k^o$. This is a measure of how large is the bias relative to the variation in the observed changes. If the effect size is small, it means it is irrelevant when compared to changes due to stochastic factors inherent in any evolutionary process.

We also calculate a multivariate error, taking all traits together. What we call the multivariate bias, $\| \Delta\bar{z}_k^o - \Delta\bar{z}_k^p \|$, is the Euclidean norm (symbolized by $\| \, . \, \|$) of the difference between the vector of predicted changes ($\Delta\bar{z}^p$) and the vector of observed changes ($\Delta\bar{z}^o$). What we call the relative multivariate bias is the multivariate bias, divided by the Euclidean norm of the vector of observed changes ($\| \Delta\bar{z}_k^o - \Delta\bar{z}_k^p \| \, / \, \| \Delta\bar{z}_k^o \|$).

MEASURING NONLINEARITY OF THE GPM

Local nonlinearity was measured based on how well a linear map describes the relationship between developmental parameters and traits. For each generation in each evolutionary simulation for the core population parameter set, we build a matrix where each column is an individual and each row a developmental

parameter, standardized by removing the mean and dividing by their standard deviation (we call this matrix U, U_{kj} is the standardized k-th developmental parameter of the j-th individual). We also have a matrix where each column is an individual and each row is the value of a standardized trait (we call this matrix Z; Z_{kj} is the standardized k-th trait of the j-th individual).

We want to find a linear transformation B, with number of rows equal to the number of traits and number of columns equal to the number of developmental parameters such that $Z = BU + e$. We want e, the error term, to be minimized. The least square solution $\tilde{B}$ is given by $\tilde{B} = ZU^T(UU^T)^{-1}$, where the operator T is the transpose, and -1 the inverse. $\tilde{B}$ allows to predict the trait values given the developmental parameters, for the best linear fit. We then measure nonlinearity as the difference between $\tilde{B}U$ and Z. If the difference is *0*, it means that the relationship between U and Z is perfectly described by the linear transformation $\tilde{B}$ (i.e., the map is linear). We calculate the nonlinearity in each trait as the root-mean-square error of the fit for the trait. We also calculate a multivariate nonlinearity as the norm of the error.

Nonlinearity in the GPM is also graphically shown using marginal parameter–phenotype maps. These maps are constructed for a given generation of a given evolutionary simulation. The mean value of each of the 21 developmental parameters in the population is calculated. We then select the two developmental parameters whose variation explains the most phenotypic variation. The values of these two developmental parameters are then tinkered around the mean value. This produces a GPM that we call marginal because it is the result of changing only two of the 21 developmental parameters. Marginal GPM show the relationship between developmental parameters and trait values.

Results

We found that in many generations there is a significant discrepancy, or prediction error, between the trait changes observed in the simulations and the trait changes predicted from the multivariate breeder's equations. Figure 2B and C shows the changes in the trait means, both observed and predicted, for two example evolutionary simulations (see also Videos S1 and S2). To minimize the prediction error due to stochastic processes such as drift, mutation, or recombination, we re-simulated each generation of every evolutionary simulation 10 times (see Material and Methods section) and calculated the mean observed change per trait and generation. Although comparatively small (shown in red in Fig. 2B and C), the estimation of the *G*-matrix also has some variation and then there is also some variation in the predicted change in each generation and evolutionary simulation.

The difference between the mean observed change and the mean predicted change is what we call the bias. This bias

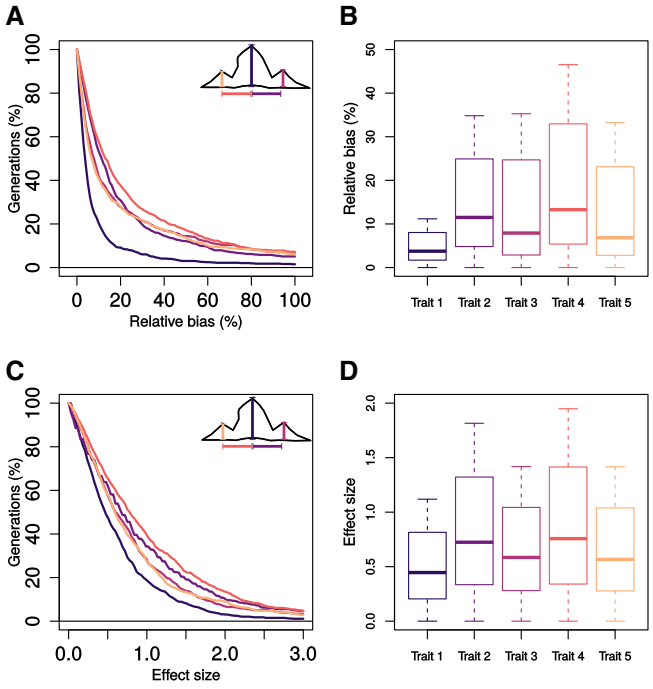

Figure 3. Distribution of bias for simulations in the core population parameter set. Panels (A) and (B) show the distribution of relative bias, calculated as the absolute value of the difference between predicted and observed change, divided by the mean of the absolute value of the two changes. Panel (A) shows a cumulative plot, showing what percentage of generations (*y*-axis) show at least a given amount of relative bias (*x*-axis), for each trait. The half-life of the curves is highly dependent on the trait (half of the generations have at least 4%, 11%, 8%, 13%, and 7% relative bias for traits 1, 2, 3, 4, and 5, respectively). Panel (B) shows the distribution of the relative bias. Traits 2 and 4 show larger relative biases that the other traits. Panels (C) and (D) show the distribution of the effect size of the bias, calculated as the absolute value of the bias divided by the standard deviation of the repetitions made to calculate the observed change (see Materials and Methods section). Similarly, the size of the effects is larger for traits 2 and 4. Panels (A) and (C) include an inset with the outline of a tooth (see Fig. 2) and the traits marked with colors. Color coding for traits is the same for all panels. Generations where the populations is in regions of the GPM where small parametric changes lead to bicuspid teeth, such as the one shown in Figure 2C, were not included in this plot because once this occurs in a generation, as we explain in the results, bias remains high evermore.

is a systematic prediction error that is not due to stochastic processes, nor to misestimation of variance components (*G*- and *P*-matrices). In the evolutionary simulation shown in Figure 2B, there is a clear prediction bias in the simulation until around generation 23. During these first 23 generations, the teeth undergo relatively large shape changes (see the example teeth in Figs. 2B and 4, and Video S1). Figure 2C shows an extreme case

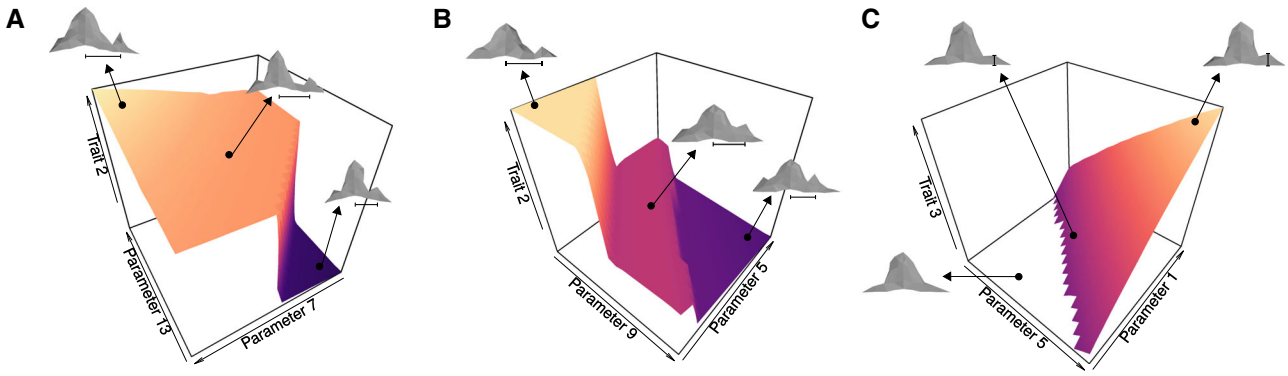

Figure 4. **Prediction bias arises when the population is in a nonlinear region of the parameter–phenotype map. The figure shows the marginal genotype–phenotype map at different time points of three different evolutionary simulations of the core population parameter set where prediction bias was found. The maps show how a trait (*z*-axis) changes when two specific developmental parameters are varied while keeping all others constant at their population mean value (*xy*-plane in the plot, see Materials and Methods section). Teeth are included for some points in the map, with a bar showing the trait being measured. Panel (A) is generation 20 of evolutionary simulation 6, where relative bias for trait 2 is 52% (see Fig. S2). Panel (B) shows generation 14 of evolutionary simulation 18, where relative bias for trait 2 is 80% (this is the simulation shown Fig. 2B). In both cases, bias in the prediction arises under these nonlinear maps, where small changes in parameter values produces relatively large changes in the measured traits. Panel (C) is for generation 15 of evolutionary simulation 5, where relative bias is 200% (this is the simulation shown Fig. 2C). In this case, when both parameters are low, one of the lateral cusps does not form. Indeed, the tooth shown for low values of both Parameter 1 and 5 is missing the posterior cusp, so there is no black bar showing the trait. Bias arises because the linear approach of multivariate breeder's equation predicts that teeth with smaller values for trait 3 can be produced when in fact they cannot.**

in which a cusp is lost during evolution. The prediction becomes very biased at around generation 10.

Bias was not exclusive to a single trait or simulation, it was common (Figs. S2 and S11 show all evolutionary simulations for the core population parameter set). Figure 3B and D shows the per-trait distribution of relative bias and effect size of the bias in the evolutionary simulations for the core population parameter set. Figure 3A and C shows how many of the simulated generations have relative biases or effects sizes larger than a given amount. Evolutionary simulations using different initial morphologies in evolution show similar relative bias (Fig. S5). The relative bias depends on the trait. Trait 1, the *y*-coordinate of the landmark located in the central cusp, show little bias across all simulations (median relative bias 3.8%, median effect size 0.45). Traits 2 and 4, the *x*-coordinates of the landmarks located in the lateral cusps, showed the most amount of bias (median relative biases: 11.8% and 13.5%; median effect sizes: 0.74 and 0.76, respectively). For trait 4, 18% of the generations have a relative bias of 50% or larger.

Bias arises from nonlinearities in the GPM. Figures 4 and S6 show explanatory examples of how specific aspects of the developmental dynamics lead to nonlinearities and then to bias. We found that bias was larger when additive variances in the direction of selection were smaller (see Fig. S7C and Table S2 for heritability estimates). This can occur either because phenotypic variation is also small or because a large part of the genetic variation is nonadditive, for example, because the GPM is nonlinear. We found bias associates with measures of nonlinearity of the GPM around the region of the developmental parameter space where the population is distributed at a given generation (see Fig. 5). Both the spread and magnitude of the bias were larger in nonlinear regions of the GPM. In these regions, small genetic changes lead to small changes in the developmental parameters but these lead to relatively large changes in some traits. As a result, when the populations move across these regions of the developmental parameter space, the change in some of the traits is different to that predicted from the linear approximation of the multivariate breeder's equation. The breeder's equation, then, over- or underestimates the change in the mean of traits. As Figure S2 shows most populations pass through regions that lead to bias as they evolve.

Some evolutionary simulations showed particularly evident prediction biases that where sustained for many generations (see Figs. 2C and 4C). These were most common when selection favored a decrease in a cusp's height (either trait 3 or 5). Such trait decreased in the evolutionary simulation until some limit value, which was always larger than zero. Development was not able to produce phenotypes with smaller values in this trait and, thus, there was no evolution beyond this limit value. That was not a global limitation of such trait but a local limitation. In other words, the model may be able to produce such phenotypes, but not from where the population is in the developmental parameter space nor from any region near it (i.e., mutation in single or few genes and recombination would not lead to further evolution

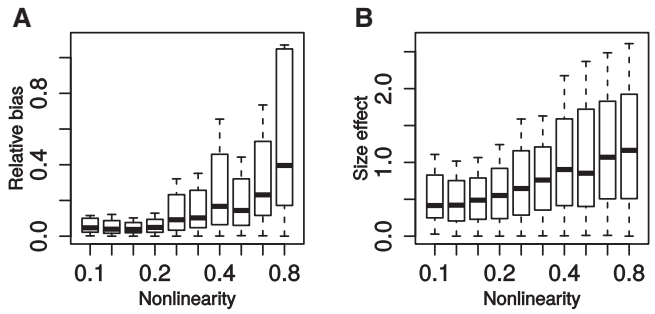

Figure 5. **Prediction bias correlates with the local nonlinearity of the genotype–phenotype map in the developmental model. Local nonlinearity was measured for each trait in each generation of all evolutionary simulations for the core population parameter set (see Materials and Methods section). Panel (A) shows the relative bias against the nonlinearity. Panel (B) shows the effect sizes for the bias against the nonlinearity. Both plots include bias and nonlinearity for all five traits. The points were binned and plotted as boxplots, with whiskers extending 0.5 times the interquartile range from the box. Both the spread and magnitude of the relative bias and the effect size of the bias increase with the nonlinearity of the GPM. This means that for linear GPMs, the prediction works very well always. For nonlinear GPMs, the prediction can show very large mismatches but can also show small biases. Generations where the populations is in regions of the GPM where small parametric changes lead to some cusps not forming, such as the one shown in Figure 2C, were not included in this plot.**

in the direction of selection). When the population is close to the limit value in a trait, individuals produce offspring with trait values higher than the limit, or offspring lacking that trait completely (i.e., individuals where a cusp does not form during development), but never offspring with trait values lower than the limit value. Individuals without one or both of the lateral cusps are given fitness 0 and not included in the estimations of G, P, and s. In the evolutionary simulations where a local trait limit is found, phenotypic variance decreases as the population mean is pushed toward the limit value. Due to mutation and recombination, however, there is always some small phenotypic and genetic variation. This phenotypic variation, however, does not accumulate at the limit trait value but spreads before it since no phenotypes exist beyond the trait limit (see Fig. S8). From this genetic variation, the multivariate breeder's equation estimates, due to its linear nature, that there should be a response to selection decreasing the trait beyond its limit (see Fig. S8). This leads to a bias since a change in the trait mean is predicted where there can be none. Although biases related to trait limits were not rare, most biases were not due to these limits. In fact, our overall estimation of bias (Figs. 3, 5, and 6) excludes this type of bias.

Population parameters such as population size, mutation rate, mutation effect size, and number of loci had a modest effect on bias (Fig. 6). We performed evolutionary simulation for the same 32 optima, but changing one population parameter from the core parameter set at a time (see Table 1). These parameters affect how fast the population moves in the trait and developmental parameter space, but do not affect the nonlinearity of the GPM itself since, as we explain in the Introduction, the nonlinearity of the GPM is determined by the developmental model alone (i.e., the mapping between genotype and developmental parameters is additive; see Fig. 1). Irrespective of the value of the population parameter set, the population is crossing the same regions of the GPM on its way to the optimum, and the prediction biases are therefore of similar relative magnitude among the population parameter sets. A notable exception is the small population size. In that condition, there is an increase in prediction errors due to both stochastic factors (drift) and misestimation of variance components (due to smaller sample sizes to estimate the parameters). This leads to larger biases.

The addition of environmental effects had a negative impact in the predictive ability of the multivariate breeder's equations (Fig. S9, see Materials and Methods section for details of the implementation). When environmental effects were simulated by adding a normally distributed random amount to each trait in each individual, bias increased only slightly. When environmental effects were simulated by adding a normally distributed random term to each locus, bias increased substantially. Essentially this increased the spread of the population in the genetic space and, consequently, in the developmental parameter space. With the population occupying a larger volume of the developmental parameter space, it is more likely that it will include highly nonlinear regions that lead to bias when applying the breeder's prediction.

Discussion

There is already some theoretical literature stating that the assumptions of the breeder's equations may not always hold (Merilä et al. 2001; Pigliucci 2006; Wilson et al. 2006; Hadfield 2008; Bonamour et al. 2017). These previous studies, however, do not directly consider the GPM. Exceptions are Rice's (Rice 2002, 2004, 2012) and Morrisey's work (Morrisey 2015). Rice's work shows that the predictions of the multivariate breeder's equation can be inaccurate under complex, nonlinear GPMs. Although this work demonstrates that inaccuracies can occur, it does not include an understanding of actual GPMs from which to evaluate how large and frequent should these inaccuracies be, as our work intends.

There is a vast literature on population genetics models with epistasis that is related to the GPM (Barton and Turelli 1989; Hansen and Wagner 2001). Hansen and Wagner (2001) introduced a multilinear model in which the effect of an allele substitution in a locus is modeled as a linear combination of locus effects and pairwise epistatic effects among loci. Using this description,

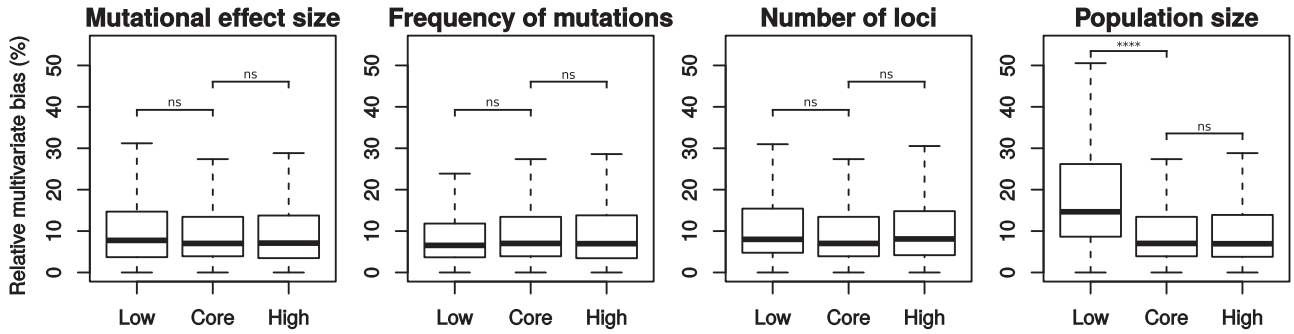

Figure 6. **Population genetic parameter has only a modest effect on the bias. The plots above show the distribution of multivariate relative bias for all 30 generations of the 32 evolutionary simulations in each of the population parameter sets. We used a bootstrap approach (10,000 resamples) to compare the medians of the multivariate relative biases of the population parameter with that of the core parameter set. (Mutational effect size: Low versus Core, *P*-value = 0.091; High versus Core, *P*-value = 0.85; Frequency of mutations: Low versus Core, *P*-value = 0.27; High versus Core, *P*-value = 0.89; Number of loci: Low versus Core, *P*-value = 0.064, High versus Core, *P*-value = 0.057; Population size: Low versus Core, *P*-value < 0.001; High versus Core, *P*-value = 0.88). Because the population parameters do not change the GPM, similar prediction biases were found in all conditions. Low population size did show much larger prediction errors than the other population parameters. This is due to the fact that there are increased errors due to stochasticity (more bias) and worse estimates of variance components (smaller sample sizes for their estimation). Generations where the populations is in regions of the GPM where small parametric changes lead to some cusps not forming, such as the one shown in Figure 2C, were not included in this plot.**

the authors develop expressions for usual quantitative genetics measures such as additive genetic variance, and most relevant to us, a modification of the breeder's equation that takes into account epistasis. Carter et al. 2005 completed this analytical work with simulation studies using the multilinear model, and showed that directional epistasis modifies the response to selection from that expected by the breeder's equation (see also Hansen et al. 2006).

Compared to these models, ours does not assume a specific pattern of epistasis among loci. Instead, such patterns arise from the dynamics of the development model. It has been argued that these epistasis models are more general than models explicitly based in development because they are not specific of any system (Hansen 2008). The characteristics of the genotype–phenotype map arising from our development model, however, are not addressable from these models (see Fig. S10). In addition, one can also argue that a single realistically complex genotype–phenotype map, such as ours, may be more informative about realistically complex genotype–phenotype maps in general than epistatic models that are not based on experimental data on how genes interact. This may be specially the case for complex phenotypes, such as morphology, that are known to arise from complex networks of interactions between genes.

Our model represents a more realistic GPM than that of the multilinear model, but it is not analytically tractable. Indeed, while working with the multilinear model, Carter et al. (2005) state that "to study truly nonlinear forms of gene interaction, we have to turn to highly specific architectures, and rely almost exclusively on computer simulations."

Among the criticisms to the predictive capacity of quantitative genetics, those of Polly (2008) are specially related to our work. Polly proposes three key consequences that development can have on morphological evolution and that evolutionary quantitative genetics cannot fully account for. We found evidence for the three of them in our simulations. First, even with continuous variation at the level of underlying developmental parameters, traits may occasionally exhibit small jumps in evolution. This occurs in our simulations and leads to bias. The posterior cusp in Figure 4A, for example, exhibits a jump in position. It is important to notice, however, that these jumps are small, and, thus, may not be perceived as trait discontinuities (in fact the range of trait variation observed in our simulations is within that observed in nature, see Fig. S3). Second, Polly argues that developmental interactions may limit the production of certain phenotypic variants in non-additive ways, leading to a bias in the prediction of the response to selection. This is mirrored in our findings of the prediction bias in certain regions of the GPM. We explicitly show that the structure of additive genetic variances and covariances in the *G*-matrix does not always summarize all associations between traits, leading to the bias in the prediction. Lastly, Polly mentions that some nonlinearities will lead to loss or appearance of structures. Evidence for this can also be found in our simulations. In some cases, such as the evolutionary simulation shown in Figures 2C and 4C, some traits are lost in some individuals (i.e., a cusp fails to appear in development). This, as explained in the Results, leads to problems in the application of the breeder's equations.

A possible criticism to our study is that we may be applying the multivariate breeder's equation in situations where its

assumptions are violated (i.e., linear GPM, see Fig. 4). It is an open question, however, how sensitive the predictions of quantitative genetics are, in practice and in theory, to departures from these assumptions. In this respect, our study can be seen as a systematic quantitative exploration of the validity of these assumptions under a realistically complex GPM. The prevailing view in quantitative genetics is that the breeder's equations are very accurate when applied to single traits (Roff 2007). When selecting for multiple traits at the same time or when selecting for some traits and studying the correlated response in other traits, however, the empirical evidence is scarce and does not always fit the expectations from the breeder's equation (Beldade and Brakefield 2002; Roff 2007; Allen et al. 2008; Hine et al. 2014). Our results suggest a simple reason for these biases and provide a theoretical nonlinear-GPM perspective on what to expect when multiple trait selection experiments finally become more common.

Our results are robust to changes in the classical population parameters such as population size and mutation rate (Fig. 6). Bias is due to the complexity of GPM, and the population parameters simply affect how fast and how widely the population moves in the trait space over generation and, consequently, on the developmental parameter space. In other words, the population parameters do not affect the overall properties of the GPM arising from the model. For example, population size determines the area, or hypervolume, of the developmental parameter space over which the population's individuals are spread in each generation but it does not affect which regions of this space have a nonlinear GPM.

It has been proposed that nonlinearities in the GPM are not relevant for the response to selection, at least in the short term (Hill et al. 2008; Hansen 2008). According to these authors the additive model will be a good local approximation to the GPM. It is argued that, even if the phenotypic effect of an allele depends nonlinearly on the genetic background (i.e., the specific combination of alleles in the other loci), sex ensures the reshuffling of alleles into different combinations in each generation. In a population, then, the effect of an allele is averaged over all backgrounds in the population. This averaging leads to a local smoothing of the GPM that would make it locally linear. However, whether such smoothing is enough to make the nonlinearity of the GPM irrelevant for short-term predictions should depend on how large are the nonlinearities, as well as on the recombination rates. Our results explicitly show that even for large populations in linkage equilibrium, the GPM is often nonlinear enough to lead to significant deviations between observed and predicted changes.

There is a long ongoing apparent discrepancy between evolutionary developmental biology and quantitative genetics (Cheverud 1984; Polly 2008; Salazar-Ciudad 2006, Hansen 2008). One approach views the GPM as simple enough, at least in practice, for trait evolution to be predictable from linear statistical approaches. The other views the GPM as highly nonlinear. Our results present a potential point of connection between these two views. The predictions of quantitative genetics would often work accurately but, very often too, they would show relatively large prediction errors that cannot be corrected by improving the estimates of quantitative genetic parameters, since they arise from the nonlinearity of the GPM.

AUTHOR CONTRIBUTIONS

I.S.C. supervised the study, I.S.C. and L.M. were associated with funding acquisition, and I.S.C. and L.M. conceptualized the study and wrote the article. L.M. and I.S.C. were associated with investigation, software, and methodology.

ACKNOWLEDGMENTS

The authors thank Tobias Uller, Asko Mäki-Tanila, Pascal Hagolani, Jukka Jernvall, Thomas F. Hansen, Juha Merilä, Heikki Helänterä, Renske Vroomans, and Jhon Alves for useful comments. The authors also thank Editor Katrina McGuigan and two anonymous reviewers for comments that greatly improved the manuscript. This research was funded by the Finnish Academy to ISC (315740 and 272280) and the Integrative Life Science graduate school to L.M.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

The source code and data are uploaded in DRYAD: https://doi.org/10.5061/dryad.9cnp5hqdr.

LITERATURE CITED

Aguirre, J. E. H., K. Mcguigan, and M. Blows. 2013. Comparing G: multivariate analysis of genetic variation in multiple populations. Heredity 112:21–29.

Alberch, P. 1982. Developmental constraints in evolutionary processes. In: Bonner J. T., (Ed.). Evolution and Development. Dahlem Konferenzen. Heidelberg. Springer. p 313–332.

Allen, C. E., P. Beldade, B. J. Zwaan, and P. M. Brakefield. 2008. Differences in the selection response of serially repeated color pattern characters: Standing variation, development, and evolution. BMC Evol. Biol 8:94.

Arnold, S. J., R. Bürger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the G-matrix. Evolution 62:2451–2461.

Barton, N. H., and M. Turelli. 1989. Evolutionary quantitative genetics: how little do we know?. Annual Review of Genetics 23:337–370.

Beldade, P. K. K., and P. M. Brakefield. 2002. Developmental constraints versus flexibility in morphological evolution. Nature 416:844.

Bonamour, S., C. Teplitsky, A. Charmantier, P.-A. Crochet, and L.-M. Chevin. 2017. Selection on skewed characters and the paradox of stasis. Evolution 71:2703–2713.

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

Cheverud, J. M. 1984. Quantitative genetics and developmental constraints on evolution by selection. J. Theor. Biol. 110:155–171.

Doroszuk, A., M. W. Wojewodzic, G. Gort, and J. E. Kammenga. 2008. Rapid divergence of genetic variance-covariance matrix within a natural population. Am. Nat. 171:291–304.

Falconer, D. S., and T. F. C., Mackay. 1996. Introduction to quantitative genetics. 4th ed. Pearson, New York, NY.

Forgacs, G., and S. A. Newman. 2005. Biological physics of the developing embryo. Cambridge Univ. Press, Cambridge, U.K.

Gilbert, S. F., and M. J. F. Barresi. 2016. Developmental biology. Sinauer Associates, Sunderland, MA.

Gjuvsland, A. B., J. O. Vik, D. A. Beard, P. J. Hunter, and S. W. Omholt. 2013. Bridging the genotype–phenotype gap: what does it take?. J. Physiol. 591:2055–2066.

Glen, C. M., M. L. Kemp, and E. O. Voit. 2019. Agent-based modeling of morphogenetic systems: advantages and challenges. PLoS Comput. Biol. 15:e1006577.

Grant, B. R., and P. R. Grant. 1993. Evolution of Darwin's finches caused by a rare climatic event. Proc. R. Soc. Lond. Ser. B Biol. Sci. 251:111–117.

Greenbury, S. F., I. G. Johnston, A. A. Louis, and S. E. Ahnert. 2014. A tractable genotype-phenotype map modelling the self-assembly of protein quaternary structure. J. R. Soc. Interface 11:20140249.

Goddard, M. E. 2001. The validity of genetic models underlying quantitative traits. Livest. Prod. Sci. 72:117–127.

Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. Proc. R. Soc. Lond. Ser. B Biol. Sci. 275:723–734.

Hansen, T. F. 2008. Macroevolutionary quantitative genetics? A comment on Polly (2008). Evol. Biol. 35:182–185.

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

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

Houle, D., D. R. Govindaraju, and S. Omholt. 2010. Phenomics: the next challenge. Nat. Rev. Genet. 11:855.

Houle, D., and K. Meyer. 2015. Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. J. Evol. Biol. 28:1542–1549.

Hine, E., K. McGuigan, and M. W. Blows. 2014. Evolutionary constraints in high-dimensional trait sets. Am. Nat. 184:119–131.

Hill, W. G., M. E. Goddard, and P. M. Visscher. 2008. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. 4:e1000008.

Hirashima, T., E. G. Rens, and R. M. H. Merks. 2017. Cellular Potts modeling of complex multicellular behaviors in tissue morphogenesis. Dev. Growth Differ. 59:329–339.

Jernvall, J., P. Kettunen, I. Karavanova, L. B. Martin, and I. Thesleff. 1994. Evidence for the role of the enamel knot as a control center in mammalian tooth cusp formation: non-dividing cells express growth stimulating Fgf-4 gene. Int. J. Dev. Biol. 38:463–469.

Jones, A. G., S. J. Arnold, and R. Burger. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution 58:1639–1654.

Jones, A. G., R. Bürger, and S. J. Arnold. 2014. Epistasis and natural selection shape the mutational architecture of complex traits. Nat. Commun. 5:3709.

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

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

McGuigan, K. 2006. Studying phenotypic evolution using multivariate quantitative genetics. Mol. Ecol. 15:883–896.

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

Meyer, K. 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B 8:815–821.

Morrissey, M. B. 2015. Evolutionary quantitative genetics of nonlinear developmental systems. Evolution 69:2050–2066.

Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the breeder's equation in observational studies of natural populations. J. Evol. Biol. 23:2277–2288.

Müller, G. B. 2007. Evo–devo: extending the evolutionary synthesis. Nat. Rev. Genet. 8:943.

Newman, S. A., and G. B. Müller. 2000. Epigenetic mechanisms of character origination. J. Exp. Zool. 288:304–317.

Oster, G., and P. Alberch. 1982. Evolution and bifurcation of developmental programs. Evolution 36:444–459.

Osterfield, M., C. A. Berg, and S. Y. Shvartsman. 2017. Epithelial patterning, morphogenesis, and evolution: Drosophila eggshell as a model. Dev. Cell 41:337–348.

Penna, A., D. Melo, S. Bernardi, M. I. Oyarzabal, and G. Marroig. 2017. The evolution of phenotypic integration: how directional selection reshapes covariation in mice. Evolution 71:2370–2380.

Pigliucci, M. 2006. Genetic variance–covariance matrices: a critique of the evolutionary quantitative genetics research program. Biol. Philos. 21:1–23.

Polly, P. D. 2008. Developmental dynamics and G-matrices: Can morphometric spaces be used to model phenotypic evolution?. Evol. Biol. 35:83.

Raspopovic, J., L. Marcon, L. Russo, and J. Sharpe. 2014. Modeling digits. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. Science 345:566–570.

Rodrigues, J. V., S. Bershtein, A. Li, E. R. Lozovsky, D. L. Hartl, and E. I. Shakhnovich. 2016. Biophysical principles predict fitness landscapes of drug resistance. Proc. Natl. Acad. Sci. USA 113:E1470–8.

Roff, D. 2000. The evolution of the G matrix: selection or drift?. Heredity 84:135–142.

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

Rolian, C. 2015. Tinkering with growth plates: a developmental simulation of limb bone evolution in hominoids. Pp. 139–165 *in* J. C. Boughner and Campbell Rolian (eds). Developmental approaches to human evolution. Wiley, Hoboken, NJ.

Rice, S. H. 2002. A general population genetic theory for the evolution of developmental interactions. Proc. Natl. Acad. Sci. USA 99:15518–15523.

———. 2004. Evolutionary theory: mathematical and conceptual foundations. Sinauer Associates, Sunderland, MA.

———. 2012. The place of development in mathematical evolutionary theory. J. Exp. Zool. B Mol. Dev. Evol. 318:480–488.

Salazar-Ciudad, I., J. Jernvall, and S. A. Newman. 2003. Mechanisms of pattern formation in development and evolution. Development 130:2027–2037.

Salazar-Ciudad, I. 2006. Developmental constraints vs. variational properties: how pattern formation can help to understand evolution and development. J. Exp. Zool. B Mol. Dev. Evol. 306:107–125.

Salazar-Ciudad, I., and J. Jernvall. 2010. A computational model of teeth and the developmental origins of morphological variation. Nature 464:583.

Salazar-Ciudad, I. 2010. Morphological evolution and embryonic developmental diversity in metazoa. Development 137:531.

Salazar-Ciudad, I., and M. Marín-Riera. 2013. Adaptive dynamics under development-based genotype–phenotype maps. Nature 497:361.

Stadler, B. M., P. F. Stadler, G. P. Wagner, and W. Fontana. 2001. The topology of the possible: formal spaces underlying patterns of evolutionary change. J. Theor. Biol. 213:241–274.

Urdy, S. 2012. On the evolution of morphogenetic models: mechano-chemical interactions and an integrated view of cell differentiation, growth, pattern formation and morphogenesis. Biol. Rev. 87:786–803.

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, D. W. Coltman, D. V. Mifsud, T. H. Clutton-Brock, and L. B. Kruuk. 2006. Environmental coupling of selection and heritability limits evolution. PLoS Biol. 4:e216.

Associate Editor: K. McGuigan
Handling Editor: D. W. Hall

Supporting Information

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

Table S1. List of developmental parameters in the tooth development model.
Table S2. Heritability estimated for all traits in generation 15 of all evolutionary scenarios in the core parameter set.
Figure. S1. List of optima used in the simulations.
Figure. S2. Results of simulations for core parameter set.
Figure S3. Changes in different parameters produce phenotypic effects of different magnitudes.
Figure S4. The method of repetitions eliminates errors due to stochasticity in a linear map.
Figure S5. Prediction bias was found in simulation starting with different initial conditions.
Figure S6. Prediction bias arises when the population is in a nonlinear region of the parameterphenotype map.
Figure S7. Distribution of additive genetic variance in the direction of selection and relationship with relative bias.
Figure S8. Bias can arise because variation does not exist beyond a trait value but the linear transformation in the breeder's equations infers that such variation should exist.
Figure S9. Including environmental effects negatively affects the performance of the breeder's equation.
Figure S10. The tooth model is not compatible with the assumptions of the multilinear model.
Figure S11. Multivariate prediction error for each generation of all 32 evolutionary simulations of the core parameter set.
Figure S12. The directions of selection in all 32 evolutionary simulations are different.
Video S1. Trait distribution in time for the 30 generations of the simulation shown in Fig. 2B.
Video S2. Trait distribution in time for the 30 generations of the simulation shown in Fig. 2C.