PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Evolutionary epidemiology: preparing for an age of genomic plenty

O. G. Pybus[1], C. Fraser[2] and A. Rambaut[3]

[1]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK
[2]Medical Research Council Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, St Mary's Campus, London W2 1PG, UK
[3]Institute of Evolutionary Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JT, UK

Introduction

Cite this article: Pybus OG, Fraser C, Rambaut A. 2013 Evolutionary epidemiology: preparing for an age of genomic plenty. Phil Trans R Soc B 368: 20120193. http://dx.doi.org/10.1098/rstb.2012.0193

One contribution of 18 to a Discussion Meeting Issue 'Next-generation molecular and evolutionary epidemiology of infectious disease'.

Subject Areas:
evolution, health and disease and epidemiology, computational biology, genomics

Keywords:
epidemiology, molecular evolution, genomics

Authors for correspondence:
O. G. Pybus
e-mail: oliver.pybus@zoo.ox.ac.uk
C. Fraser
e-mail: c.fraser@imperial.ac.uk
A. Rambaut
e-mail: a.rambaut@ed.ac.uk

The fields of infectious disease epidemiology and molecular evolution have a surprising amount in common. At the most fundamental level they aim to describe and explain basic biological processes of transmission and loss, of pathogens and parasites in one case, and of genetic polymorphisms in the other. Both disciplines are rigorously quantitative and are underpinned by a mature framework of dynamical mathematical models. These frameworks were derived logically from first principles and survived mostly intact as empirical data of sufficient accuracy to examine them became available: a situation far more common in the physical sciences than in biology. Furthermore, stochastic models are common in mathematical epidemiology and molecular evolution, and progress in both fields has been accelerated in recent decades by the rapid and sustained growth in computer processing power and concomitant advances in methods of statistical inference.

These similarities may explain, at least in part, why research at the interface of evolution and epidemiology is flourishing so strongly, demonstrated by the diverse and vibrant research published here. The contributors to this volume are drawn from speakers at the Royal Society Discussion Meeting of 14–15 May 2012, titled 'Next-generation molecular and evolutionary epidemiology of infectious disease' and from speakers at the subsequent satellite meeting on the same topic held at the Kavli Royal Society International Centre. Much of the work presented at the meetings could be said to be 'strongly' inter-disciplinary, in that it aimed to identify or use common frames of reference by which concepts and models from both fields can be quantitatively melded. This is distinct from the more commonly encountered 'weak' inter-disciplinarity, by which a shared problem is investigated from multiple perspectives using methods that are not, or cannot, be integrated into a formal mathematical and inferential framework.

The link between the epidemiology and evolution of infectious disease agents runs deep, especially for those pathogens that evolve rapidly. Since the evolutionary and ecological dynamics of rapidly evolving pathogens occur on approximately the same time-scale, they must be studied conjointly to be properly understood. During the course of a single outbreak or epidemic season, mutations are generated de novo and can spread through an infectious disease population, creating a reciprocal link between the polymorphisms carried by a particular pathogenic organism and its propensity for onward transmission in a heterogeneous host population. The term 'phylodynamic' is often used to describe studies that aim to characterize the joint evolutionary and epidemiology behaviour of infectious diseases, particularly those that incorporate tools from the field of phylogenetics [1]. The term originated from—and is most commonly associated with—studies of RNA viruses, whose mutation rates can be more than one million times faster than their metazoan hosts [2]. However, in the last few years, it has become increasingly clear that the same perspective can be applied readily to other groups of pathogens with lower rates of mutation, including DNA viruses [3,4] and some bacteria [5,6].

We deliberately omitted 'phylodynamic' from the title of the Discussion Meeting for two reasons: first, because we wanted to highlight and explore a

wide range of infectious organisms, not just RNA viruses; and second, because not all important research at the interface of evolution and epidemiology uses phylogenetic methods. Despite this, many of the contributions to this issue do concern RNA viruses, which partly reflects the research interests of the organizers but also the large amount of research directed towards the RNA viruses that represent many of the most important infectious diseases of humans and livestock.

The idea for this Discussion Meeting grew directly from a collaboration among the three organizers that formed during the initial discovery and emergence of pandemic H1N1 'swine-origin' human influenza A in spring 2009. This pandemic was the first major outbreak of the post-genomic era, during which large numbers of whole viral genomes were generated and shared online in real time as the pandemic unfolded. Although viral genetic information was also generated during the epidemic of severe acute respiratory syndrome in 2003, it was done so on a smaller scale and not routinely placed in the public domain, and consequently had much less epidemiological impact. The plentiful and timely publication of influenza genomes during the 2009 H1N1 pandemic meant that traditional surveillance epidemiology could be undertaken concurrently with evolutionary analyses, providing two independent sources of information about the virus' dynamical behaviour. The results of this collaboration were included in the WHO Rapid Pandemic Assessment Report [7], which notably included side-by-side estimates of the basic reproductive number R_0 of the outbreak obtained from multiple epidemiological sources and from coalescent analysis of viral genomes. The ability to cross-validate estimates of key parameters using fundamentally different methods and data greatly increased confidence in those estimates, as individually each is associated with substantial uncertainty.

The 2009 influenza A pandemic proved that the joint evolutionary and epidemiological investigation of rapidly evolving pathogens was both feasible and can provide useful and timely information for public health and epidemic control decisions. As such it marked the end of phylodynamic's infancy and prompted us to organize the Royal Society meetings to explore what the 'next generation' of evo-epidemiological approaches will look like and what they might be able to achieve. The phrase 'next-generation' of course also alludes to the astonishing impact of high-throughput sequencing technologies on infectious disease research. It would not be unrealistic to predict that within a decade the whole genomes of all cases of an emerging pathogen could be sequenced during the course of an outbreak.

However, the greatest obstacles to realizing the full public health potential of pathogen genomics will be perhaps social, not technological. For example, how can considerations of patient privacy be balanced against the potential public benefit of open publication of pathogen genomes and their crucially valuable metadata, such as sampling location? The ethical, political and legal implications of data from which it may be possible to infer routes of transmission at a very granular level (albeit only probabilistically) deserve careful consideration. The scientific and public health value of such data is without doubt, so every effort must be made to report and share it in a manner that will still allow for robust and creative analysis. It is also increasingly clear that human pathogens have a truly global reach and cannot be fully understood by studies limited in spatial or temporal range. The benefit to the wider community of timely and open access to globally aggregated genomic information on pathogens cannot be underestimated.

The contributions to this issue span a wide range of questions and approaches, with several themes recurring. A common goal is the development of theoretical models that formally integrate evolutionary and ecological processes in order to explain the distribution of pathogen genetic variation through time and space [8,9], to enable the estimation of epidemiological properties from genetic data [10–12], and to help predict the emergence of new epidemics [13,14]. Another common strand is the development of quantitative methods that can combine, either loosely or tightly, heterogeneous sources of data. These combinations include spatial and genetic incidence data [15], viral protein structures and their nucleotide sequences [16,17], genetic diversity and immunological assay data [18], and information for disease mapping gleaned from online social networks [19]. The challenges and opportunities arising from the development of high-throughput sequencing technologies are explored for bacteria [20] and for viruses [21]. Other contributions provide new insights into the interaction of evolutionary processes at different biological scales, particularly those within and between infected hosts [14,22], into the dynamics of pathogen adaptation to new host species [17,23] and into the evolutionary trajectory of a newly emerged human pandemic of A/H5N1 influenza [24].

This issue points to a vigorous future for the field and show how datasets that combine genetic, spatial, immunological and social information can transform our understanding of epidemic dynamics. While the notion of a 'data-deluge' has quickly become both a platitude and a truism, its impact in epidemiology has too often been restricted to high-resolution yet essentially narrative descriptions of individual epidemic or evolutionary histories. The work highlighted here demonstrates that theory and computation can—and should—form the organizing principles for next-generation evolutionary epidemiology.

References

1. Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC. 2004 Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**, 327–332. (doi:10.1126/science.1090727)
2. Pybus OG, Rambaut A. 2009 Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10**, 540–550. (doi:10.1038/nrg2583)
3. Firth C, Kitchen A, Shapiro B, Suchard MA, Holmes EC, Rambaut A. 2010 Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol. Biol. Evol.* **27**, 2038–2051. (doi:10.1093/molbev/msq088)
4. Kerr PJ, Ghedin E, Depasse JV, Fitch A, Cattadori IM, Hudson PJ, Tscharke DC, Read AF, Holmes EC. 2012 Evolutionary history and attenuation of myxoma virus on two continents. *PLoS Pathog* **8**, e1002950. (doi:10.1371/journal.ppat.1002950)
5. Harris SR *et al*. 2010 Evolution of MRSA during hospital transmission and intercontinental spread.

Science **327**, 469–474. (doi:10.1126/science.1182395)

6. Croucher NJ *et al*. 2011 Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434. (doi:10.1126/science.1198545)
7. Fraser C *et al*. 2009 Pandemic potential of a novel strain of influenza A (H1N1): early findings. *Science* **324**, 1557–1561. (doi:10.1126/science.1176062)
8. Duke-Sylvester SM, Biek R, Real LA. 2013 Molecular evolutionary signatures reveal the role of host ecological dynamics in viral disease emergence and spread. *Phil. Trans. R. Soc. B* **368**, 20120194. (doi:10.1098/rstb.2012.0194)
9. Viboud C, Nelson MI, Tan Y, Holmes EC. 2013 Contrasting the epidemiological and evolutionary dynamics of influenza spatial transmission. *Phil. Trans. R. Soc. B* **368**, 20120199. (doi:10.1098/rstb.2012.0199)
10. Dearlove B, Wilson DJ. 2013 Coalescent inference for infectious disease: meta-analysis of hepatitis C. *Phil. Trans. R. Soc. B* **368**, 20120314. (doi:10.1098/rstb.2012.0314)
11. Stadler T, Bonhoeffer S. 2013 Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Phil. Trans. R. Soc. B* **368**, 20120198. (doi:10.1098/rstb.2012.0198)
12. Frost SDW, Volz EM. 2013 Modelling tree shape and structure in viral phylodynamics. *Phil. Trans. R. Soc. B* **368**, 20120208. (doi:10.1098/rstb.2012.0208)
13. Yuan H-Y, Koelle K. 2013 The evolutionary dynamics of receptor binding avidity in influenza A: a mathematical model for a new antigenic drift hypothesis. *Phil. Trans. R. Soc. B* **368**, 20120204. (doi:10.1098/rstb.2012.0204)
14. Park M, Loverdo C, Schreiber SJ, Lloyd-Smith JO. 2013 Multiple scales of selection influence the evolutionary emergence of novel pathogens. *Phil. Trans. R. Soc. B* **368**, 20120333. (doi:10.1098/rstb.2012.0333)
15. Cybis GB, Sinsheimer JS, Lemey P, Suchard MA. 2013 Graph hierarchies for phylogeography. *Phil. Trans. R. Soc. B* **368**, 20120206. (doi:10.1098/rstb.2012.0206)
16. Meyer A, Dawson E, Wilke C. 2013 Cross-species comparison of site-specific evolutionary-rate variation in influenza hemagglutinin. *Phil. Trans. R. Soc. B* **368**, 20120334. (doi:10.1098/rstb.2012.0334)
17. Bhatt S *et al*. 2013 The evolutionary dynamics of influenza A virus adaptation to mammalian hosts. *Phil. Trans. R. Soc. B* **368**, 20120382. (doi:10.1098/rstb.2012.0382)
18. Wikramaratna PS, Sandeman M, Recker M, Gupta S. 2013 The antigenic evolution of influenza: drift or thrift? *Phil. Trans. R. Soc. B* **368**, 20120200. (doi:10.1098/rstb.2012.0200)
19. Hay S *et al*. 2013 Global mapping of infectious disease. *Phil. Trans. R. Soc. B* **368**, 20120250. (doi:10.1098/rstb.2012.0250)
20. Croucher NJ, Harris SR, Grad YH, Hanage WP. 2013 Bacterial genomes in epidemiology present and future. *Phil. Trans. R. Soc. B* **368**, 20120202. (doi:10.1098/rstb.2012.0202)
21. Watson SJ, Welkers MRA, Depledge DP, Coulter E, Breuer JM, de Jong MD, Kellam P. 2013 Viral population analysis and minority variant detection using short read next-generation sequencing. *Phil. Trans. R. Soc. B* **368**, 20120205. (doi:10.1098/rstb.2012.0205)
22. Orton RJ *et al*. 2013 Observing micro-evolutionary processes of viral populations at multiple scales. *Phil. Trans. R. Soc. B* **368**, 20120203. (doi:10.1098/rstb.2012.0203)
23. Faria NR, Suchard MA, Rambaut A, Streicker DG, Lemey P. 2013 Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints. *Phil. Trans. R. Soc. B* **368**, 20120196. (doi:10.1098/rstb.2012.0196)
24. Boni MF, Nguyen TD, de Jong MD, van Doorn HR. 2013 Virulence attenuation during an influenza A/H5N1 pandemic. *Phil. Trans. R. Soc. B* **368**, 20120207. (doi:10.1098/rstb.2012.0207)