CellPress

Review

Forecasting Epidemiological and Evolutionary Dynamics of Infectious Diseases

Sylvain Gandon,[1,*] Troy Day,[2] C. Jessica E. Metcalf,[3] and Bryan T. Grenfell[3]

Mathematical models have been powerful tools in developing mechanistic understanding of infectious diseases. Furthermore, they have allowed detailed forecasting of epidemiological phenomena such as outbreak size, which is of considerable public-health relevance. The short generation time of pathogens and the strong selection they are subjected to (by host immunity, vaccines, chemotherapy, etc.) mean that evolution is also a key driver of infectious disease dynamics. Accurate forecasting of pathogen dynamics therefore calls for the integration of epidemiological and evolutionary processes, yet this integration remains relatively rare. We review previous attempts to model and predict infectious disease dynamics with or without evolution and discuss major challenges facing the development of the emerging science of epidemic forecasting.

The Promise of Forecasting Natural Disasters

Part of the promise of scientific research into the fundamental drivers of natural processes is that a deeper understanding of mechanism will improve our ability to forecast. The most immediate example of our ability to predict the dynamics of our environment is perhaps weather forecasting. It has long been recognized that, in principle, the laws of physics and thermodynamics can be used to numerically predict dynamics occurring in the atmosphere [1]. Nevertheless, in practice early attempts at prediction were overwhelmed by the complexity resulting from multiple nonlinear interactions. Yet, the science of weather forecasting has made tremendous strides over the past decades as a result of both refined theory and the availability of more detailed data [2]. It is important to note that not all predictions are based on a detailed mechanistic understanding of underlying processes. For instance, the dynamics of cloud formation remains particularly difficult to model because the processes involved occur at a scale below the current meteorological grid size. Yet, a statistical description of the processes occurring at this scale can be combined with mechanistic models at a larger spatial scale to generate remarkably accurate predictions. For example, the trajectories of hurricanes can now be predicted several days in advance (Figure 1). This improved ability to forecast the position in space and time of massive storms undoubtedly carries major benefits for public safety and the economy.

Here we discuss some of the potential and challenges of forecasting another type of natural disaster: infectious disease epidemics. Like hurricanes, epidemics are dynamical systems that can be modeled mathematically [3–5]. Their dynamics are driven by transmission and recovery events as well as by changing external conditions, including those arising from public-health interventions (e.g., vaccination, drug treatment). Furthermore, because most pathogens have the ability to adapt rapidly, evolutionary change can also affect disease epidemics. In these situations, forecasting infectious disease dynamics requires a good understanding of both the underlying ecology and the evolutionary processes driving these adaptations.

Trends

Long-term monitoring of infectious disease dynamics allows the estimation of key parameters of epidemiological models. These models can be used to forecast future epidemics and to implement effective public-health control measures.

Many pathogens exhibit extensive genetic variation and so can readily adapt to control measures like drugs and vaccines. We review recent attempts to combine epidemiology and evolution to predict the evolutionary trajectories of pathogens.

Inspired by the success of weather forecasting we discuss the current limits of the predictive power of evolutionary epidemiology. The development of the emerging science of epidemic forecasting requires better integration of mathematical epidemiology, population genetics, statistics, and numerical computation.

[1]CEFE UMR 5175, CNRS–Université de Montpellier–Université Paul-Valéry Montpellier–EPHE, 1919 route de Mende, 34293 Montpellier cedex 5, France
[2]Department of Biology, Queen's University, Kingston, Canada
[3]Department of Ecology and Evolutionary Biology and Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA

*Correspondence: sylvain.gandon@cefe.cnrs.fr (S. Gandon).

 http://dx.doi.org/10.1016/j.tree.2016.07.010

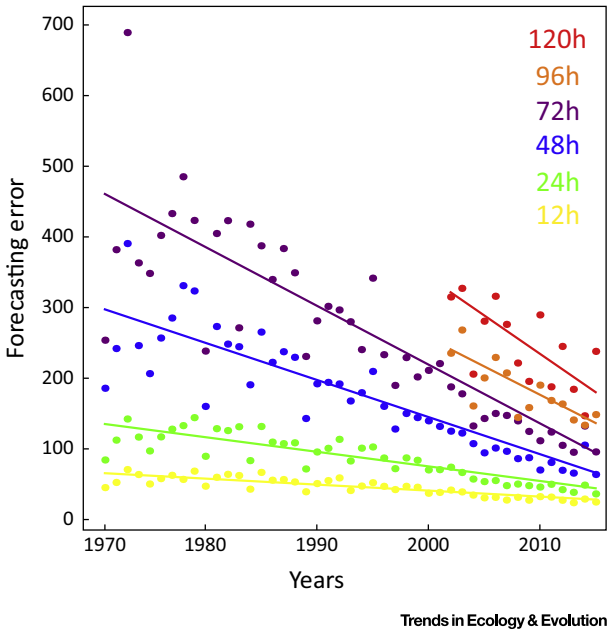

Figure 1. The Improving Science of Weather Forecasting. Annual average official track errors (distance in nautical miles between predictions and actual trajectories of tropical cyclones) for the Atlantic basin from 1970–2014 with least-squares trend lines superimposed. Dots refer to predictions made 12 h, 24 h, 48 h, 72 h, 96 h, or 120 h in advance. Data obtained from http://www.nhc.noaa.gov/verification/.

Here we review previous attempts to predict infectious disease dynamics. We start by reviewing models that predict the epidemiology of infectious diseases under the assumption that evolution is a much slower dynamical process than epidemiology. We show that this assumption is often not tenable and we then review attempts to model both the epidemiology and the evolutionary dynamics of pathogens, outlining major directions for further development.

Forecasting the Epidemiology of Infection Diseases

Although the first epidemiological models were developed long ago [6–9], the field of mathematical epidemiology has grown enormously during the past half-century [3–5]. Over this time research has provided a general mathematical framework that can describe the dynamics of a wide diversity of animal and plant pathogens. Many of the epidemiological models that have been developed rely on the *a priori* assumption that the pathogen population is monomorphic; that is, there are no differences in transmission, pathogenicity, or any other life history features of the pathogens circulating within the population. When mutation is relatively rare and/or largely deleterious, the adaptive potential of the pathogen remains very low and evolution can be safely neglected when describing epidemiological dynamics.

Measles is arguably the best example to illustrate the predictive power of mathematical epidemiology. This highly contagious human disease is caused by a horizontally transmitted RNA virus. Infected individuals usually recover after a few days and acquire lifelong immunity, but the case fatality rate may be high in situations such as low-resource contexts or in immunocompromised individuals [10]. An inexpensive and safe vaccine has been available since the 1960s and provides prolonged immunity. Despite the high mutation rate of RNA viruses, measles viruses have not evolved escape strategies against host immunity. Previous studies have shown that the epidemiology of the disease can be accurately described with the following

set of differential equations:

$$\dot{S} = \theta - \delta S - \frac{\beta}{N} SI$$
$$\dot{I} = \frac{\beta}{N} SI - (\delta + \gamma) I \quad [1]$$
$$\dot{R} = \gamma I - \delta R$$

This simple compartmental model tracks the dynamics of susceptible, infected, and recovered hosts (S, I, and R respectively). The total population size N may vary across time depending on θ, the influx of new susceptibles, and δ, the natural death rate of the host. The spread of the pathogen is governed by β, the transmission rate, and γ, the recovery rate. Whether the number of infected individuals grows is governed by $r = \beta S/N - (\delta + \gamma)$, which is the *per capita* rate of change of the infected population (a per unit of time equivalent of the effective reproductive ratio R_e). If $r > 0$, the number of infections increases; and if $r < 0$, it decreases.

The epidemiological dynamics depend heavily on the density of susceptible hosts; this density is determined by the depletion of susceptibles from previous epidemics (which immunize a fraction of the population) and the influx of new susceptible individuals (resulting from birth or immigration). Seasonal forcing of the transmission rate β is also an important driver of measles dynamics. Accurate estimates of the seasonal variation in β can be obtained from the discretized dynamics of susceptible and infected individuals [11–13] (Box 1). This very simple model sheds light on an array of historical patterns of measles dynamics. For instance, the sudden increase in the influx of susceptibles following World War II (from the baby boom), led to a transitory shift from a biannual to an annual periodicity of measles epidemiological dynamics (Box 1). These models provide quantitative predictions under different control measures and constitute an effective tool to optimize public-health strategies. For instance, Takahashi *et al.* [14] quantified the consequences of a drop in vaccination coverage against measles in West Africa due to the Ebola epidemic. These predictions can be used to make practical recommendations to prevent future measles epidemics in the area.

Nevertheless, several additional factors may alter the predictive power of these models. First, this epidemiological model does not always generate simple periodic dynamics. For instance, strong seasonality can push the system into a chaotic dynamical regime [15–17]. Second, when host population size becomes very small, demographic stochasticity can have a major impact on epidemiology [18,19] and this simple deterministic model is expected to lose its predictive power [20]. Third, the life-history traits of the pathogen may change over time. The model described in Box 1 allows transmission rates to vary periodically throughout the year but they are assumed to remain constant across years. Thus, although predictable patterns are the norm for measles dynamics (Box 1), the key result of the above work has been an understanding of limits on predictability, based on remarkably simple models, illustrated by the impact of relatively modest modifications, for example in the shape of the seasonal forcing function on periodicity [21], or magnitude of transmission on the impact of stochasticity [22]. Similar arguments on quantifying limits on prediction also arise in the forecasting of–often highly nonlinear and chaotic–weather dynamics [2,23,24].

Forecasting the Evolution of Infectious Diseases

As mentioned above, antigenic evolution of measles (i.e., vaccine-escape mutations) has been limited so far and thus is not a challenge for the forecasting of measles dynamics. Other pathogens, however, may exhibit rapid changes in pathogen population composition that yield dramatic perturbations of epidemiological dynamics. For example, the antigenic drift occurring in influenza A provides a marked contrast to measles. Axelsen *et al.* [25] showed that accurately describing influenza dynamics in Israel requires accounting for occasional modification of the antigenic epitopes of the virus, which allows the pathogen to escape immune recognition by

Box 1. Measles Dynamics: The Old Faithful of Epidemiology

Measles infection is associated with completely sterilizing immunity. This leads to strong feedbacks in dynamics, as infection is inevitably associated with susceptible depletion (infected individuals are subsequently immune for life), which in turn restricts spread. Such strong feedbacks, combined with little additional complexity–for example, there is no evidence of substantial immune escape following natural infection–strengthen our ability to forecast incidence.

Detailed understanding of the driving mechanisms of the dynamics of measles (susceptible depletion and replenishment via births) allows us to construct appropriate models that can be used to make quantitative forecasts regarding the magnitude of the incidence (Figure I). The time-series susceptible–infected–recovered (TSIR) model provides a mechanistic bridge between theoretical models and empirical data. For measles, the duration of the transition from infection to recovery and lifelong immunity is 2 weeks. To fit the model, measles incidence data are accordingly aggregated into biweekly time steps. The analysis of the TSIR is based around taking logs to linearize the discrete time dynamics $I_{t+1} = \beta_s I_t S_t/N_t$ (where I_t reflects number of cases, N_t is the total population size, and S_t reflects the number of susceptibles in each biweek obtained via susceptible reconstruction, which comprises depleting susceptibles by infection and increasing them by birth rate in each time step). This is then fitted using linear regression to estimate the seasonal fluctuation in transmission β_s, where the index s indicates the 26 biweeks in each year across the whole time series. Forward simulation simply reverses this process but using a negative binomial model where $I_{t+1} \sim \text{NegBin}(\beta_s I_t S_t/N_t, I_t)$ and starting values are defined during the original fitting process.

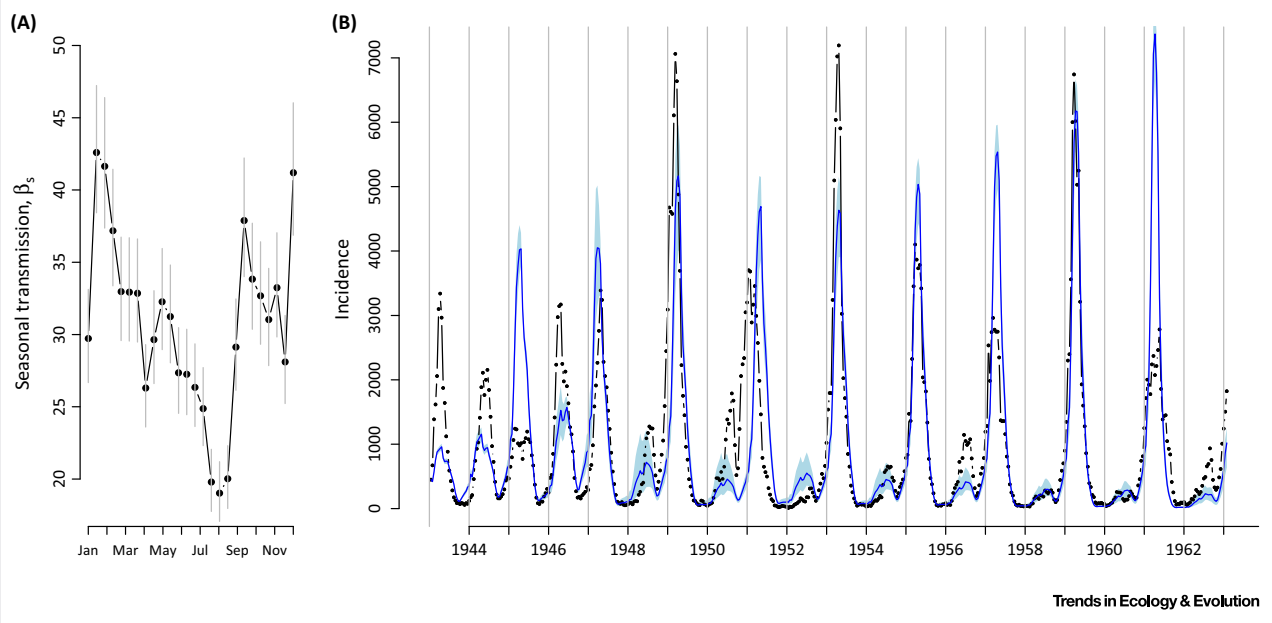

Figure I. Estimation and Simulation from a Time-Series Susceptible–Infected–Recovered (TSIR) Model. (A) Estimated seasonal transmission β_s (y-axis) over the course of the year (x-axis) indicating a pattern that broadly reflects school-term times; vertical lines show 95% confidence intervals. (B) Associated data on measles incidence in London (black points) and quartiles of predicted incidence from stochastic simulations based around these estimates (light blue polygons; medians shown by blue lines). This model can successfully capture the transition from annual dynamics after World War II and during the baby boom to biennial dynamics thereafter, when the birth rate fell.

recovered individuals. Where such processes are at play, forecasting the dynamics of infectious disease cannot focus solely on disease incidence: pathogens such as influenza challenge the predictive power of epidemiological models because they require the inclusion of an evolutionary dimension in forecasts.

Taking evolution into account in epidemiological models might involve a least two different approaches. First, it might involve statistical models describing pathogen changes without a formal description of the underlying processes leading to these evolutionary changes. For instance, the influenza model of Axelsen *et al.* [25] better captures the epidemiological dynamics only after taking into account evolutionary change occurring over years. However, this model does not explicitly model this evolution and consequently cannot provide epidemiological predictions across years. Second, taking into account evolution might involve a mechanistic description of evolutionary processes if different pathogen strains are allowed to circulate

simultaneously. In this case it is important to describe how these different strains interact with each other. In the following we distinguish two extreme cases that cover very different biological situations: complete cross-immunity and partial cross-immunity.

Life-History Evolution with Complete Cross-immunity

We start with a situation where all of the pathogen genotypes interact very strongly. For instance, a host currently or previously infected with strain *i* cannot be infected by another strain, *j*. This considerably simplifies the life cycle, because all pathogens are competing for a common resource, *S*, the population of susceptible hosts. Consequently, the dynamics of multistrain models with complete cross-immunity can be formalized using the following extension of the dynamical system (Equation 1):

$$\begin{aligned}\dot{S} &= \theta - \delta S - \frac{\overline{\beta}}{N} SI \\ \dot{I}_i &= \frac{\beta_i}{N} SI_i - (\delta + \gamma_i) I_i \\ \dot{R} &= \overline{\gamma} I - \delta R\end{aligned} \quad [2]$$

where I_i is the density of hosts infected with strain *i* (one of the *n* different strains circulating in the population) and $I = \sum_{i=1}^{n} I_i$ is the total density of infected hosts. Each strain may have different life-history traits β_i and γ_i and overbars denote averages over all strains circulating in the population. In this situation it is possible to track the evolution of the parasite population and to recover classical population genetics results (Box 2). This framework has been used to study and predict the evolution of key life-history traits of pathogens such as virulence and resistance to chemotherapy.

Box 2. From Epidemiology to Quantitative Genetics and Back

The change in the frequency $p_i = I_i/I$ of strain *i* provides an array of important insights and can be expressed as:

$$\dot{p}_i = p_i(r_i - \bar{r}) = p_i(1 - p_i)s_i \quad [I]$$

where $r_i = \beta_i S/N - (\delta + \gamma_i)$ is the Malthusian growth rate of strain *i* and $\bar{r} = \sum_{i=1}^{n} r_i p_i$ is the average Malthusian growth rate of the total pathogen population: $\dot{I} = \bar{r} I$. The parameter s_i is the classical selection coefficient of population genetics and is equal to [29,98]:

$$s_i = r_i - \sum_{j=1, j \neq i}^{n} \frac{r_j p_j}{1 - p_i} \quad [II]$$

It is also possible to recover a classical quantitative genetics formulation of the evolution of the two pathogen life-history traits of this model:

$$\dot{\bar{\mathbf{z}}} = \mathbf{G} \begin{pmatrix} S/N \\ -1 \end{pmatrix} \quad [III]$$

where $\bar{\mathbf{z}} = (\overline{\beta}, \overline{\gamma})^T$ is the average mean pathogen phenotype, **G** is the genetic variance–covariance matrix and $(S/N, -1)^T$ is the selection gradient [99,100]. In other words, the evolution of pathogen life history is constrained by the availability of genetic variation, by possible covariation among traits (i.e., life-history tradeoffs), and by the direction of selection. It is important to note that the direction of the selection gradient changes with the density *S* of susceptible hosts. Selection is driven by the epidemiological state of the population and, in return, epidemiology is driven by the evolution of mean phenotypes (see Equation 2). The dynamics of the host and pathogen populations thus results from the interplay between epidemiology and evolution. Note that the model in Equation 2 describes a very simple scenario, but this framework can be generalized to encompass more complex life cycles [29,73,98].

When genetic variance is very limited, evolutionary processes become much slower than epidemiological ones. In this case we recover the classical assumptions of adaptive dynamics and the direction of evolution is given by maximization of the basic reproductive ratio:

$$R_0 = \beta/(\delta + \gamma) \quad [IV]$$

Virulence Evolution

Theoretical predictions of virulence evolution are often based on assumed genetic constraints between virulence and other life-history traits. There is good experimental support for these genetic constraints, but detailed quantification of genetic tradeoffs remains notoriously elusive [26]. Furthermore, most virulence evolution models are based on the adaptive dynamics approach and generate only long-term predictions (Box 2) rather than the shorter timescales more likely to be directly observed. Consequently, most of these predictions remain qualitative and are difficult to test using short-term experiments. In principle, however, strengthening of the integration between theory and experimental testing is increasingly within reach. For instance, Berngruber *et al.* [27] studied the effect of epidemiological perturbations on virulence evolution occurring in bacteriophage chemostats. This experimental system allows the prevalence of the infection and the frequency of two virus strains to be tracked through time. The match between the predictions of a simple deterministic model and the transient dynamics observed in the chemostats shows that, in controlled laboratory conditions, it is possible to predict the epidemiology and evolution of a pathogen (Figure 2). Further work is needed to generate more quantitative predictions and to find ways to estimate genetic constraints in wild pathogens. In a more applied context, it is particularly important to predict both the short-term and the long-term life-history adaptations of pathogens following human interventions like vaccination [28,29]. Detailed studies of life-history variation among different pathogen genotypes in different types of host remain rare [30–32] but are essential for generating such predictions.

Drug Resistance Evolution

Drug resistance may be viewed as another important life-history trait that allows pathogens to survive in treated hosts. Many theoretical models have been developed to help us understand the interplay between several evolutionary forces on the evolution of drug resistance at different spatial scales. For instance, drug resistance is expected to increase when drugs are used, but if drug resistance carries intrinsic fitness costs this trend may be reversed in the absence of the drug. Across European countries qualitative predictions regarding drug consumption and

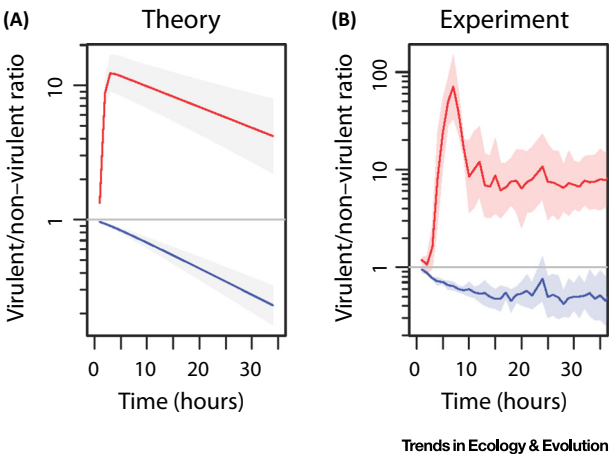

Figure 2. Theoretical and Experimental Evolution of Bacteriophage Virulence during an Epidemic. (A) Theoretical predictions for the change in the ratio of virulent (λcI857) to avirulent (λ) phage among lysogenic bacteria. The initial value of the λcI857:λ ratio in the provirus was 1:1 and two initial prevalence values were considered: 1% (red) and 100% (blue). Ten thousand simulations were performed, allowing some variation over the phenotypic values of the two virus strains (see Berngruber *et al.* [27] for details). The gray envelopes show the 95% confidence intervals among all simulation runs and colored lines show the median of the log-transformed data. (B) Change in the ratio of virulent (λcI857) to avirulent (λ) phage among lysogenic bacteria. As in (A), the initial value of the λcI857:λ ratio in the provirus was 1:1 and competition was started from two initial prevalence values: 1% (red) and 100% (blue). The lines are the means of four chemostat experiments and the envelopes show the 95% confidence intervals of the log-transformed data (see Berngruber *et al.* [27] for details).

frequency of drug resistance are generally well supported by empirical studies [33]. Within countries, as expected, reduction in drug use has led to substantial drops in drug resistance after several years [34–36]. At the scale of hospitals, interventions aimed at controlling the spread of drug resistance have often led to much faster declines in drug resistance [36]. This is expected from the fast turnover of patients typically occurring in hospitals [37]. In other words, the migration of sensitive bacteria acquired from the community can also be a potent evolutionary force at this spatial scale. Studying drug resistance evolution at the scale of the individual host requires additional evolutionary forces like mutation, within-host competition between bacterial genotypes, and demographic stochasticity [38,39].

Hence, at each spatial scale predictions require the estimation of different parameters. This information is often lacking and, in particular, estimates of the fitness cost of drug resistance remain scarce. Luciani *et al.* [40], however, used approximate Bayesian computation and molecular data to infer rates of evolution and transmission cost associated with drug resistance in *Mycobacterium tuberculosis* across different countries. These estimations allowed them to generate predictions for the evolution of drug resistance under different treatment scenarios. The ability to make detailed quantitative predictions is key to identifying optimal therapeutic strategies with these drugs and similar inference studies on other pathogens are urgently needed to manage drug resistance at different spatial scales.

Antigenic Evolution with Partial Cross-immunity

Many infections generate only partial cross-immunity against other pathogens. These situations are intrinsically more complex because each strain has access to a different number of hosts and one needs to keep track of the number of the different types of hosts to understand and predict the dynamics of the pathogen population. With partial cross-immunity the above epidemiological model can be modified in the following way:

$$\begin{aligned}\dot{S} &= \theta - \delta S - \frac{\overline{\beta}}{N} SI \\ \dot{I}_i &= \frac{\beta_i}{N} I_i (S + \sigma_{ij} R_j) - (\delta + \gamma_i) I_i \\ \dot{R}_i &= \gamma_i I_i - \delta R_i\end{aligned} \qquad [3]$$

where the parameter σ_{ij} governs the level of cross-immunity to strain i induced by previous exposure to strain j. This is just one model among a wide diversity of partial cross-immunity models where previous exposure could, for instance, affect transmission and/or recovery rates [41]. From the perspective of model development, partial cross-immunity is rapidly overwhelming because of the multitude of possible scenarios. Under some assumptions (e.g., symmetric interactions, effects on transmission rates only) it is possible to simplify and fully characterize the dynamics of these models [42]; however, knowing whether specific pathogens follow one scenario or another is itself a challenge [43].

The analysis of long-term epidemiological time series that keep track of the circulation of different strains can help to illuminate these multiple strain dynamics. For instance, Koelle *et al.* [44] examined the oscillations between two serotypes of cholera (Inaba and Ogawa) in Bangladesh over 40 years. These two serotypes have very similar phenotypes (severity of the infection, survival in the environment) but have distinct antigenic determinants. Koelle *et al.* [44] showed that a high level of cross-immunity with demographic stochasticity could yield sustained cycles among those serotypes. Reich *et al.* [45] analyzed the circulation of the four antigenically distinct serotypes of dengue virus in Thailand over 40 years and inferred the level of short-term cross-immunity among those different strains. In dengue, accurate prediction of the circulation of these four serotypes is particularly relevant given that the more serious manifestation of the disease (hemorrhagic fever) might occur more frequently on reinfection with a different serotype [46]. In both of these analyses, the availability of long-term datasets is key to the reliable development of

multistrain epidemiological models–and, of course, the higher the number of circulating strains the more difficult such inference becomes [43]. One needs either to adopt simplifying assumptions or to reduce the ambition of the predictions. For instance, more than 90 different serotypes of *Streptococcus pneumoniae* can induce pneumococcal disease. Because the pneumococcal conjugate vaccine targets a subsample of seven of these serotypes, many studies have pooled all of the non-vaccine serotypes (NVTs) into one compartment. This simplification allowed these models to track serotype replacement following vaccination campaigns [47,48]. These different studies demonstrate the predictive power of multistrain models based on the analysis of longitudinal studies tracking the diversity of pathogens over time. Such longitudinal studies remain relatively rare but the monitoring of the epidemiology and evolution of influenza virus provides a notable exception.

Influenza Evolution and Optimization of Vaccination Design

The Spanish flu pandemic of 1918–1920 revealed the potential threat associated with the emergence of new genetic variants in influenza. Since this pandemic, the molecular evolution of influenza viruses has been monitored at a global scale and these data streams have been invaluable in unveiling its intricate dynamics. As well as pandemic threats, the build-up of immunity in human populations continually selects for new antigenic variants of seasonal influenza, and the associated process of antigenic drift yields very specific patterns of molecular evolution. In particular, the phylogeny of the hemagglutinin gene of influenza A has a distinct shape that supports the idea that this gene is under very strong selection [44,49,50]. The rapid turnover of influenza strains and the dominance of a limited number of antigenic variants at any given point in time have stimulated research on the predictability of influenza evolution. Understanding this evolution has important practical implications because it could help us design more effective vaccines. Influenza vaccines are regularly updated to keep up with the evolution of the virus. The use of a predictive model of influenza evolution could greatly help in designing vaccine with the variants that are most likely to spread successfully in the following seasons.

Various attempts have been made to predict this antigenic evolution. Bush *et al.* [51] identified 18 codons in the immunogenic part of hemagglutinin under strong positive selection. Plotkin *et al.* [52] showed that strains that differ at these sites can be aggregated into antigenic clusters. Mutations leading to the emergence of a new antigenic cluster are likely to seed the next epidemics. The identification of relevant antigenic variation may be used to improve the selection of influenza strains used to develop the next flu vaccine [52]. Luksza and Lässig [53] improved this method with a more detailed estimation of the fitness of circulating variants. This estimation of fitness accounts for the occurrence of both deleterious (non-epitope) and beneficial (epitope) mutations in the virus genome as well as the epidemiological state of the host. Interestingly, their model allows epidemiology to feedback on fitness; strains that have been circulating for longer build up a higher level of herd immunity and thus have fewer transmission opportunities (Figure 3). The optimal vaccine strains predicted by this model better track the evolution of influenza than the actual vaccine strains recommended in the Northern Hemisphere from 1994 to 2012 [53]. More recently, several studies improved these predictions by combining information based on hemagglutination inhibition (HI) assays and genetic information and the shape of genealogical trees [54–56]. The predictive power of these models is encouraging, but once again their implementation illustrates the need for accurate time series as well as additional phenotypic (e.g., HI) and genetic information on the circulation of multiple strains.

Accounting for the Effects of Mutations

Most of the multistrain models described above are built on the standing genetic variation of the pathogen population and do not take into account the influx of new pathogen genotypes by mutation. Yet mutation (as well as horizontal gene transfer, recombination, and/or reassortment

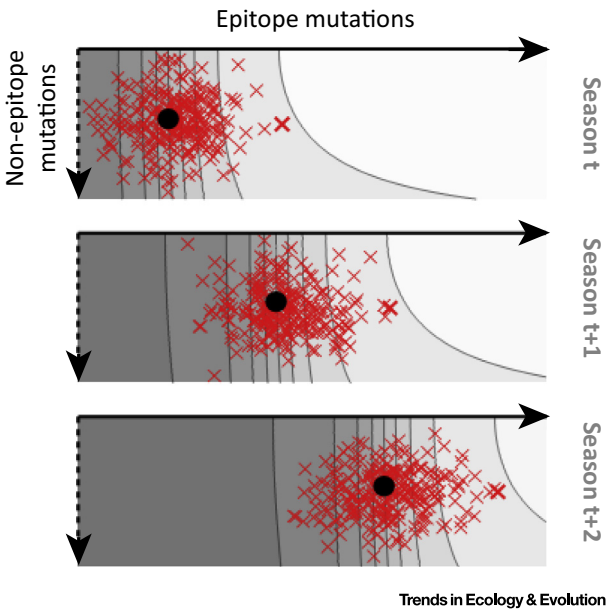

Figure 3. Influenza Evolution and Vaccine Optimization. Schematic description of influenza virus evolution over three consecutive seasons. Each panel presents the position of the current circulating strain (black dot) on a 2D fitness landscape. The first dimension (horizontal axis) is the antigenic phenotype (e.g., number of epitope mutations) and the second dimension (vertical axis) measures the effect of deleterious mutations on other parts of the genome (e.g., number of non-epitope mutations); the shades of gray indicate fitness measures associated with each phenotype (lighter shades refer to higher fitness). Each season, new strains are generated by mutation (red crosses) and the strain with the highest fitness (bold red crosses) is most likely to seed the next season's epidemic. Note how the emergence and the spread of new strains constantly modifies the fitness landscape.

in some microbes) is key because it fuels the evolution of the pathogen population. Models of adaptation show how the dynamics of adaption depends on the interplay between the influx of new mutations, selection, and genetic drift [57–59]. When selection is strong relative to the input of mutation, the population evolves through the independent fixation of isolated adaptive substitutions. By contrast, when the input of mutation is large relative to selection several different genotypes may circulate and interfere with one another. In this scenario, the balance between mutation, selection, and genetic drift may result in the maintenance of genetic variation. This heritable variation is the fuel of population adaptation and may be used to determine the instantaneous rate of adaptation of the population. Gerrish and Sniegowski [60] developed an algorithm to forecast the fitness and phenotypes of a population in a constant environment several generations in the future. The algorithm derives from a description of evolution as an unclosed hierarchy of dynamical equations for the cumulants of the distribution of fitness. A major value of the above models of adaption is the identification of the type of information needed to develop predictive models of infectious disease evolution.

To account for the influx of new mutations one first needs good estimates of the mutation rate. Fortunately, estimates of mutation rates are available for a broad range of microorganisms [61,62]. Second, it is important to take into account the phenotypic effects of these mutations. The large number of possible mutations and their interactions may generate an intrinsic limit to long-term evolutionary predictions [63]. However, short-term predictions may be feasible with good approximate descriptions of the distribution of the fitness effects of these mutations. Recently, several fitness-landscape models have been developed to derive expectations for the distribution of fitness effects of new mutations [64,65]. For instance, the geometric model of adaptation provides successful predictions of the distributions of the fitness effects of mutations of microbes [66,67]. In particular, this model successfully captures the change of these distributions when the population is increasingly close to the fitness optimum [68,69]. Interestingly the distribution of fitness effects of mutations may differ among different drugs. For instance, Chevereau *et al.* [70] studied the effects of mutations on resistance in *Escherichia coli* against nitrofurantoin is substantially lower than against other drugs. As expected by the theory, the dynamics of adaptation are much slower against this drug.

Theoretical models also show the impact of population size on the evolutionary process. First, small populations cannot maintain a large amount of genetic variability and consequently adaptation results from the sequential fixation of independent beneficial mutations. Second, small populations allow deleterious mutations to become fixed. Although microbial populations can reach large sizes they are often subject to strong bottlenecks and selective sweeps. For instance, the antigenic evolution of influenza is characterized by selective sweeps of specific antigenic mutants driven by the buildup of immunity in the host population. Because of the limited amount of intrasegment recombination in influenza [71], each selective sweep is often associated with hitchhiking by several deleterious mutations. As indicated by Koelle and Rasmussen [72], the accumulation of deleterious mutations can affect the rate of antigenic evolution and the shape of the influenza phylogenetic tree. Further theoretical work is required to develop a comprehensive analytical framework accounting for the interplay between demographic stochasticity and the evolutionary dynamics of pathogens.

Models of adaptation are also often based on the simplifying assumption that the mean fitness is entirely governed by the genetic composition of the evolving population [57–60]. Yet the environment (i.e., the density and composition of the host population as well as the abiotic environment) is also a dynamical variable that can change dramatically from one generation to the next. This is particularly true for influenza (where the change in the immune status of individuals drives antigenic evolution) but applies more broadly to host–parasite interactions [72–74]. Taking into account these epidemiological feedbacks could further improve the predictive power of evolutionary epidemiology theory.

Outstanding Questions

Is it possible to identify the factors limiting the predictability of pathogen dynamics? Are these factors intrinsic properties of specific pathogens or can the limits be pushed back with more and/or better data?

Is it possible to develop predictive numerical methods that can help us design more effective vaccines against seasonal influenza?

Can we use forecasting models to optimize the use of antimicrobial drugs at different spatial scales?

Is it possible to infer pathogen life-history traits like transmission rates and recovery rates from phylogenetic trees based on neutral genetic variation?

How do demographic and environmental stochasticity affect the predictive power of numerical models of pathogen evolution?

Concluding Remarks

In our rapidly changing world, forecasting ecological dynamics has an expanding role in policy and management [75–77]. In particular, being able to predict infectious disease dynamics is necessary to generate useful recommendations for public-health management and interventions [78–85]. Most epidemiological models are based on the simplifying assumption that pathogens do not evolve. It is clear, however, that taking into account this evolution may help improve predictions of epidemiological dynamics for a broad array of pathogens relevant to a diverse set of public-health issues. Being able to predict the evolution of important pathogen traits may be particularly relevant in settings including drug resistance, virulence, and vaccine escape. Here we have reviewed recent attempts to predict the epidemiological and evolutionary dynamics of pathogens.

Evolutionary biology provides a framework for both shaping our intuition and probing the data to better understand and predict the effect of alternative public-health interventions on the evolution of pathogens. Classically, studies have been based on adaptive dynamics-derived approaches and thus make predictions in the very long term. Some of the most exciting studies we reviewed here mix epidemiological and evolutionary dynamics to capture the transient evolutionary dynamics of pathogens. To date, most predictions coupling epidemiology and evolution remain qualitative but, in principle, more quantitative predictions are feasible. Two factors are currently limiting this advance.

First, parameterizing multistrain models requires long-term studies by ‘epidemiological detectives’ monitoring the circulation of many different pathogen genotypes [86]. An increasing number of epidemiological studies provide genetic information on the pathogens. Real-time tracking of virus evolution is feasible in influenza [87] and could be generalized to other pathogens. To predict the competitive abilities of different genotypes additional information is needed about the phenotypes of these genotypes. Some phenotypes, like drug resistance or antigenicity, may be accessible through *in vitro* experiments, although translation to *in vivo* is always challenging. Furthermore, the development of an international surveillance network is

necessary to monitor the dynamics of pathogens with these specific phenotypes at different spatial scales [88]. Other phenotypes like virulence and transmission are more difficult to estimate because they often require *in vivo* experiments. Molecular epidemiology and phylodynamics may provide an alternative to infer the relative fitness of different pathogen genotypes or key life-history phenotypes (transmission, recovery, virulence) of different clades [40,55,56,89–91] and thus bypass phenotypic assays (although this comes with limitations in terms of understanding the mechanisms). However, further theoretical development is required to obtain a robust statistical framework to infer variations of life-history traits from phylogenetic trees [55,92]. In addition, more appropriate sampling may need to be developed to extract population-level information from sequence data [93].

Second, modeling the interplay between epidemiology and evolution raises several theoretical challenges. For instance, the management of multidrug resistance requires models with multiple loci [94]. Other difficulties remain to be addressed, like the effects of demographic stochasticity, environmental stochasticity, and spatial structure on the evolutionary epidemiology of pathogens. Like weather forecasting, it seems likely that numerical models should be able to incorporate the increasing amount of information that is becoming available on a growing number of pathogens [95]. Interestingly, the success of weather forecasting despite the complex nonlinear dynamics of meteorological phenomena may inspire new ways to cope with our ignorance regarding the mechanistic details of epidemiological processes. Elaborate ensemble forecasting methods have been developed to incorporate uncertainties in initial conditions and model formulations [2,23,24]. The development of such a probabilistic approach of forecasting allows us to evaluate the predictability of different meteorological events in different geographical areas. For instance, the climate in the tropics is strongly determined by the underlying sea-surface temperature and is weakly affected by the initial conditions of the atmosphere [96]. This allows us to predict large-scale seasonal circulation and rainfall several seasons in advance. It would be particularly useful to evaluate the predictability of the epidemiological and evolutionary dynamics of different infectious diseases. Like the tropical weather, the dynamics of some pathogens may be more predictable because they are forced by potent environmental factors (e.g., influx of susceptible hosts in measles). By contrast, other pathogens may be sensitive to a plethora of complex forces that may reduce the forecast horizon [76]. A major advantage of numerical weather forecasting is the possibility of testing predictions with a constant flux of observations so that success and failure can be used to continuously improve predictive skills. Similarly, future progress in forecasting the epidemiology and evolution of infectious diseases hinges on the development of inference methods, evolutionary epidemiology theory, and epidemiological surveillance networks to allow validation of past predictions.

As J.B.S Haldane said in his collection of writings *Adventures of a Biologist*, 'No scientific theory is worth anything unless it enables us to predict something which is actually going on. Until that is done, theories are a mere game of words, and not such a good game as poetry' [97]. Previous attempts to predict the dynamics of infectious diseases indicate that combining epidemiological and evolutionary models is often necessary to elucidate and forecast the complexity of pathogen dynamics. In particular these models may be used to explore the factors that limit the predictability of pathogen dynamics (e.g., chaotic dynamics, behavioral changes affecting epidemiology, lack of information on the amount of genetic variation in key life-history traits). Identifying these limiting factors will be particularly useful for exploring new ways to improve the forecasting of infectious disease dynamics (see Outstanding Questions).

Acknowledgments

The authors thank Trevor Bedford, Marc Choisy, Guillaume Martin, and Sébastien Lion for many inspiring discussions.

References

1. Richardson, L.F. (1922) *Weather Prediction by Numerical Process,* Cambridge University Press
2. Bauer, P. *et al.* (2015) The quiet revolution of numerical weather prediction. *Nature* 525, 47–55
3. Anderson, R.M. and May, R.M. (1991) *Infectious Diseases of Humans,* Oxford University Press
4. Diekmann, O. and Heesterbeek, J.A.P. (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation,* Wiley
5. Keeling, M.J. and Rohani, P. (2008) *Modelling Infectious Diseases,* Princeton University Press
6. Bernoulli, D. (1766) Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. In *In Histoire de l'Académie Royale des Sciences: Mémoires de Mathématiques et de Physique*, pp. 1–45, Imprimerie Royale (in French)
7. Ross, R. (1915) Some *a priori* pathometric equations. *BMJ* 1, 546-447
8. Kermack, W.O. and McKendrick, A.G. (1927) A contribution to the mathematical theory of epidemics. *Proc. R. Soc. A* 115, 700–721
9. MacDonald, G. (1956) Epidemiological basis of malaria control. *Bull. World Health Organ.* 15, 613–626
10. Wolfson, L.J. *et al.* (2009) Estimates of measles case fatality ratios: a comprehensive review of community-based studies. *Int. J. Epidemiol.* 38, 195–205
11. Finkenstädt, B.F. and Grenfell, B.T. (2000) Time series modelling of childhood diseases: a dynamical systems approach. *J. R. Stat. Soc. Ser. C Appl. Stat.* 49, 187–205
12. Bjornstad, O.N. *et al.* (2002) Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* 72, 169–184
13. Grenfell, B.T. *et al.* (2002) Dynamics of measles epidemics: scaling noise, determinism, and predictability with the TSIR model. *Ecol. Monogr.* 72, 185–202
14. Takahashi, S. *et al.* (2015) Reduced vaccination and the risk of measles and other childhood infections post-Ebola. *Science* 347, 1240–1242
15. Olsen, L. and Schaffer, W. (1990) Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science* 249, 499–504
16. Ferrari, M.J. *et al.* (2008) The dynamics of measles in sub-Saharan Africa. *Nature* 451, 679–684
17. Dalziel, B.D. *et al.* (2016) Persistent chaos of measles epidemics in the prevaccination United States caused by a small change in seasonal transmission patterns. *PLoS Comput. Biol.* 12, e1004655
18. Jansen, V.A.A. *et al.* (2003) Measles outbreaks in a population with declining vaccine uptake. *Science* 301, 804
19. Roy, M. *et al.* (2014) Epidemic cholera spreads like wildfire. *Sci. Rep.* 4, 3710
20. Caudron, Q. *et al.* (2015) Predictability in a highly stochastic system: final size of measles epidemics in small populations. *J. R. Soc. Interface* 12, 20141125
21. Keeling, M.J. *et al.* (2001) Seasonally forced disease dynamics explored as switching between attractors. *Physica D* 148, 317–335
22. Rozhnova, G. *et al.* (2013) Characterizing the dynamics of rubella relative to measles: the role of stochasticity. *J. R. Soc. Interface* 10, 20130643
23. Epstein, E.S. (1969) Stochastic dynamic prediction. *Tellus* 21, 739–759
24. Leith, C.E. (1974) Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.* 102, 409–418
25. Axelsen, J.B. *et al.* (2014) Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9538–9542
26. Lynch, M. and Walsh, B. (1998) *Genetics and Analysis of Quantitative Traits,* Sinauer Associates
27. Berngruber, T.W. *et al.* (2013) Evolution of virulence in emerging epidemics. *PLoS Pathog.* 9, e1003209
28. Gandon, S. *et al.* (2001) Imperfect vaccines and the evolution of pathogen virulence. *Nature* 414, 751–756
29. Gandon, S. and Day, T. (2007) The evolutionary epidemiology of vaccination. *J. R. Soc. Interface* 4, 803–817
30. Mackinnon, M.J. and Read, A.F. (2003) Effects of immunity on relationships between growth rate, virulence and transmission in semi-immune hosts. *Parasitology* 126, 103–112
31. Barclay, V.C. *et al.* (2012) The evolutionary consequences of blood-stage vaccination on the rodent malaria *Plasmodium chabaudi*. *PLoS Biol.* 10, e1001368
32. Read, A.F. *et al.* (2015) Imperfect vaccination can enhance the transmission of highly virulent pathogens. *PLoS Biol.* 13, e1002198
33. Goossens, H. *et al.* (2005) Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet* 365, 579–587
34. Seppälä, H. *et al.* (1997) The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland. Finnish Study Group for Antimicrobial Resistance. *N. Engl. J. Med.* 337, 441–446
35. Austin, D.J. *et al.* (1999) The relationship between the volume of antimicrobial consumption in human communities and the frequency of resistance. *Proc. Natl. Acad. Sci. U.S.A.* 96, 1152–1156
36. Lipsitch, M. (2001) The rise and fall of antimicrobial resistance. *Trends Microbiol.* 9, 438–444
37. Lipsitch, M. *et al.* (2000) The epidemiology of antibiotic resistance in hospitals: paradoxes and prescriptions. *Proc. Natl Acad. Sci. U.S.A.* 97, 1938–1943
38. Wargo, A.R. *et al.* (2007) Competitive release and facilitation of drug resistant parasites following therapeutic chemotherapy in a rodent malaria model. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19914–19919
39. Day, T. and Read, A.F. (2016) When does high-dose antimicrobial chemotherapy prevent the emergence of resistance? *PLoS Comput. Biol.* 12, e1004689
40. Luciani, F. *et al.* (2009) The epidemiological fitness cost of drug resistance in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14711–14715
41. Kucharski, A.J. *et al.* (2015) Capturing the dynamics of pathogens with many strains. *J. Math. Biol.* 72, 1–24
42. Gog, J.R. and Grenfell, B.T. (2002) Dynamics and selection of many-strain pathogens. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17209–17214
43. Shrestha, S. *et al.* (2011) Statistical inference for multi-pathogen systems. *PLoS Comput. Biol.* 7, e1002135
44. Koelle, K. *et al.* (2006) Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* 314, 1898–1903
45. Reich, N.G. *et al.* (2013) Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. *J. R. Soc. Interface* 10, 20130414
46. Mongkolsapaya, J. *et al.* (2003) Original antigenic sin and apoptosis in the pathogenesis of dengue hemorrhagic fever. *Nat. Med.* 9, 921–927
47. Weinberger, D.M. *et al.* (2011) Serotype replacement in disease following pneumococcal vaccination: a discussion of the evidence. *Lancet* 378, 1962–1973
48. Croucher, N.J. *et al.* (2013) Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat. Genet.* 45, 656–663
49. Ferguson, N.M. *et al.* (2003) Ecological and immunological determinants of influenza evolution. *Nature* 422, 428–433
50. Bedford, T. *et al.* (2012) Canalization of the evolutionary trajectory of the human influenza virus. *BMC Biol.* 10, 38
51. Bush, R.M. *et al.* (1999) Predicting the evolution of human influenza A. *Science* 286, 1921–1925

52. Plotkin, J.B. *et al.* (2002) Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6263–6268
53. Luksza, M. and Lässig, M. (2014) A predictive fitness model for influenza. *Nature* 507, 57–61
54. Bedford, T. *et al.* (2014) Integrating influenza antigenic dynamics with molecular evolution. *Elife* 3, e01914
55. Neher, R.A. *et al.* (2014) Predicting evolution from the shape of genealogical trees. *Elife* 3, e03568
56. Neher, R.A. *et al.* (2016) Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6263–6268
57. Gerrish, P. and Lenski, R. (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* 102-103, 127–144
58. Desai, M.M. and Fisher, D.S. (2007) Beneficial mutation selection balance and the effect of genetic linkage on positive selection. *Genetics* 176, 1759–1798
59. Good, B.H. *et al.* (2012) Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4950–4955
60. Gerrish, P.J. and Sniegowski, P.D. (2012) Real time forecasting of near-future evolution. *J. R. Soc. Interface.* 9, 2268–2278
61. Drake, J.W. *et al.* (1998) Rates of spontaneous mutation. *Genetics* 148, 1667–1686
62. Holmes, E.C. (2009) *The Evolution and Emergence of RNA Viruses,* Oxford University Press
63. Day, T. (2011) Computability, Gödel's incompleteness theorem, and an inherent limit on the predictability of evolution. *J. R. Soc. Interface* 9, 624–639
64. Orr, H.A. (2005) The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* 6, 119–127
65. Martin, G. and Lenormand, T. (2006) A general multivariate extension of Fisher's geometric model and the distribution of mutation fitness effects across species. *Evolution* 60, 893–907
66. Kassen, R. and Bataillon, T. (2006) Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nat. Genet.* 38, 484–488
67. Tenaillon, O. (2015) The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* 45, 179–201
68. Martin, G. *et al.* (2007) Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* 39, 555–560
69. Couce, A. and Tenaillon, O. (2015) The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* 6, 99
70. Chevereau, G. *et al.* (2015) Quantifying the determinants of evolutionary dynamics leading to drug resistance. *PLoS Biol.* 13, e1002299
71. Boni, M.F. *et al.* (2008) Homologous recombination is very rare or absent in human influenza A virus. *J. Virol.* 82, 4807–4811
72. Koelle, K. and Rasmussen, D.A. (2015) The effects of a deleterious mutation load on patterns of influenza A/H3N2s antigenic evolution in humans. *Elife* 4, e07361
73. Gandon, S. and Day, T. (2009) Evolutionary epidemiology and the dynamics of adaptation. *Evolution* 63, 826–838
74. Luo, S. and Koelle, K. (2013) Navigating the devious course of evolution: the importance of mechanistic models for identifying eco-evolutionary dynamics in nature. *Am. Nat.* 181, S58–S75
75. Clark, J.S. *et al.* (2001) Ecological forecasts: an emerging imperative. *Science* 293, 657–660
76. Petchey, O.L. *et al.* (2015) The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.* 18, 597–611
77. Mouquet, N. *et al.* (2015) Predictive ecology in a changing world. *J. Appl. Ecol.* 52, 1293–1310
78. Woolhouse, M. (2011) How to make predictions about future infectious disease risks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2045–2054
79. Shaman, J. and Karspeck, A. (2012) Forecasting seasonal outbreaks of influenza. *Proc. Natl. Acad. Sci. U.S.A.* 109, 20425–20430
80. Shaman, J.A. *et al.* (2013) Real-time influenza forecasts during the 2012-2013 season. *Nat. Commun.* 4, 2837
81. Rodó, X. *et al.* (2013) Climate change and infectious diseases: can we meet the needs for better prediction? *Clim. Change* 118, 625–640
82. Corley, C.D. *et al.* (2014) Disease prediction models and operational readiness. *PLoS ONE* 9, e91989
83. Heesterbeek, H. *et al.* (2015) Modeling infectious disease dynamics in the complex landscape of global health. *Science* 347, aaa4339
84. Roy, M. *et al.* (2015) Predictability of epidemic malaria under non-stationary conditions with process-based models combining epidemiological updates and climate variability. *Malar. J.* 14, 419
85. Yang, W. *et al.* (2015) Forecasting influenza epidemics in Hong Kong. *PLoS Comput. Biol.* 11, e1004383
86. Hilborn, R. and Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data,* Princeton University Press
87. Neher, R.A. and Bedford, T. (2015) nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics* 31, 3546–3548
88. Grundmann, H. (2014) Towards a global antibiotic resistance surveillance system: a primer for a roadmap. *Ups. J. Med. Sci.* 119, 87–95
89. Grenfell, B.T. *et al.* (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303, 327–332
90. Stadler, T. (2011) Inferring epidemiological parameters on the basis of allele frequencies. *Genetics* 188, 663–672
91. Luksza, M. *et al.* (2014) *Epidemiological and evolutionary analysis of the 2014 Ebola virus outbreak.* arxiv:1411.1722
92. Stadler, T. (2011) Inferring speciation and extinction processes from extant species data. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16145–16146
93. Stack, J.C. *et al.* (2010) Protocols for sampling viral sequences to study epidemic dynamics. *J. R. Soc. Interface* 7, 1119–1127
94. Day, T. and Gandon, S. (2012) Evolutionary epidemiology of multilocus drug resistance. *Evolution* 66, 1582–1587
95. Pybus, O.G. *et al.* (2013) Evolutionary epidemiology: preparing for an age of genomic plenty. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120193
96. Shukla, J. (1998) Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* 282, 728–731
97. Haldane, J.B.S. (1937) *Adventures of a Biologist,* Harper & Bros
98. Day, T. and Gandon, S. (2007) Applying population-genetic models in theoretical evolutionary epidemiology. *Ecol. Lett.* 10, 876–888
99. Lande, R. (1976) Natural-selection and random genetic drift in phenotypic evolution. *Evolution* 30, 314–334
100. Day, T. and Proulx, S.R. (2004) A general theory for the evolutionary dynamics of virulence. *Am. Nat.* 163, E40–E63