The why, what and how of predicting evolution across biology: from disease to biotechnology to biodiversity

Abstract

Evolution has traditionally been a historical field of study and predicting evolution into the future has long been considered challenging or even impossible. However, evolutionary predictions are increasingly being made and used in many situations in medicine, agriculture, biotechnology and conservation biology. Because every field uses their own language and makes predictions from their background, researchers are not always aware of the breadth of evolutionary predictions. Evolutionary predictions may be used for several purposes such as to prepare for the future, to try and change the course of evolution or simply to determine how well we understand an evolutionary system. Exactly what aspect of an evolving population we want to predict, such as the most common genotype, average or individual fitness, or population size, depends on the situation. In addition, there are many uses of evolutionary predictions that may not be recognized as such. Therefore, the main goal of this review is to increase awareness of methods and data that are used to make these predictions in different fields, by showing the breadth of situations in which evolutionary predictions are made. We describe how evolutionary predictions are highly diverse, but nevertheless share a common structure described by the predictive scope, horizon, precision and risk. Then, by using examples ranging from SARS-CoV2 and influenza to CRISPR-based gene drives and sustainable product formation by microorganisms, we discuss the methods for predicting evolution, factors that affect predictability, and how predictions can be used to prevent unwanted evolution or promote beneficial evolution. We hope that this review will increase collaboration between fields by creating a common language for evolutionary predictions.

Keywords
Evolution, prediction, models, population genetics, disease modeling, evolutionary control, SARS-CoV2, gene drive, influenza, predictability

Running head: Predicting evolution across biology

The why, what and how of predicting evolution across biology: from disease to biotechnology to biodiversity **1**

Abstract **1**

1. Introduction **3**

1. Motivating questions 3

2. The scientific basis of evolutionary predictions 4
3. Why predict evolution? 5
4. What do we want to predict? 6

2. What is an evolutionary prediction? 7
1. A conceptual model of evolutionary predictions 7
2. Applying the model of evolutionary predictions to influenza strains 9

3. How can we predict evolution? 10
1. Methods of predicting evolution 10
2. Experimental evolution 13
3. Using the mutational and fitness landscape 13
4. (Microbial) metabolic and growth models 14
5. Population-genetic models 15
6. Quantitative genetics and the breeder's equation 15
7. Epidemiological models (SIR models) 16
8. Species distributions across space and environmental conditions 16
9. Multi-scale evolutionary modeling 17
10. Machine learning 18

4. Predictability 18
1. What makes evolution more or less predictable? 18
2. Mutational supply and mutation bias 19
3. Distribution of fitness effects (DFE) of mutations. 20
4. Epistasis 21

5. Evolutionary control 21
1. Evolutionary control needs predictions 21
2. Health measures during epidemics influence evolution 22
3. Preventing or reversing antibiotic resistance in bacterial pathogens 22
4. Insect resistance to transgenic plants 23
5. Prevention of resistance to gene drives 23
6. Preventing extinction by promoting evolution 24

Box 1: Examples 25
1. SARS-CoV-2 25
2. Sustainable product formation by microorganisms 25
3. Adaptation of natural populations 25
4. CRISPR-based Gene drives 26
5. Influenza 26

6. Conclusion and outlook 26

7. Data Archiving Statement 27

8. Bibliography **28**

1. Introduction

1. Motivating questions

Important questions for battling diseases, developing biotechnology and protecting biodiversity include: “Which SARS-Cov2 strains will be most prevalent a month from now?”, “When and where will SARS-Cov2 mutants that escape the vaccine arise, threatening vaccine efficacy?”, “Which patient will be cured of cancer and which patient will see the tumor come back, but now resistant to chemotherapy?”, “Can we use gene drive systems to get rid of (vectors for) dangerous diseases or will they evolve resistance to the gene drive?”, “How fast will a strain engineered for ethanol production evolve and lose its efficiency during prolonged fermentation?” and “Which endangered species will go extinct and which will adapt successfully to their changing environment?”.

Answering these questions requires the ability to **predict the course of evolution into the future**. In addition, some of these situations would have us trying to influence the course of evolution. While some fields have been working for many years on predicting and influencing evolution, either explicitly or implicitly, for other fields this is a new endeavor. We argue that predicting and trying to influence evolution is more common than you may think, but it is not always easy to recognize because the jargon used in different fields is varied. For example, in the influenza virus literature, there are many articles on predicting evolution and these predictions are used to design vaccines for the next influenza season (Łuksza & Lässig 2014, Morris *et al.* 2018, Barrat-Charlaix *et al.* 2020, Hayati *et al.* 2020). On the other hand, in the literature on the evolution of drug resistance in HIV, predictions are described differently. For instance, a large study on 6500 HIV patients in British Columbia (Rocheleau *et al.* 2018) determined which patient characteristics were most predictive of within-host drug resistance evolution. However, the words “prediction” and “evolution” are never used in the paper. The main goal of this review is to show the breadth of situations in which evolutionary predictions are being made. We also aim to bring diverse aspects of evolutionary forecasting together, showcase the strengths and weaknesses of different approaches, and form a platform from which future studies could launch research into evolutionary predictions.

Note that throughout this paper, we focus on predictions of how populations will evolve, that is how the genetic and phenotypic makeup of populations will be different in the future (Nosil *et al.* 2020), rather than predictions about the evolution of new species, or predictions about the past. We thus generally take a more applied approach to predicting evolution compared to some of the existing literature on predicting evolution (Gould 1990, Morris 2003).

2. The scientific basis of evolutionary predictions

What is the basis upon which we can make sound predictions about evolution? Many hypotheses, theories and models can be used to make evolutionary predictions. Arguably the basis of all evolutionary predictions is Darwin's theory of evolution by natural selection which states that if populations of entities manifest heritable variance in fitness, then these populations will adapt to their environment. For example, we can predict that if we treat bacteria with antibiotics, and if these bacteria harbor genetic variation for antibiotic susceptibility, then the bacterial population will adapt to that challenge and become resistant. We can also recreate this scenario as an experiment in the lab and see whether our prediction holds true.

In addition to the basics of Darwin's theory, there are many extensions that make Darwin's theory more precise and quantitative. For example, our understanding of the polygenic nature of quantitative traits has aided in developing tools such as the "breeder's equation" (Fig. 3E) and "genomic selection", facilitating selective breeding strategies in order to deliver particular (predicted) outcomes in animal husbandry and agriculture (Cooper *et al.* 2014, Masuka *et al.* 2017).

For other situations we need more extensive population genetic models to include forces that can distort the expected impact of selection. Population genetic theory includes random genetic drift, migration, recombination and mutation and it can deal with the stochasticity of these forces. However, population genetic theory only works well for predicting changes in allele frequencies or population sizes if it is known which mutations have an effect on fitness and what these effects are (though see the extensive literature on adaptive walks (Gillespie 1983, Orr 1998)). Yet, inferring the distribution of fitness effects of new mutations is difficult because of the sheer number of possible mutations and because this distribution may differ between genetic backgrounds and between different environments. This endeavor relies either on the availability of large data sets capturing the genetic variation in natural populations (e.g. (Eyre-Walker & Keightley 2007, Tataru *et al.* 2017) or on extensive measurements of fitness effects of selected mutations in the laboratory (e.g., (Fowler & Fields 2014, Cote-Hammarlof *et al.* 2021).

An additional complicating factor is that populations impact their environment. In many situations, we therefore have to consider both evolutionary and ecological dynamics. For example, the fate of endangered species critically relies on the abundance of pollinators, predators, and prey and other members of their ecological community, while these populations are in turn affected by the endangered species in question (Govaert *et al.* 2019). A core challenge in predicting evolution is thus to integrate ecological and evolutionary forecasting to describe eco-evolutionary dynamics.

Predicting evolution has long been considered challenging, or even impossible. Fundamental difficulties of predicting evolution include the inherent stochasticity of mutation, reproduction and environment, and the unknowns of genotype-phenotype and phenotype-fitness maps which, together, determine a fitness landscape (Wright 1932, de Visser & Krug 2014, Fragata *et al.* 2019). In addition, eco-evolutionary feedback loops make long-term predictions challenging.

These aspects of evolving populations will limit accuracy and predictions will therefore always be probabilistic and provisional especially for predictions further into the future. Short-term and microevolutionary predictions may be most achievable (Lässig *et al.* 2017).

3. Why predict evolution?

There are different reasons why we are interested in predicting evolution, which we have organized in three main categories (Fig. 1). The first reason is to be prepared for the future. A key example here is seasonal influenza. In the spring of any year, vaccines are produced for the next fall (for the Northern hemisphere). To make sure the vaccine is as effective as possible, it is necessary to predict which strains will be most common in the next influenza season.

The second reason to predict the course of evolution is to choose actions that influence the direction or (more often) speed of evolution – also referred to as evolutionary control. For example, treatment regimes may be chosen with different combinations of antibiotics that, together, reduce the risk of resistance evolution or that guide evolution to low fitness types that are less likely to spread in antibiotic-free environments. We devote a section of this review to evolutionary control.

Third, we use predictions for experimental systems to test our fundamental assumptions of evolving systems. Most work in experimental evolution falls in this category. A prediction here is similar to a hypothesis. These experiments can focus on the speed of adaptation, the distribution of fitness effects of new and existing mutations, repeatability of evolutionary outcomes and causes of repeatability. These experiments force us to define the necessary information to predict evolution and determine reasons for failure, and they allow us to test the limits of the generality of predictions.

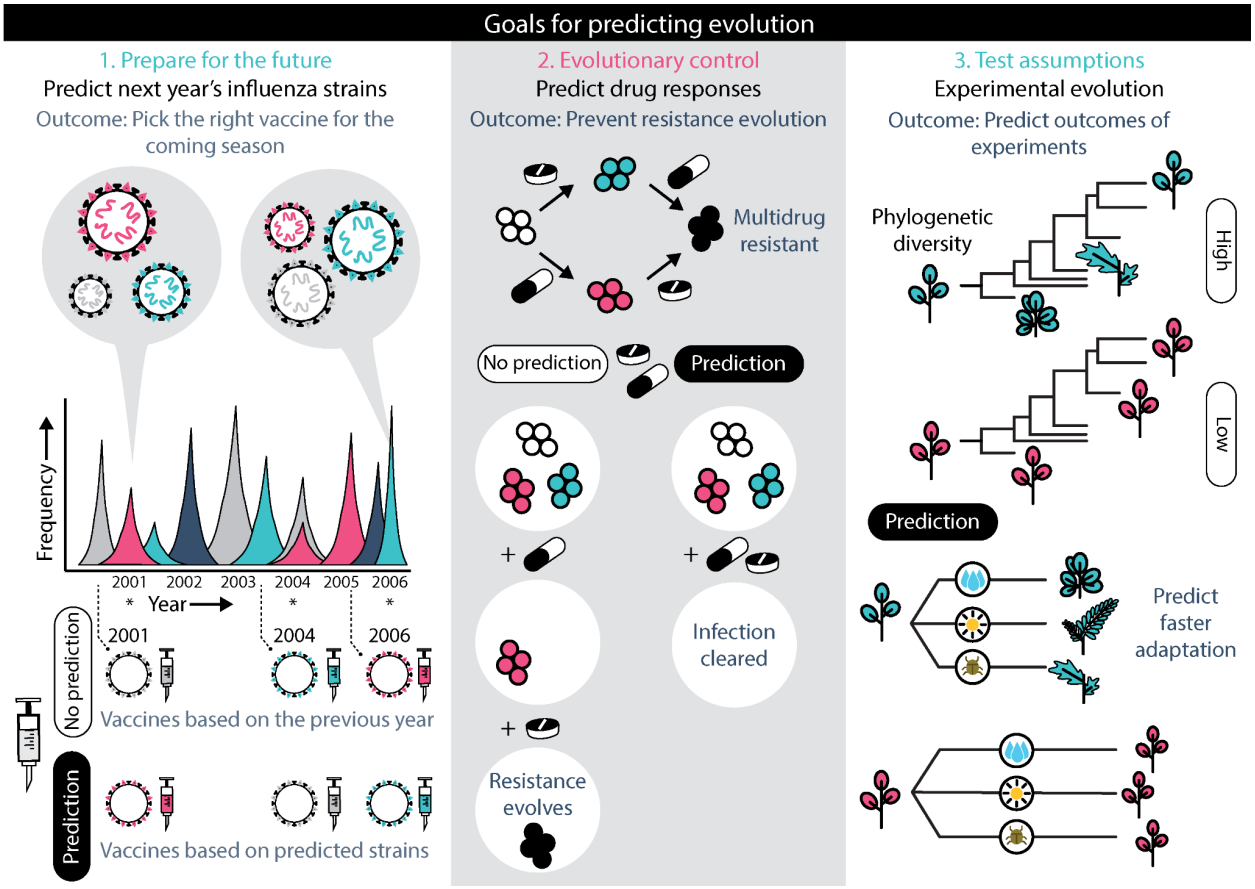


Figure 1. Why do we need predictions? *(1) To be prepared for future outbreaks, we aim to match vaccines with the most common influenza strains each year. (2) To have control over evolutionary outcomes and design treatment strategies that prevent evolution of resistance from happening in pathogens. (3) To test hypotheses of evolution for a better fundamental understanding of evolving systems. Based on their phylogenetic history we can predict how species respond to a given treatment. These predictions can be tested with experimental evolution approaches.In this review we focus on predicting evolution for goals (1) and (2).*

4. *What* do we want to predict?

When making evolutionary predictions, we can focus on many different aspects of the future state of a population. Here it may be useful to briefly compare evolutionary predictions to weather and climate predictions, which have many dimensions as well. Sometimes we care about *whether or not* it will rain tomorrow; whereas other times, when we worry about flooding, we care about exactly *how much* rain will fall in the next 24 hours. Also, on most days, wind speeds may not be mentioned in a weather forecast, but when a hurricane is arriving, wind speeds are suddenly crucial to prepare for the impact of the hurricane. Evolutionary predictions are similarly diverse. Evolutionary predictions can be about different population variables (e.g. majority phenotype, average fitness, identity of fixed mutations, allele frequencies, population size) and usually include a time component (Will this population be extinct one year from now? When will a new variant dominate the SARS-Cov2 pandemic?).

We sometimes focus on *which strains* will be most common next year (e.g., influenza). Other times, we want to predict *whether* or *when* drug resistance will evolve in a virus or bacterial infection in a patient. In these cases, we might want predictions at the level of the genotype or at the level of the phenotype. In experimental evolution, the goal may be to predict a certain *phenotype* (e.g. cell size) or the *average fitness* of a population after some amount of time, or we may want to predict *which genes* will acquire new mutations or confer increased fitness. With increasing interest in engineering microbial communities, interest turns to predicting the evolution of *interacting populations*, and we may want to predict population sizes in these microbial communities. In conservation biology, the focus is also usually on predicting future *population sizes*. For all of these cases, besides predicting the most likely outcome, sometimes the *probability* of a *certain outcome* (*e.g., extinction*) is most important.

2. What is an evolutionary prediction?

1. A conceptual model of evolutionary predictions

An evolutionary prediction is a statement about the future state of an evolving population. Communities of populations evolve too, but we focus here mainly on populations. Predictions can be focused on either the genotypic or the phenotypic level. For instance, at the genotypic level we can predict the frequency of influenza variants in the next influenza season (Łuksza & Lässig 2014), or we can predict the mutational targets of *E. coli* responding to various environmental pressures (Wang *et al.* 2018). At the phenotypic level, we can predict the shape of Darwin's finches beaks after a drought (Grant & Grant 2002) or the repeated evolution of equivalent spider ecomorphs over macroevolutionary timescales (Gillespie et al. 2018). Predictions operate across a range of timescales – from a single generation to millions of years. In this paper we focus on evolutionary predictions that are forward-looking in the sense that they concern future events; but we can also predict aspects of the past (hindcasting) based on the state of a population further in the past. Hindcasting can be used to directly test models' predictive power, for example, to test whether models can predict which influenza mutations will fix or be lost (Barrat-Charlaix *et al.* 2020).

These examples illustrate how evolutionary predictions are highly diverse in their topics, but at an abstract level, they nevertheless share a common structure (Fig. 2). All evolutionary predictions result from a model — including conceptual, verbal, mechanistic, statistical, computational or mathematical models — that allows for a projection of the state of the evolving system beyond the input that is provided.

The model for all evolutionary predictions starts from describing the current state of the evolving system, and incorporates the total prior scientific knowledge of relevance (e.g. facts and mechanisms, evolutionary processes, etc.). This is the input of the model. The assumptions that are made, and that have not (yet) been proven to be true can also form input or constraints for the model. The output of the model describes the state of the evolving system in the future. The extent to which the projection of the future state overlaps with the current state of the evolving

system describes how risky or safe the predictions are. The part of the output that overlaps with existing knowledge can be used to validate the model.

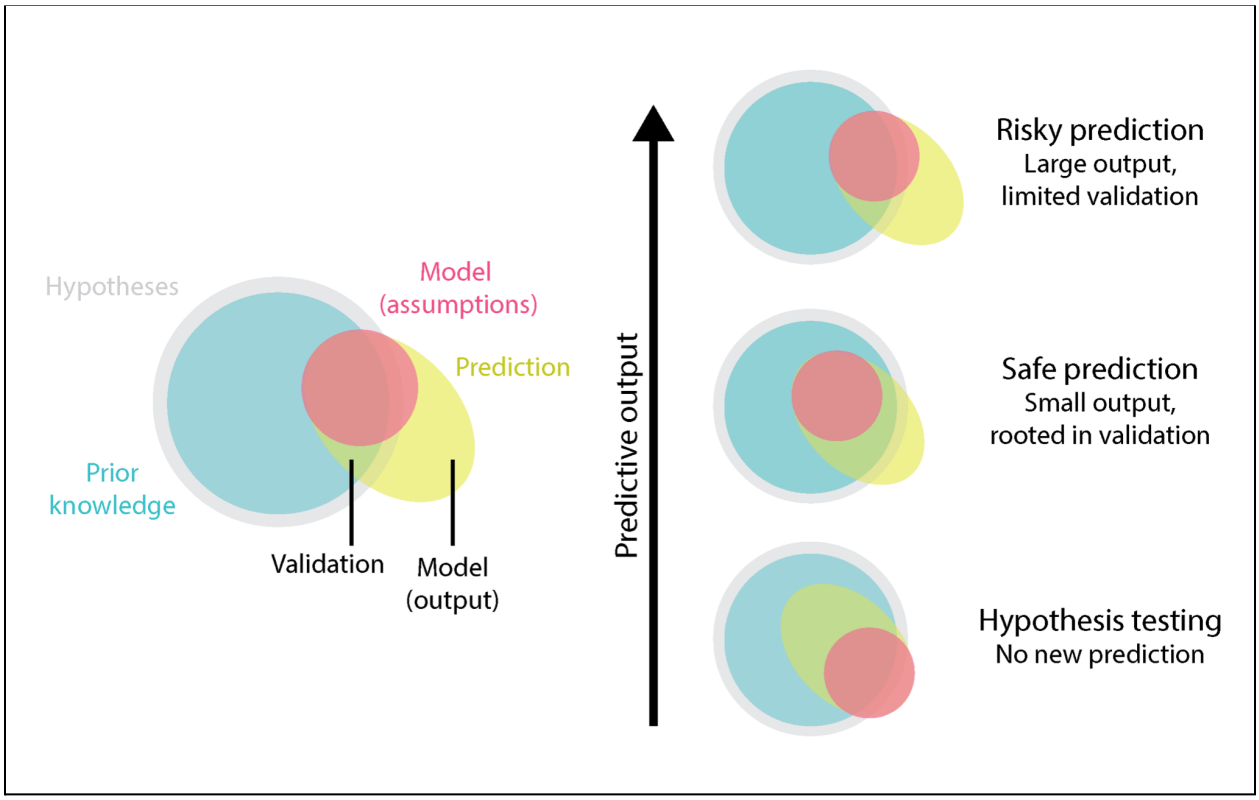


Figure 2. Shared abstract structure of all (evolutionary) predictions. *Starting from prior knowledge (blue circle), models can be formalized (pink circle) which may project into the unknown (here, visualized by the yellow ellipse), and may use assumptions (grey aura). If pointing outwards, the yellow area of the ellipse constitutes a prediction, whereas the green area is used for validation of the model and its assumptions. The model assumptions and output will determine the orientation of the ellipse, and to what extent it projects into the unknown. The predictive output will range from the most risky predictions of more fundamental and serendipitous nature, to safer predictions which will be more suitable for predicting more practical matters such as potential pathogen escape mutants. Hypothesis testing here is the case where predictions are made about results that are known, to test if the model that we have of a process explains the observations.*

When we describe predictions, we can consider various attributes. First, we should consider which attributes of a population we want to predict, or what is the **predictive scope**: Is the prediction about genotypes, phenotypes or population sizes? Are we trying to predict the average of a population, or the distribution of a trait? Are we predicting the evolutionary path or the outcome?

Second, we need to consider the temporal scope of a prediction, the **predictive horizon**: We can predict the state of a population at a given time point (in the near or far future), or we can predict the time point at which a given state is reached. Generally, predictions of a system's trajectory are more precise for the near future, carry most information on an intermediate time

scale, and get fuzzy when too far out. This is relevant especially when predictions are needed to decide on actions, such as which vaccine to manufacture. For example, predicting influenza strain frequencies a year into the future is optimal for information gain (Lassig et al. 2017).

Third, we can consider the level of detail of a prediction or **predictive precision**. A prediction about the direction of an effect is less detailed than one that also includes its magnitude, rate or trajectory. For instance, predicting that microbes will evolve to consume a novel food source is less detailed than predicting that the evolution will occur within 100 generations, via a given sequence of mutations (trajectory) (Lind *et al.* 2019). In a sense, predictions are a type of hypothesis and if they are too general they cannot be falsified.

A fourth attribute of a prediction may be the a priori likelihood (absence of surprisingness) of the prediction, the **predictive risk**. All else being equal, predicting something which is a priori unlikely, given background knowledge, is harder and possibly more interesting.

2. Applying the model of evolutionary predictions to influenza strains

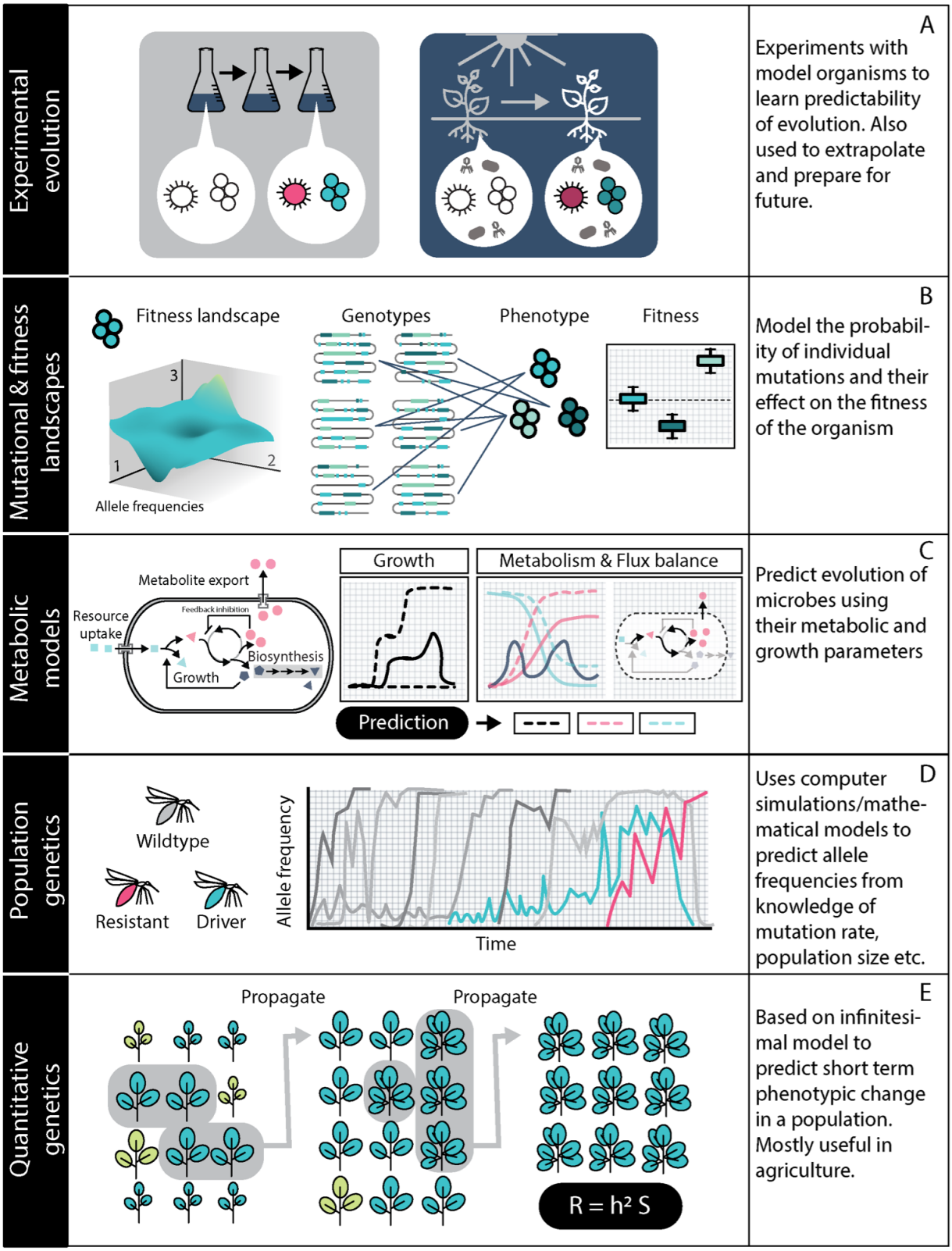
Let's apply our conceptual model of predictions (Fig. 2) and the four attributes described above to one of the best known examples of evolutionary predictions: the prediction about seasonal influenza strains in the next season, which is used to design the next vaccine. First of all, the prior knowledge (blue circle in Fig. 2) is the current state of the population as inferred by observed frequencies of influenza strains in humans and other hosts, and the trajectory of these frequencies in the recent past. The model assumptions (grey) can include mathematical descriptions of how the frequency trajectories can be extrapolated into the future (taking into account mutation, drift and selection). The outcome of the model (yellow) is the predicted frequencies of the major strains during the next influenza season, or possibly only the list of the most prevalent clades. At an even more coarse-grained level, we may only be interested in predicting the antigenic phenotypes of the future virus population. All of these choices place predictions at different points on the line between safe and informative, and between realistic and over-ambitious.

In terms of the four attributes, the predictive scope is the genetic or antigenic composition of the population, the predictive horizon is a time scale 6-12 months into the future. The predictive precision is high in terms of the time scale (we need to know the common strains for the next season, not earlier or later than that) and also in terms of genetics, as we are predicting the exact strains that will be most common. The a priori likelihood is bounded when we consider only existing strains, as there may only be a small number of possible outcomes when only few strains currently exist in the population. However, the exact predictive risk depends on whether a common or uncommon strain is predicted to be prevalent the next season. Over longer time scales *de novo* mutations become important. *De novo* mutations are hard to predict, which is one reason why predictability is reduced over longer time scales.

3. How can we predict evolution?

1. Methods of predicting evolution

If we establish that we want to predict evolution, and what such a prediction entails, the next question is *how* we can predict evolution. The most straightforward prediction comes from observations of repeatability: we have seen evolution take a certain path in similar conditions (either controlled or natural), so we expect to observe the same evolutionary process again. Although this might be one of the safest predictions, often we want to predict evolution under conditions or over timescales that we have not encountered before, which means that we have to extrapolate from prior observations by incorporating different types of models. These models could include a lot of data (e.g., machine learning models) or more mechanistic knowledge of traits or molecular details (e.g., metabolic growth models). There are many different approaches and methods to predict evolution, and we describe some of those here. Note that phylogenetic models are not described here but feature in chapter 5 and box 1.


Experimental evolution
A
Experiments with model organisms to learn predictability of evolution. Also used to extrapolate and prepare for future.
Mutational & fitness landscapes
Fitness landscape
Genotypes
Phenotype
Fitness
3
1
2
Allele frequencies
B
Model the probability of individual mutations and their effect on the fitness of the organism
Metabolic models
Metabolite export
Feedback inhibition
Resource uptake
Biosynthesis
Growth
Growth
Metabolism & Flux balance
Prediction
C
Predict evolution of microbes using their metabolic and growth parameters
Population genetics
Wildtype
Resistant
Driver
Allele frequency
Time
D
Uses computer simulations/mathematical models to predict allele frequencies from knowledge of mutation rate, population size etc.
Quantitative genetics
Propagate
Propagate
R = h² S
E
Based on infinitesimal model to predict short term phenotypic change in a population. Mostly useful in agriculture.

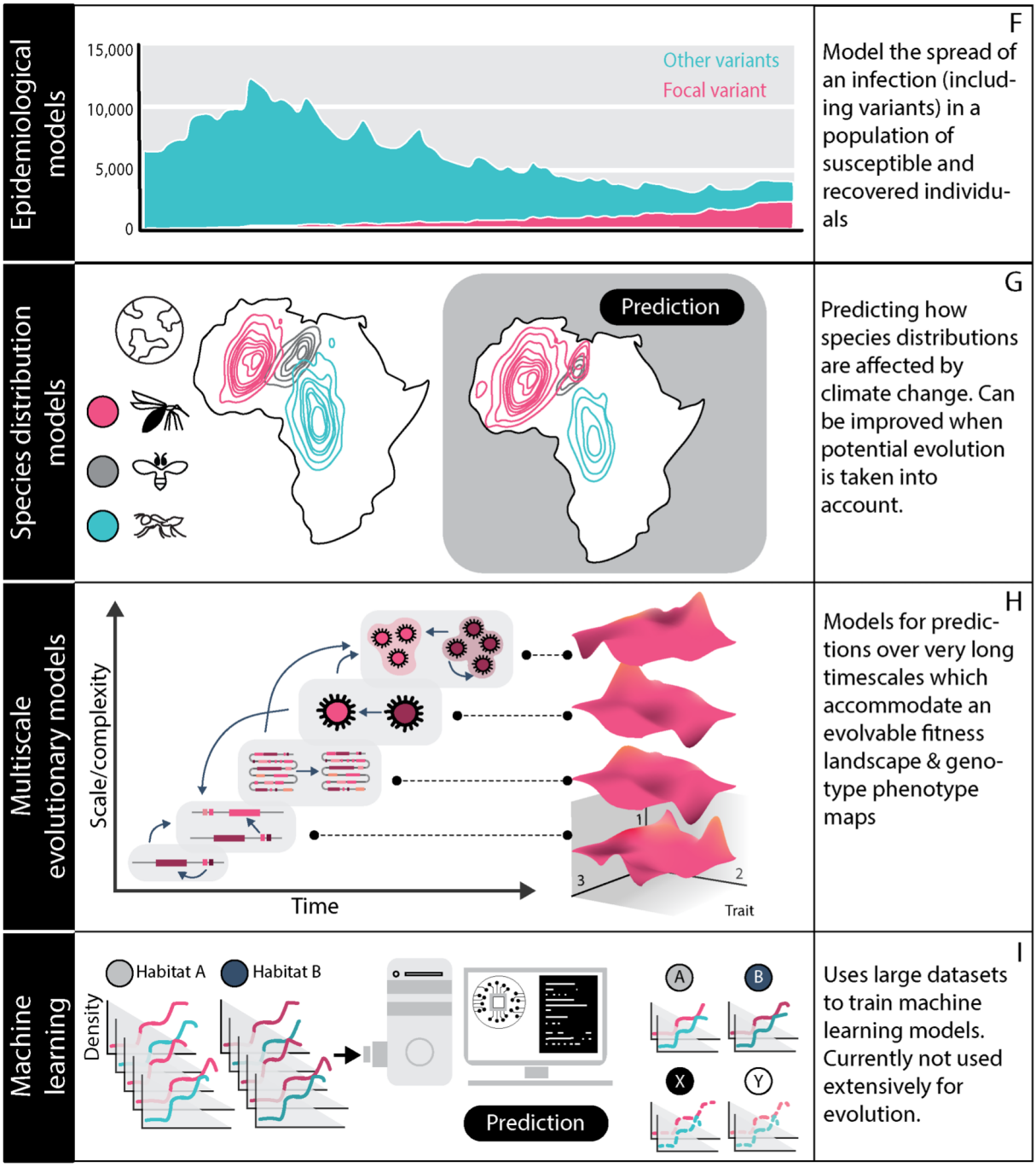


Figure 3. Selection of methods that are used to predict evolution.

2. Experimental evolution

One main method of evolutionary predictions is seeing how evolution occurs under the conditions of interest, in order to extrapolate from these observations. There are two different types of environments that can be used for these observations: laboratory studies and field studies; and those experiments can be started from a clonal population, restricting evolution to *de novo* mutations, or a genetically diverse population.

Laboratory studies of experimental evolution are used extensively for the more curiosity driven predictions, and perhaps the most famous one is the long term evolution experiment with *E. coli* (Wiser *et al.* 2013, Blount *et al.* 2018). Analysis of the fitness trajectories and of the genomic response, by way of resequencing of evolved populations, observed in that experiment and others, have revealed some common rules of microbial adaptation. For instance, (i) fitness improvement is faster in maladapted genotypes (Couce & Tenaillon 2015), (ii) beneficial mutation supply is large, such that often multiple beneficial mutations coexist and compete in a population (Lang *et al.* 2013), (iii) in most environments mutatios with large fitness benefits are only found in a few genes (Tenaillon *et al.* 2012, Lind *et al.* 2017), which lead to high convergence at the gene level, and (iv) a change in mutation rate can easily be selected for in the course of a adaptation (Sniegowski *et al.* 1997). Interestingly, these observations, while made mostly *in vitro*, were recovered when applied to experiments in more natural conditions such as the mouse gut (Barroso-Batista *et al.* 2014, Lescat *et al.* 2017) and have been found in natural conditions (Zhao *et al.* 2019).

Laboratory experimental evolution is diverse, it is commonly used for fast-evolving systems that are easily kept in the laboratory, such as for antibiotic resistance evolution (Kawecki *et al.* 2012, Remigi *et al.* 2019), and larger organisms such as *Drosophila* (Burke & Rose 2009) and *Caenorhabditis* (Teotónio *et al.* 2017). Interactions with the environment (eco-evolutionary interactions) can be included by considering aspects of natural conditions in the laboratory, such as the role of environmental spatial structure (Nadell *et al.* 2016) or the natural biotic context (Zandbergen *et al.* 2021).

Field studies are more appropriate for conditions or organisms that cannot be studied in the lab, for example because habitats are too large. In such cases, the effect of environmental conditions on evolution can be studied with reciprocal transplant experiments (Edwards 2015) or observations of “natural field experiments”, such as sticklebacks that have moved from marine environments to fresh water repeatedly (Jones *et al.* 2012).

3. Using the mutational and fitness landscape

Though currently still out of reach, one day, it may be possible to predict the next evolutionary step for a population using detailed knowledge of the mutation and fitness landscape for the population in a given environment. To know what variation is available to a population, we need to know the single step mutational landscape (the probability of each mutation to occur) and the effect of these mutations on the fitness of the organism (fitness landscape). Currently, limited

information on the mutational landscape might already inform evolutionary predictions (e.g., knowledge of epistasis for key mutations can predict evolutionary pathways (Salverda *et al.* 2011)).

Obtaining the single step mutational landscape with the accompanying fitness effects is a hard problem given the sheer scale of it. Even if we only focus on a single mutational step, and assume that the environment is stable, the number of possible mutations is very high (~>10^9 in eukaryotes) which makes it hard to determine the fitness effect for each mutation. Several approaches have been used to try and tackle parts of this problem, by focussing on small genome viruses (Tisthammer *et al.* 2020), one gene (Lee *et al.* 2018), on a small metabolic pathway (Kemble *et al.* 2020) or using mutation accumulation experiments (Sane *et al.* 2018). For any prediction to be relevant, we need to take into account changing environments and multiple mutations and their interactions (epistasis), although there is some indication that environmental change does not change the fitness landscape completely (Vos *et al.* 2018).
Nevertheless, progress has been made through a combination of two approaches: the statistical approach where measured fitness effects are correlated with specific mutations (top-down) (Wang *et al.* 2018) and a mechanistic approach to predict by reconstructing the genotype-phenotype-fitness map (bottom-up) (de Vos *et al.* 2015). For instance, all the epistatic interactions between mutations affecting the expression of two genes in a linear metabolic pathway were resolved with a mechanistic perspective taking into account the flux in the pathway, the toxicity of the intermediate metabolite and the protein expression cost (Kemble *et al.* 2020). At a more integrated level, the well-established polarity network in budding yeast was predictive of mutation effects (Daalman & Laan 2020). Both approaches have their own advantages and disadvantages, and in most cases both approaches will need to be combined to make the best prediction.

4. (Microbial) metabolic and growth models

In many environments, selection on microorganisms is related to metabolism and growth, and therefore we can predict their evolution using metabolic and growth models. These models come in different levels of detail. A simple phenomenological model can predict fitness increase (Wiser *et al.* 2013) and a more detailed model with intracellular ‘macro’-reactions can be used to predict antibiotic resistance evolution (Pinheiro *et al.* 2021), or overflow metabolism (Molenaar *et al.* 2009, Wortel *et al.* 2016). Much more detailed models include genome-scale metabolic models (GEMs, current status in (Gu *et al.* 2019)), metabolism and expression models (ME models, (O’Brien *et al.* 2013)), resource balance analysis models (RBA models, (Goelzer *et al.* 2015)) and the most extensive whole cell models (Karr *et al.* 2012). Flux Balance analysis (Orth *et al.* 2010) can be used to predict fluxes that lead to optimal growth, constrained by measured maximal fluxes, in GEMs, and therefore as a prediction of evolution (in *E. coli* (Schuetz *et al.* 2007) and *Lactobacillus plantarum* (Teusink *et al.* 2009)). To improve predictions, proteome and kinetic constraints can be added (Sánchez *et al.* 2017, Chen & Nielsen 2021), or detailed descriptions of protein-metabolite dynamics can be incorporated at the expense of decreasing the model size (Wortel *et al.* 2018). GEMs can also be used to predict how knock-outs of

specific enzymes affect growth and product formation (Du *et al.* 2018), such as in the example in Box 1.

5. Population-genetic models

Population-genetic models are models that keep track of the genetic status (often at one or a few loci) of an entire population. Population-genetic models are used to predict the evolutionary dynamics of gene-drive systems. Box 1 has more information about gene drive systems, but briefly, these systems are designed to aid the rapid spread of a new allele in a population (Fig. 3D). Unckless *et al.* used a population-genetic model to predict the evolutionary dynamics over time of a CRISPR/Cas9 gene drive system (Unckless *et al.* 2017) . In this model the frequencies of three alleles at one locus are followed over time. The alleles are (1) the **wildtype allele** (which is susceptible to the gene drive system), (2) the **driver allele** (which can convert the wildtype allele into a driver allele and may come with a fitness cost, in particular when the gene drive is used to control a population) and (3) a **resistant allele**. The resistant allele can be created from the wildtype allele by mutation and cannot be converted to the driver allele. The model follows a population of individuals that is initially largely wildtype, and in which a driver allele spreads (due to its non-Mendelian inheritance). If a resistant allele is present in the population (from standing genetic variation or due to new mutations), it becomes highly beneficial once the driver allele is common in the population. Ultimately, the model predicts when the resistant allele will stop the spread of the gene drive construct.

One interesting complexity which was included in the model by Unckless *et al.* is that the CRISPR/Cas9 gene drive system itself is mutagenic, because the cuts it makes in chromosomes are often repaired by the non-homologous end-joining pathway which leads to mutations. The population-genetic model showed that if we want gene drive systems to be successful, we need to either reduce the rate at which resistance alleles are created, or we need to drastically increase the cost of resistance. Both of these methods have now successfully been used in experimental populations (Noble *et al.* 2017, Champer *et al.* 2020).

6. Quantitative genetics and the breeder's equation

The field of quantitative genetics was developed in close association with animal and plant breeders (Fig. 3E). In situations where the focus is on altering one or a few phenotypes, over short timescales and in relatively controlled environments, it has achieved great predictive success. For example, plant-breeding programs that are designed to improve crop yield focus on creating variable populations of hybrids and selecting for plants with the highest yield (under normal or stressed conditions) (Cooper *et al.* 2014, Gaffney *et al.* 2015, Masuka *et al.* 2017). Thanks to quantitative genetics, it is well understood that faster improvements will be attained when there is either more genetic variation to start with, or stronger selection. Specific predictions are achieved through the application of general statistical tools, often under simple assumptions of polygenicity and additivity (the infinitesimal model), to large sample sizes (Barton *et al.* 2017).

While application of the ‘Breeder’s equation’ to predict the response of single traits, from their heritability and a defined selection differential, has yielded encouraging results (Walsh & Lynch 2018), p. 607), multivariate selection often fails, even in the controlled environment of the laboratory (Milocco & Salazar‐Ciudad 2020, Rouzic *et al.* 2020) reviewed in (Roff 2007)). In the wild, our aspirations must currently be limited to a number of exceptional, closely-studied systems such as the red deer on the Isle of Rum in Scotland (Bonnet *et al.* 2019), Soay sheep on St Kilda island (Clutton-Brock *et al.* 1991), and bird populations such as great and blue tits (Charmantier *et al.* 2008, 2014, 2016).

7. Epidemiological models (SIR models)

Classical SIR models are used to model the spread of an infectious agent in a population with susceptible, infected and recovered individuals. When we include the possibility of the pathogen to evolve (e.g., changing virulence or infection probabilities), we can predict the spread of an evolving infectious agent (Gordo *et al.* 2009). This type of model has been applied to modeling of influenza evolution (Boni *et al.* 2006).

In the most simple case, evolution in SIR models can be modeled by including two or more versions of the pathogen with different parameters such as the replication number or transmission probability. Such models are used for the SARS-CoV-2 pandemic to predict case numbers in the near future and the spread of new variants. In early 2021, two main virus variants affected the predictions of cases in the near future in Europe: the “wildtype” strain and the alpha variant. In many locations, a simultaneous decline of the wildtype virus and increase of the alpha variant resulted in an initial reduction of total cases, followed by an increase (Fig. 3F). When enough sequencing data was available and R_0 was known for both variants, this U-shaped pattern of cases over time could accurately be predicted. When predictions are made further into the future, they need to account for the possibility that new, currently unknown, variants will emerge (Cobey 2020, Day *et al.* 2020, Kissler *et al.* 2020).

8. Species distributions across space and environmental conditions

Forecasting biodiversity responses to climate change are generally done through species distribution models, which include niche, envelope and bioclimatic models (Waldvogel *et al.* 2020). These models have been used in so-called rewilding, i.e. conservation efforts that include ecological restoration and reintroduction of predators and keystone species. However, species distribution models usually do not include intraspecific variation, adaptive plasticity and evolutionary potential (Jay *et al.* 2012, Fitzpatrick & Edelsparre 2018), and therefore greatly underestimate species range dynamics. Alternative models that include genomic data and evolutionary responses have been developed to predict the (potential) range expansion of *Aedes aegypti* mosquitoes transmitting dengue virus (Kearney *et al.* 2009), to predict coral adaptation to future ocean warming (Bay *et al.* 2017) and to predict future population declines of yellow warblers (*Setophaga petechia*) and to guide effective mitigation efforts for these birds (Bay *et al.* 2018). Using genomic data from a species combined with community-level modeling,

the genomic variation of natural populations that is related to environmental variables can be obtained (Fitzpatrick & Keller 2015). This information can then be used to predict how vulnerable populations are to environmental change, such as climate change.

9. Multi-scale evolutionary modeling

Predictions to test fundamental assumptions of evolving systems often require predictions over very long timescales. These predictions are complicated by the fact that the local mutational landscape changes with the accumulation of many new mutations, and the genotype-phenotype map changes as evolutionary innovations occur. Studying long-term evolution therefore requires models in which the genotype-phenotype map can evolve. An evolvable genotype-phenotype map can be achieved by including more than one level of organization in a model and allowing for the evolution of traits at multiple spatiotemporal scales, leading to multi-scale evolutionary models.

Examples of multi-scale models are: the coupling of large gene regulatory networks to tissue-level patterning, which have been used for hindcasting the order of evolutionary innovations in bilateral animals (Vroomans *et al.* 2016), and estimating the likelihood that mutations increase morphological complexity (Hagolani *et al.* 2021); models of genome evolution (Cuypers & Hogeweg 2012), which predict that genomically complex ancestors primarily adapted through gene loss during major radiation events (Deutekom et al. 2019); and agent based models with rudimentary genomes, which predict emergent selective forces that can drive major evolutionary transitions (Colizzi *et al.* 2020). Multi-scale models can also generate predictions for clinically relevant evolutionary problems, such as HIV evolution (Doekes *et al.* 2017) or tumor progression (Szabó & Merks 2017).

A different scale on which selection can act, is at the level of the community (Doulcier *et al.* 2020, Chang *et al.* 2021). Predicting evolution at the level of communities is a newly emerging field with possible applications in medicine, biotechnology and agriculture. A promising application is for microbial communities that degrade hazardous compounds.

Multi-scale models are also used to study the predictability of evolution. This may either be done directly, e.g., to determine under what conditions evolution is predictable (Meijer *et al.* 2020), and show that although evolutionary paths and detailed characteristics can be hard to predict, in some cases population level traits are predictable (van Dijk *et al.* 2019). More indirectly, they can predict properties of the evolutionary landscape that affect predictability (see Chapter 4). E.g., common predictions of multi-scale models are that ruggedness and smoothness coexist in high dimensional fitness landscapes (e.g. arising from many-genes interactions), and the distribution of fitness effects is biased to high neutrality and high lethality, but with few intermediately deleterious mutants (Hogeweg 2012).

10. Machine learning

In cases where large amounts of data on repeated evolutionary trajectories in the past are available, machine learning approaches are likely to become increasingly important for predicting evolution. However, depending on the particular machine learning approach taken, these methods may or may not increase our understanding of the underlying evolutionary forces. An important limitation for using machine learning to make predictions is that it requires data from a very similar situation because the predictions can usually not be extended to novel situations beyond the data set used for training, in contrast to mechanistic models. One interesting example of machine learning for evolutionary predictions is a study on experimental evolution using the integration of numerous evolve and resequence experiments with *E. coli*. The researchers developed a predictor of the genes that will be modified in the course of adaptation, depending on the *E. coli* strain and the selection pressure (Wang *et al.* 2018). The model could predict around one third of the mutational targets in a new but similar evolve and resequence experiment.

Machine learning methods have also been used to predict the somatic evolution of cancer (Caravagna *et al.* 2018, Gerhauser *et al.* 2018) and success of influenza virus variants (Hayati *et al.* 2020). A particularly promising future direction is the ability for machine learning methods to combine increasingly complex cancer genomic data with data on transcriptome, epigenome and advanced imaging to guide precision medicine (Gerstung *et al.* 2020).

4. Predictability

1. What makes evolution more or less predictable?

Some evolutionary systems are more predictable than others. Broadly speaking, the predictability of evolution increases with the strength and efficiency of selection relative to genetic drift, it increases with increasing population size and it decreases with the noise introduced by the availability of variants or mutations (that is, the number of directions evolution could move towards). It is worth imagining a most predictable scenario, when only some mutations are likely to happen in a population (mutation bias), and only one of those mutations is highly beneficial (narrow distribution of fitness effects, evolution can only go one direction). If, in addition, the population is large, the environment is constant in space and time, and the mutation supply is high, then the only available beneficial mutation is guaranteed to occur and increase in frequency.

In this predictable scenario, the fitness increase of the population, and possibly the increase in population size, can be accurately predicted with a deterministic model. In addition, the mutational pathway is also highly predictable because only one strongly beneficial mutation is available. In reality, there are many reasons why predictability is usually much lower: the mutation supply may be low (stochastic waiting time for a successful beneficial mutation), multiple mutations with different beneficial effects may occur (wider distribution of fitness

effects), fitness effects may be influenced by interactions with other mutations (linkage, epistasis and pleiotropic effects), the population may be small (allele frequencies affected by drift), and the environment may change. Predictability may also be affected by sex, recombination (or lack thereof), species interactions and feedback loops, though we will not discuss this in detail.

The predictability of evolution can be studied most easily in experimental evolution studies, and many such studies have indeed focused on the predictability (or repeatability) of evolution (e.g., (Bull & Molineux 2008, Sackman *et al.* 2017, Lind 2019, Schenk *et al.* 2021). Predictability (in the sense of repeatability) is also studied in the context of parallel or convergent evolution in natural systems. Examples of this approach include the emergence of drug-resistant pathogens in different patients (Lieberman *et al.* 2011, Feder, Pennings, *et al.* 2021), and herbicide resistance in different species or geographic areas (Kreiner *et al.* 2019, Hawkins & Fraaije 2021).

2. Mutational supply and mutation bias

We first consider the influence of mutational supply and mutation bias. Mutational supply is the total number of mutations that occur in a generation or unit of time in a population, determined by the population size and the mutation rate. Mutation bias on the other hand is the difference in mutation rates at classes of genomic sites. In general, high mutational supply and strong mutational bias both increase predictability of evolution, both in terms of the mutational path taken and in terms of fitness increase (Stoltzfus & Yampolsky 2009, Storz 2016, Storz *et al.* 2019).

When the mutational supply is high, that is many mutations occur every generation, then the mutation with the highest selection coefficient will likely fix in the population either because it starts increasing in frequency first or because it outcompetes other mutations in a process called clonal interference (Gerrish & Lenski 1998, Schenk *et al.* 2021). For example, (Feder, Pennings, *et al.* 2021) show that in 19 of 20 patients infected with HIV and treated with a single drug (3TC), drug resistance in these viral populations evolves quickly through exactly the same mutation (M184V) in the reverse transcriptase gene, even though the M184I mutation usually occurs first (Keulen *et al.* 1997).

When mutational supply is weaker, mutation biases play an important role. Mutational biases exist in every genome. For example, mutation rates may be higher at CpG sites in eukaryotes (Duncan & Miller 1980), and transition mutations can be up to ten times more common than transversion mutations in viral genomes (Mansky & Temin 1995). Theory predicts that in the absence of selection (i.e., under neutrality), mutation bias is the only driver of parallel evolution (Kimura 1983). But even when selection occurs, a strong mutational bias reduces the spectrum of mutations available for selection, and should therefore increase predictability. Somewhat counter-intuitively, when selection is very strong for multiple possible mutations, mutation bias is again as important as it is under neutrality (Stoltzfus 2021). An elegant study on adaptation to high altitudes in birds found parallel evolution, in part due to mutation bias at CpG sites (Storz *et*

al. 2019). In a wide range of taxa, such biases at least partly explain parallel genetic adaptation (Stern & Orgogozo 2008, Bailey *et al.* 2017, 2018, Stoltzfus & McCandlish 2017).

In sufficiently small populations where beneficial mutations are substituted one by one (i.e. the strong-selection, weak-mutation regime), variation in both the mutation supply rate and fitness effects of mutations contribute to the probability that a mutation occurs and fixes in the population (Orr 2005, Storz 2016). This is because when highly beneficial mutations are under-sampled due to the existing mutational bias, other smaller-effect but more frequent mutations may fix instead. Such a pattern was observed in replicate evolving populations of bacteriophage genotypes (Sackman *et al.* 2017) where the mutation with the largest fitness effect was not the one that reached fixation most often, because its mutation rate was lower than that of other mutations with smaller fitness effects. In larger populations, when mutational supply is high so that multiple beneficial alleles are present simultaneously (i.e., the clonal interference regime), selection bias is expected to dominate over mutation bias, and fix the most beneficial mutations largely independent of their mutation rate (Szendro *et al.* 2013, Bailey *et al.* 2017, Pennings *et al.* 2020, Pinheiro *et al.* 2021).

3. Distribution of fitness effects (DFE) of mutations.

The distribution of fitness effects of new mutations tells us what percentage of mutations have what fitness effect. While most mutations are deleterious or neutral in most populations, it's the distribution of beneficial fitness effects that is usually of greatest importance in the dynamics of adapting populations. If there exists only one or a few potential mutations with a large fitness effect, it makes the mutational path of an evolving population more predictable; but at the same time, if the mutational supply is low, having only a few potential large effect mutations means that the waiting time for one of these mutations to occur may be long, making the increase in fitness per time unit unpredictable (Orr 2005).

The effects of the shape of the DFE on predictability have largely been addressed in theoretical work assuming strong-selection weak-mutation conditions, with a focus on beneficial mutations driving adaptive evolution. Using extreme value theory, it has been shown that regardless of the specific shape of the entire DFE (i.e. including deleterious, neutral, and beneficial mutations), there will always be many more small than large-effect mutations. This reduces predictability because the more numerous small-effect mutations may collectively have a similar fixation probability compared to the small set of large-effect mutations (Joyce *et al.* 2008).

Studies that directly test the fitness effects of beneficial mutations and characterize their distribution tend to find exponential or exponential-like right-truncated distributions (reviewed in (Bataillon & Bailey 2014). A study on bacteriophages (Sackman *et al.* 2017) experimentally quantified the fitness effects of new mutations and then used the shape of the quantified DFE and number of beneficial mutations to predict the probability of parallel evolution (eqn 37 in (Joyce *et al.* 2008)), comparing those estimates to observed measures of parallel evolution within the same system. Encouragingly, these authors found that (1) their estimates of the probability of parallel evolution fell within the 95% confidence intervals for observed parallel

evolution measures, and (2) including the DFEs shape parameter improved estimates, providing support for the idea that DFEs are important drivers of evolutionary predictability.

4. Epistasis

Our discussion above implicitly assumed that the effects of mutations are independent of each other, and that adaptation primarily occurs via single, large-effect, beneficial mutations. However, in reality, mutations are not independent and the interactions between mutations (epistasis) introduce ruggedness in fitness landscapes. Generally speaking, epistasis reduces predictability, because even if fitness effects are measured in one genetic background we don't know the effects in another background (Milton & Tokuriki 2016).
However, there are interesting nuances; e.g., when epistatic interactions change the sign of mutational effects from beneficial to deleterious or vice versa, a condition referred to as sign epistasis (Weinreich *et al.* 2005). A seminal study on the antibiotic resistance enzyme TEM-1 β-lactamase showed that sign epistasis can strongly reduce the number of mutational pathways along which a population can evolve towards higher fitness (Weinreich *et al.* 2006). In this sense, sign epistasis increases the predictability of evolutionary trajectories. However, sign epistasis can also lead to fitness landscapes with multiple peaks (Poelwijk *et al.* 2011), which means that populations can end up moving towards different fitness peaks depending on which mutation fixes first, thus decreasing the predictability of evolution. Moreover, biophysical models of high-dimensional fitness landscapes (with many loci) show that epistasis and neutral pathways can coexist (Huynen *et al.* 1996, Ciliberti *et al.* 2007). This results in low long-term predictability at the sequence level, but higher predictability of some features of the evolutionary trajectory, such as the presence of punctuated equilibria and the evolution of mutational robustness.

5. Evolutionary control

1. Evolutionary control needs predictions

Evolutionary control is the alteration of an evolutionary process with a specific purpose. Control can either suppress evolution, e.g., to prevent pathogens evolving drug resistance, or facilitate evolution, e.g., to increase the range of a species to avoid extinction. There are general measures to achieve these goals, but measures can be more targeted if we can predict their effects on evolution.

Using predictions for control requires an extension of the scope of the prediction: we have to predict not only the evolutionary process under natural conditions, but also its response to specific control interventions. Important applications of evolutionary control are medical interventions against evolving human pathogens such as HIV (Lässig & Mustonen 2020). The case of HIV shows that predictions don't have to be very precise in order to allow some control. When triple-drug therapy became available for HIV patients in the 1990s, the prediction – based on a mathematical model – was that it would slow down the evolution of drug resistance,

reducing the progression to AIDS and saving lives. Even though the model that was used wasn't entirely correct, the high level predictions held, and many lives were saved (Perelson & Nelson 1999, Rocheleau *et al.* 2018, Feder, Harper, *et al.* 2021).
In this section we will describe several examples of evolutionary control, where predictions are used or could prove beneficial to improve control measures.

2. Health measures during epidemics influence evolution

For epidemics, public health interventions that aim to reduce transmission and therefore infections are aimed to reduce morbidity and mortality. However, they also reduce the amount of replicating pathogen in the population and can therefore change the evolutionary dynamics of the pathogen. For example, in the case of SARS-Cov2, it is thought that public health measures like physical distancing and vaccination will reduce the population size of the virus and therefore the rate at which new strains emerge. This means that vaccination campaigns and physical distancing rules can be seen, at least in part, as an evolutionary control measure.

3. Preventing or reversing antibiotic resistance in bacterial pathogens

Increasing rates of antibiotic resistance threaten the efficacy of this mainstay of treatment for bacterial disease. Because the discovery and development of novel antimicrobial agents lags behind the rate of resistance evolution, newer approaches that focus on antimicrobial stewardship have emerged whose aim is to reverse or prevent resistance evolution to existing drugs (Read & Woods 2014, Nichol *et al.* 2015, Perron *et al.* 2015, Andersson *et al.* 2020). These ideas, which can be implemented for individuals or at the population level in a hospital or agricultural setting, fundamentally rely on an accurate, empirical understanding of antimicrobial resistance evolution and spread. Two broad categories of evolutionary predictions to inform therapeutic decisions can be envisioned: one to avoid a specific outcome and another to promote one.

For example, several models and experiments have supported the intuitive predictions that combination therapies decrease the likelihood of resistance evolution compared to monotherapy. First, combination therapies increase the rate of pathogen decline, limiting the time window for *de novo* resistance mutations to occur. Second, they require resistance mutations in multiple targets (though see (Feder, Harper, *et al.* 2021). Combination therapies encompass a wide range of approaches: using multiple antimicrobial agents simultaneously (e.g. antibiotic-antibiotic, antibiotic-adjuvant, antibiotic-phage or phage cocktails); antibiotic mixing, that is, random allocation of different antibiotics for different patients in the same hospital ward; and implementing temporally alternating therapies including antibiotic cycling (population level) and sequential therapy (individual level) (Sarraf-Yazdi *et al.* 2012, Abel zur Wiesch *et al.* 2014, Nichol *et al.* 2015, Yen & Papin 2017, Tyers & Wright 2019).

More interesting are approaches that would be used to drive a particular outcome. These are based largely on known epistatic or pleiotropic effects of resistance mutations. Resistance

mutations are typically associated with fitness costs, and several authors have promoted treatment strategies designed to maximize these costs, thus maximizing the probability that these strains are replaced once antibiotic selection is relaxed (Andersson & Hughes 2010). Interactions between resistance mutations can also be used to exploit drug synergies, thereby driving faster rates of population decline in the pathogen. Finally, a recent strategy is based on the idea that resistance to a given drug pleiotropically increases susceptibility (i.e. causes collateral sensitivity) to a second drug (Sommer *et al.* 2017). Knowledge of collateral sensitivity could help choose drugs to be used sequentially: if resistance to one drug evolves, it concomitantly increases the efficacy of the other one. Note however that epistatic effects (when a mutation's effect depends on the genetic background) can make it harder to predict, and thus use, collateral sensitivity (Barbosa *et al.* 2019, Hernando-Amado *et al.* 2020).

4. Insect resistance to transgenic plants

Evolutionary predictions have been used to guide the deployment of the most successful transgenic plants designed to protect crops against insect damage. Cotton, maize, and other crops have been transformed with the genes for insecticidal protein toxins from the bacterium *Bacillus thuringiensis* (Bt) and have been remarkably successful as an alternative to sprays with chemical insecticides (Tabashnik *et al.* 2013). From the beginning, the evolution of insect resistance to Bt toxins was anticipated and deployment strategies were deliberately designed to minimize its spread. For example, toxins were chosen for which resistance was known to be recessive, and for which standing genetic variation for resistance was shown to be low. This was done because population genetic models predict that evolution will proceed more slowly for recessive alleles or when alleles are rare. Population genetic models of natural selection were also used to predict how soon resistance would evolve, based on assumptions about the genetics of resistance, the strength of selection, pest dispersal among Bt and non-Bt crops, and other genetic and ecological factors (Roush *et al.* 1998). Strategies based on these predictions were mandated by governmental regulatory agencies, supported by the agricultural biotechnology industries responsible for the transgenic plants, and adopted by growers. Retrospective analysis of the efficacy of these strategies has shown that they were more successful when more of the underlying assumptions were valid in the field, and when grower compliance was high (Tabashnik *et al.* 2013). This is probably the most successful test of the performance of evolutionary predictions in modern agriculture.

5. Prevention of resistance to gene drives

Gene drive systems use CRISPR/Cas9-based genetic constructs to force the spread of a trait into a population (see Box 1). These systems are being developed to, one day, change wild populations (e.g., make mosquitoes resistant to malaria), but the target species can evolve to become resistant to the drive allele. Population genetic methods are used to predict how fast resistance would evolve against such a system and to design resistance-proof systems. One gene drive system is a so-called homing drive, in which the gene drive constructs cleave target sequences that are then repaired by copying the gene drive construct (using homologous recombination), which leads to an increase in the frequency of the gene drive construct in the

population. Resistance to such gene drive systems can evolve when the cut that is made by the gene drive construct is repaired by the non-homologous end-joining pathway. This is because end-joining often leads to changes in the target sequence, which means that the drive construct can no longer cleave that sequence (Champer *et al.* 2020, Gomulkiewicz *et al.* 2021). When resistance has thus evolved, drive construct can no longer spread and may go down in frequency if it comes with a fitness cost.

Because the evolution of resistance to a gene drive element is a key obstacle to deploying gene drive systems, research is focused on evolutionary control by reducing the rate of evolution of resistance (Unckless *et al.* 2017). The susceptibility of a gene drive construct to resistance can be reduced by (a) increasing the number of sites at which the gene drive construct can cut (Champer *et al.* 2020) or by (b) introducing a cost to resistance by targeting the gene drive to an essential gene (Noble *et al.* 2017). Predictions have been validated by laboratory experiments: several groups have shown that gene drives with multiple guide RNAs (which target multiple sites), or that target an essential gene, can spread in a population for much longer before resistance evolves (Kyrou *et al.* 2018, Champer *et al.* 2020, Kandul *et al.* 2021). In an important example of successful evolutionary control – at least in the lab – researchers from Imperial College London were able to create a gene drive system that targeted an important fertility gene (*doublesex*) in 2018. With this gene drive system no resistance evolved and the experimental mosquito populations all went extinct as hoped and predicted (Kyrou *et al.* 2018).

6. Preventing extinction by promoting evolution

One case where we want to promote evolution is to rescue a species from extinction. It is widely thought that lack of genetic variation increases extinction risk. For example, the Tasmanian devil had very low genetic variation and its population size was severely reduced by an infectious cancer (Hendricks *et al.* 2017). At the same time, loss of habitat, loss of dispersal opportunities and decrease in population size can lead to lower genetic variation in a species leading to an “extinction vortex” (Olivieri *et al.* 2015). The main method of promoting adaptation to a changed environment is therefore by increasing genetic diversity, as genetic diversity has shown to be beneficial for adaptation and rescue (Hughes *et al.* 2008, Agashe 2009, Agashe *et al.* 2011). Maintaining genetic variation also plays a role when breeding programs are used to rescue a population (Ebenhard 1995). Increasing genetic diversity in endangered populations, termed genetic rescue, is a promising intervention, but whether it predictably leads to increases in population sizes to prevent extinction remains to be seen (Bell 2017).

Box 1: Examples

1. SARS-CoV-2

SARS-CoV-2, a coronavirus that was first detected infecting humans in late 2019, is a source for evolutionary predictions at the phenotypic and genetic level. At first, the main aim was to predict total case numbers. Next, when multiple strains arose, the dynamics of different strains became important (see Fig. 3F). In the longer term, and especially when vaccines came into play, it became important to predict how the phenotype and also the genotype of the virus changes. It is common for viruses to adapt further after a host switch. Using SIR models, predictions for traits under selection can be made. The strongest selection is expected to be on increased transmission, whereas weaker selection is predicted on other disease life-history traits, such as the lengthening of the pre-symptomatic case, decrease of asymptomatic disease (because asymptomatic disease seems to account for lower number of new cases) and decreased virulence. The conclusion is that direct selection on virulence is very weak, so if this trait changes it is most likely an effect of pleiotropy. But the pleiotropic effects are not well understood which makes it hard to predict how virulence will evolve. At the genotypic level, analyzing different coronaviruses, (Armijos-Jaramillo *et al.* 2020) found largely purifying selection on the spike protein (which the virus uses to enter host cells). However, they also identify a few positions where they expect potentially beneficial mutations to occur, information that can be used to assess if observed mutations are indeed adaptive and could perhaps be taken into account for vaccine development.

The extremely high global prevalence of this virus as well as the recent host change and change in selective pressures (such as with vaccines) as well as the unprecedented scale of data that is available, make it possible to quickly test the accuracy of many predictions. As the new variants in 2021 mostly increased transmission, the phenotypic predictions can be deemed accurate, although the lack of a timeline in the predictions make it difficult to define accuracy. For the genetic changes, the predicted positions where mutations were expected indeed seems to be borne out in the new variants (Harvey *et al.* 2021).

2. Sustainable product formation by microorganisms

Microorganisms can produce (bio)chemical compounds sustainably. The major challenge is that compound production diverts resources from growth, and therefore the organisms are likely to evolve reduced product formation. We can predict in which cases evolutionary loss of product formation is least likely, for example when product formation is coupled to biomass production. Computational techniques that use genome-scale metabolic models can also predict which gene knock-outs couple product formation to biomass, which should lead to increased stability of product formation (Du *et al.* 2018). These predictions were applied for formate production by cyanobacteria. Experimental validation showed that under growth coupling, no decrease of product formation occurred within a month, whereas without growth coupling formate production decreased after days (Du *et al.* 2019).

3. Adaptation of natural populations

Ideally, we want to be able to predict which species or populations can adapt to changing conditions and which will be threatened with extinction, to plan our conservation efforts most efficiently. These types of predictions are difficult to make, but four approaches are taken to enable forecasting of future population states. (1) Apply a certain selection pressure in the lab and observe the adaptive potential (e.g. under size selective harvesting (Uusi-Heikkilä *et al.* 2015) or higher temperatures (Kellermann *et al.* 2009, Morgan *et al.* 2020)). (2) Similarly, species can be transplanted to different habitats and their adaptation monitored ((Colautti & Barrett 2013) and see (Edwards 2015) for a review). (3) We can determine if ongoing evolution has already led to adaptive change (e.g. a recent metaanalysis of evolutionary adaptation of birds to climate change showed that it is unlikely that adaptation will rescue populations (Radchuk *et al.* 2019); and contrary to predictions it does not seem that Atlantic Cod has adapted to fishing (Pinsky *et al.* 2021)). (4) We can use species distribution models along with genomic information (Hoffmann *et al.* 2015). Predictions for the last type have been made for the dwarf birch (Borrell *et al.* 2020), oaks (Rellstab *et al.* 2016) and the yellow warbler (Bay *et al.* 2018).

Approach (1) is difficult or impossible and time-consuming for many species and it remains to be seen how well these predictions can be extrapolated to natural systems. The latter problem is reduced in approach (2). Approach (3) can only be applied for changes that are already happening and needs a large amount of long-term data on individuals. Finally, approach (4) has usually yielded very weak correlations. Therefore, predicting which species may adapt to novel environmental conditions (e.g. rapid climate change) remains a big challenge.

4. CRISPR-based Gene drives

Gene drives or gene drive constructs are a special type of segregation distorter that use the CRISPR/Cas9 gene editing technology. Under Mendelian expectations, each gamete has a 50% chance to carry the allele that came from one parent and 50% chance to carry the allele that came from the other parent. One type of gene drives are called homing gene drives. They use a kind of copy-paste mechanism that distorts the 50-50 rule and can be present in many more gametes, which then leads to offspring that inherits the gene drive allele at a rate much greater than 50%. This works as follows: when the gene drive construct is initially present in the heterozygous state in a cell, it can cleave a genomic target site in the chromosome that doesn't carry the construct. This cut then induces the cell to repair the damage by copying the drive sequence into the damaged chromosome. The result is that the cell now has two copies of the gene drive allele. In this way, the gene drive construct can rapidly spread in the germline and therefore in the population. When the gene drive system is linked to an allele of interest (the "payload"), the transmission bias forces the spread of that allele. While CRISPR/Cas gene drives are not yet used outside the laboratory, there are plans to use this technology in mosquitoes and other species that cause harm to humans.

Evolutionary predictions at two different levels are of interest here: 1) how fast will the drive allele spread in the population, and 2) when will resistance to the gene drive evolve and spread? The first is a fairly straightforward application of existing population genetic and population dynamic models, with an additional parameter for non-Mendelian segregation. However, in early experiments, the populations almost always became resistant to the gene drive element. The second level of prediction is therefore possibly more important: how fast will resistance evolve (Unckless *et al.* 2017, Dhole *et al.* 2020, Gomulkiewicz *et al.* 2021)?

5. Influenza

One of the best known examples of evolutionary predictions is predictions of which influenza strains will be common the next season, as a basis for vaccine development. Two main methods, that can also be combined, are used to make these predictions: genealogical trees and molecular properties of the virus (Morris *et al.* 2018). The first method uses data from recent clinical samples and makes a phylogenetic tree of these strains. The "bushy" parts that have a lot of recent diversification are the expected genotypes that will dominate next year (Neher *et al.* 2014). The second method uses specific molecular details of virus attachment to narrow down the entire genotype-fitness map to a short stretch of base pairs. From this short section of the HA protein, the physics of protein stability and binding to the human antibodies sum up to the fitness of the virus particle (Łuksza & Lässig 2014). Even though these methods are in use, success rates are still limited and further improvement would have a large impact on the effectiveness of influenza vaccines. Such improvement can come from a better understanding of the genotype-phenotype map for virus-antibody interactions, such that antigenic evolution can be predicted better from sequence data.

---

6. Conclusion and outlook

Evolutionary predictions are used in many fields, including infectious disease, biotechnology and conservation biology. In some cases, the use of evolutionary models that include mutations, selection and drift is very explicit (such as in the gene-drive example or in the influenza vaccine predictions), whereas in other cases evolution may only be implicitly included in population size predictions (such as for the SARS-CoV2 epidemic or predicting extinction risk for endangered

species). In this review, we have shown how evolutionary predictions are used in many biological sub fields. Because researchers in different subfields use different languages, it is not always obvious that they are, in essence, doing the same thing: predicting the future state of an evolving population. Predictions can be improved when researchers can learn and be inspired by results from other subfields, but for this to happen, we need to use a common language. This review is meant as a start to bridge those research communities. We believe that those who work on preventing evolution (in biotechnology, agriculture and health) and also those who work on promoting or steering evolution (such as in conservation biology and biotechnology) could benefit from much more extensive communication.

Most of the theory we described in this review was featured in chapter 4 on predictability. This reflects the relative early stage of the field, where the researchers involved focus on understanding under which conditions evolution is more or less predictable, and which factors drive predictability. While this has led to many relevant insights, these theoretical insights are still far away from the applications where predictions are needed. We believe that the field could benefit from a stronger connection between theory and applications. Specifically, the efforts that are underway in the area of influenza research (using data and theory to predict influenza strains to design the best vaccine) could be replicated in other situations. For example, models could be made to predict drug resistance in a hospital over the next year, given the current state and parameters such as antibiotic use. Also, evolutionary models could more explicitly be used to predict population sizes for red-listed species a year from now, or to predict which tumors will recur with resistance. By applying evolutionary methods to real life situations, we will discover the gaps in our knowledge and contribute to making evolutionary predictions more accurate and useful.

Finally, we expect that increasing access to genome-centered information, and the use of modern statistical techniques, including machine learning, will improve evolutionary predictions in amenable taxa over the next few years. However, there will be a continued need to develop mechanistic models of evolution for various systems, for at least two fundamental reasons: (i) to extrapolate predictions to conditions other than those used to parameterize specific models, and (ii) to further our understanding of how evolution works. While the increasing access to high-resolution phenotype and genotype data make it tempting to include all these details in such models, more coarse-grained mechanistic models may allow more powerful predictions. We hope that together, improved collaboration, a shared language, and new combinations of methods will lead to further maturation of the field, leading to evolutionary predictions becoming mainstream in areas such as infectious disease, conservation biology and biotechnology.

7. Data Archiving Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

8. Bibliography

Abel zur Wiesch, P., R. Kouyos, S. Abel, W. Viechtbauer, and S. Bonhoeffer. 2014. Cycling empirical antibiotic therapy in hospitals: meta-analysis and models. PLoS Pathog. 10: e1004225.

Agashe, D. 2009. The stabilizing effect of intraspecific genetic variation on population dynamics in novel and ancestral habitats. Am. Nat. 174: 255–267.

Agashe, D., J. J. Falk, and D. I. Bolnick. 2011. Effects of Founding Genetic Variation on Adaptation to a Novel Resource. Evolution 65: 2481–2491.

Andersson, D. I., N. Q. Balaban, F. Baquero, P. Courvalin, P. Glaser, U. Gophna, R. Kishony, S. Molin, and T. Tønjum. 2020. Antibiotic resistance: turning evolutionary principles into clinical reality. FEMS Microbiol. Rev. 44: 171–188.

Andersson, D. I., and D. Hughes. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance? Nat. Rev. Microbiol. 8: 260–271.

Armijos-Jaramillo, V., J. Yeager, C. Muslin, and Y. Perez-Castillo. 2020. SARS-CoV-2, an evolutionary perspective of interaction with human ACE2 reveals undiscovered amino acids necessary for complex stability. Evol. Appl. 13: 2168–2178.

Bailey, S. F., F. Blanquart, T. Bataillon, and R. Kassen. 2017. What drives parallel evolution? BioEssays 39: e201600176.

Bailey, S. F., Q. Guo, and T. Bataillon. 2018. Identifying Drivers of Parallel Evolution: A Regression Model Approach. Genome Biol. Evol. 10: 2801–2812.

Barbosa, C., R. Römhild, P. Rosenstiel, and H. Schulenburg. 2019. Evolutionary stability of collateral sensitivity to antibiotics in the model pathogen Pseudomonas aeruginosa P. J. Wittkopp, C. Pal, B. Csorgo, C. MacLean, and P. Johnsen (Eds.). eLife 8: e51481.

Barrat-Charlaix, P., J. Huddleston, T. Bedford, and R. A. Neher. 2020. Limited predictability of amino acid substitutions in seasonal influenza viruses. bioRxiv 2020.07.31.231100.

Barroso-Batista, J., A. Sousa, M. Lourenço, M.-L. Bergman, D. Sobral, J. Demengeot, K. B. Xavier, and I. Gordo. 2014. The First Steps of Adaptation of Escherichia coli to the Gut Are Dominated by Soft Sweeps. PLOS Genet. 10: e1004182.

Barton, N. H., A. M. Etheridge, and A. Véber. 2017. The infinitesimal model: Definition, derivation, and implications. Theor. Popul. Biol. 118: 50–73.

Bataillon, T., and S. F. Bailey. 2014. Effects of new mutations on fitness: insights from models and data. Ann. N. Y. Acad. Sci. 1320: 76–92.

Bay, R. A., R. J. Harrigan, V. L. Underwood, H. L. Gibbs, T. B. Smith, and K. Ruegg. 2018. Genomic signals of selection predict climate-driven population declines in a migratory bird. Science 359: 83–86.

Bay, R. A., N. H. Rose, C. A. Logan, and S. R. Palumbi. 2017. Genomic models predict successful coral adaptation if future ocean warming rates are reduced. Sci. Adv. 3: e1701413.

Bell, G. 2017. Evolutionary Rescue. Annu. Rev. Ecol. Evol. Syst. 48: 605–627.

Blount, Z. D., R. E. Lenski, and J. B. Losos. 2018. Contingency and determinism in evolution: Replaying life's tape. Science 362: eaam5979.

Boni, M. F., J. R. Gog, V. Andreasen, and M. W. Feldman. 2006. Epidemic dynamics and antigenic evolution in a single season of influenza A. Proc. R. Soc. B Biol. Sci. 273: 1307–1316.

Bonnet, T., M. B. Morrissey, A. Morris, S. Morris, T. H. Clutton-Brock, J. M. Pemberton, and L. E. B. Kruuk. 2019. The role of selection and evolution in changing parturition date in a red deer population. PLOS Biol. 17: e3000493.

Borrell, J. S., J. Zohren, R. A. Nichols, and R. J. A. Buggs. 2020. Genomic assessment of local adaptation in dwarf birch to inform assisted gene flow. Evol. Appl. 13: 161–175.

Bull, J. J., and I. J. Molineux. 2008. Predicting evolution from genomics: experimental evolution of bacteriophage T7. Heredity 100: 453–463.

Burke, M. K., and M. R. Rose. 2009. Experimental evolution with Drosophila. Am. J. Physiol.-Regul. Integr. Comp. Physiol. 296: R1847–R1854.

Caravagna, G., Y. Giarratano, D. Ramazzotti, I. Tomlinson, T. A. Graham, G. Sanguinetti, and A. Sottoriva. 2018. Detecting repeated cancer evolution from multi-region tumor sequencing data. Nat. Methods 15: 707–714.

Champer, J., E. Yang, E. Lee, J. Liu, A. G. Clark, and P. W. Messer. 2020. A CRISPR homing gene drive targeting a haplolethal gene removes resistance alleles and successfully spreads through a cage population. Proc. Natl. Acad. Sci. 117: 24377–24383.

Chang, C.-Y., J. C. C. Vila, M. Bender, R. Li, M. C. Mankowski, M. Bassette, J. Borden, S. Golfier, P. G. L. Sanchez, R. Waymack, X. Zhu, J. Diaz-Colunga, S. Estrela, M. Rebolleda-Gomez, and A. Sanchez. 2021. Engineering complex communities by directed evolution. Nat. Ecol. Evol. 1–13.

Charmantier, A., C. Doutrelant, G. Dubuc-Messier, A. Fargevieille, and M. Szulkin. 2016. Mediterranean blue tits as a case study of local adaptation. Evol. Appl. 9: 135–152.

Charmantier, A., D. Garant, and L. E. B. Kruuk. 2014. Quantitative Genetics in the Wild. OUP Oxford.

Charmantier, A., R. H. McCleery, L. R. Cole, C. Perrins, L. E. B. Kruuk, and B. C. Sheldon. 2008. Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. Science 320: 800–803.

Chen, Y., and J. Nielsen. 2021. Mathematical modeling of proteome constraints within metabolism. Curr. Opin. Syst. Biol. 25: 50–56.

Ciliberti, S., O. C. Martin, and A. Wagner. 2007. Robustness Can Evolve Gradually in Complex Regulatory Gene Networks with Varying Topology. PLOS Comput. Biol. 3: e15.

Clutton-Brock, T. H., O. F. Price, S. D. Albon, and P. A. Jewell. 1991. Persistent Instability and Population Regulation in Soay Sheep. J. Anim. Ecol. 60: 593–608.

Cobey, S. 2020. Modeling infectious disease dynamics. Science 368: 713–714.

Colautti, R. I., and S. C. H. Barrett. 2013. Rapid Adaptation to Climate Facilitates Range Expansion of an Invasive Plant. Science 342: 364–366.

Colizzi, E. S., R. M. Vroomans, and R. M. Merks. 2020. Evolution of multicellularity by collective integration of spatial information R. E. Goldstein and A. M. Walczak (Eds.). eLife 9: e56349.

Cooper, M., C. D. Messina, D. Podlich, L. R. Totir, A. Baumgarten, N. J. Hausmann, D. Wright, and G. Graham. 2014. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65: 311.

Cote-Hammarlof, P. A., I. Fragata, J. Flynn, D. Mavor, K. B. Zeldovich, C. Bank, and D. N. A. Bolon. 2021. The Adaptive Potential of the Middle Domain of Yeast Hsp90. Mol. Biol. Evol. 38: 368–379.

Couce, A., and O. A. Tenaillon. 2015. The rule of declining adaptability in microbial evolution experiments. Front. Genet. 6. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2015.00099/full [Accessed July 5, 2021].

Cuypers, T. D., and P. Hogeweg. 2012. Virtual Genomes in Flux: An Interplay of Neutrality and Adaptability Explains Genome Expansion and Streamlining. Genome Biol. Evol. 4: 212–229.

Daalman, W. K.-G., and L. Laan. 2020. Predicting an epistasis-rich genotype-phenotype map with a coarse-grained bottom-up model of budding yeast polarity. bioRxiv 2020.11.09.374363.

Day, T., S. Gandon, S. Lion, and S. P. Otto. 2020. On the evolutionary epidemiology of SARS-CoV-2. Curr. Biol. 30: R849–R857.

Dhole, S., A. L. Lloyd, and F. Gould. 2020. Gene Drive Dynamics in Natural Populations: The

Importance of Density Dependence, Space, and Sex. Annu. Rev. Ecol. Evol. Syst. 51: 505–531.

van Dijk, B., J. Meijer, T. D. Cuypers, and P. Hogeweg. 2019. Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives. BMC Evol. Biol. 19: 201.

Doekes, H. M., C. Fraser, and K. A. Lythgoe. 2017. Effect of the Latent Reservoir on the Evolution of HIV at the Within- and Between-Host Levels. PLOS Comput. Biol. 13: e1005228.

Doulcier, G., A. Lambert, S. De Monte, and P. B. Rainey. 2020. Eco-evolutionary dynamics of nested Darwinian populations and the emergence of community-level heredity W. Shou, P. J. Wittkopp, W. Shou, S. Mitri, and A. Sanchez (Eds.). eLife 9: e53433.

Du, W., J. A. Jongbloets, C. van Boxtel, H. P. Hernández, D. Lips, B. G. Oliver, K. J. Hellingwerf, and F. B. Dos Santos. 2018. Alignment of microbial fitness with engineered product formation: obligatory coupling between acetate production and photoautotrophic growth. Biotechnol. Biofuels 11: 1–13.

Du, W., J. A. Jongbloets, M. Guillaume, B. van de Putte, B. Battaglino, K. J. Hellingwerf, and F. Branco dos Santos. 2019. Exploiting Day-and Night-Time Metabolism of Synechocystis sp. PCC 6803 for Fitness-Coupled Fumarate Production around the Clock. ACS Synth. Biol. 8: 2263–2269.

Duncan, B. K., and J. H. Miller. 1980. Mutagenic deamination of cytosine residues in DNA. Nature 287: 560–561.

Ebenhard, T. 1995. Conservation breeding as a tool for saving animal species from extinction. Trends Ecol. Evol. 10: 438–443.

Edwards, C. E. 2015. Looking to the future of conservation genetics: The case for using quantitative genetic experiments to estimate the ability of rare plants to withstand climate change. Am. J. Bot. 102: 1011–1013.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8: 610–618.

Feder, A. F., K. N. Harper, C. J. Brumme, and P. S. Pennings. 2021. Understanding patterns of HIV multi-drug resistance through models of temporal and spatial drug heterogeneity G. H. Perry, D. Gifford, and K. Atkins (Eds.). eLife 10: e69032.

Feder, A. F., P. S. Pennings, and D. A. Petrov. 2021. The clarifying role of time series data in the population genetics of HIV. PLOS Genet. 17: e1009050.

Fitzpatrick, M. C., and S. R. Keller. 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. 18: 1–16.

Fitzpatrick, M. J., and A. H. Edelsparre. 2018. The genomics of climate change. Science 359: 29–30.

Fowler, D. M., and S. Fields. 2014. Deep mutational scanning: a new style of protein science. Nat. Methods 11: 801–807.

Fragata, I., A. Blanckaert, M. A. Dias Louro, D. A. Liberles, and C. Bank. 2019. Evolution in the light of fitness landscape theory. Trends Ecol. Evol. 34: 69–82.

Gaffney, J., J. Schussler, C. Löffler, W. Cai, S. Paszkiewicz, C. Messina, J. Groeteke, J. Keaschall, and M. Cooper. 2015. Industry-Scale Evaluation of Maize Hybrids Selected for Increased Yield in Drought-Stress Conditions of the US Corn Belt. Crop Sci. 55: 1608–1618.

Gerhauser, C. et al. 2018. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell 34: 996-1011.e8.

Gerrish, P. J., and R. E. Lenski. 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102: 127.

Gerstung, M. et al. 2020. The evolutionary history of 2,658 cancers. Nature 578: 122–128.

Gillespie, J. H. 1983. A simple stochastic gene substitution model. Theor. Popul. Biol. 23: 202–215.

Goelzer, A., J. Muntel, V. Chubukov, M. Jules, E. Prestel, R. Nölker, M. Mariadassou, S. Aymerich, M. Hecker, P. Noirot, D. Becher, and V. Fromion. 2015. Quantitative prediction of genome-wide resource allocation in bacteria. Metab. Eng. 32: 232–243.

Gomulkiewicz, R., M. L. Thies, and J. J. Bull. 2021. Evading resistance to gene drives. Genetics 217. Available at: https://doi.org/10.1093/genetics/iyaa040 [Accessed July 7, 2021].

Gordo, I., M. G. M. Gomes, D. G. Reis, and P. R. A. Campos. 2009. Genetic Diversity in the SIR Model of Pathogen Evolution. PLOS ONE 4: e4876.

Gould, S. J. 1990. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton & Company.

Govaert, L., E. A. Fronhofer, S. Lion, C. Eizaguirre, D. Bonte, M. Egas, A. P. Hendry, A. D. B. Martins, C. J. Melián, J. A. M. Raeymaekers, I. I. Ratikainen, B.-E. Saether, J. A. Schweitzer, and B. Matthews. 2019. Eco-evolutionary feedbacks—Theoretical models and perspectives. Funct. Ecol. 33: 13–30.

Grant, P. R., and B. R. Grant. 2002. Unpredictable Evolution in a 30-Year Study of Darwin's Finches. Science 296: 707–711.

Gu, C., G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee. 2019. Current status and applications of genome-scale metabolic models. Genome Biol. 20: 121.

Hagolani, P. F., R. Zimm, R. Vroomans, and I. Salazar-Ciudad. 2021. On the evolution and development of morphological complexity: A view from gene regulatory networks. PLOS Comput. Biol. 17: e1008570.

Harvey, W. T., A. M. Carabelli, B. Jackson, R. K. Gupta, E. C. Thomson, E. M. Harrison, C. Ludden, R. Reeve, A. Rambaut, S. J. Peacock, and D. L. Robertson. 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nat. Rev. Microbiol. 19: 409–424.

Hawkins, N. J., and B. A. Fraaije. 2021. Contrasting levels of genetic predictability in the evolution of resistance to major classes of fungicides. Mol. Ecol.

Hayati, M., P. Biller, and C. Colijn. 2020. Predicting the short-term success of human influenza virus variants with machine learning. Proc. R. Soc. B Biol. Sci. 287: 20200319.

Hendricks, S., B. Epstein, B. Schönfeld, C. Wiench, R. Hamede, M. Jones, A. Storfer, and P. Hohenlohe. 2017. Conservation implications of limited genetic diversity and population structure in Tasmanian devils (Sarcophilus harrisii). Conserv. Genet. Print 18: 977–982.

Hernando-Amado, S., F. Sanz-García, and J. L. Martínez. 2020. Rapid and robust evolution of collateral sensitivity in Pseudomonas aeruginosa antibiotic-resistant mutants. Sci. Adv. 6: eaba5493.

Hoffmann, A., P. Griffin, S. Dillon, R. Catullo, R. Rane, M. Byrne, R. Jordan, J. Oakeshott, A. Weeks, L. Joseph, P. Lockhart, J. Borevitz, and C. Sgrò. 2015. A framework for incorporating evolutionary genomics into biodiversity conservation and management. Clim. Change Responses 2: 1.

Hogeweg, P. 2012. Toward a Theory of Multilevel Evolution: Long-Term Information Integration Shapes the Mutational Landscape and Enhances Evolvability. *In* O. S. Soyer (Ed.) Evolutionary Systems Biology. Advances in Experimental Medicine and Biology. pp. 195–224, Springer, New York, NY. Available at: https://doi.org/10.1007/978-1-4614-3567-9_10 [Accessed June 18, 2021].

Hughes, A. R., B. D. Inouye, M. T. J. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. Ecol. Lett. 11: 609–623.

Huynen, M. A., P. F. Stadler, and W. Fontana. 1996. Smoothness within ruggedness: the role of neutrality in adaptation. Proc. Natl. Acad. Sci. 93: 397–401.

Jay, F., S. Manel, N. Alvarez, E. Y. Durand, W. Thuiller, R. Holderegger, P. Taberlet, and O. François.

2012. Forecasting changes in population genetic structure of alpine plants in response to global warming. Mol. Ecol. 21: 2354–2368.

Jones, F. C. et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484: 55–61.

Joyce, P., D. R. Rokyta, C. J. Beisel, and H. A. Orr. 2008. A General Extreme Value Theory Model for the Adaptation of DNA Sequences Under Strong Selection and Weak Mutation. Genetics 180: 1627–1643.

Kandul, N. P., J. Liu, J. B. Bennett, J. M. Marshall, and O. S. Akbari. 2021. A confinable home-and-rescue gene drive for population modification C. Desplan and P. J. Wittkopp (Eds.). eLife 10: e65939.

Karr, J. R., J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. 2012. A Whole-Cell Computational Model Predicts Phenotype from Genotype. Cell 150: 389–401.

Kawecki, T. J., R. E. Lenski, D. Ebert, B. Hollis, I. Olivieri, and M. C. Whitlock. 2012. Experimental evolution. Trends Ecol. Evol. 27: 547–560.

Kearney, M., W. P. Porter, C. Williams, S. Ritchie, and A. A. Hoffmann. 2009. Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito Aedes aegypti in Australia. Funct. Ecol. 23: 528–538.

Kellermann, V., B. van Heerwaarden, C. M. Sgrò, and A. A. Hoffmann. 2009. Fundamental Evolutionary Limits in Ecological Traits Drive Drosophila Species Distributions. Science 325: 1244–1246.

Kemble, H., C. Eisenhauer, A. Couce, A. Chapron, M. Magnan, G. Gautier, H. L. Nagard, P. Nghe, and O. Tenaillon. 2020. Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway. Sci. Adv. 6: eabb2236.

Keulen, W., N. K. Back, A. van Wijk, C. A. Boucher, and B. Berkhout. 1997. Initial appearance of the 184Ile variant in lamivudine-treated patients is caused by the mutational bias of human immunodeficiency virus type 1 reverse transcriptase. J. Virol. 71: 3346–3350.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge Available at: https://www.cambridge.org/core/books/neutral-theory-of-molecular-evolution/0FF60E9F47915B17FFA2620C49400632 [Accessed June 1, 2021].

Kissler, S. M., C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch. 2020. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science 368: 860–868.

Kreiner, J. M., D. A. Giacomini, F. Bemm, B. Waithaka, J. Regalado, C. Lanz, J. Hildebrandt, P. H. Sikkema, P. J. Tranel, D. Weigel, J. R. Stinchcombe, and S. I. Wright. 2019. Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. Proc. Natl. Acad. Sci. 116: 21076–21084.

Kyrou, K., A. M. Hammond, R. Galizi, N. Kranjc, A. Burt, A. K. Beaghton, T. Nolan, and A. Crisanti. 2018. A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. Nat. Biotechnol. 36: 1062–1066.

Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500: 571–574.

Lässig, M., and V. Mustonen. 2020. Eco-evolutionary control of pathogens. Proc. Natl. Acad. Sci. 117: 19694–19704.

Lässig, M., V. Mustonen, and A. M. Walczak. 2017. Predicting evolution. Nat. Ecol. Evol. 1: 0077.

Lee, J. M., J. Huddleston, M. B. Doud, K. A. Hooper, N. C. Wu, T. Bedford, and J. D. Bloom. 2018. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proc. Natl. Acad. Sci. 115: E8276–E8285.

Lescat, M., A. Launay, M. Ghalayini, M. Magnan, J. Glodt, C. Pintard, S. Dion, E. Denamur, and O. Tenaillon. 2017. Using long-term experimental evolution to uncover the patterns and determinants of molecular evolution of an Escherichia coli natural isolate in the streptomycin-treated mouse gut. Mol. Ecol. 26: 1802–1817.

Lieberman, T. D., J.-B. Michel, M. Aingaran, G. Potter-Bynoe, D. Roux, M. R. Davis, D. Skurnik, N. Leiby, J. J. LiPuma, J. B. Goldberg, A. J. McAdam, G. P. Priebe, and R. Kishony. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. 43: 1275–1280.

Lind, P. A. 2019. Repeatability and Predictability in Experimental Evolution. *In* P. Pontarotti (Ed.) Evolution, Origin of Life, Concepts and Methods. pp. 57–83, Springer International Publishing, Cham. Available at: https://doi.org/10.1007/978-3-030-30363-1_4 [Accessed November 3, 2020].

Lind, P. A., A. D. Farr, and P. B. Rainey. 2017. Evolutionary convergence in experimental Pseudomonas populations. ISME J. 11: 589–600.

Lind, P. A., E. Libby, J. Herzog, and P. B. Rainey. 2019. Predicting mutational routes to new adaptive phenotypes P. J. Wittkopp (Ed.). eLife 8: e38822.

Łuksza, M., and M. Lässig. 2014. A predictive fitness model for influenza. Nature 507: 57–61.

Mansky, L. M., and H. M. Temin. 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. 69: 5087–5094.

Masuka, B., G. N. Atlin, M. Olsen, C. Magorokosho, M. Labuschagne, J. Crossa, M. Bänziger, K. V. Pixley, B. S. Vivek, A. von Biljon, J. Macrobert, G. Alvarado, B. M. Prasanna, D. Makumbi, A. Tarekegne, B. Das, M. Zaman-Allah, and J. E. Cairns. 2017. Gains in Maize Genetic Improvement in Eastern and Southern Africa: I. CIMMYT Hybrid Breeding Pipeline. Crop Sci. 57: 168–179.

Meijer, J., B. van Dijk, and P. Hogeweg. 2020. Contingent evolution of alternative metabolic network topologies determines whether cross-feeding evolves. Commun. Biol. 3: 1–13.

Milocco, L., and I. Salazar-Ciudad. 2020. Is evolution predictable? Quantitative genetics under complex genotype-phenotype maps. Evolution 74: 230–244.

Molenaar, D., R. van Berlo, D. de Ridder, and B. Teusink. 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol. Syst. Biol. 5: 323.

Morgan, R., M. H. Finnøen, H. Jensen, C. Pélabon, and F. Jutfelt. 2020. Low potential for evolutionary rescue from climate change in a tropical fish. Proc. Natl. Acad. Sci. 117: 33365–33372.

Morris, D. H., K. M. Gostic, S. Pompei, T. Bedford, M. Łuksza, R. A. Neher, B. T. Grenfell, M. Lässig, and J. W. McCauley. 2018. Predictive Modeling of Influenza Shows the Promise of Applied Evolutionary Biology. Trends Microbiol. 26: 102–118.

Morris, S. C. 2003. Life's Solution: Inevitable Humans in a Lonely Universe. Cambridge University Press.

Nadell, C. D., K. Drescher, and K. R. Foster. 2016. Spatial structure, cooperation and competition in biofilms. Nat. Rev. Microbiol. 14: 589–600.

Neher, R. A., C. A. Russell, and B. I. Shraiman. 2014. Predicting evolution from the shape of genealogical trees G. McVean (Ed.). eLife 3: e03568.

Nichol, D., P. Jeavons, A. G. Fletcher, R. A. Bonomo, P. K. Maini, J. L. Paul, R. A. Gatenby, A. R. A. Anderson, and J. G. Scott. 2015. Steering Evolution with Sequential Therapy to Prevent the Emergence of Bacterial Antibiotic Resistance. PLoS Comput. Biol. 11.

Noble, C., J. Olejarz, K. M. Esvelt, G. M. Church, and M. A. Nowak. 2017. Evolutionary dynamics of CRISPR gene drives. Sci. Adv. 3: e1601964.

Nosil, P., S. M. Flaxman, J. L. Feder, and Z. Gompert. 2020. Increasing our ability to predict

contemporary evolution. Nat. Commun. 11: 5592.

O'Brien, E. J., J. A. Lerman, R. L. Chang, D. R. Hyduke, and B. Ø. Palsson. 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9: 693.

Olivieri, I., J. Tonnabel, O. Ronce, and A. Mignot. 2015. Why evolution matters for species conservation: perspectives from three case studies of plant metapopulations. Evol. Appl. 9: 196–211.

Orr, H. A. 1998. THE POPULATION GENETICS OF ADAPTATION: THE DISTRIBUTION OF FACTORS FIXED DURING ADAPTIVE EVOLUTION. Evol. Int. J. Org. Evol. 52: 935–949.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6: 119–127.

Orth, J. D., I. Thiele, and B. Ø. Palsson. 2010. What is flux balance analysis? Nat. Biotechnol. 28: 245–248.

Pennings, P. S., C. B. Ogbunugafor, and R. Hershberg. 2020. Reversion is most likely under high mutation supply, when compensatory mutations don't fully restore fitness costs. bioRxiv 2020.12.28.424568.

Perelson, A. S., and P. W. Nelson. 1999. Mathematical Analysis of HIV-1 Dynamics in Vivo. SIAM Rev. 41: 3–44.

Perron, G. G., R. F. Inglis, P. S. Pennings, and S. Cobey. 2015. Fighting microbial drug resistance: a primer on the role of evolutionary biology in public health. Evol. Appl. 8: 211–222.

Pinheiro, F., O. Warsi, D. I. Andersson, and M. Lässig. 2021. Metabolic fitness landscapes predict the evolution of antibiotic resistance. Nat. Ecol. Evol. 1–11.

Pinsky, M. L., A. M. Eikeset, C. Helmerson, I. R. Bradbury, P. Bentzen, C. Morris, A. T. Gondek-Wyrozemska, H. T. Baalsrud, M. S. O. Brieuc, O. S. Kjesbu, J. A. Godiksen, J. M. I. Barth, M. Matschiner, N. C. Stenseth, K. S. Jakobsen, S. Jentoft, and B. Star. 2021. Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. Proc. Natl. Acad. Sci. 118. Available at: https://www.pnas.org/content/118/15/e2025453118 [Accessed April 15, 2021].

Poelwijk, F. J., S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans. 2011. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J. Theor. Biol. 272: 141–144.

Radchuk, V. et al. 2019. Adaptive responses of animals to climate change are most likely insufficient. Nat. Commun. 10: 3109.

Read, A. F., and R. J. Woods. 2014. Antibiotic resistance management. Evol. Med. Public Health 2014: 147.

Rellstab, C., S. Zoller, L. Walthert, I. Lesur, A. R. Pluess, R. Graf, C. Bodénès, C. Sperisen, A. Kremer, and F. Gugerli. 2016. Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. Mol. Ecol. 25: 5907–5924.

Remigi, P., C. Masson-Boivin, and E. P. C. Rocha. 2019. Experimental Evolution as a Tool to Investigate Natural Processes and Molecular Functions. Trends Microbiol. 27: 623–634.

Rocheleau, G., C. J. Brumme, J. Shoveller, V. D. Lima, and P. R. Harrigan. 2018. Longitudinal trends of HIV drug resistance in a large Canadian cohort, 1996–2016. Clin. Microbiol. Infect. 24: 185–191.

Roff, D. A. 2007. A Centennial Celebration for Quantitative Genetics. Evolution 61: 1017–1032.

Roush, R. T., G. P. Fitt, N. W. Forrester, and J. C. Daly. 1998. Resistance management for insecticidal transgenic crops: theory and practice. *In* M. P. Zalucki, R. Drew, and G. G. White (Eds.) Pest Management - Future Challenges. University of Queensland Press, Australia.

Rouzic, A. L., C. Renneville, A. Millot, S. Agostini, D. Carmignac, and É. Édeline. 2020. Unidirectional response to bidirectional selection on body size II. Quantitative genetics. Ecol. Evol. 10: 11453–11466.

Sackman, A. M., L. W. McGee, A. J. Morrison, J. Pierce, J. Anisman, H. Hamilton, S. Sanderbeck, C. Newman, and D. R. Rokyta. 2017. Mutation-Driven Parallel Evolution during Viral Adaptation. Mol. Biol. Evol. 34: 3243–3253.

Salverda, M. L. M., E. Dellus, F. A. Gorter, A. J. M. Debets, J. van der Oost, R. F. Hoekstra, D. S. Tawfik, and J. A. G. M. de Visser. 2011. Initial Mutations Direct Alternative Pathways of Protein Evolution. PLOS Genet. 7: e1001321.

Sánchez, B. J., C. Zhang, A. Nilsson, P. Lahtvee, E. J. Kerkhoven, and J. Nielsen. 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5572397/ [Accessed April 16, 2021].

Sane, M., J. J. Miranda, and D. Agashe. 2018. Antagonistic pleiotropy for carbon use is rare in new mutations. Evolution 72: 2202–2213.

Sarraf-Yazdi, S., M. Sharpe, K. M. Bennett, T. L. Dotson, D. J. Anderson, and S. N. Vaslef. 2012. A 9-Year retrospective review of antibiotic cycling in a surgical intensive care unit. J. Surg. Res. 176: e73–e78.

Schenk, M. F., M. P. Zwart, S. Hwang, P. Ruelens, E. Severing, J. Krug, and J. A. G. M. de Visser. 2021. Population size mediates the contribution of high-rate and large-benefit mutations to parallel evolution. bioRxiv 2021.02.02.429281.

Schuetz, R., L. Kuepfer, and U. Sauer. 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol. Syst. Biol. 3: 119.

Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. Evolution of high mutation rates in experimental populations of E. coli. Nature 387: 703–705.

Sommer, M. O. A., C. Munck, R. V. Toft-Kehler, and D. I. Andersson. 2017. Prediction of antibiotic resistance: time for a new preclinical paradigm? Nat. Rev. Microbiol. 15: 689–696.

Stern, D. L., and V. Orgogozo. 2008. The Loci of Evolution: How Predictable Is Genetic Evolution? Evolution 62: 2155–2177.

Stoltzfus, A. 2021. Mutation, Randomness, and Evolution. Oxford University Press Available at: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=Q9fzhu4AAAAJ&sortby=pubdate&citation_for_view=Q9fzhu4AAAAJ:p2g8aNsByqUC [Accessed February 4, 2022].

Stoltzfus, A., and D. M. McCandlish. 2017. Mutational Biases Influence Parallel Adaptation. Mol. Biol. Evol. 34: 2163–2172.

Stoltzfus, A., and L. Y. Yampolsky. 2009. Climbing Mount Probable: Mutation as a Cause of Nonrandomness in Evolution. J. Hered. 100: 637–647.

Storz, J. F. 2016. Causes of molecular convergence and parallelism in protein evolution. Nat. Rev. Genet. 17: 239–250.

Storz, J. F., C. Natarajan, A. V. Signore, C. C. Witt, D. M. McCandlish, and A. Stoltzfus. 2019. The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. Philos. Trans. R. Soc. B Biol. Sci. 374: 20180238.

Szabó, A., and R. M. H. Merks. 2017. Blood vessel tortuosity selects against evolution of aggressive tumor cells in confined tissue environments: A modeling approach. PLOS Comput. Biol. 13: e1005635.

Szendro, I. G., J. Franke, J. A. G. M. de Visser, and J. Krug. 2013. Predictability of evolution depends nonmonotonically on population size. Proc. Natl. Acad. Sci. 110: 571–576.

Tabashnik, B. E., T. Brévault, and Y. Carrière. 2013. Insect resistance to Bt crops: lessons from the first billion acres. Nat. Biotechnol. 31: 510–521.

Tataru, P., M. Mollion, S. Glémin, and T. Bataillon. 2017. Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data. Genetics 207: 1103–1119.

Tenaillon, O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut.

2012. The Molecular Diversity of Adaptive Convergence. Science 335: 457–461.

Teotónio, H., S. Estes, P. C. Phillips, and C. F. Baer. 2017. Experimental Evolution with Caenorhabditis Nematodes. Genetics 206: 691–716.

Teusink, B., A. Wiersma, L. Jacobs, R. A. Notebaart, and E. J. Smid. 2009. Understanding the Adaptive Growth Strategy of Lactobacillus plantarum by In Silico Optimisation. PLOS Comput. Biol. 5: e1000410.

Tisthammer, K. H., W. Dong, J. B. Joy, and P. S. Pennings. 2020. Assessing in vivo mutation frequencies and creating a high-resolution genome-wide map of fitness costs of Hepatitis C virus. bioRxiv 2020.10.01.323253.

Tyers, M., and G. D. Wright. 2019. Drug combinations: a strategy to extend the life of antibiotics in the 21st century. Nat. Rev. Microbiol. 17: 141–155.

Unckless, R. L., A. G. Clark, and P. W. Messer. 2017. Evolution of Resistance Against CRISPR/Cas9 Gene Drive. Genetics 205: 827–841.

Uusi-Heikkilä, S., A. R. Whiteley, A. Kuparinen, S. Matsumura, P. A. Venturelli, C. Wolter, J. Slate, C. R. Primmer, T. Meinelt, S. S. Killen, D. Bierbach, G. Polverino, A. Ludwig, and R. Arlinghaus. 2015. The evolutionary legacy of size-selective harvesting extends from genes to populations. Evol. Appl. 8: 597–620.

de Visser, J. A. G. M., and J. Krug. 2014. Empirical fitness landscapes and the predictability of evolution. Nat. Rev. Genet. 15: 480–490.

de Vos, M. G. J., A. Dawid, V. Sunderlikova, and S. J. Tans. 2015. Breaking evolutionary constraint with a tradeoff ratchet. Proc. Natl. Acad. Sci. 112: 14906–14911.

Vos, M. G. J. de, S. E. Schoustra, and J. A. G. M. de Visser. 2018. Ecology dictates evolution? About the importance of genetic and ecological constraints in adaptation. EPL Europhys. Lett. 122: 58002.

Vroomans, R. M. A., P. Hogeweg, and K. H. W. J. ten Tusscher. 2016. In silico evo-devo: reconstructing stages in the evolution of animal segmentation. EvoDevo 7: 14.

Waldvogel, A.-M., B. Feldmeyer, G. Rolshausen, M. Exposito-Alonso, C. Rellstab, R. Kofler, T. Mock, K. Schmid, I. Schmitt, T. Bataillon, O. Savolainen, A. Bergland, T. Flatt, F. Guillaume, and M. Pfenninger. 2020. Evolutionary genomics can improve prediction of species' responses to climate change. Evol. Lett. 4: 4–18.

Walsh, B., and M. Lynch. 2018. Evolution and Selection of Quantitative Traits. Oxford University Press.

Wang, X., V. Zorraquino, M. Kim, A. Tsoukalas, and I. Tagkopoulos. 2018. Predicting the evolution of Escherichia coli by a data-driven approach. Nat. Commun. 9: 3562.

Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science 312: 111–114.

Weinreich, D. M., R. A. Watson, and L. Chao. 2005. Perspective: Sign Epistasis and Genetic Costraint on Evolutionary Trajectories. Evolution 59: 1165–1174.

Wiser, M. J., N. Ribeck, and R. E. Lenski. 2013. Long-Term Dynamics of Adaptation in Asexual Populations. Science 342: 1364–1367.

Wortel, M. T., E. Bosdriesz, B. Teusink, and F. J. Bruggeman. 2016. Evolutionary pressures on microbial metabolic strategies in the chemostat. Sci. Rep. 6: 29503.

Wortel, M. T., E. Noor, M. Ferris, F. J. Bruggeman, and W. Liebermeister. 2018. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. PLOS Comput. Biol. 14: e1006010.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc. 6th Int. Congr. Genet. 1: 356–366.

Yen, P., and J. A. Papin. 2017. History of antibiotic adaptation influences microbial evolutionary dynamics during subsequent treatment. PLOS Biol. 15: e2001586.

Zandbergen, L. E., T. Halverson, J. K. Brons, A. J. Wolfe, and M. G. J. de Vos. 2021. The Good and the Bad: Ecological Interaction Measurements Between the Urinary Microbiota and Uropathogens. Front. Microbiol. 12. Available at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.659450/full [Accessed July 5, 2021].

Zhao, S., T. D. Lieberman, M. Poyet, K. M. Kauffman, S. M. Gibbons, M. Groussin, R. J. Xavier, and E. J. Alm. 2019. Adaptive Evolution within Gut Microbiomes of Healthy People. Cell Host Microbe 25: 656-667.e8.