Five challenges in evolution and infectious diseases

C.J.E. Metcalf[1,2,*], R. Birger[1], S. Funk[3], R.D. Kouyos[4], J.O. Lloyd-Smith[2,5] , V.A.A. Jansen[6]

[1]Dept of Ecology and Evolutionary Biology, Princeton University, Princeton NJ, USA
[2] Division of Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States of America
[3]Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene & Tropical Medicine, London, UK
[4]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland
[5]Dept of Ecology and Evolutionary Biology, UCLA, Los Angeles, USA
[6]School of Biological Sciences, Royal Holloway University of London, Egham, UK

*corresponding author: cmetcalf@princeton.edu

Abstract
Evolution is a key aspect of the biology of many pathogens, driving processes ranging from immune escape to changes in virulence. Because evolution is inherently subject to feedbacks, and because pathogen evolution plays out at scales ranging from within-host to between-host and beyond, evolutionary questions provide special challenges to the modelling community. In this article, we provide an overview of five challenges in modelling the evolution of pathogens and their hosts, and point to areas for development, focussing in particular on the issue of linking theory and data.

Introduction
Evolution is the change in gene frequencies resulting from selection (where genes with greater reproductive contributions to future generations spread within populations), mutation, recombination or re-assortment (where genetic material is exchanged between chromosomes), or drift. Evolution plays an important role in the dynamics of many infectious diseases. Vaccine escape in influenza, drug-resistance in HIV, and virulence evolution in Marek's disease are all examples of evolutionary processes. Developing models that accurately describe pathogen evolution is inherently challenging because of the complexity of pathogen life cycles and the difficulty in characterizing the (dynamic) fitness landscapes driving pathogen evolution. Ultimately, the pathogen's genotype, together with the characteristics of the host, determines both how disease is caused and how much of the pathogen is emitted by the host. Once emitted, pathogens must infect new hosts. How much transmission is realised also depends on the physical environment, the host's behaviour and population structure, as well as the distribution of the disease in the population. To understand pathogen evolution we need to integrate from the genotype, and span these levels, encompassing stochastic processes such as transmission bottlenecks (see, [1]). This requires the integration of knowledge from

various fields: molecular biology, microbiology, medicine and epidemiology to name a few (Figure 1).

Here, we outline five challenges of modelling evolution that reflect this interaction across scales. We start by detailing the most basic and general challenge of all, that of characterizing fitness. Next, we address challenges for modelling how pathogens shape each other's evolution (coinfection) and the related topic of how pathogens shape host immune diversity; and the classic evolutionary problem of what forces allow maintenance of pathogen diversity (coexistence). Finally, we discuss how modelling can help us understand how mechanisms of pathogen replication influence the generation of genetic variation, upon which selection acts.

Challenge 1: Defining and measuring fitness for pathogens across scales

If we know how fitness changes with changes in the genes in the pathogen, and how it does so across scales (Figure 1), we can make informed statements about selection and adaptation. Fitness is generally defined as the reproductive contribution of an individual to the next generation, in a particular environment. Pathogens will experience different such environments over the course of an infection: for instance, they will have to overcome the host's defences, colonise the host, withstand attacks of the immune system, and accomplish transmission and infection. The components of fitness can vary over such a cycle (and indeed the cycle often involves numerous pathogen generations), and to calculate fitness, an appropriate average has to be taken over this path, integrating information across various scales.

Although defining fitness of pathogens is straightforward in principle, linking this definition to attainable data in order to quantify fitness is not. Researchers have typically broken the evolutionary cycle apart to focus on particular levels of selection – for example, distinguishing within-host fitness (describing the growth of the pathogen population within an infected individual) and between-host fitness (describing transmission of infection to new host individuals). This has the benefit of corresponding to clear biological differences, as well as quantities that can be measured (although the path to building back across scales to fitness is not obvious). However, even with the process broken down into more manageable parts, there are still considerable barriers to defining scale-specific fitness components (see Challenge 2 in [1] for more complexities related to within-host fitness), and there is no general relationship between fitness at the within-host scale and the number of new hosts infected [2].

The challenges inherent to even the (apparently contained) problem of measurement of the reproduction of individual pathogens within-host has led to the development of a range of *in vitro* systems designed to quantify variation in rates of pathogen replication in different contexts. Inevitably, these estimated pathogen replication rates reflect only one aspect of fitness at an *in vivo* scale. Key modelling challenges include providing further innovations in linking *in vitro* data-streams to *in vivo* measurements of aspects of fitness, such as viral titre kinetics or the outcome of competition assays [3] (see also Challenge 7, [4]), and accounting for the fact that

the genotype to phenotype to fitness map is likely to be context-dependent, and the within-host fitness landscape may change (see also Challenge 3, [1]; and challenge 6 [4] on the challenges of developing genotype to phenotype maps). In particular, the fitness of a genotype will often depend on the frequency of all other genotypes, as a result of immune system activity. Machine-learning and modelling approaches can be used to bridge the *in-vitro* and *in-vivo* levels [5], but given the nature of the underlying *in-vitro* data such approaches currently neglect crucial within-host fitness determinants such as the immune system. This poses the challenge of finding novel ways to parameterize the activity of the host's immune system from data (e.g., [6]) and incorporate it into the models for pathogen fitness, and the associated (and shifting) fitness landscapes.

Beyond the individual host, other instances of population structure (e.g., age groups or host species) may influence pathogen evolution. If these host classes additionally compete or otherwise interact, this will affect the pathogen's evolution, and any evolutionary outcomes are likely to depend on the details of this interaction. The next generation matrix approach is useful for these types of systems [7], as are approaches that renormalize the system to describe group-level reproduction [8], but few general mathematical principles are known, and furthermore, parameterizing such models given available data remains challenging [9] (see also [10]).

Bringing together all these various threads to estimate an all-encompassing fitness value for any particular pathogen genotype is a major challenge, and would be even if all the data were available. Even though conceptual and mathematical frameworks for dealing with such multi-scale processes have been developed [2, 11], such calculations can be extremely cumbersome, and their interpretation complex, particularly when evolution operates at different scales, as is the case for pathogens.

Challenge 2: Developing models to capture the impact of co-infection on the evolutionary process

For many pathogens, infection by multiple strains or other pathogens may have little or no epidemiological impact – the key distinction is simply whether a host is infected, or not (as for measles, for instance). However, there are pathogens for which coinfection alters pathogen dynamics, and this can have two major impacts on evolutionary processes. First, coinfection can lead to genetic exchange between co-infecting pathogens (especially for viruses and bacteria) that may be essential to immune escape, or host jumps. Such exchanges may occur both among pathogens of the same species (e.g. homologous recombination) or among pathogens of different species (transformation, transduction, conjugation) (see Challenge 5, [4] for more details). Second, coinfection may be associated with within-host competition (e.g., mouse malaria parasites may compete for red blood cells), or facilitation (e.g., helminth mediated immune-suppression might increase microparasite within-host growth rates), both of which can alter fitness, and thus evolutionary outcomes. The resulting dynamics are analogous to individuals competing within metapopulations,

and can be formulated theoretically using kin selection or multilevel selection formalisms [11], and these models can be extended to encompass within-patch dynamics [12]. However, such methods assume known costs and benefits to the interaction between competing coinfecting pathogens. This implies developing adequate within-host models, a tremendous challenge given sparse data on the complex nature of pathogen-pathogen interactions, as mediated by immune response, resource competition and treatment (see also [1]). Where sufficient elements of the biology are known, but the specifics of coinfection interactions remain unclear, models can be deployed to explore the outcomes of different interactions in terms of measurable quantities to identify key mechanisms [13]. These frameworks can then be used to prioritize experimental directions that inform expected evolutionary trajectories. A key area for research of this kind is on how drug resistance spreads in the face of different treatment regimes in the context of co-infection [3, 14].

Theoretical challenges also remain: for some co-infecting pathogens, the order of infection can affect prognosis, and whether infections are sequential, co-transmitted, or super-infecting can change dynamics (see also Challenge 5, [1]). While incorporating such effects into models can be straightforward, realistically describing interdependent processes such as immunosuppression or cross-reactive immunity can be complicated. Modelling evolution under coinfection presents a special theoretical challenge for bacterial communities, because of the genetic exchange between species through mobile genetic elements such as plasmids and phages (see also below, challenge 5). The development of models that can capture community assembly, invasibility, competition, and immune interactions within the bacterial microbiome may be an especially rich area for modelling, in particular with reference to exploiting the potential of meta-genomic and meta-transcriptomic data. Very few models have considered, e.g., the dynamics of traits within a community of pathogens that are exchanging genetic material (see [15, 16]).

Challenge 3: Modelling how pathogen characteristics shape the evolution of host immune diversity

Host-parasite interactions reflect an inherently co-evolutionary process. Despite this, the research focus to date has been somewhat one-sided. Models of selection pressures on the parasite (e.g., evolution of virulence, etc.) are widespread [17]. Research into the host's potential for coevolution [18] in terms of diversity of immune responses also has a relatively long history (e.g., [19]). However, key features of host biology such as immunopathology [20], host variability in tolerance [21] or in susceptibility [22], and structure across host immune recognition loci [23] are just starting to be considered. This bias reflects in part the difference of time-scales in operation – parasites' generation times are generally so much shorter than hosts' that a focus on parasite evolution alone seems justified. A key challenge is identifying when, or for what host traits, this assumption is no longer valid; and subsequently, identifying what parasite community features select for particular host responses. For example, what pathogen characteristics might select for a "dangerous" immune response, i.e., over-investment and risk of immune pathology?

There is also considerable opportunity to develop models that explore the mechanistic basic for results obtained in comparative immunology studies; e.g., correlations between promiscuity and white blood cell counts across primates [24]. Macro-evolutionary forces operating at broad scales and long timescales that act to shape host features will affect within-host dynamics, in particular those linked to aspects such as immunopathology. The role of feedbacks in this process is another very interesting question for which modelling might provide insights. For example, if changes in host ecology shift host longevity, this is likely to alter selection on pathogen virulence (assuming some form of virulence-transmission trade-off), which might then feed back onto altering selection on host longevity.

Challenge 4: Understanding maintenance of pathogen diversity.

A range of mechanisms can maintain pathogen diversity [25]. At the most basic level, neutral processes can maintain diversity, via a balance of mutation – creating diversity – and genetic drift – destroying diversity. Selective mechanisms will also play a role, with life-history trade-offs (e.g. between within-host replication and transmissibility) allowing stable coexistence; and temporal and spatial fluctuations in selection, allowing temporal or spatial niche separation and coexistence. For example, antibiotic consumption varies across individuals, age-classes, institutions (hospital versus community), regions and countries [26], which creates a mosaic of selective pressures that may allow coexistence of resistant and sensitive strains. Similarly, selection for antigenic escape varies as a function of differences in host genetics or exposure history (previous infections and vaccinations).

A key challenge is determining the relative contribution of neutral and selective mechanisms. For example, the diversity of HIV-1 has been regarded primarily as a result of the demography and geography of viral spread [25]. However, several studies have also found adaptive substitutions in HIV-1 in particular populations of humans [27] and it is currently unclear whether selective or neutral processes dominate in shaping the diversity of HIV-1. Similarly, Cobey et al. [28] argued that the diversity of *Streptococcus pneumoniae* can be explained by the interplay of niche and neutral effects. Building on these types of analyses to encompass a broader array of pathogens is a key direction for future research.

Another complication is that selective pressure may vary across scales (e.g., from within-host to between-host, see Challenge 1), but which scale is key to maintenance of diversity remains poorly resolved for most pathogens. For example, many bacterial species (such as *Staphylococcus aureus*) colonize different anatomical sites, which may impose different selective requirements [29]. Assessing the impact of this within-patient heterogeneity at the population level requires understanding the cross-scale dynamics of these pathogens, and models have a key role to play in linking the relevant mechanisms. Similarly, it has been argued for HIV that selection at the within-host and between host-level act in opposite directions [30]. Such a trade-off can help to maintain diversity, but the actual quantitative contribution of this mechanism remains to be determined.

Mathematical models have made a range of important predictions about the determinants of pathogen diversity. Much of the current concern about the wide spread use of antibiotics, and subsequent emergence or spread of resistance are at least implicitly based on evolutionary models. Likewise, the evolutionary consequences of imperfect vaccines were, at first, theoretical prediction [31]. More specific questions have also been tackled – such as how the strain distribution of *Streptococcus pneumoniae* responds to vaccination [28, 32], or the potential for coexistence of drug resistant and drug-sensitive strains in specific settings [33]. The challenging goal of making more detailed predictions relating to how much and when evolution will take place will require integrating the increasingly available data on selective forces (antibiotic consumption, host genetic variation, exposure to pathogens) and host demography into epidemiological models. Quantitative predictive models will also require much better estimates of pathogen fitness (see Challenge 1, especially with the application of in vitro systems to in vivo inference) and mutational pathways, with particular focus on compensatory mutations, fitness at different levels (within-host vs. transmission, see Challenge 1), and the interplay of selective and stochastic effects.

Finally, for some pathogens it is unclear why diversity is not larger than observed (HIV at the within-host scale, influenza at the population scale [25]). Modelling studies can play an important role in proposing mechanisms to explain this pattern [34], which can spur targeted empirical work to test these hypotheses.

Challenge 5. The impact of genetic systems for pathogen evolution

The operation of inheritance in bacteria is highly complex. Unrelated individuals may exchange DNA ('horizontal gene transfer') and many individuals carry plasmids (DNA that is physically separate of the organisms' genomic DNA, and can be replicated independently from it) that can often carry key genes such as those linked to antibiotic resistance. The environment that a gene will have experienced in its evolutionary past will depend on the route through which it is passed on: a gene on a plasmid will have been in different organisms and environments than a gene on a chromosome. Therefore the fitness, and selective forces on a gene depend on the details of the genetics. For many pathogens (viruses, as well as bacteria, etc.) related complexities emerge from the process of recombination (see [4] for more details). A challenge here is to develop models not just of individual pathogens, but also of individual genes, which take the genetic architecture into account.

Even for the comparatively straightforward process of mutation, the basic biology of pathogen replication can influence evolutionary dynamics strongly. The molecular- and cellular-scale mechanisms by which pathogens replicate their genomes and create new infectious particles are subjects of intensive study in microbiology and virology, but have been largely ignored in dynamic models of pathogen evolution. Recent theoretical models have shown that the within-host emergence of new viral strains is strongly affected by whether offspring virions are released via budding or bursting [35], by different mechanisms of genome replication [36], and by how

genomes and proteins are mixed in virion assembly [37]. One important challenge is to test and expand this developing body of theory by comparison with experimental data, particularly given new opportunities arising from deep sequencing data. Current models have only scratched the surface of microbiological knowledge, so there is scope to include many more details – one key task for modellers will be to determine if, when, and how these details impact evolutionary processes at higher scales, and to assess whether the cost in additional model complexity is worth bearing. For instance, recent progress on modelling viral replication mechanisms (e.g. [38]) opens interesting opportunities for cross-scale modelling to explore potential impacts on evolutionary dynamics. Do certain replication mechanisms confer greater phenotypic robustness to mutations, or higher propensity for adaptive evolution? Alternatively, could the details of pathogen replication present unrecognized barriers to adaptation such as the delayed expression of beneficial phenotypes (e.g. [37])? Conversely, better models of genome replication and viral packaging will advance the important goal of achieving better estimates of viral mutation rates [39].

A better understanding of the genetic systems of pathogens will be crucial for our understanding of how pathogen populations can move on fitness landscapes (see Challenge 1). For example fitness valleys can represent a barrier to adaptation a low mutation rates [40] but not at high mutation rates. Similarly the probability of evolutionary rescue of pathogen populations (e.g. in the context of antimicrobial therapy or vaccination) may crucially depend on mutation rates [41].

Conclusion

Many of the challenges in modelling pathogen evolution that we introduced here revolve around questions of quantifying fitness. We focused particularly on biological complexities and uncertainties, the impact of coinfection, and evolutionary mechanisms that create and shape diversity in hosts and pathogens. However, the effects of evolution on pathogen dynamics are vast, and potential modelling challenges reflect this. Other papers within this Special Issue tackle questions arising in the context of emergence of novel pathogens [42], vaccine escape [43], or in extending the use of phylodynamics [4]. Progress in tackling these challenges has the potential to contribute to a broad array of highly applied questions, including management of drug resistance, improvement of clinical care, and reconciling individual and population goals for public health in the context of pathogen evolution.

Figure 1: Illustration of the scales at which epidemiological dynamics and broad outlines of which fields pertain to which scales. Note that the scales are illustrative rather than exact.

Acknowledgements: This work emerged from discussion at the Isaac Newton Institute, Cambridge; and was funded by the Bill and Melinda Gates Foundation (CJEM), the National Science Foundation (EF-0928690; JLS), the Medical Research Council (SF) and the RAPIDD program of the Science & Technology Directorate,

Department of Homeland Security and the Fogarty International Center, National Institutes of Health (CJEM, JLS).

References

1. Gog, J., et al., *Seven challenges in modelling pathogen dynamics within-host and across scales.* Epidemics.
2. Park, M., et al., *Multiple scales of selection influence the evolutionary emergence of novel pathogens.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2013. **368**(1614): p. 20120333.
3. Huijben, S., et al., *Aggressive chemotherapy and the selection of drug resistant pathogens.* PLoS pathogens, 2013. **9**(9): p. e1003578.
4. Frost, S.D., et al., *Challenges in Phylodynamics.* Epidemics.
5. Kouyos, R.D., et al., *Assessing predicted HIV-1 replicative capacity in a clinical setting.* PLoS pathogens, 2011. **7**(11): p. e1002321.
6. Metcalf, C., et al., *Partitioning regulatory mechanisms of within-host malaria dynamics using the effective propagation number.* Science, 2011. **333**(6045): p. 984-988.
7. Diekmann, O., J. Heesterbeek, and M. Roberts, *The construction of next-generation matrices for compartmental epidemic models.* Journal of the Royal Society Interface, 2010. **7**(47): p. 873-885.
8. Ball, F., D. Mollison, and G. Scalia-Tomba, *Epidemics with two levels of mixing.* The Annals of Applied Probability, 1997: p. 46-89.
9. Funk, S., et al., *Identifying transmission cycles at the human-animal interface: the role of animal reservoirs in maintaining Gambiense Human African Trypanosomiasis.* PLoS computational biology, 2013. **9**(1): p. e1002855.
10. Buhnerkempe, et al., *Ten challenges in modelling multiple hosts and multiple infectious agents.* Epidemics.
11. Lion, S., V.A. Jansen, and T. Day, *Evolution in structured populations: beyond the kin versus group debate.* Trends in ecology & evolution, 2011. **26**(4): p. 193-201.
12. Jansen, V.A. and R. Vitalis, *The evolution of dispersal in a Levin's type metapopulation model.* Evolution, 2007. **61**(10): p. 2386-2397.
13. Mideo, N., et al., *Causes of variation in malaria infection dynamics: insights from theory and data.* The American Naturalist, 2011. **178**(6): p. E174.
14. Kouyos, R.D., et al., *The path of least resistance: aggressive or moderate treatment? .* Proceedings of the Royal Society B: Biological Sciences, In press.
15. Morton, E.R., et al., *Non-additive costs and interactions alter the competitive dynamics of co-occurring ecologically distinct plasmids.* Proceedings of the Royal Society B: Biological Sciences, 2014. **281**(1779): p. 20132173.
16. San Millan, A., K. Heilbron, and R.C. MacLean, *Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations.* The ISME journal, 2013.

17. Alizon, S., et al., *Virulence evolution and the trade‐off hypothesis: history, current state of affairs and the future.* Journal of evolutionary biology, 2009. **22**(2): p. 245-259.
18. Little, T.J., et al., *The coevolution of virulence: tolerance in perspective.* PLoS pathogens, 2010. **6**(9): p. e1001006.
19. Borghans, J.A., J.B. Beltman, and R.J. De Boer, *MHC polymorphism under host-pathogen coevolution.* Immunogenetics, 2004. **55**(11): p. 732-739.
20. Day, T., A.L. Graham, and A.F. Read, *Evolution of parasite virulence when host responses cause disease.* Proceedings of the Royal Society B: Biological Sciences, 2007. **274**(1626): p. 2685-2692.
21. Best, A., A. White, and M. Boots, *Maintenance of host variation in tolerance to pathogens and parasites.* Proceedings of the National Academy of Sciences, 2008. **105**(52): p. 20786-20791.
22. Boots, M., et al., *The importance of who infects whom: the evolution of diversity in host resistance to infectious disease.* Ecology letters, 2012. **15**(10): p. 1104-1111.
23. Penman, B.S., et al., *Pathogen selection drives nonoverlapping associations between HLA loci.* Proceedings of the National Academy of Sciences, 2013. **110**(48): p. 19645-19650.
24. Nunn, C.L., J.L. Gittleman, and J. Antonovics, *Promiscuity and the primate immune system.* Science, 2000. **290**(5494): p. 1168-1170.
25. Grenfell, B.T., et al., *Unifying the epidemiological and evolutionary dynamics of pathogens.* Science, 2004. **303**: p. 327-332.
26. Cars, O., S. Mölstad, and A. Melander, *Variation in antibiotic use in the European Union.* The Lancet, 2001. **357**(9271): p. 1851-1853.
27. Kawashima, Y., et al., *Adaptation of HIV-1 to human leukocyte antigen class I.* Nature, 2009. **458**(7238): p. 641-645.
28. Cobey, S. and M. Lipsitch, *Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes.* Science, 2012. **335**(6074): p. 1376-1380.
29. Klein, E., D.L. Smith, and R. Laxminarayan, *Community-associated methicillin-resistant Staphylococcus aureus in outpatients, United States, 1999-2006.* Emerging infectious diseases, 2009. **15**(12).
30. Lythgoe, K.A., L. Pellis, and C. Fraser, *Is HIV short-sighted? Insights from a multistrain nested model.* Evolution, 2013. **67**(10): p. 2769-2782.
31. Gandon, S., et al., *Imperfect vaccination: some epidemiological and evolutionary consequences.* Proceedings of the Royal Society of London. Series B: Biological Sciences, 2003. **270**(1520): p. 1129-1136.
32. Bottomley, C., et al., *A mathematical model of serotype replacement in pneumococcal carriage following vaccination.* Journal of The Royal Society Interface, 2013. **10**(89): p. 20130786.
33. Kouyos, R., E. Klein, and B. Grenfell, *Hospital-community interactions foster coexistence between methicillin-resistant strains of staphylococcus aureus.* PLoS pathogens, 2013. **9**(2): p. e1003134.
34. Koelle, K., et al., *Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans.* Science, 2006. **314**(5807): p. 1898-1903.

35. Pearson, J.E., P. Krapivsky, and A.S. Perelson, *Stochastic theory of early viral infection: continuous versus burst production of virions.* PLoS computational biology, 2011. **7**(2): p. e1001058.
36. Loverdo, C., et al., *Influence of viral replication mechanisms on within-host evolutionary dynamics.* Evolution, 2012. **66**(11): p. 3462-3471.
37. Loverdo, C. and J.O. Lloyd‐Smith, *Intergenerational phenotypic mixing in viral evolution.* Evolution, 2013. **67**(6): p. 1815-1822.
38. Sardanyés, J., et al., *Dynamics of alternative modes of RNA replication for positive-sense RNA viruses.* Journal of The Royal Society Interface, 2012. **9**(69): p. 768-776.
39. Thébaud, G., et al., *The relationship between mutation frequency and replication strategy in positive-sense single-stranded RNA viruses.* Proceedings of the Royal Society B: Biological Sciences, 2010. **277**(1682): p. 809-817.
40. Levin, B.R., V. Perrot, and N. Walker, *Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria.* Genetics, 2000. **154**(3): p. 985-997.
41. Kirkpatrick, M. and S. Peischl, *Evolutionary rescue by beneficial mutations in environments that change in space and time.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2013. **368**(1610): p. 20120082.
42. Lloyd-Smith, J.O., et al., *Ten challenges in modelling the emergence of novel pathogens.* Epidemics, In prep.
43. Metcalf, C.J.E., et al., *Seven challenges in mdoeling vaccine preventable diseases.* Epidemics, in prep.