Supplementary Figure 1. 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 (visualized by the yellow ellipse), using 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.

As an example we apply the conceptual model to the prediction about seasonal influenza strains in the next season, which is used to design the next vaccine. Prior knowledge (blue circle) 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, only the list of the most prevalent clades, or the antigenic phenotypes of the future virus population.

In terms of the four attributes described in the main text, 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 time (we need to know the common strains for the next season, not earlier or later than that) and genetics, as we are predicting the exact strains. 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.
Methods for making evolutionary predictions

1. Experimental evolution

A straightforward method for evolutionary predictions is creating the conditions of interest (in the lab or a different natural environment), observe (a lack of) evolution, and extrapolate from these observations into the future.

Laboratory experimental evolution is diverse; it is commonly used for fast-evolving microbial systems (e.g. antibiotic resistance evolution (Kawecki et al. 2012, Remigi et al. 2019), but also larger organisms such as *Drosophila* (Burke & Rose 2009) and *Caenorhabditis* (Teotónio et al. 2017). Experimental evolution can also include interactions with the environment (i.e. eco-evolutionary interactions), 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).

Experimental evolution can also be used to study the predictability of evolution, 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. 2022).

2. 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). 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), mutational scanning of one gene (Lee et al. 2018), or on a small metabolic pathway (Kemble et al. 2020). Mutation accumulation experiments can also be used to understand the mutational landscape although they are unlikely to capture all mutations (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.

3. (Microbial) metabolic and growth models

In many environments, selection on microorganisms is differential population growth of alternative genotypes, and therefore metabolism and growth are directly linked to fitness. This means we can use metabolic and growth models to predict evolution. The most well-known predictions are growth rates and metabolic adaptations in new environments (e.g. in E. coli (Schuetz et al. 2007) and Lactobacillus plantarum (Teusink et al. 2009)). These predictions use genome-scale metabolic models (Gu et al. 2019), which predict metabolic enzymes from genome information, constrain uptake fluxes with experimental data
and optimise for growth rate with flux balance analysis (Orth et al. 2010). Genome-scale metabolic models are also used for biotechnological applications, for example to predict how knock-outs of specific enzymes affect growth and product formation (see the Box in the main text)

Microbial growth and metabolism 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 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. 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).

4. 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 (Hartl 2020). Population genetic models can include mutation, fitness, reproduction, recombination and other parameters, with both deterministic and stochastic forces. These models are used widely to predict allele frequencies, e.g. incidence of deleterious recessive alleles in human populations (e.g. due to mutation/selection balance), or of conditionally beneficial alleles (e.g. sickle cell anaemia alleles in areas with Malaria parasites). Additionally, they can be used as a null model, to detect signatures of selection within population genomics, or deviations from panmixia (e.g. population differentiation).

Population-genetic models predict the evolutionary dynamics of allelic variants of genes. They use population-genetic theory, to describe the frequencies of alternative alleles (at one or multiple loci) over time. New alleles can be created from the wildtype allele by mutation, or may get introduced through immigration. Alleles at multiple sites can be recombined in following generations.

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...
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. Moreover, the mutations of interest, which are the beneficial ones, are only a small fraction of the possible mutations. Using population genetic theory to predict evolution 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)).

Both Mendelian and non-Mendelian inheritance can be incorporated into the models. The models follow a population of individuals, to assess the separate or combined effects of mutation, selection, drift, non-random mating and migration, to calculate expectations for which allele(s) will spread, stabilise in frequency or disappear. In the absence of any of the evolutionary processes, allele frequencies are expected to remain largely constant and genotype frequencies can be accurately predicted, as described by the Hardy-Weinberg principle. If a beneficial allele is present in the population (from standing genetic variation, immigration or due to new mutations), it is expected to become more common in the population due to natural selection, and the models can predict the rate of change and the eventual equilibrium frequency to which the allele frequencies in the population will evolve. Ultimately, these models specify the particulars of the various deterministic and stochastic processes that operate concurrently, and how these processes contribute to the resulting evolutionary dynamics.

One interesting case study of population genetic models for predicting evolution is the evolutionary dynamics of gene drive systems (Unckless et al. 2017). This is explained in more detail in Box 1. In this model the frequencies of three alleles 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, which can be created from the wildtype allele by mutation, and cannot be converted to the driver allele. An additional layer of complexity that 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).
5. Quantitative genetics and the breeder’s equation

The field of quantitative genetics was developed in close association with animal and plant breeders. 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), though see (Beldade et al. 2002, Bolstad et al. 2015). 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).

6. 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 modelling of influenza evolution (Boni et al. 2006).
In the most simple case, evolution in SIR models can be modelled 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 (see figure). 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).

7. 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). With models that include genomic data from a species, genomic variation can be related to environmental variables (Fitzpatrick & Keller 2015). This information can then be used to predict how vulnerable populations are to environmental change, such as climate change.
8. Multi-scale evolutionary modelling

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 organisation 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 tumour 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 Section 2 of main text). 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).
9. 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).

References


Doulcier, G., A. Lambert, S. De Monte, and P. B. Rainey. 2020. Eco-evolutionary dynamics of


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’


