# Laplace's demon in biology: models of evolutionary prediction

Zachariah Gompert[1*], Samuel M. Flaxman[2], Jeffrey L. Feder[3],
Luis-Miguel Chevin[4], Patrik Nosil[4]

1. Department of Biology, Utah State University, Logan, UT 84322, USA
2. EBIO, University of Colorado Boulder, Boulder, CO 80309, USA
3. Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA
4. CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier, France

∗ Corresponding author; e-mail: zach.gompert@usu.edu.

*Running title*: Laplace's demon in biology

*Manuscript type*: Perspective

## *Author Contributions*

ZG and PN conceived and designed the study. ZG and L-MC analyzed the analytical models. ZG conducted the simulations. ZG and PN drafted the initial version of the manuscript and all authors contributed substantially to later versions of the manuscript.

## *Acknowledgments*

This research was funded by a grant from National Science Foundation (NSF) (DEB-1844941) to Z.G., from NSF (DEB-1638997) and the USDA-NIFA program (2015-67013-23289) to J.L.F., and from the European Research Council (EE-Dynamics 770826, https://erc.europa.eu/) to P.N. The support and resources from the Center for High Performance Computing at the University of Utah are also gratefully acknowledged.

## *Data Accessibility*

Simulation data supporting the results from this paper are available from `GitHub`, `https://github.com/zgompert/LaplaceDemonSims/`

# Laplace's demon in biology: models of evolutionary prediction

## Abstract

Our ability to predict natural phenomena can be limited by incomplete information. This issue is exemplified by 'Laplace's demon', an imaginary creature proposed in the 18th century, who knew everything about everything, and thus could predict the full nature of the universe forward or backward in time. Quantum mechanics, among other things, has cast doubt on the possibility of Laplace's demon in the full sense, but the idea still serves as a useful metaphor for thinking about the extent to which prediction is limited by incomplete information on deterministic processes versus random factors. Here we use simple analytical models and computer simulations to illustrate how data limits can be captured in a Bayesian framework, and how they influence our ability to predict evolution. We show how uncertainty in measurements of natural selection, or low predictability of external environmental factors affecting selection, can greatly reduce predictive power, often swamping the influence of intrinsic randomness caused by genetic drift. Thus, more accurate knowledge concerning the causes and action of natural selection is key to improving prediction. Fortunately, our analyses and simulations show quantitatively that reasonable improvements in data quantity and quality can meaningfully increase predictability.

*Keywords*: prediction, selection, genetic drift, environmental stochasticity, randomness, determinism, simulation models

## Introduction

Prediction is a critical component of the sciences, and a major theme in evolutionary biology. For example, instances of repeated, parallel evolution in response to similar environmental pressures can provide evidence of predictable evolution by natural selection (e.g., Campbell-Staton et al., 2020; Chaturvedi et al., 2018; Chevin et al., 2010b; Colosimo et al., 2005; Ferris et al., 2021; Haenel et al., 2019; Martin and Orgogozo, 2013; Rêgo et al., 2019; Soria-Carrasco et al., 2014; Stuart et al., 2022). In contrast, idiosyncratic outcomes can indicate constraints on the power of selection. At the core of this issue is the extent to which evolution is driven by deterministic versus random processes (Lenormand et al., 2009; Nosil et al., 2020; Sober, 1984). Resolving this question concerning the predictability of evolution is not only of great basic scientific interest but also has practical implications for forecasting organismal responses to natural and human-induced environmental change, the planning of plant and animal breeding programs, and the design of medicines and strategies to combat the spread of disease. In the end, all predictions are really probability distributions with breadth (e.g., variance) reflecting our uncertainty about the underlying processes. What we want to know here is how much one can shrink those distributions through gaining a better understanding of natural selection.

Here, we focus on the ability to forecast evolutionary dynamics, that is trait values or allele frequencies, over time. There are two main classes of explanation for limits in the ability of scientists to predict evolution (Nosil et al., 2020). First, predictability can be limited by random evolutionary processes, described as the 'random limits' hypothesis in Nosil et al. (2020). The key mechanisms underlying the random limits hypothesis are stochastic changes in allele frequency due to genetic drift and the random nature of mutation (Sober, 1984; Wright, 1931). Second, even evolution driven by natural selection–a deterministic factor conditional on the environment–could exhibit low predictability, due to measurement error, and limited data and models that in turn lead to poor understanding of selection and trait variation (the 'data limits' hypothesis hereafter) (Marques et al., 2018; Nosil et al., 2020, 2018; Reimchen, 1995). Under the data limits hypothesis, the assumption is that with better data and better analysis, evolution by natural selection can be better predicted. Limits to our understanding of evolution by selection can occur because environmental sources of selection, such as climatic conditions or predator abundance, fluctuate in ways that are difficult to predict (Chevin et al., 2022; Grant and Grant, 2002; Lenormand et al., 2009; Nosil et al., 2018). And even if the environment is constant or can be predicted, limited information about how environmental factors affect resource and trait distributions, and thus selection, can make evolution less predictable. We recently treated these topics in a publication that helped lay a conceptual foundation for studying the causes of variation in our ability to predict evolution (Nosil et al., 2020). This previous study was purely verbal and thus did not allow us to quantify the diverse causes of variable predictability, or to begin to resolve their relative contributions and extent to which they could be ameliorated. These are our main goals here.

The data limits hypothesis is exemplified by a thought experiment proposed by 18th century mathematician Pierre-Simon Laplace. This experiment, now called Laplace's demon, posits an imaginary creature with unlimited computational powers, who knew everything about everything (i.e., the position and velocity of all the particles in the universe), and thus could predict the full nature of the universe forward or backward in time from the laws of Newtonian physics. By analogy, we are interested in the extent to which prediction of evolution can be improved by increased knowledge stemming from data quantity and quality.

In contrast, random processes impose fundamental limits on prediction, even for an all-knowing intellect like Laplace's demon, as true randomness persists even after accounting for all causally relevant parameters affecting a process. Notably, the only proposed source in the universe of absolute randomness in this sense is the collapse of wave functions in quantum mechanics (most clearly in spontaneous wave-function collapse theories but also in a Bayesian context with branching in Everettian quantum mechanics) (Maudlin, 2019; Wallace, 2012). Because these quantum processes can directly impact mutation, mutation likely includes a random component in this absolute sense. For example, quantum transitions between keto and enol forms of guanine, which bond with cytosine or thymine, respectively, can cause mutations during replication (Carroll, 2020; Kimsey et al., 2015, 2018). Genetic drift, on the other hand, is random in a more limited sense (Sober, 1984). Variation in survival and fecundity that gives rise to drift presumably has some deterministic causes that could be known, but these causes are not the genotype or phenotype of the organism. Thus, change due to drift can be productively viewed as random with respect to genotype and phenotype, and thus as random with respect to evolution by natural selection.

Finally, selection depends on the interactions of genes, phenotypes, and the environment, and the environment often varies or fluctuates in space and time. There is a long and productive history of modeling components of environmental variability as random processes (e.g., Chevin, 2019; Gillespie, 1991; Lenormand et al., 2009; Ohta, 1972; Sæther and Engen, 2015). Nonetheless, most environmental variation is unlikely to be random in an absolute sense, but rather reflects some combination of limited knowledge (uncertainty), chaos, and complexity that can make aspects of environmental change indistinguishable from random processes (Sugihara et al., 1990). Indeed, even systems that are deterministic may be fundamentally unpredictable in their dynamics. Perhaps one of the best-known cases comes from physics, where chaotic dynamics arise in three-body gravitating systems. These systems can be so dependent on initial conditions that measurement precision would need to be at or below the Planck length (an extremely small scale at which quantum gravity becomes relevant) to allow for reliable long-term predictions (Liao, 2013, 2014). It is unclear how often or to what extent environmental variability falls into this category of fundamental unpredictability. However, in at least some cases, some prediction about future environments on some scales is possible, and such predictions can likely be improved (if not perfected) with additional data and better models or computational power. Our focus here is on this aspect of environmental variability, which we place within the data limits hypothesis (we return to the topic of environmental variability and randomness in the Discussion).

Herein, we begin by considering constraints on predicting evolution in analytical models that include drift and uncertainty in the strength of selection. We then use simulations to go beyond these analytical models and jointly consider multiple sources of uncertainty–environmental variation, genetic drift, selection, and genetic architecture–and how predictability varies over time. We do so via two case studies that are motivated by empirical systems, but our results do not correspond one-to-one to any particular biological system (i.e., our approach is not a model-fitting endeavor). Thus, the concepts invoked here are grounded in empirical reality but apply to a wide range of environmental factors, traits, and taxa (as we discuss below). We quantify predictability based on the variability among different probable evolutionary trajectories. The rationale for this choice is that evolution is more predictable when probable evolutionary trajectories (i.e., alternative, reasonable outcomes from replaying the tape of life) are more similar. We specifically consider the precision (reciprocal of the variance) in allele frequencies or trait values as a metric of predictability. In this context, higher precision denotes higher predictability (this avoids comparisons to some assumed, "actual" evolutionary trajectory, which is especially problematic when incorporating genetic drift).

We emphasize Bayesian inference and prediction throughout our quantitative treatment of predicting evolution. We do this because Bayesian methods provide a clear, probabilistic framework for quantifying and updating uncertainty (Figure 1A). Prior probability distributions describe initial uncertainty about relevant demographic or evolutionary parameters, such as the strength of selection. These prior distributions are then updated repeatedly by observations and experiments generating increasingly more informative posterior probability distributions, that is reduced uncertainty about model parameters. Posterior distributions are the basis for probabilistic predictions of evolutionary trajectories. Importantly, with Bayesian methods it is often possible to accurately estimate the effect of additional data on shrinking the uncertainty in posteriors, and thus on increasing predictive power for evolutionary trajectories. We take just such an approach with both the analytical models and case studies here. This information can guide decisions about which additional experiments or studies might be most useful for increasing scientists' ability to accurately predict evolution.

## Analytical arguments and case-study simulations

### *Analytical arguments*

We begin by showing quantitatively and concretely how random genetic drift and uncertainty in selection combine to place limits on our ability to predict evolution under simple conditions. We do this in an explicitly Bayesian framework. First, consider a single locus evolving for one generation by random drift and (uncertain) selection. We can approximate the expected allele frequency $p$ in the subsequent generation $t+1$ as $p_{t+1} \approx p + p(1-p)s$, where $s$ is the selection coefficient (defined from relative fitnesses as $w_{AA} = 1+2s$, $w_{Aa} = 1+s$ and $w_{aa} = 1$, assuming additivity) (Walsh and Lynch, 2018). Genetic drift and uncertainty in selection (i.e., imperfect knowledge of the value of $s$) each contribute additively to the variance around this expectation. Specifically, the variances caused by drift and uncertainty in selection are $\frac{p(1-p)}{2Ne}$ and $(p(1-p))^2 \operatorname{var}(s)$, respectively, with $N_e$ denoting the effective population size and $\operatorname{var}(s)$ the variance of a Bayesian (prior or posterior) distribution for $s$. Uncertain selection should impose a greater limit on predictability than drift when $\operatorname{var}(s)p(1-p) > \frac{1}{2Ne}$ (Chevin, 2019; Ohta, 1972) (this relationship was derived for the case where the variance in $s$ is caused by a randomly fluctuating environment, but applies equally well to the case here of uncertainty in a fixed value of $s$).

Whereas $N_e$ is an intrinsic property of the system, $\operatorname{var}(s)$ depends on data and a statistical model. And importantly, the magnitude of $\operatorname{var}(s)$ declines with more precise estimates of selection. We can make this explicit with an example. Assume the relative fitnesses of alternative homozygotes are estimated from a release-recapture experiment with equal initial release frequencies (we focus on an experiment without heterozygotes for mathematical simplicity). Let $y$ denote the number of $AA$ individuals recaptured out of $n$ recaptures. If we assume a binomial sampling distribution (likelihood) with a conjugate beta prior on the binomial parameter (denoted $\pi$ to avoid confusion with $p$, the allele frequency), the posterior distribution has a known form of $\Pr(\pi | y, n) \sim \text{beta}(\alpha = a_0 + y, \beta = b_0 + n - y)$. Here, $a_0$ and $b_0$ denote prior sample sizes, which could reflect past experiments or could be set to low values to denote prior ignorance; $y$ and $n$ depend on the sample size of the current experiment. This implies a variance for $s$ of $\frac{\alpha(\alpha+\beta-1)}{(\beta-2)(\beta-1)^2}\frac{1}{4}$ (Johnson et al., 1995). We illustrate the corresponding precisions in allele frequency (i.e., the predictability of allele frequency) over a range of values of effective population sizes (random limits) and sample sizes (data limits) in Figure 1B (also see Figure S1). Importantly, our results show that drift and uncertain selection can place comparable limits on the precision of predictions, and that increasing the experimental sample size (reducing the data limits) has a more pronounced effect when $N_e$ is not too small.

Second, consider the evolution of a quantitative trait by random drift and (uncertain) selection. The expected change in the mean trait value in one generation is given by the breeder's equation, $R = h^2 S$ where $S$ is the selection differential and $h^2$ is the trait's heritability $\left(\frac{\sigma_g^2}{\sigma_z^2}\right)$. Again, drift and uncertainty in selection contribute variances around this expectation of $\sigma_g^2/(2N_e)$ and $(h^2)^2\text{var}(S)$, respectively (Lande, 1976). Here, $\sigma_g^2$ is the additive genetic variance and $\text{var}(S)$ comes from a Bayesian probability distribution that depends on data and a model. We can again make the latter explicit with an example. Assume knowledge of $S$ comes from regressing fitness (or a component of fitness) on standardized trait values (e.g., Lande and Arnold, 1983). If we model the data with a normal sampling distribution (likelihood) and a normal prior on the standardized selection differential (and for simplicity assume that the residual variance is known), the posterior distribution on $S$ (here, the standardized selection differential which is equivalent to the standardized regression coefficient or selection gradient) is normal with $\text{var}(S) = \frac{\sigma^2}{n_0+n}$. Here, $\sigma^2$ is the residual trait variance, and $n_0$ and $n$ are prior and actual sample sizes for the experiment. We show the corresponding precisions for the predicted mean trait values with different effective population sizes (random limits) and sample sizes (data limits) in Figure 1C (also see Figure S1). As for the single locus case, effective population size and experimental sample size have similar effects on our ability to predict evolution, and once again only the latter is (partially) under the control of scientists.

Additional constraints on predictability occur when selection depends in an uncertain way on the environment or on an uncertain future environment. For example, if the selection differential $S$ is a linear function of the environment $x$, such that $S = a + bx$, then the total variance caused by fluctuating and poorly predicted selection alone is $\text{var}(h^2 S) = (h^2)^2[\text{var}(a) + E(b)^2\text{var}(x) + E(x)^2\text{var}(b)]$, which may be quite large if environmental fluctuations are substantial and poorly predicted (large $\text{var}(x)$), or if there is uncertainty in selection ($\text{var}(a)$) or in how it varies with the environment ($\text{var}(b)$) (a similar argument could be made for selection on a single locus). Uncertainty in genetic architecture further inflates this variance by making $h^2$ a random variable.

Additional limits to prediction occur when considering the genetic loci underlying a quantitative trait. In such cases, the genetic effects on traits are often only ascribed probabilistically, and causal variants are often not even known; instead we detect genetic variants in linkage disequilibrium with putative causal ones. Jointly considering these different sources of uncertainty is beyond the reach of the simple analytical arguments laid out in this paragraph. Instead, we now turn to simulation-based case studies to begin to explore the relative, quantitative importance of different sources of uncertainty on predicting evolution where these complexities can be modeled explicitly and jointly under realistic conditions. These simulations also allow us to examine the decline in predictability over time, that is from a few to tens of generations.

## *Overview and motivation of case studies*

The two case studies concern selection that varies in space and time, but for distinct reasons that should make the first case more predictable than the second. The case studies also introduce uncertainty in the genetic basis of the trait under selection, again with differences that should make the first case more predictable than the second. The first involves predator behavior and the evolution of anti-predator traits, motivated by long-term studies of the evolution of cryptic coloration in stick insects (Nosil et al., 2018). This scenario incorporates frequency-dependent selection, which was shown to increase the predictability of evolution under some conditions, even in an unpredictable environment (Chevin et al., 2022). The second involves climatic variation and the evolution of trophic traits, motivated by long-term studies of beak size evolution in Darwin's finches (Grant and Grant, 2002). We quantify uncertainty by computing the variance in evolutionary outcomes among replicate simulations under each scenario (e.g., with and without genetic drift, with and without uncertainty in natural selection, variable sample sizes, and genetic architectures). Thus, the results tackle another famous thought experiment posed by Stephen J. Gould on the extent to which repeatedly 'replaying the tape of life' would yield similar evolutionary outcomes (Gould, 1990). Our simulations show how data limits can strongly mediate the extent to which scientists can predict evolution, and how modest increases in the size or scale of experiments can meaningfully reduce these data limits, with the goal of motivating progress towards making evolution a more predictive science.

## *Case study 1: predation and frequency-dependent selection*

Predation affects most organisms and is a common and general source of natural selection (Meyer and Kassen, 2007; Reimchen, 1995; Svensson and Friberg, 2007). Predation can cause negative frequency-dependent selection (NFDS) when predators focus on more common prey types. In such cases, the fitness of a phenotype is expected to fluctuate because it depends on its frequency in the population and involves feedbacks with predator choice. This has been documented, for example, in cichlids, guppies, stickleback, and stick insects (Bolnick and Stutz, 2017; Hori, 1993; Hughes et al., 2013; Nosil et al., 2018; Olendorf et al., 2006). Moreover, evolutionary dynamics and equilibrium outcomes in predator-prey systems and under NFDS have received considerable theoretical attention (Abrams, 2000; Abrams et al., 2008; Chevin et al., 2022). Such systems might represent cases where evolution is easier to predict, especially when selection is primarily a function of the current state (phenotype frequency) of a focal population. Thus, we first consider predictability in the context of predation and NFDS.

We used data concerning NFDS on color pattern in the stick insect *Timema cristinae* to help guide our choice of parameters for our illustrative model (Nosil et al., 2018). We emphasize that our goal is not to fit a model for the *T. cristinae* system, but rather to ensure that we use biologically relevant parameters and data-based levels of uncertainty. *Timema cristinae* exhibits striped and unstriped color-pattern morphs. Striped morphs are more cryptic on one host plant (*Adenostoma*), and thus generally favored on this host by selection from visual predators such as birds and lizards (Nosil, 2004; Nosil and Crespi, 2006; Sandoval, 1994a,b). However, experimental and observational data show that striped individuals are less fit when they become very common compared to when they are rarer, thus demonstrating NFDS (Nosil et al., 2018).

We quantified the effects of limited knowledge of selection and genetic drift on the ability to predict evolutionary trajectories under NFDS (Figure 2A). We did this through three comparisons. First, for uncertainty in selection, we are referring to what happens when one predicts the future course of evolutionary change when one has exact knowledge of, for example, selection coefficients for traits under NFDS at different frequencies versus less exact estimates, reflected in a Bayesian probability distribution for these values. In the case of uncertainty in selection, we eliminate the randomness due to drift by simulating an infinite population size. Second, in the drift analysis, we consider having exact information for all the relevant variables affecting selection, but simulate finite-sized populations. Thus, the course of selection is determined by our exact knowledge of how selection would work, plus a degree of randomness introduced by genetic drift in a finite population size. Finally, in drift plus uncertainty in selection simulations, the above uncertainty in selection is incorporated into the runs along with finite population size, introducing the effects of random drift.

The logic then is that by comparing the three types of simulations, the relative effects of drift versus uncertainty in selection can be resolved with respect to their effects on predicting evolutionary change. This can further help identify cases where one can, through increased sampling and rigor in experimental design, decrease the variance in the Bayesian probability distributions around these point estimates to see how much this effort, which is at least empirically tractable to some degree, can improve our ability to forecast evolution.

We assumed the existence of two morphs or phenotypes, denoted '$A$' (e.g., striped) and '$B$' (e.g., unstriped). We further assumed that the $A$ phenotype was advantageous when its population frequency was not too high (less than a critical value), with relative fitness values of $w_A$ and $w_B$ for the $A$ and $B$ morphs, and with $w_A > w_B$. However, when $A$ was very common (above a critical value) we assumed selection favored the $B$ morph with $w_B^* > w_A^*$ (here $w^*$ denotes relative fitness when morph $A$ is very common). We based the magnitude of uncertainty in selection on the sample sizes and results from a *T. cristinae* release-recapture field experiment. In this experiment, 500 *T. cristinae* stick insects were released in two treatments, one with an initial stripe frequency of 20%, and one with an initial stripe frequency of 80%. Survivors were then collected to estimate the strength of selection for or against stripe under the two different treatments (see

Nosil et al., 2018). In both treatments, survival was measured on the host plant *Adenostoma*. The initial stripe frequencies for the two treatments were chosen based on our expectations that they would be on different sides of the critical point where NFDS would alternately favor the striped versus non-striped form.

For our current purposes, we obtained Bayesian estimates of survival probabilities assuming a binomial likelihood for recapture in each treatment and with uninformative beta priors on the recapture probabilities (the beta priors set both shape parameters, i.e., prior sample sizes $a_0$ and $b_0$, to 0). This resulted in closed-form posterior distributions for absolute fitness (survival probabilities) of beta(52, 48) (mean = 0.52) and beta(62, 338) (mean = 0.155) for the $A$ (striped) and $B$ (unstriped) morphs, respectively (these values come from the experiment with stripe at an initial frequency of 20%) (Nosil et al., 2018). The means of these distributions were used to calculate relative fitness values in the case where selection was assumed known (1.0 and 0.3, respectively). We knew less about fitness when the striped ($A$) morph was above the critical point (in the original experiment the 80% stripe treatment resulted in nearly equal recapture rates for both morphs). To approximate this, we halved the sample sizes, then flipped the recapture rates and shifted the counts slightly to construct probability distributions for this case: beta(31+10, 169-10) (mean = 0.186) and beta(26-2.5, 24+2.5) (mean = 0.49) for striped ($A$) and unstriped ($B$). We further assumed the critical value where selection switched between favoring $A$ versus favoring $B$ was known to fall between 0.7 and 0.9 (frequency of $A$) and took on any value within that range with equal probability (i.e., we assumed a uniform probability distribution constrained by the previous experiment). We used 0.85 for cases of known selection. Thus, for each simulation of evolution, we either used these point estimates (selection known) or sampled fitness values from these four beta distributions (i.e., posterior distributions), and a value for the critical point from U(0.7, 0.9)(selection uncertain). As a comparison, an additional set of simulations were conducted to assess predictability with weak, but uncertain selection. For this, survival values were sampled from beta(50.5, 49.5) and beta(49.5, 50.5) for the favored and less fit morph, respectively (here the expected relative fitness values are 1.0 and 0.98, respectively).

We considered trait heritabilities of 0.8 or 1 (consistent with Comeault et al., 2016; Lindtke et al., 2017), and an initial frequency for morph $A$ of 50%. We incorporated genetic drift by binomial sampling, such that $p_{t+1} \sim \text{binomial}(E[p_{t+1}], 2N_e)$. Here, $p_{t+1}$ is the frequency of morph $A$ in the next generation, $E[p_t + 1]$ is the expected frequency given the current frequency ($p_t$), selection and the trait heritability, and $N_e$ is the effective population size, which we set to 110 diploid individuals. This value comes from an empirical estimate of the variance effective population size in a *T. cristinae* population (population code FHA, 43°30.958′ N, 119°48.050′ W), which is based on genome-wide allele frequency change (Nosil et al., 2018). We conducted 100 simulations of evolution incorporating uncertainty in selection, genetic drift, or both, and measured the effect of each factor on predictability based on the precision (median across 100 generations) in evolutionary trajectories. In most cases, the model rapidly fell into a stable, equilibrium oscillation, consistent with patterns of change associated with NFDS documented in nature (Figure 3 and Nosil et al., 2018) (such stable oscillations are a specific outcome of using a step function for NFDS; compare to Chevin et al., 2022). These simulations were written in `R` and are available via GitHub (`https://github.com/zgompert/LaplaceDemonSims/`).

Uncertainty in selection and genetic drift both caused variability in evolutionary trajectories affecting predictability (i.e., increased variability among probable, simulated evolutionary trajectories) (Figure 3). Predictability was highest for the first few generations, but then quickly declined and remained relatively constant from about five to 100 generations (i.e., for the remainder of the simulated time; Figure S2). We thus focus on the mean predictability (precision) when comparing sources of uncertainty.

Importantly, uncertainty in selection did not lead to erroneous qualitative predictions, as a stable oscillation in stripe frequency was always predicted (this is expected given the step function assumed for NFDS, unlike in Chevin et al., 2022). However, limited knowledge of selection did cause considerable quantitative uncertainty in the evolutionary trajectory (i.e., in the pattern and characteristics of the oscillations), and this was greater than the uncertainty caused by genetic drift (Figure 3F). For example, with $h^2$ = 0.8, the median precision across generations was two times higher with genetic drift and known selection (median precision = 277.7) than with no genetic drift and uncertainty in selection (median precision = 128.4) (higher values indicate higher precision in predicting evolution). Moreover, adding drift to the latter case (i.e., both drift and uncertainty in selection) did not markedly lower precision (median precision = 118.3). Similar results were observed for $h^2 = 1$ (median precision = 192.2, 96.5, and 92.3, respectively). These results, with associated measures of variability and uncertainty, are depicted in detail in Figures S3 and S4.

These effects of uncertainty in selection versus drift on predictability were not restricted to a scenario of strong selection. For example, even with weak selection (1% difference in expected survival probabilities), drift had less of an effect on predictability than did uncertainty in selection (median precision across generations of 36.2 versus 5.8) (Figure S4). This perhaps counter-intuitive result arose because weak NFDS combined with uncertainty in selection resulted in transient directional selection (rather than strongly fluctuating) for or against either phenotype being among the set of probable selection models. In such cases, directional selection, even if transient, was consistent enough to fix one or the other morph, resulting in a high variance (low precision) in evolutionary trajectories.

Whereas genetic drift is a property of the effective size of any finite population, predictability can be increased by better knowledge of selection. To examine the effects of such knowledge, we conducted additional simulations assuming two or five times larger sample sizes for the release-recapture experiment (and consequent reduction in uncertainty in selection) and a decrease of the range of possible values where the transition occurred from selection favoring morph $A$ to $B$ (at a frequency of $A$ of 0.8-0.9). With strong selection and $h^2 = 0.8$, doubling the sample size essentially doubled the predictability. Thus, with a two times larger sample size the predictability with uncertainty in selection (precision = 242.4) was almost as high as the level observed for just drift (precision = 277.7; precision with both uncertainty in selection and drift = 214.6), suggesting quite meaningful increases in predictability would be possible with only a reasonable increase in effort (Figure 3D-F). Similar results were observed with $h^2 = 1$ or with weak selection (Figures S3D-F and S4D-F), but increasing the sample size five-fold had little additional effect on improving predictability (Figure S5).

## *Case study 2: climatic variability and trophic evolution*

Temporal variation in climatic conditions can cause the direction and magnitude of selection on a trait to vary in time. Climate and weather can themselves be agents of selection or can indirectly cause selection by affecting resource availability, predators, competitors, etc., which then act as agents of selection. Temporally fluctuating selection caused by climatic variability is likely to be general, as it has been documented in numerous species (Bergland et al., 2014; Busoms et al., 2018; Reimchen and Nosil, 2004; Rudman et al., 2022; Siepielski et al., 2009, 2017; de Villemereuil et al., 2020). Perhaps the best-known example comes from Darwin's finches, where variation in rainfall on Daphne Major has been shown to affect the relative abundances of small versus large seeds, which in turn exerts selection on beak size in *Geospiza fortis* (Boag and Grant, 1981; Grant and Grant, 2014). Such cases are of particular interest for predicting evolution because they include both uncertainty in climatic conditions and uncertainty in selection. Selection has been described as unpredictable in the case of *G. fortis* (Grant and Grant, 2002), not because we do not understand selection (i.e., selection is known to be exerted by seed size distributions), but rather because we cannot predict climatic fluctuations or how these affect the seed size distribution. Because of this dual complexity and the potential generality of such conditions, we consider climatic variability and trophic evolution as our second case study. Here, we also consider the effects of uncertainty in the detailed genetic basis of the selected trait. As for the case of NFDS, we used general empirical knowledge (in this case from the finch system) to parameterize our illustrative models, but without the aim of fitting specific models to the finch system. Rather, we use this scenario to understand the extent to which evolution can be predicted without directly measuring selection in all generations, but instead relying on known (current generation) or projected (future generations) environmental data. This is important to assess because environmental data can be simpler to gather than measurements of selection, so the prediction process could be simplified by first estimating the relationship between selection and the environment (environmental sensitivity of selection) over a few generations, and then combining this with environmental projections to predict evolutionary dynamics (Chevin et al., 2010a).

We either assumed climatic conditions were known, or incorporated uncertainty in such conditions, as would occur when trying to project future environments based on past time series (Figure 2B). An observed annual rainfall time series on Daphne Major, which included data from 1973 to 2012, was used for cases where we assumed climate conditions were known (Grant and Grant, 2014). Uncertainty in climatic conditions was modeled by sampling from this time series with replacement (i.e., here we do not adopt a Bayesian approach). Alternative, (simple) model-based approaches to account for climate uncertainty failed to capture the salient features of the actual time series, especially the extreme variability (i.e., the extreme values relative to, e.g., a normal or exponential distribution; results not shown).

We assumed that, conditional on climatic conditions, the nature of selection was either known or uncertain. We considered two sources of uncertainty: the link between climatic conditions and resource, and between resource and selection differential (Figure 2B). For resource abundance, we first used the data from Daphne Major to infer the relationship between rainfall and resource (the relative abundance of small seeds versus large seeds) (Grant and Grant, 2014). To do this, we regressed resource abundance on five year cumulative rainfall in a model that included linear and quadratic effects (five-year rainfall performed better, i.e., higher $r^2$, than sums over fewer years). Point estimates for these parameters were: intercept = $1.57 \times 10^{-1}$, rain linear = $6.5 \times 10^{-4}$, and rain quadratic = $-6.36 \times 10^{-7}$ ($r^2 = 0.498$, $P = 0.011$). These were used when selection was assumed known. In cases where selection was not known, we incorporated uncertainty in this relationship by sampling regression coefficients from Gaussian distributions centered on the point estimates, and with standard deviations equal to the standard errors of the coefficients (SEs: intercept = $1.47 \times 10^{-1}$, rain linear = $1.84 \times 10^{-1}$, rain quadratic = $3.13 \times 10^{-7}$). From a Bayesian perspective, using the standard errors in this way is analogous to placing flat priors on the regression coefficients and treating the residual variance as known (Congdon, 2007). In terms of the resource-selection link, with selection known we assumed the selection differential $S = a + bx$, with $a$ = 0.18, $b$ = -0.81 and $x$ = resource abundance. These values assume resource abundance has been centered (mean = 0) and are approximately equal to the extremes of selection observed in *G. fortis*. When selection was uncertain, values for $a$ and $b$ were sampled from normal distributions with standard deviations of 0.1 and means of 0.18 and -0.81, respectively. Thus, we treat these distributions as Bayesian probability distributions for uncertain $a$ and $b$; uncertainty is encoded by the non-zero standard deviations, which we set to 0.1 (we chose these values in the absence of pertinent information but consider alternative values that reflect an increase in data below).

We further assumed the trophic trait was moderately heritable ($h^2 \approx 0.5$), with $\sim$27 causal variants with a normal effect size distribution (inspired by genetic mapping results from Chaves et al., 2016) (Figure 2B). More specifically, we created 1415 loci potentially affecting the trait. We assigned Bayesian probabilities of effect/association to each locus sampling from uniform distributions: U(0.1,0.6) (15 loci), U(0.05,0.1) (100 loci), U(0.01,0.05) (300 loci) and U(0.001,0.01) (1000 loci). Thus, a small number of loci had high probabilities of association and many had much lower probabilities of association (the expectation is 27.2 associated loci). Phenotypic effects were assigned to the loci by sampling from a standard normal distribution, and allele frequencies for each bi-allelic locus were drawn from a beta distribution, beta(0.6,0.6) (this gives a U-shaped distribution of allele frequencies). Then, for each simulation of evolution, the subset of causal variants was determined by sampling loci according to their probabilities of association (analogous to posterior inclusion probabilities from Bayesian polygenic models for genomic prediction, see, e.g., Gompert, 2021; Zhou et al., 2013). Sampled loci were assigned their respective effect sizes, and other loci were assigned an effect size of 0.

We modeled the evolution of expected trait values (akin to the genome-estimated breeding values) by calculating the effect of phenotypic selection on expected allele frequency change across loci. The selection differential denoted the expected phenotypic change. We then approximated the selection on each allele as $s_i = w_i - 1 \approx b_i \frac{S}{\sigma_z^2}$, where $b_i$ is the average excess of locus $i$, $\sigma_z^2$ denotes the phenotypic variance and was set to 2 to give a heritability of about 0.5 (the exact value varied based on the specific causal loci), and $\frac{S}{\sigma_z^2}$ equals the selection gradient $\beta$ (Gompert, 2021; Kimura and Crow, 1978; Walsh and Lynch, 2018). This approximation assumes the trait remains normally distributed, effect sizes are small, and causal loci are unlinked. The expected change in allele frequency is then given by $\Delta p_i = p_i s_i$ (Kimura and Crow, 1978). Genetic drift was incorporated by binomial sampling around this expectation. We used an effective population size of 60 (Grant and Grant, 1992), ran 100 simulations for each set of conditions, and ran each simulation for 35 generations. We tracked allele frequencies and expected trait values (genome-estimated breeding values). These simulations were written in `C++` with the Gnu Scientific Library (code available via GitHub; `https://github.com/zgompert/LaplaceDemonSims/`) (Galassi et al., 2003). Predictability was measured as the precision in expected trait values or allele frequencies.

Uncertainty in selection, climatic conditions, and genetic architecture, as well as genetic drift, all acted to limit the predictability of the evolutionary time series for the expected value of the trophic trait (i.e., the mean genome-estimated breeding value) (Figure 4), and predictability (i.e., precision across replicates) again declined over time (Figure S6). Of these factors, genetic drift had the smallest effect, resulting in the highest predictability for models that included drift alone (median precision 383.6; Figure 4A). The other factors caused much larger reductions in predictability (other factors in combination or isolation gave median precisions ranging between 4.2 and 25.4; Figure 4).

Likewise, genetic drift often had only a modest effect on predictions for the evolutionary trajectories of individual alleles (Figure 5, S7). The magnitude of the relative effects of drift and uncertainty in selection depended in part on the probability that a locus was associated with (i.e., caused variation in) the trophic trait. In cases where an association between a locus and the trait was uncertain but likely (i.e., inclusion probability of $\sim$10-50%), drift and uncertainty in selection had similar effects on precision (i.e., predictability) (e.g., Figure 5A-H). The main factor causing variability in trajectories was instead uncertainty in genetic architecture. In particular, distinct sets of trajectories were evident for simulations where the locus was versus was not assumed to be associated with the trait (also compare Figures 5 and S5). This effect was reduced with lower probabilities of association (Figure S9). Indeed, the main exception to the pattern of drift having less effect than uncertainty in selection occurred for loci with very low probabilities of association (e.g., Figure 5M-P). In such cases, drift was the bigger cause of poor predictability, as almost all probable trajectories included no selection on these loci (because they were not actually associated with the trait), and thus uncertainty in selection was of minimal relevance for evolutionary dynamics.

As with the first case study, we conducted additional simulations to determine the effect of increased information about selection on our ability to predict evolution. Here, we focused on the effect of increased information about the link between the environment and selection. Specifically, we considered a three-fold increase in the number of experiments used to determine the relationship between resources and selection differentials, which would decrease the standard deviations for the Bayesian probability distributions on the intercept ($a$) and slope ($b$) from 0.1 to 0.058. This increase in the number of experiments increased the predictability for the expected value of the trophic trait with the precision increasing from 25.4 to 188.2, again suggesting that a feasible increase in effort can result in quite notable gains in predictability (Figure S8). Higher predictability remained (though to a lesser extent) when uncertainty in selection was combined with drift or uncertainty in weather, but was less evident when uncertainty in genetic architecture was included.

## *Summary considerations and moving beyond these case studies*

We think several practical messages emerge from our analytical arguments and case studies. First, we can improve our ability to predict evolution by obtaining better estimates of selection within an environment; our results suggest that sample sizes on the order of the effective population size (or a bit bigger) represent a reasonable (and often feasible) goal. Here, we refer to the variance effective population size, which is often much smaller (e.g., $\sim \frac{1}{10}$) than the local census population size (e.g., Frankham, 1995; Gompert et al., 2021; Waples, 2022). Beyond that effort, our results suggest that measuring selection in additional environments will be more productive for improving predictions of evolution. This of course adds considerable complexity and work, but perhaps starting with relevant environmental extremes (e.g., wettest versus driest habitats, low versus high trait frequencies, etc.) would allow one to at least place bounds on the extent to which environmental variation is associated with variation in selection in a given system.

Second, and unsurprisingly, predicting evolution at the genetic level was easier for a highly heritable trait (stripe) than for a less heritable polygenic trait (beak size). This is likely to hold in general, and very large genetic mapping studies will likely be necessary for precise predictions of allele frequency change for polygenic traits. Often it will be more profitable and more reasonable to make predictions at the level of expected breeding values (polygenic scores), either using pedigrees and classic quantitative genetic methods or genetic-marker based genomic prediction methods. Importantly, we did not consider cases where the trait value itself is affected by the environment (i.e., where there is plasticity) (e.g., Crozier et al., 2011); in such cases, predicting evolutionary change (albeit perhaps not phenotypic patterns) will likely require larger genetic mapping or quantitative genetic studies spanning multiple environments.

Lastly, the case studies we considered were based on empirical work that combines long-term monitoring of populations with selection experiments and genetic mapping or traditional quantitative genetics. We think that such combined approaches are critical for making and testing evolutionary predictions (e.g., Wade and Kalisz, 1990). Several other systems have similar features, such as Soay sheep on St. Kilda island (Ashraf et al., 2021; Clutton-Brock and Pemberton, 2004; Johnston et al., 2013), great tits (Garant et al., 2004; Gienapp et al., 2019; Husby et al., 2011), collared flycatchers (Merilä et al., 2001), threespine stickleback fish (Marques et al., 2018; Reimchen and Nosil, 2002, 2004; Reimchen, 1995), and Edith's Checkerspot butterflies (Ehrlich et al., 2004, 1975; Parmesan and Singer, 2022). We think that these and other long-term studies (reviewed in, e.g., Clutton-Brock and Sheldon, 2010) can provide further tests of the predictability of evolution, and that our work here can generate hypotheses for how to improve predictability in these systems.

# Discussion

The analytical arguments and case studies we considered illustrate how data limitations and uncertainty in selection, including that caused by low environmental predictability, can substantially affect the predictability of phenotypic and genetic evolutionary change, much beyond the influence of random evolutionary processes, in this case genetic drift. Moreover, our results are likely conservative, as we only modeled a subset of possible uncertainties in selection and considered modest population sizes, where the potential for genetic drift is substantial. Despite this, uncertainty in selection generally had much more pronounced effects on reducing predictability than did genetic drift, especially in the two case studies. This is consistent with theoretical expectations that uncertainty in selection should dominate when $\text{var}(s)p(1-p) > \frac{1}{2Ne}$ (Chevin, 2019; Ohta, 1972).

These results suggest that progress towards predicting evolution can be made with empirical and analytical effort, because the largest limit does not come from intrinsic properties of the population. For example, our results show how larger sample sizes can increase the predictability of evolution; in case study 1, doubling the sample size made the Bayesian probability distribution for selection more precise and thereby doubled our ability to predict evolutionary dynamics (Figure 3). Similar gains in predictability were observed in simulations with better knowledge of additional factors affecting evolutionary dynamics, such as trait genetic architectures, weather and climate, and the environmental causes and ecological consequences of selection (Figure 4). The only notable exception to this pattern comes from the analytical models, which show that increasing sample sizes has minimal effect when the effective population size is very low (i.e., below 50) (Figure 1). Consequently, while perfect prediction as envisioned by Laplace's demon is forever out of reach due to the action of largely random processes of genetic drift and mutation, our results suggest that even modest improvements in data quantity and quality can meaningfully increase predictability (i.e., we will not get an omniscient demon, but at least we can have a serviceable, low-level imp that throws loaded dice).

Still, improving the predictability of evolution is not without its challenges. First, collecting sufficient data for meaningful prediction is not a trivial task. For example, gathering adequate time-series data may require multiple scientists' entire careers, and very large sample sizes can be required for genetic mapping of traits, especially in the presence of gene-gene or gene-environment interactions (Wang et al., 2005; Wei et al., 2014). With that said, evolution of quantitative traits can be successfully predicted without knowing the effects of individual genes using the animal model and pedigrees or kinship matrixes inferred from genetic marker data (e.g., Bonnet et al., 2022; Charmantier et al., 2014; Meuwissen et al., 2001; Walsh and Lynch, 2018). Still, this methods require extensive data and even with detailed (e.g., pedigree-based) observational work, it can often be difficult to determine whether or to what extent changes in phenotypes (or breeding values) reflect selection or only random drift (e.g., Hadfield et al., 2010; Pigeon et al., 2016).

Second, pleiotropy could confound predictions, especially at the genetic level, if mutations affecting a favored trait have additional effects on other traits and genetic effects or selection on these other traits has not been measured (e.g., Gromko, 1995; Saltz et al., 2017). Third, the precision of predictions cannot likely be increased indefinitely by collecting more data (e.g., Figure 1). Many ecologically relevant environmental variables (such as temperature) include sufficient noise (Halley, 1996; Ruokolainen et al., 2009; Vasseur and Yodzis, 2004) that, from a practical perspective, they can only be predicted in a probabilistic sense. The envelope breadth of evolutionary predictions in response to such environmental variables will be bounded below by these practical limits. Improving underlying physical models may increase environmental predictability to some extent, but it cannot realistically be expected to fully explain environmental variability.

Moreover, our ability to predict evolution can be further compromised when systems exhibit extreme sensitivity to initial conditions, for example leading to chaotic dynamics (e.g., Costantino et al., 1997). Evolutionary theory has raised the possibility of chaos in evolutionary dynamics (Doebeli and Ispolatov, 2014; Gavrilets and Hastings, 1995), including in a changing environment (Chevin et al., 2022; Rego-Costa et al., 2018), but its actual existence and prevalence remains to be investigated empirically. Another situation where initial conditions are critical to the outcome is when evolution occurs on rugged adaptive landscapes caused by strong epistatic interactions (Kauffman et al., 1993; Nosil et al., 2020), such that randomly occurring mutations or slight differences in standing genetic variation may lead populations into different, irreversible evolutionary paths (e.g., Park et al., 2022). Finally, organisms may perceive environmental fluctuations as random, regardless of whether they really are, and this may select for specific biological mechanisms such as bet hedging, making evolution difficult to predict even when the causes of environmental variation can be deciphered (e.g., Crean and Marshall, 2009; Simons, 2014).

Another critical consideration is that temporal scale is important for assessing our ability to predict evolution. As our simulations showed, the variance of predictions increases with time when selection estimates are imprecise or the environment is only partly predictable. The time scale of predictability is determined by the patterns of environmental fluctuations, the sensitivity to initial conditions (i.e., the Lyapunov exponent of chaotic dynamics), and the genetic architecture of responses to selection. We focused on short-term predictability of evolutionary or ecological dynamics, where selection acted on standing genetic variation. We expect evolution from standing variation to be more predictable than evolution from new mutations (e.g., Blount et al., 2008; Colosimo et al., 2005; Haenel et al., 2019; Rêgo et al., 2019) (but see, e.g., Chan et al., 2010). This is because evolution from standing variation removes a major source of randomness, that is mutation (Barrett and Schluter, 2008; Lenormand et al., 2009). Moreover, the evolutionary fate of new mutations, even those favored by selection, is greatly impacted by genetic drift (Kimura, 1983). And in general, rare and difficult to predict events contribute more to evolution on longer time-scales (e.g., Blount et al., 2018; Gould, 1990). Likewise, many-to-one mapping of form to function suggests that multiple genotype or trait combinations can result in functionally equivalent phenotypes, further limiting evolution's long-term predictability (e.g., Wainwright et al., 2005). Thus, predictions in evolutionary biology may always have a limited time horizon.

In conclusion, our analyses and simulations show that data limits can profoundly curtail our ability to predict evolution. We show that optimizing data collection towards increasing the precision of selection estimates or their dependence on the environment can meaningfully improve our ability to predict evolution, at least on shorter time-scales, but we temper this by noting that this is a non-trivial undertaking and that fundamental limits to predictability will remain. Perfect precision will be impossible or at least impractical, but also often unnecessary. Chaos may increase uncertainty, and to the best of our knowledge some processes are either truly random (mutation) or at least random with respect to genotype and phenotype (genetic drift), and others will be treated as effectively random at least beyond their time window of predictability (environmental stochasticity). We cannot have the perfectly predictive model of causal determinism in evolution conjured by Laplace's demon, but modest increases in data can still lead to quantifiably more robust predictions. As such increases could benefit basic and applied science, this is what we are after.

# References

Abrams, P. A., 2000. The evolution of predator-prey interactions: theory and evidence. Annual Review of Ecology and Systematics 31:79–105.

Abrams, P. A., C. Rueffler, and G. Kim, 2008. Determinants of the strength of disruptive and/or divergent selection arising from resource competition. Evolution 62:1571–1586.

Ashraf, B., D. C. Hunter, C. Bérénos, P. A. Ellis, S. E. Johnston, J. G. Pilkington, J. M. Pemberton, and J. Slate, 2021. Genomic prediction in the wild: A case study in Soay sheep. Molecular Ecology .

Barrett, R. D. and D. Schluter, 2008. Adaptation from standing genetic variation. Trends in Ecology & Evolution 23:38–44.

Bergland, A. O., E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov, 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. PLoS Genetics 10:e1004775.

Blount, Z. D., C. Z. Borland, and R. E. Lenski, 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proceedings of the National Academy of Sciences 105:7899–7906.

Blount, Z. D., R. E. Lenski, and J. B. Losos, 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362.

Boag, P. T. and P. R. Grant, 1981. Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galapagos. Science 214:82–85.

Bolnick, D. I. and W. E. Stutz, 2017. Frequency dependence limits divergent evolution by favouring rare immigrants over residents. Nature 546:285–288.

Bonnet, T., M. B. Morrissey, P. de Villemereuil, S. C. Alberts, P. Arcese, L. D. Bailey, S. Boutin, P. Brekke, L. J. Brent, G. Camenisch, et al., 2022. Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals. Science 376:1012–1016.

Busoms, S., P. Paajanen, S. Marburger, S. Bray, X.-Y. Huang, C. Poschenrieder, L. Yant, and D. E. Salt, 2018. Fluctuating selection on migrant adaptive sodium transporter alleles in coastal *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences 115:E12443–E12452.

Campbell-Staton, S. C., K. M. Winchell, N. C. Rochette, J. Fredette, I. Maayan, R. M. Schweizer, and J. Catchen, 2020. Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. Nature Ecology & Evolution 4:652–658.

Carroll, S. B., 2020. A Series of Fortunate Events: Chance and the Making of the Planet, Life, and You. Princeton University Press.

Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, et al., 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302–305.

Charmantier, A., D. Garant, and L. E. Kruuk, 2014. Quantitative genetics in the wild. OUP Oxford.

Chaturvedi, S., L. K. Lucas, C. C. Nice, J. A. Fordyce, M. L. Forister, and Z. Gompert, 2018. The predictability of genomic changes underlying a recent host shift in Melissa blue butterflies. Molecular Ecology 27:2651–2666.

Chaves, J. A., E. A. Cooper, A. P. Hendry, J. Podos, L. F. De León, J. A. Raeymaekers, W. O. MacMillan, and J. A. C. Uy, 2016. Genomic variation at the tips of the adaptive radiation of Darwin’s finches. Molecular Ecology 25:5282–5295.

Chevin, L.-M., 2019. Selective sweep at a QTL in a randomly fluctuating environment. Genetics 213:987–1005.

Chevin, L.-M., Z. Gompert, and P. Nosil, 2022. Frequency dependence and the predictability of evolution in a changing environment. Evolution Letters 6:21–33.

Chevin, L.-M., R. Lande, and G. M. Mace, 2010a. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biology 8:e1000357.

Chevin, L.-M., G. Martin, and T. Lenormand, 2010b. Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. Evolution 64:3213–3231.

Clutton-Brock, T. and B. C. Sheldon, 2010. Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. Trends in Ecology & Evolution 25:562–573.

Clutton-Brock, T. H. and J. M. Pemberton, 2004. Soay Sheep: Dynamics and Selection in an Island Population. Cambridge University Press.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley, 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933.

Comeault, A. A., C. F. Carvalho, S. Dennis, V. Soria-Carrasco, and P. Nosil, 2016. Color phenotypes are under similar genetic control in two distantly related species of *Timema* stick insect. Evolution 70:1283–1296.

Congdon, P., 2007. Bayesian Statistical Modelling. John Wiley & Sons.

Costantino, R. F., R. Desharnais, J. M. Cushing, and B. Dennis, 1997. Chaotic dynamics in an insect population. Science 275:389–391.

Crean, A. J. and D. J. Marshall, 2009. Coping with environmental uncertainty: dynamic bet hedging as a maternal effect. Philosophical Transactions of the Royal Society B: Biological Sciences 364:1087–1096.

Crozier, L. G., M. D. Scheuerell, and R. W. Zabel, 2011. Using time series analysis to characterize evolutionary and plastic responses to environmental change: a case study of a shift toward earlier migration date in sockeye salmon. The American Naturalist 178:755–773.

Doebeli, M. and I. Ispolatov, 2014. Chaos and unpredictability in evolution. Evolution 68:1365–1373.

Ehrlich, P. R., I. Hanski, et al., 2004. On the wings of checkerspots: a model system for population biology. Oxford University Press.

Ehrlich, P. R., R. R. White, M. C. Singer, S. W. McKechnie, and L. E. Gilbert, 1975. Checkerspot butterflies: A historical perspective: Long-term studies of *Euphydryas* butterffies have revealed much about the biology of natural populations. Science 188:221–228.

Ferris, K. G., A. S. Chavez, T. A. Suzuki, E. J. Beckman, M. Phifer-Rixey, K. Bi, and M. W. Nachman, 2021. The genomics of rapid climatic adaptation and parallel evolution in North American house mice. PLoS Genetics 17:1–25. URL `https://doi.org/10.1371/journal.pgen.1009495`.

Frankham, R., 1995. Effective population size/adult population size ratios in wildlife: a review. Genetics Research 66:95–107.

Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Rossi, 2003. GNU scientific library. Reference Manual. Edition 1.4, for GSL Version 1.4 .

Garant, D., L. E. Kruuk, R. H. McCleery, and B. C. Sheldon, 2004. Evolution in a changing environment: a case study with great tit fledging mass. The American Naturalist 164:E115–E129.

Gavrilets, S. and A. Hastings, 1995. Intermittency and transient chaos from simple frequency-dependent selection. Proceedings of the Royal Society of London. Series B: Biological Sciences 261:233–238.

Gienapp, P., M. P. Calus, V. N. Laine, and M. E. Visser, 2019. Genomic selection on breeding time in a wild bird population. Evolution Letters 3:142–151.

Gillespie, J. H., 1991. The Causes of Molecular Evolution. Oxford University Press On Demand.

Gompert, Z., 2021. A population-genomic approach for estimating selection on polygenic traits in heterogeneous environments. Molecular Ecology Resources 21:1529–1546.

Gompert, Z., A. Springer, M. Brady, S. Chaturvedi, and L. K. Lucas, 2021. Genomic time-series data show that gene flow maintains high genetic diversity despite substantial genetic drift in a butterfly species. Molecular Ecology 30:4991–5008.

Gould, S. J., 1990. Wonderful life: the Burgess Shale and the nature of history. WW Norton & Company.

Grant, P. R. and B. R. Grant, 1992. Demography and the genetically effective sizes of two populations of darwin's finches. Ecology 73:766–784.

———, 2002. Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707–711.

———, 2014. 40 years of evolution: Darwin's finches on Daphne Major Island. Princeton University Press.

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

Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. Kruuk, 2010. The misuse of BLUP in ecology and evolution. The American Naturalist 175:116–125.

Haenel, Q., M. Roesti, D. Moser, A. D. MacColl, and D. Berner, 2019. Predictable genome-wide sorting of standing genetic variation during parallel adaptation to basic versus acidic environments in stickleback fish. Evolution Letters 3:28–42.

Halley, J. M., 1996. Ecology, evolution and 1f-noise. Trends in Ecology & Evolution 11:33–37.

Hori, M., 1993. Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. Science 260:216–219.

Hughes, K. A., A. E. Houde, A. C. Price, and F. H. Rodd, 2013. Mating advantage for rare males in wild guppy populations. Nature 503:108–110.

Husby, A., M. E. Visser, and L. E. Kruuk, 2011. Speeding up microevolution: the effects of increasing temperature on selection and genetic variance in a wild bird population. PLoS biology 9:e1000585.

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1995. Continuous univariate distributions, volume 2, vol. 289. John wiley & sons.

Johnston, S. E., J. Gratten, C. Berenos, J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, and J. Slate, 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature 502:93–95.

Kauffman, S. A. et al., 1993. The origins of order: Self-organization and selection in evolution. Oxford University Press, USA.

Kimsey, I. J., K. Petzold, B. Sathyamoorthy, Z. W. Stein, and H. M. Al-Hashimi, 2015. Visualizing transient watson–crick-like mispairs in dna and rna duplexes. Nature 519:315–320.

Kimsey, I. J., E. S. Szymanski, W. J. Zahurancik, A. Shakya, Y. Xue, C.-C. Chu, B. Sathyamoorthy, Z. Suo, and H. M. Al-Hashimi, 2018. Dynamic basis for dg• dt misincorporation via tautomerization and ionization. Nature 554:195–201.

Kimura, M., 1983. The Neutral Theory of Molecular. Cambridge University Press.

Kimura, M. and J. F. Crow, 1978. Effect of overall phenotypic selection on genetic change at individual loci. Proceedings of the National Academy of Sciences 75:6168–6171.

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

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

Lenormand, T., D. Roze, and F. Rousset, 2009. Stochasticity in evolution. Trends in Ecology & Evolution 24:157–165.

Liao, S., 2013. On the numerical simulation of propagation of micro-level inherent uncertainty for chaotic dynamic systems. Chaos, Solitons & Fractals 47:1–12.

———, 2014. Physical limit of prediction for chaotic motion of three-body problem. Communications in Nonlinear Science and Numerical Simulation 19:601–616.

Lindtke, D., K. Lucek, V. Soria-Carrasco, R. Villoutreix, T. E. Farkas, R. Riesch, S. R. Dennis, Z. Gompert, and P. Nosil, 2017. Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. Molecular Ecology 26:6189–6205.

Marques, D. A., F. C. Jones, F. Di Palma, D. M. Kingsley, and T. E. Reimchen, 2018. Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. Nature Ecology & Evolution 2:1128–1138.

Martin, A. and V. Orgogozo, 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67:1235–1250.

Maudlin, T., 2019. Philosophy of Physics: Quantum Theory, vol. 33. Princeton University Press.

Merilä, J., L. Kruuk, and B. Sheldon, 2001. Cryptic evolution in a wild bird population. Nature 412:76–79.

Meuwissen, T. H., B. J. Hayes, and M. Goddard, 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Meyer, J. R. and R. Kassen, 2007. The effects of competition and predation on diversification in a model adaptive radiation. Nature 446:432–435.

Nosil, P., 2004. Reproductive isolation caused by visual predation on migrants between divergent environments. Proceedings of the Royal Society of London. Series B: Biological Sciences 271:1521–1528.

Nosil, P. and B. J. Crespi, 2006. Experimental evidence that predation promotes divergence in adaptive radiation. Proceedings of the National Academy of Sciences 103:9090–9095.

Nosil, P., S. M. Flaxman, J. L. Feder, and Z. Gompert, 2020. Increasing our ability to predict contemporary evolution. Nature Communications 11:1–6.

Nosil, P., R. Villoutreix, C. F. de Carvalho, T. E. Farkas, V. Soria-Carrasco, J. L. Feder, B. J. Crespi, and Z. Gompert, 2018. Natural selection and the predictability of evolution in *Timema* stick insects. Science 359:765–770.

Ohta, T., 1972. Fixation probability of a mutant influenced by random fluctuation of selection intensity. Genetics Research 19:33–38.

Olendorf, R., F. H. Rodd, D. Punzalan, A. E. Houde, C. Hurt, D. N. Reznick, and K. A. Hughes, 2006. Frequency-dependent survival in natural guppy populations. Nature 441:633–636.

Park, Y., B. P. Metzger, and J. W. Thornton, 2022. Epistatic drift causes gradual decay of predictability in protein evolution. Science 376:823–830.

Parmesan, C. and M. C. Singer, 2022. Mosaics of climatic stress across species' ranges: tradeoffs cause adaptive evolution to limits of climatic tolerance. Philosophical Transactions of the Royal Society B 377:20210003.

Pigeon, G., M. Festa-Bianchet, D. W. Coltman, and F. Pelletier, 2016. Intense selective hunting leads to artificial evolution in horn size. Evolutionary Applications 9:521–530.

Rêgo, A., F. J. Messina, and Z. Gompert, 2019. Dynamics of genomic change during evolutionary rescue in the seed beetle *Callosobruchus maculatus*. Molecular Ecology 28:2136–2154.

Rego-Costa, A., F. Debarre, and L.-M. Chevin, 2018. Chaos and the (un)predictability of evolution in a changing environment. Evolution 72:375–385.

Reimchen, T. and P. Nosil, 2002. Temporal variation in divergent selection on spine number in threespine stickleback. Evolution 56:2472–2483.

———, 2004. Variable predation regimes predict the evolution of sexual dimorphism in a population of threespine stickleback. Evolution 58:1274–1281.

Reimchen, T. E., 1995. Predator-induced cyclical changes in lateral plate frequencies of *Gasterosteus*. Behaviour Pp. 1079–1094.

Rudman, S. M., S. I. Greenblum, S. Rajpurohit, N. J. Betancourt, J. Hanna, S. Tilk, T. Yokoyama, D. A. Petrov, and P. Schmidt, 2022. Direct observation of adaptive tracking on ecological time scales in ¡i¿drosophila¡/i¿. Science 375:eabj7484. URL `https://www.science.org/doi/abs/10.1126/science.abj7484`.

Ruokolainen, L., A. Lindén, V. Kaitala, and M. S. Fowler, 2009. Ecological and evolutionary dynamics under coloured environmental variation. Trends in Ecology & Evolution 24:555–563.

Sæther, B.-E. and S. Engen, 2015. The concept of fitness in fluctuating environments. Trends in Ecology & Evolution 30:273–281.

Saltz, J. B., F. C. Hessel, and M. W. Kelly, 2017. Trait correlations in the genomics era. Trends in ecology & evolution 32:279–290.

Sandoval, C. P., 1994a. Differential visual predation on morphs of *Timema cristinae* (Phasmatodeae: Timemidae) and its consequences for host range. Biological Journal of the Linnean Society 52:341–356.

———, 1994b. The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking-stick *Timema cristinae*. Evolution 48:1866–1879.

Siepielski, A. M., J. D. DiBattista, and S. M. Carlson, 2009. It's about time: the temporal dynamics of phenotypic selection in the wild. Ecology Letters 12:1261–1276.

Siepielski, A. M., M. B. Morrissey, M. Buoro, S. M. Carlson, C. M. Caruso, S. M. Clegg, T. Coulson, J. DiBattista, K. M. Gotanda, C. D. Francis, et al., 2017. Precipitation drives global variation in natural selection. Science 355:959–962.

Simons, A., 2014. Playing smart vs. playing safe: the joint expression of phenotypic plasticity and potential bet hedging across and within thermal environments. Journal of Evolutionary Biology 27:1047–1056.

Sober, E., 1984. The Nature of Selection: Evolutionary Theory in Philosophical Focus. MIT Press.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, and P. Nosil, 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738–742.

Stuart, K. C., W. B. Sherwin, J. J. Austin, M. Bateson, M. Eens, M. C. Brandley, and L. A. Rollins, 2022. Historical museum samples enable the examination of divergent and parallel evolution during invasion. Molecular Ecology .

Sugihara, G., B. T. Grenfell, and R. M. May, 1990. Distinguishing error from chaos in ecological time series. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 330:235–251.

Svensson, E. I. and M. Friberg, 2007. Selective predation on wing morphology in sympatric damselflies. The American Naturalist 170:101–112.

Vasseur, D. A. and P. Yodzis, 2004. The color of environmental noise. Ecology 85:1146–1152.

de Villemereuil, P., A. Charmantier, D. Arlt, P. Bize, P. Brekke, L. Brouwer, A. Cockburn, S. D. Côté, F. S. Dobson, S. R. Evans, et al., 2020. Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. Proceedings of the National Academy of Sciences 117:31969–31978.

Wade, M. J. and S. Kalisz, 1990. The causes of natural selection. Evolution 44:1947–1955.

Wainwright, P. C., M. E. Alfaro, D. I. Bolnick, and C. D. Hulsey, 2005. Many-to-one mapping of form to function: a general principle in organismal design? Integrative and Comparative Biology 45:256–262.

Wallace, D., 2012. The Emergent Multiverse: Quantum Theory According to the Everett Interpretation. Oxford University Press.

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

Wang, W. Y., B. J. Barratt, D. G. Clayton, and J. A. Todd, 2005. Genome-wide association studies: theoretical and practical concerns. Nature Reviews Genetics 6:109–118.

Waples, R. S., 2022. What is Ne, anyway? Journal of Heredity URL `https://doi.org/10.1093/jhered/esac023`. Esac023.

Wei, W.-H., G. Hemani, and C. S. Haley, 2014. Detecting epistasis in human complex traits. Nature Reviews Genetics 15:722–733.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16:97.

Zhou, X., P. Carbonetto, and M. Stephens, 2013. Polygenic modeling with bayesian sparse linear mixed models. PLoS Genetics 9:e1003264.

# Figures

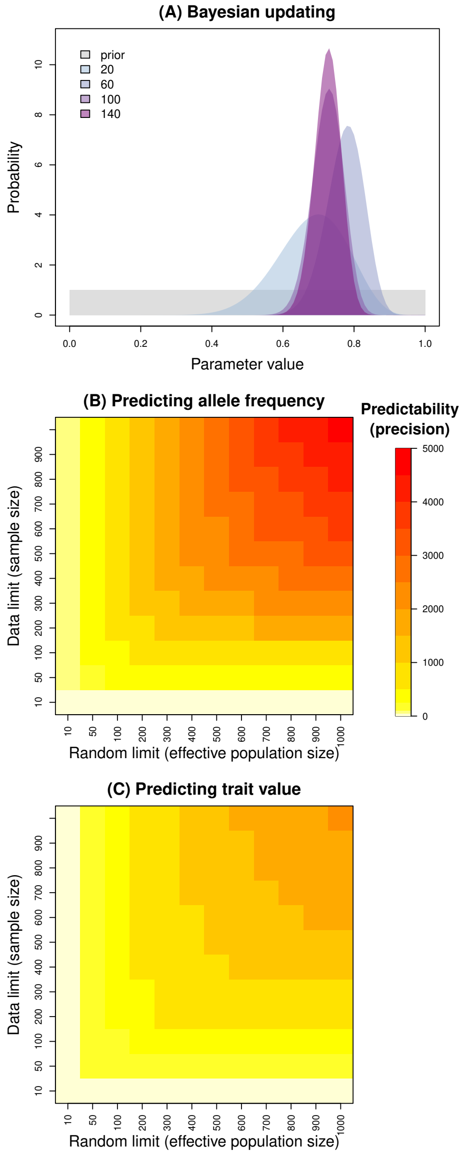


Figure 1: Conceptual overview of Bayesian updating of uncertainty and summary of predictability from simple analytical models. Panel (A) illustrates the effect of increasing the sample size of an experiment on a Bayesian posterior distribution (the example here assumes a binomial likelihood and conjugate beta prior distribution). Colors denote posteriors based on different sample sizes. Heat maps in panels (B) and (C) show the precision (reciprocal of the variance) in the allele frequency (B) or mean trait value (C) following after one generation of evolution by drift and selection. Darker colors indicate higher precision (predictability). Results are shown over a range of effective population sizes and experimental sample sizes with selection coefficients (B) or selection differentials (C) of 0.1. The initial allele frequency in (B) is 0.5 and the trait heritability in (C) is 0.5. Results with other levels of genetic variation are shown in Figure S1.

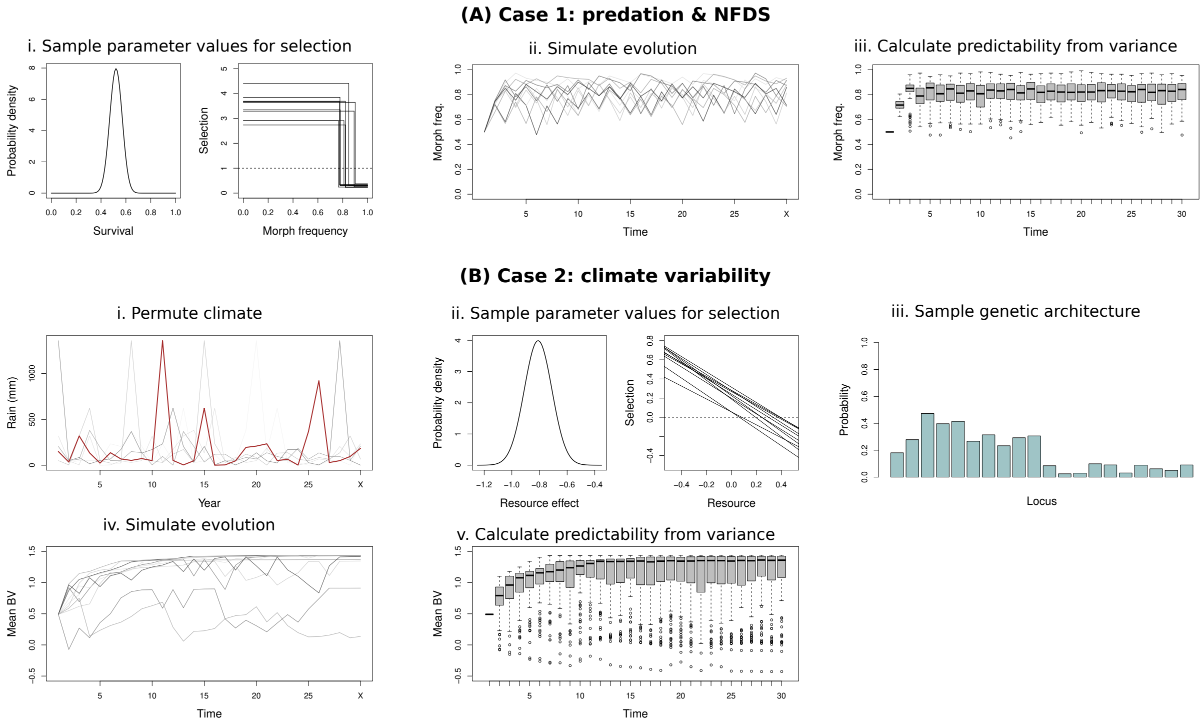

(A) Case 1: predation & NFDS
i. Sample parameter values for selection
Probability density
Survival
Selection
Morph frequency
ii. Simulate evolution
Morph freq.
Time
iii. Calculate predictability from variance
Morph freq.
Time
(B) Case 2: climate variability
i. Permute climate
Rain (mm)
Year
ii. Sample parameter values for selection
Probability density
Resource effect
Selection
Resource
iii. Sample genetic architecture
Probability
Locus
iv. Simulate evolution
Mean BV
Time
v. Calculate predictability from variance
Mean BV
Time

Figure 2: Illustrative overview of the case studies and simulations for considering the predictability of evolution. In this figure, we provide an overview of our procedure for simulating evolution in for case studies one (A) and two (B). In case study 1 (A), we conduct the following three in silico simulation steps. (i) Survival probabilities for selection functions are sampled from the appropriate probability distributions (e.g., Bayesian posterior distributions). Alternative, sampled NFDS functions are shown. (ii) Given the relevant parameter values, evolution by selection or selection and drift is simulated multiple times (lines denote alternative possible evolutionary trajectories). (iii) Based on these trajectories, we calculate the variance in morph frequencies in each generation. We convert these values to a metric of predictability by computing the precision (1/variance) and summarizing this statistic across time steps. In case study 2 (B), we conduct the following five in silico simulation steps. (i) A climate (annual rainfall) time series is first resampled with replacement (the red line denotes the original time series and the gray lines show examples of re-sampled data). Given the climate time series, resource abundance is determined by first sampling parameters for a linear regression that relates climate to resource. (ii) Second, regression parameters for the linear model for the selection differential are sampled from the appropriate probability distributions (e.g., Bayesian posterior distributions). Alternative, sampled linear functions for selection differentials are shown. (iii) Next, causal genetic loci for the selected trait are sampled based on their probabilities of association, which are high for a small number of loci, but low for most. (iv) Given the relevant parameter values, evolution by selection or selection and drift is simulated multiple times, and (v) based on these trajectories, we calculate the variance in breeding values or allele frequencies in each generation, which we convert to precision as described above.

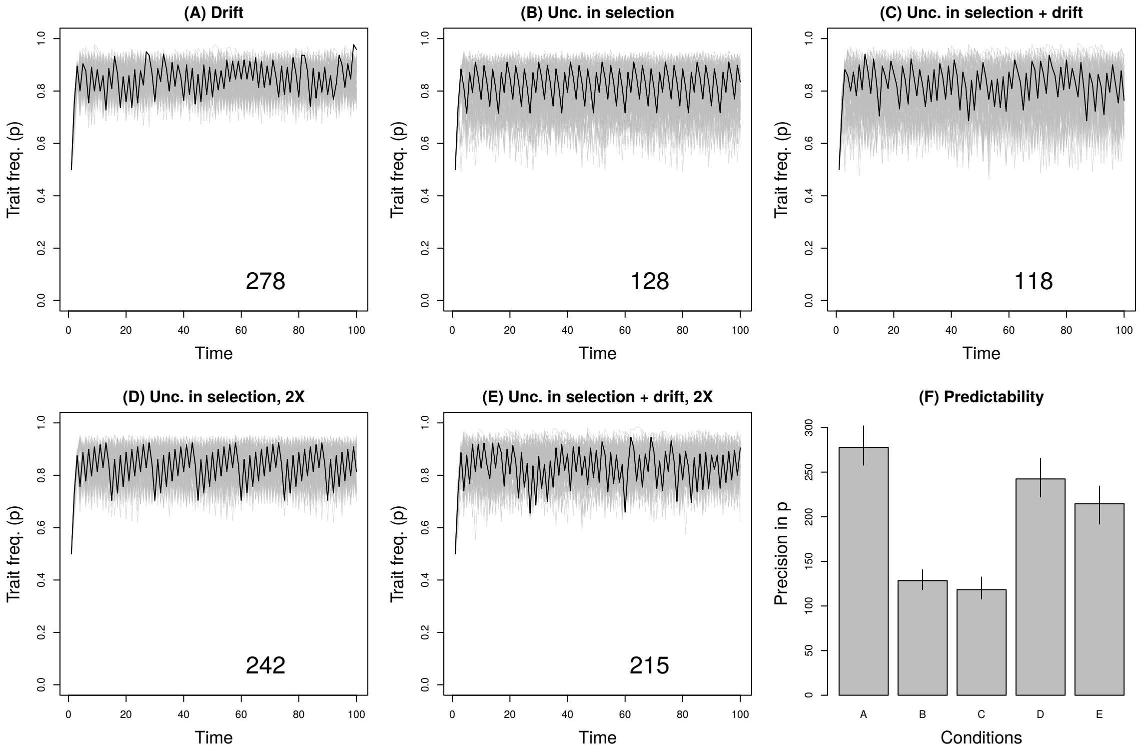


Figure 3: Case study 1 with negative frequency-dependent selection (NFDS). Gray lines show 100 probable evolutionary trajectories through time (in generations) under NFDS. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), and genetic drift and uncertainty in selection (C). Panels (D) and (E) show the effect of uncertainty in selection and drift plus uncertainty in selection given a hypothetical, larger experiment. In each case, an arbitrary trajectory is shown in black to make the dynamics more clear and to emphasize the fact that one of the possible trajectories would be realized in a given biological system. Panel (F) summarizes the predictability of evolution across each of the five conditions shown (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in trait (morph) frequencies (denoted p) across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Median precision for each condition is also reported in each panel as a numerical inset. Results are shown for a heritability of 0.8. Similar results with heritability of 1.0 are shown in Figure S3.

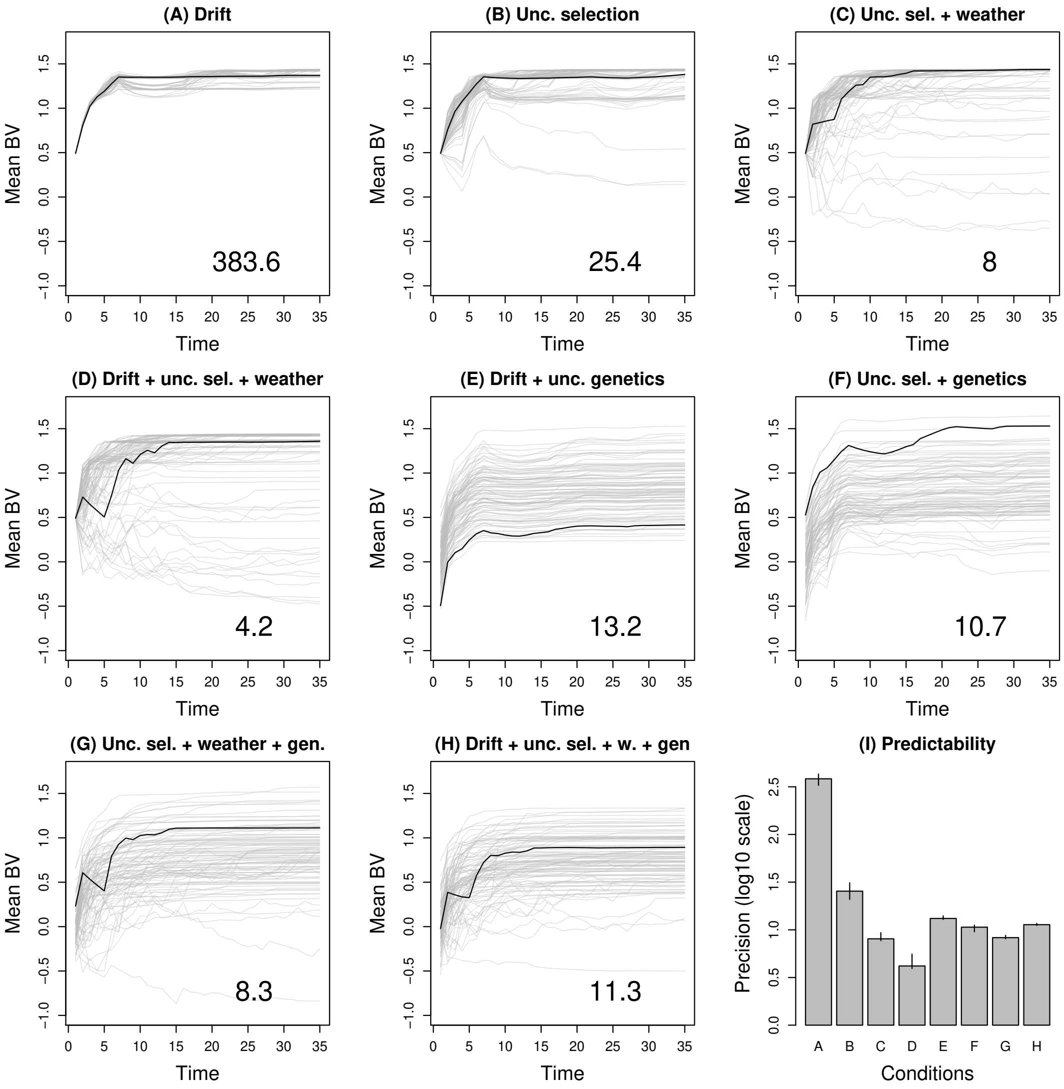


Figure 4: Case study 2 with climatic variation and breeding values. Gray lines show 100 probable evolutionary trajectories of the expected breeding value (BV) of a trophic trait under temporally-fluctuating selection. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), uncertainty in climatic conditions and selection (sel.) (C), and genetic drift and uncertainty in climatic conditions and selection (E). Panels (E-H) combine these effects with uncertainty in genetics (gen.) (i.e., in which a subset of 1415 genetic loci affect the trophic trait). In each case, an arbitrary trajectory is shown in black to make the dynamics more clear and to emphasize the fact that one of the possible trajectories would be realized in a given biological system. Panel (I) summarizes the predictability of evolution across each of the eight conditions shown (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in the mean (expected) BV across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Median precision for each condition is also reported in each panel as a numerical inset.

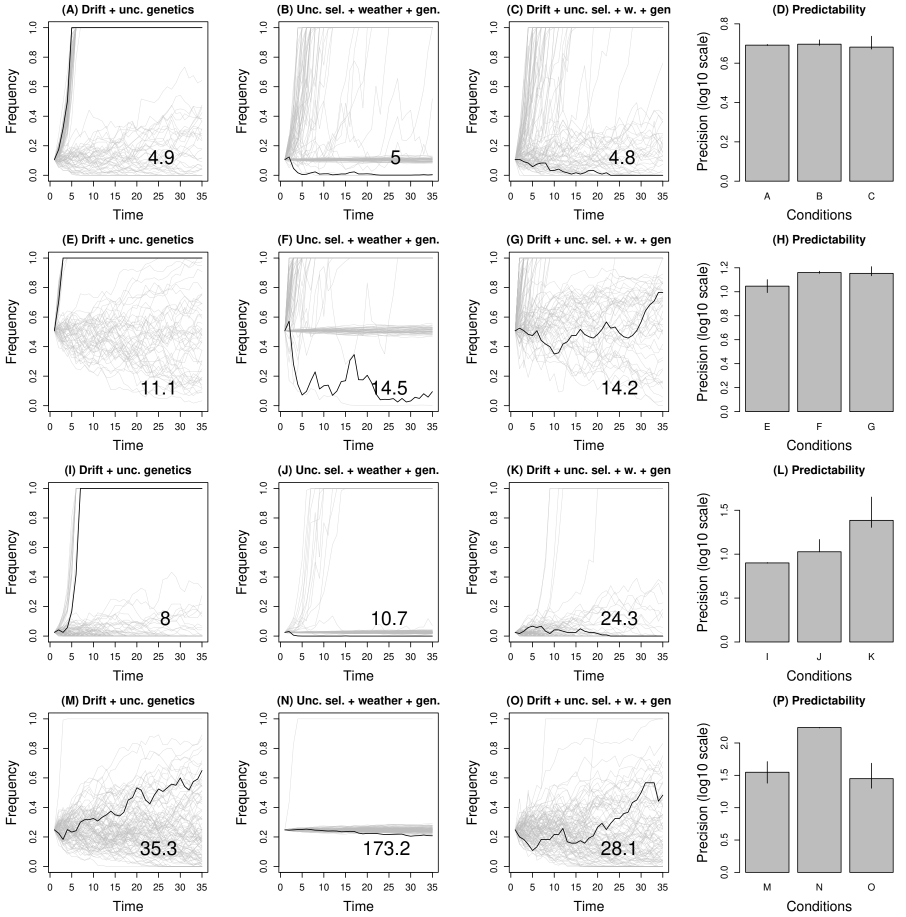


Figure 5: Case study 2 with climatic variation and consideration of genetic architecture and individual loci. Gray lines show 100 probable evolutionary trajectories of allele frequencies for loci potentially under temporally-fluctuating selection. Each row in this multi-panel figure corresponds with one of the 1415 genetic loci. The probabilities of trait association (e.g., Bayesian posterior inclusions probabilities or PIPs) are: 0.557 (panels A-D), 0.554 (panels E-H), 0.098 (panels I-L) and 0.007 (panels M-P). All results shown assume uncertainty in the trait genetic architecture. Additional sources of variability in each panel are genetic drift (A, E, I, M), uncertainty (Unc.) in climatic conditions (w.) and selection (B, F, J, N), and genetic drift and uncertainty in climatic conditions and selection (sel.) (C, G, K, 0). In each case, an arbitrary trajectory is shown in black to make the dynamics more clear and to emphasize the fact that one of the possible trajectories would be realized in a given biological system. Panels (D, H, L, P) summarize the predictability of evolution across each of the three conditions shown for each locus (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in the allele frequency across probable trajectories. Bars show the median (across generations) and vertical lines denote the 25th and 75th percentiles. Median precision for each condition is also reported in each panel as a numerical inset. Compare to Figure S7, which shows the same loci with genetic architecture known.

# Supplemental Figures

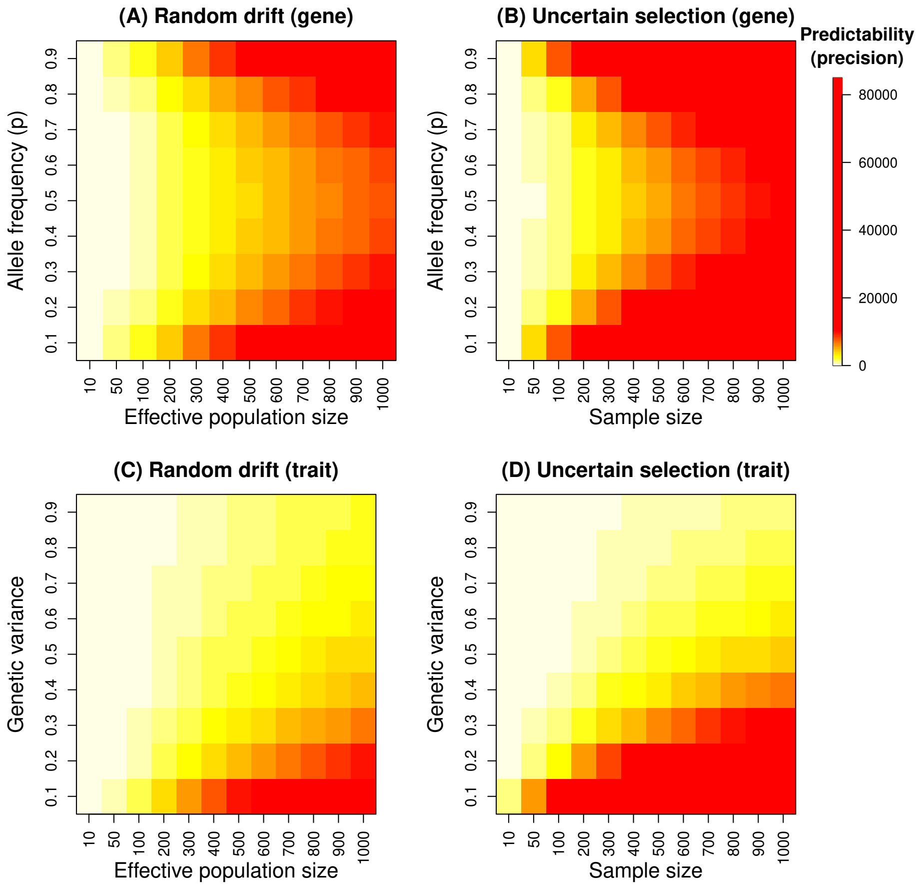


Figure S1: Predictability of evolution by random genetic drift (panels A and C) or uncertain selection (panels B and D) for an allele frequency (A and B) or mean trait value (C and D). Heat maps show the precision (reciprocal of the variance) in the allele frequency (A and B) or mean trait value (C and D) following after one generation of evolution by drift and selection. Darker colors indicate higher precision (predictability). Results are shown over a range of effective population sizes or experimental sample sizes and a range of initial allele frequencies or additive genetic variances (we assume a trait variance of 1 so these are identical to the trait heritabilities). Panels (A) and (C) assume no selection, whereas the selection coefficient or differential in (B) and (D) are both 0.1.

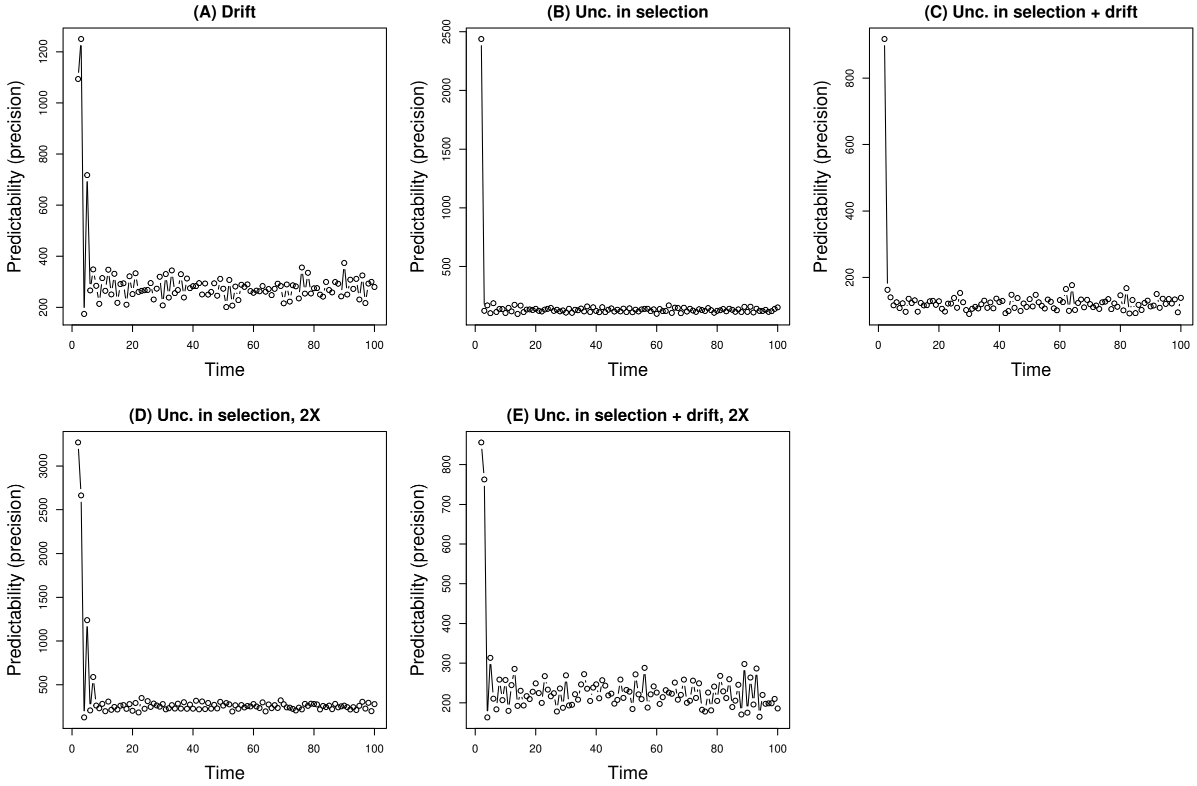


Figure S2: Precision over time in case study 1 with negative frequency-dependent selection (NFDS). Points connected by lines show the precision (reciprocal of the variance) in evolutionary trajectories over time under NFDS. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), and genetic drift and uncertainty in selection (C). Panels (D) and (E) show the effect of uncertainty in selection and drift plus uncertainty in selection given a hypothetical, larger experiment.

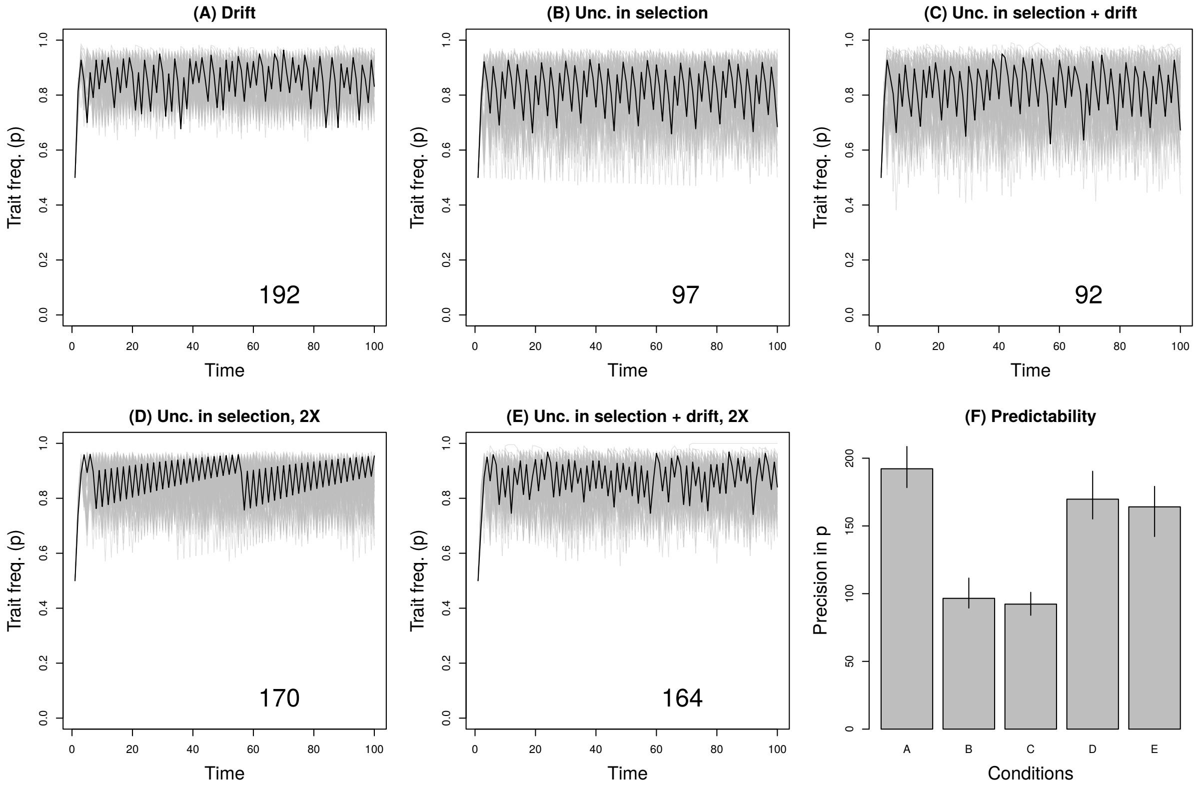


Figure S3: Case study 1 with NFDS, $h^2$ = 1.0. Gray lines show 100 probable evolutionary trajectories under NFDS. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), and genetic drift and uncertainty in selection (C). Panels (D) and (E) show the effect of uncertainty in selection and drift plus uncertainty in selection given a hypothetical, larger experiment. In each case, an arbitrary trajectory is shown in black to make the dynamics more clear and to emphasize the fact that only one trajectory would be realized in a given biological system. Panel (F) summarizes the predictability of evolution across each of the five conditions shown (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in trait (morph) frequencies (denoted p) across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Results are shown for a heritability of 1.0.

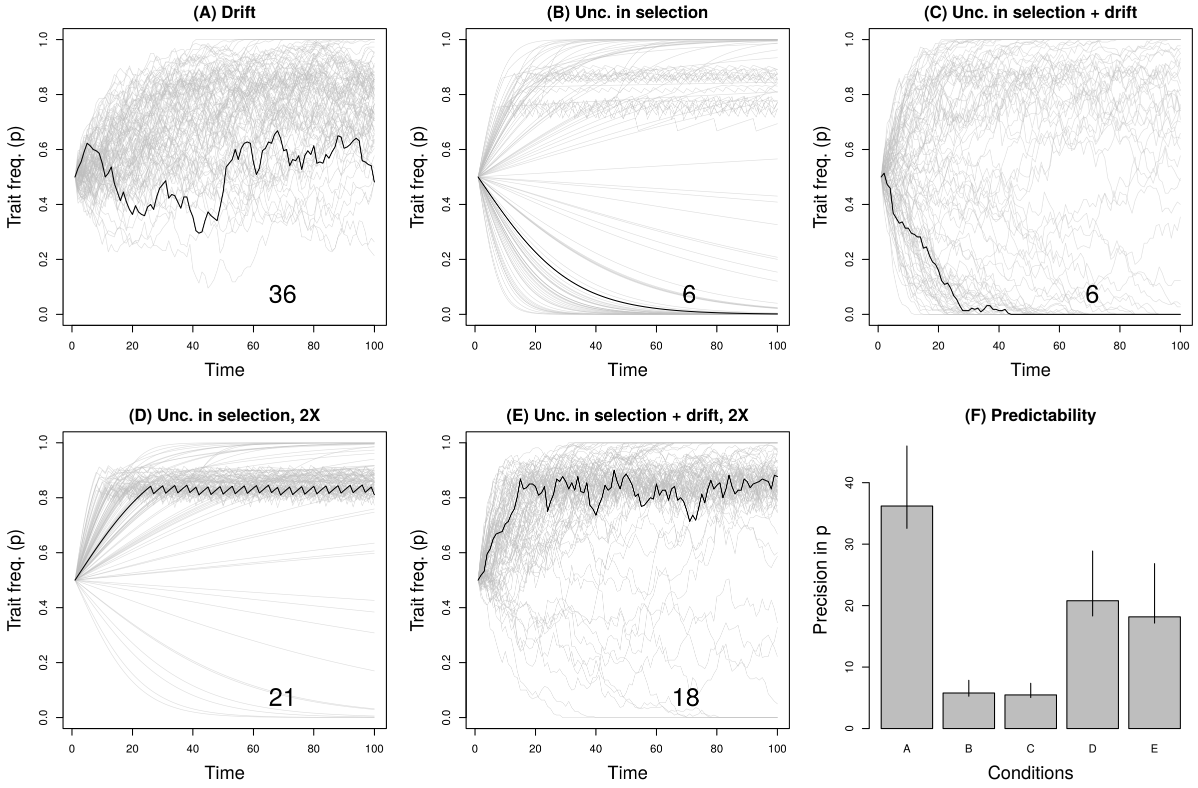


Figure S4: Case study 1 with NFDS, $h^2$ = 1.0, weak selection. Gray lines show 100 probable evolutionary trajectories under NFDS. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), and genetic drift and uncertainty in selection (C). Panels (D) and (E) show the effect of uncertainty in selection and drift plus uncertainty in selection given a hypothetical, larger experiment. In each case, an arbitrary trajectory is shown in black to make the dynamics clearer and to emphasize the fact that only one trajectory would be realized in a given biological system. Panel (F) summarizes the predictability of evolution across each of the five conditions shown (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in trait (morph) frequencies (denoted p) across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Results are shown for a heritability of 1.0, and with weak selection (1% difference in survival probabilities between morphs).

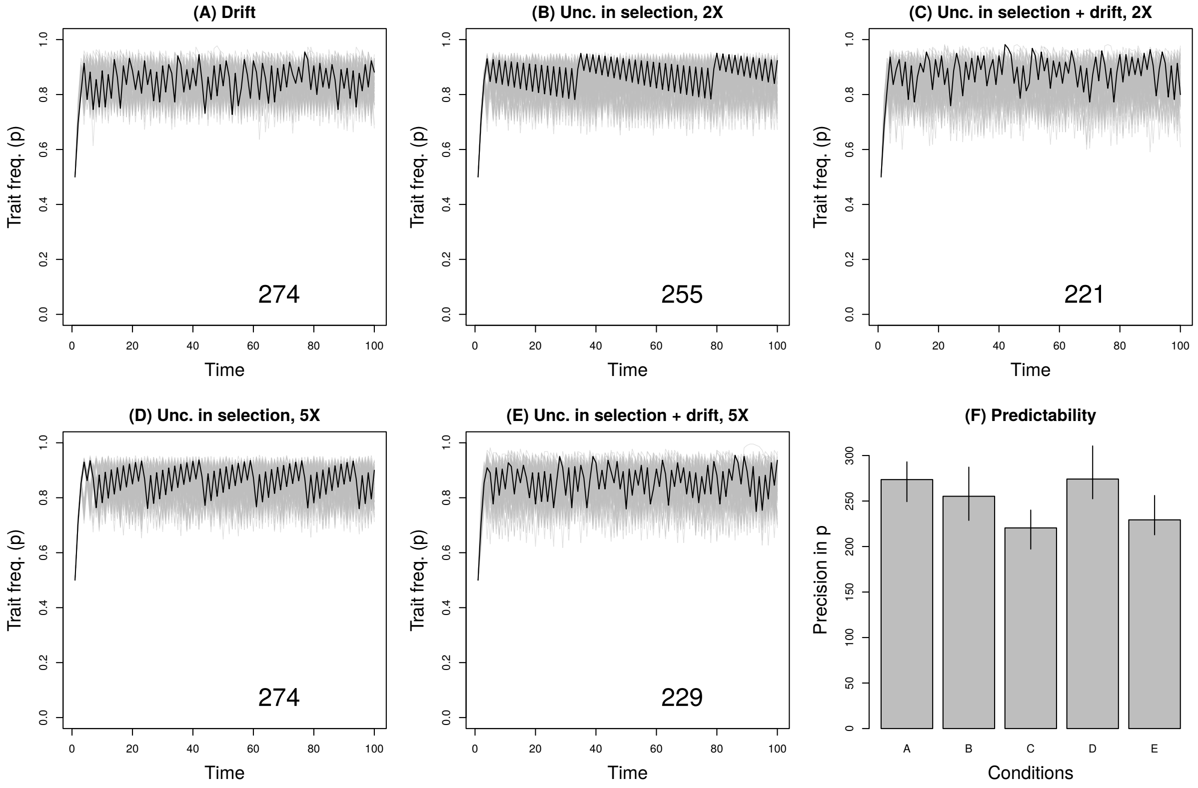


Figure S5: Case study 1 with NFDS, increased sample size. Gray lines show 100 probable evolutionary trajectories under NFDS. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), and genetic drift and uncertainty in selection (C). Panels (B) and (C) show the effect of uncertainty in selection and drift plus uncertainty in selection given a hypothetical, doubling of the experiment, whereas panes (D) and (E) show the effect of hypothetically increasing the sample size five-fold. In each case, an arbitrary trajectory is shown in black to make the dynamics clearer and to emphasize the fact that only one trajectory would be realized in a given biological system. Panel (F) summarizes the predictability of evolution across each of the five conditions shown (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in trait (morph) frequencies (denoted p) across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Results are shown for a heritability of 0.8.

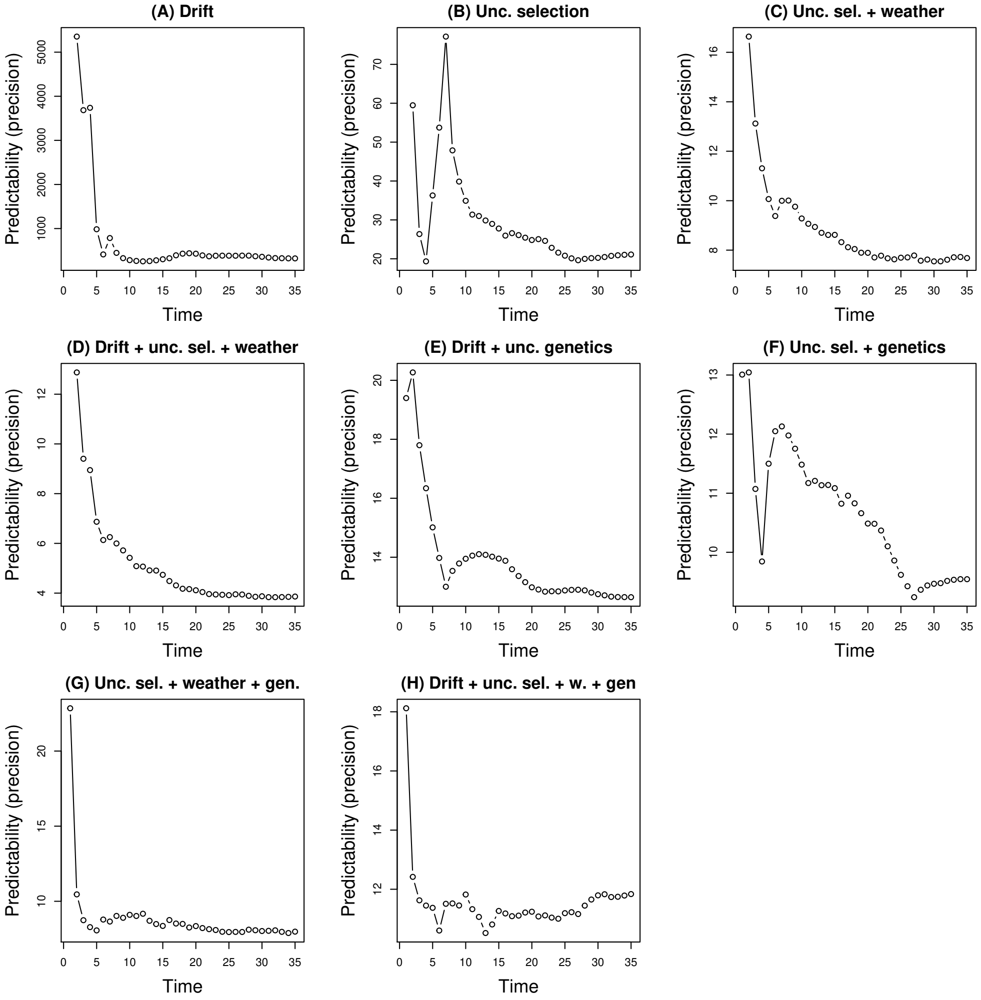


Figure S6: Precision over time in case study 2 with climatic variation and breeding values. Points connected by lines show the precision (reciprocal of the variance) in evolutionary trajectories of the expected breeding value (BV) of a trophic trait over time under temporally fluctuating selection. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), uncertainty in climatic conditions and selection (sel.) (C), and genetic drift and uncertainty in climatic conditions and selection (E). Panels (E-H) combine these effects with uncertainty in genetics (gen.) (i.e., in which a subset of 1415 genetic loci affect the trophic trait).

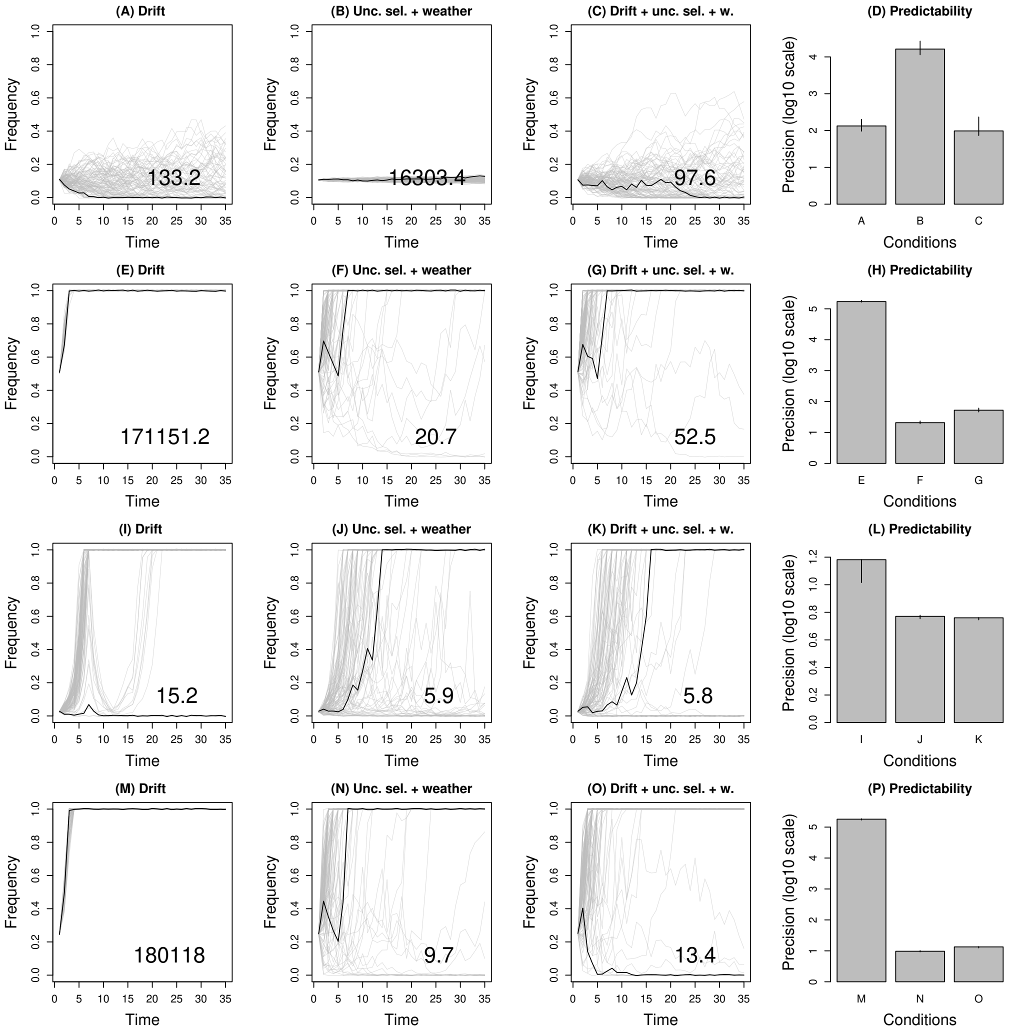


Figure S7: Case study 2 with climatic variation and no uncertainty in genetic architecture. Gray lines show 100 probable evolutionary trajectories of allele frequencies for loci. Each row in this multi-panel figure corresponds with one of the 1415 genetic loci. Here, we assume trait associations were certain, with loci in all but panels A-D being associated with the trait. Sources of variability in each panel are genetic drift (A, E, I, M), uncertainty (Unc.) in climatic conditions and selection (sel.) (B, F, J, N), and genetic drift and uncertainty in climatic conditions (w.) and selection (C, G, K, 0). In each case, an arbitrary trajectory is shown in black to make the dynamics more clear and to emphasize the fact that one of the possible trajectories would be realized in a given biological system. Panels (D, H, L, P) summarize the predictability of evolution across each of the three conditions shown for each locus (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in the allele frequency across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Median precision for each condition is also reported in each panel as a numerical inset. Compare to Figure 5, which shows the same loci with uncertainty in genetic architecture.

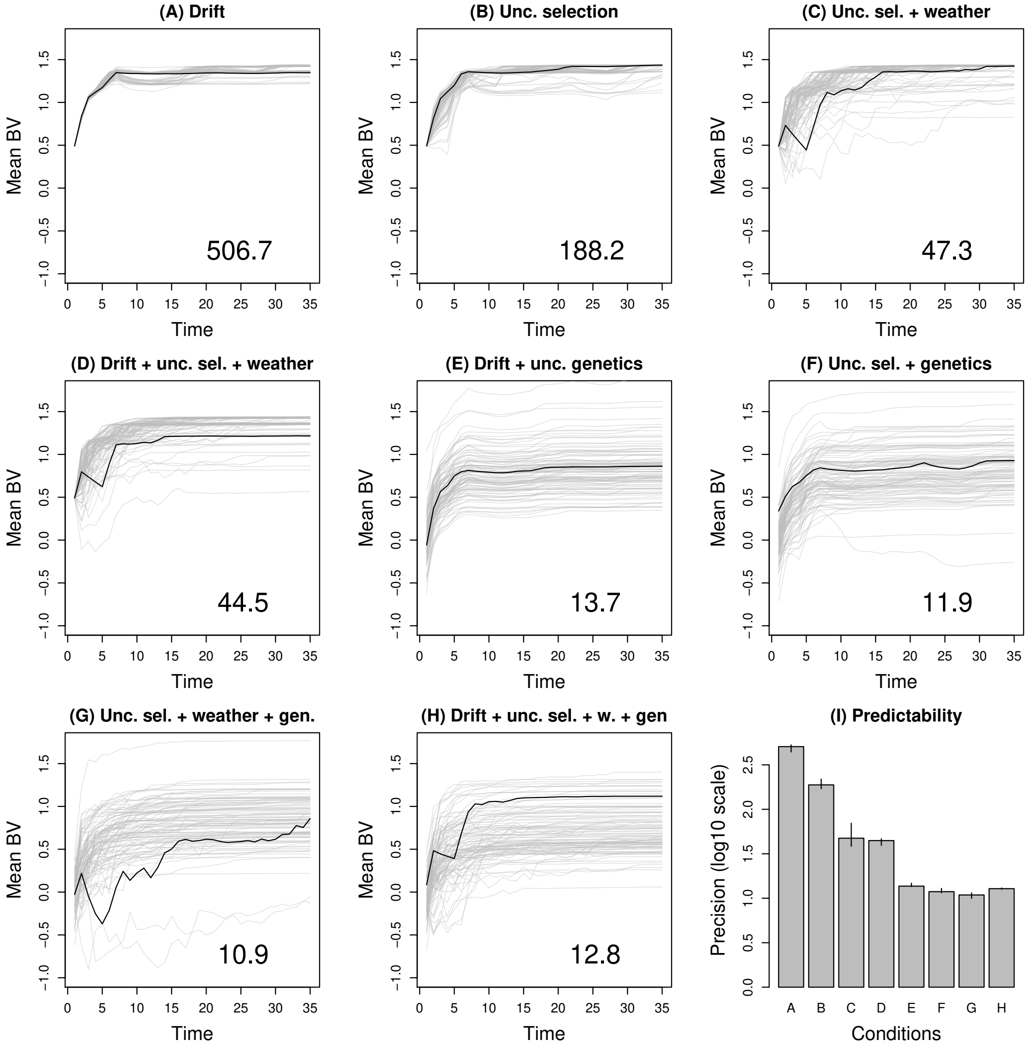
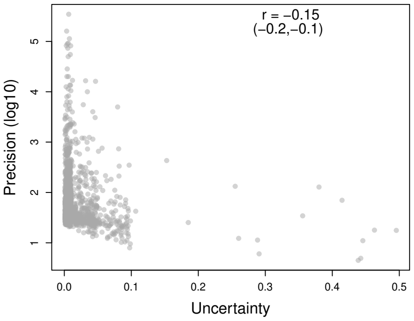
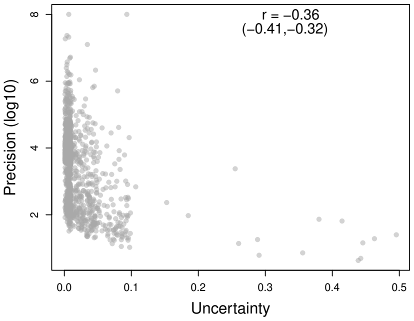


Figure S8: Case study 2 with climatic variation and breeding values, increased sample size. Gray lines show 100 probable evolutionary trajectories of the expected breeding value (BV) of a trophic trait under temporally fluctuating selection. Sources of variability in each panel are genetic drift (A), uncertainty (Unc.) in selection (B), uncertainty in climatic conditions and selection (sel.) (C), and genetic drift and uncertainty in climatic conditions and selection (E). Here, uncertainty in selection is reduced relative to the main results (compare to 4). Panels (E-H) combine these effects with uncertainty in genetics (gen.) (i.e., in which a subset of 1415 genetic loci affect the trophic trait). In each case, an arbitrary trajectory is shown in black to make the dynamics more clear and to emphasize the fact that one of the possible trajectories would be realized in a given biological system. Panel (I) summarizes the predictability of evolution across each of the eight conditions shown (labeled by their panel letter in this figure). Predictability is measured by the precision (reciprocal of the variance) in the mean (expected) BV across probable trajectories. Bars shown the median (across generations) and vertical lines denote the 25th and 75th percentiles. Median precision for each condition is also reported in each panel as a numerical inset.

(A) Drift + unc. genetics



(B) Unc. sel. + weather + gen.



(C) Drift + unc. sel. + w. + gen

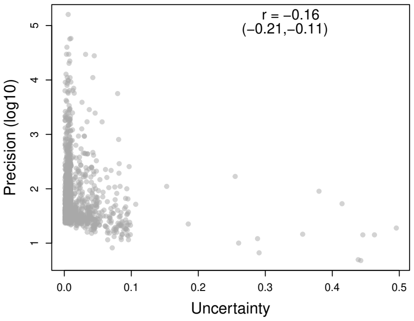


Figure S9: Case study 2 with climatic variation and consideration of genetic architecture and individual loci. Here, we show how the degree of uncertainty in whether a locus was under selection affects precision at the level of locus-specific allele frequency change. All results shown assume uncertainty (unc.) in the trait genetic architecture; additional sources of variability in each panel are genetic drift (A), uncertainty in climatic conditions and selection (sel.) (B), and genetic drift and uncertainty in climatic conditions (w.) and selection (C). In each panel, each point denotes the precision (median over generations on a $\log_{10}$ scale) for one of 1415 genetic loci as a function of the uncertainty in the genotype-phenotype association. This latter value is the minimum of the probability of association or 1 minus the probability of association (i.e., a 0 would indicate perfect association or, as is closer to the case here, a perfect lack of association). The Pearson correlation ($r$) between this measure of uncertainty in genetic architecture and the $\log_{10}$ precision is reported along with the 95% confidence intervals (all $P < 0.0001$). These negative correlations indicate that precision was lower for loci with less certain associations, especially in the absence of genetic drift (as indicated by the larger negative correlation in panel B).