

Institution: University of Oxford
Unit of Assessment: UOA5
Title of case study: <p style="text-align: center;"><i>BEAST</i> and phylogenetic inference in viral disease epidemiology</p>
<p>1. Summary of the impact</p> <p>Research at the University of Oxford into molecular evolution led to the development of <i>BEAST</i>, a powerful suite of computer programs for evolutionary analysis. Viral genome sequences from infected populations can be analysed to infer both viral population history and epidemiological parameters. This approach has been used to track and predict the transmission and evolution of pathogens, particularly viral infections of humans such as influenza and HIV. <i>BEAST</i> was used alongside traditional epidemiological methods by the World Health Organization to rapidly assess and identify the origins of the 2009 H1N1 'Swine Flu' pandemic; immediate recommendations for necessary international action followed. This approach is now widely adopted by health protection agencies and health ministries around the world and is being applied to understand viral diseases of both humans and animals.</p>
<p>2. Underpinning research</p> <p>Phylogenetic trees are used to represent the evolutionary relationships among organisms, based upon similarities and differences in their physical and/or genetic characteristics. By the early 1990s, phylogenetic trees also began to play a role in molecular epidemiology, where they were used to understand the forces that shape patterns of viral genetic diversity.</p> <p>In 1995 Professor Eddie Holmes and colleagues in the Department of Zoology at the University of Oxford showed for the first time that phylogenetic trees could be used to do more than simply map the genetic relationships and evolutionary history of viruses. They could also trace the dynamics of viral transmission within populations, and show whether transmission rates were constant, declining or increasing. Prior to this, the transmission rates were identified using conventional epidemiological techniques. Using the fact that different rates of viral population growth leave different genetic 'signatures', this new research used computer and graphical analyses to investigate the growth of HIV-1 and HCV (hepatitis C virus) populations. In particular, the analyses suggested that HCV had coexisted with human populations far longer than HIV-1 and had undergone an explosion in transmission within the previous 50 years, probably marking a transition from endemic to epidemic state¹.</p> <p>Since 1995, the Oxford Virus Evolution Group in the Department of Zoology has made a unique contribution to developing and applying methods of phylogenetic analysis, fuelled by the exponential increase both in the availability of pathogen genome sequence data and in computing capability. In 2000, Professor Oliver Pybus and colleagues developed a more formal and mathematically rigorous approach to interpreting genome information by introducing the 'skyline plot', a non-parametric estimate of demographic history (in essence, a plot of how many infections there are through time)². The approach quantified a wide range of epidemic scenarios, and using this technique it became possible to see, for example, that populations of some HIV strains increased more rapidly than others (including prior to the date of the virus' discovery). As a result the work began to attract serious interest from epidemiologists.</p> <p>Crucially, a 2001 paper showed that it was possible to estimate R_0, a key parameter used by epidemiologists to characterise the transmissibility of a virus or disease, from genome data. This provided a completely independent source of information to estimate R_0 that was distinct from the traditional but resource-intensive methods of tracing cases in the field. Using these new methods,</p>

the research found significant differences in epidemic behaviour among HCV subtypes, and suggested that these were largely the result of subtype-specific transmission routes. The methods were especially suitable for rapidly evolving viruses that do not induce lifelong immunity, since the R_0 values of such viruses cannot be estimated from the average age at first infection³.

In 2003, three pieces of software (*GENIE*, *TipDate* and *MEPI*) written individually by Oliver Pybus, Andrew Rambaut, and Alexei Drummond were combined by the Oxford Virus Evolution Group into the single framework of *BEAST* ('Bayesian Evolutionary Analysis by Sampling Trees'), a ground-breaking piece of software which used Bayesian Markov chain Monte Carlo (MCMC) sampling procedures to analyse molecular sequences. *BEAST* was released in June 2003 and is now open-access and in worldwide use; the website supporting this software has been accessed >500,000 times. *BEAST* was used to create the Bayesian Skyline Plot, an improved version of the plot described in² that now included credibility intervals for the estimated effective population size at every point in time. Pybus and colleagues used the new plot to analyse two datasets previously investigated using alternative methods (HCV in Egypt and mitochondrial DNA of Beringian bison). The new method revealed previously undetected demographic signatures, demonstrating its ability to uncover demographic trends over ecological, paleontological and evolutionary time spans⁴. Subsequent applications have been made in a broad range of fields, from molecular anthropology and ancient DNA to conservation genetics and epidemiology.

The first demonstration that the technique could be applied to human influenza was published in 2008 in a collaboration between Pybus at the University of Oxford and researchers at five other universities and laboratories. An analysis of 1,302 complete viral genomes using the *BEAST* software suggested that new influenza lineages were seeded from a persistent influenza reservoir, possibly in the tropics, to sink populations in temperate regions⁵.

3. References to the research

1. Holmes EC, Nee S, Rambaut A, Garnett G, Harvey PH. (1995) Revealing the history of infectious disease epidemics through phylogenetic trees. *Phil Trans R Soc Lond B* 349: 33-40. Available from: <http://www.jstor.org/stable/56121> **First paper to use phylogenetic trees to investigate rates of virus population growth.**
2. Pybus OG, Rambaut A, Harvey PH. (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155: 1429-1437. Available from: <http://www.genetics.org/content/155/3/1429.full> **Paper introducing the concept of skyline plots to display the demographic information contained in reconstructed genealogies.**
3. Pybus OG, Charleston M, Gupta S, Rambaut A, Holmes EC, Harvey PH. (2001) The epidemic behaviour of the hepatitis C virus. *Science* 292: 2323-2325. doi: 10.1126/science.1058321 **First estimate of R_0 from pathogen gene sequences.**
4. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology & Evolution* 22: 1185-1192. doi: 10.1093/molbev/msi103 **Paper introducing the Bayesian Skyline Plot, a new method for estimating past population dynamics through time from a sample of molecular sequences.**
5. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. (2008) The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453: 615-619. doi: 10.1038/nature06945 **First paper applying the new phylogenetic methods to the evolution of human influenza.**

Funding for research: This research was supported from 1997-2002 by grants totalling ~ £1.3M from the Wellcome Trust and the Royal Society.

4. Details of the impact

The 'phylodynamic' techniques created and developed by the Oxford Virus Evolution Group provided a completely new source of information about the transmission parameters of diseases, independent of traditional epidemiological methods of information-gathering through personal interviews, clinical diagnosis and mathematical analysis. The technique has had a significant impact on the way that current pandemics are assessed and dealt with, particularly in relation to influenza and HIV. In addition the *BEAST* software that originated at Oxford University has become a standard tool worldwide for the study of virus evolution and is increasingly applied to understand viral (and now also bacterial) disease in humans and animals.

In 2009, on the basis of the reputation established through the research described above, Christophe Fraser (the lead author of WHO's Rapid Pandemic Assessment report) invited Oliver Pybus (Oxford) and Andrew Rambaut (Edinburgh; formerly part of the Oxford Virus group) to lead the evolutionary phylogenetic component of WHO's urgent investigation into the arrival of influenza A (H1N1) – 'Swine Flu'. The new technique of determining transmission rates using genome data was used in parallel with established epidemiological methods, and the transmissibility parameter R_0 was estimated by both methods; notably the confidence limits for the results obtained by the two methods overlapped. The team was able to show within a week that the virus had been circulating in humans for months, and within 30 days had produced a comprehensive report about the potential effects of the pandemic⁶. A further analysis used *BEAST* to investigate the origins of the new strain of influenza in more detail (in terms of both geography and timescales)⁷. WHO used these reports to help inform its ongoing recommendations for international precautions and preparations. This was the first time that phylogenetic estimates of R_0 had been derived concurrently with traditional methods, and the fact that WHO used the new approach as a key part of their official response to Swine Flu is a clear indication that they considered it to be as valuable and informative as conventional epidemiology.

Subsequent take-up of the phylodynamic approach to understanding the epidemiology of human and animal viral disease has been wide-reaching⁸⁻¹². The approach was used by the (former) UK Health Protection Agency (UKHPA) as part of their evaluation of the spread of Swine Flu both to, and within, the UK in 2009⁸. Pybus and others were able to map the spread and persistence of the H1N1 virus in the UK and show that multiple independent invasions had taken place; some of these had occurred before the invasion date inferred by traditional epidemiological methods. Since phylogenetics has the ability to distinguish the ancestry of viruses, the study was also able to show that geographically-linked outbreaks did not necessarily share the same origin. Subsequently, UKHPA's successor, Public Health England, has used phylodynamics approaches in a real time assessment of the origins, spread and transmission potential of MERS coronavirus⁸.

BEAST software, and the phylodynamic approach is also used by many non-UK based government agencies investigating the spread of disease in humans. For example:

- (i) The Chinese government's Centre for Disease Control and Prevention has used it to assess and aid in the control of the current HIV epidemic in China. Studies using this approach have enabled a clearer delineation of the origins, timescales, spatial spread and risk population structure of HIV in China, and revealed that the origins of the HIV epidemic are much more complex than previously thought. These studies have thereby informed public health decisions about how the virus can be tackled⁹.
- (ii) The Japanese National Institute of Infectious Disease has used the phylodynamic approach to track the transmission and spread of HIV and influenza in Japan and neighbouring Asian countries¹⁰.
- (iii) The Brazilian Ministry of Health, attracted by the improved resolution the methods offer for immunological surveillance, has used *BEAST* and related analytical approaches to study the spread of a range of viral diseases including dengue fever, oropouche fever and rabies¹¹.

Finally, the analytical methods developed at Oxford and the *BEAST* software have been applied by

the UK's Animal Health and Veterinary Laboratories Agency (AHVLA) to study the epidemiology, at a European-wide scale, of a number of viral diseases of economic importance in animals. These include avian and swine influenza, and most recently, the Schmallenberg virus which is an emerging vector-borne virus infecting a range of livestock species¹².

5. Sources to corroborate the impact

6. Fraser C, et al. (The WHO Rapid Pandemic Assessment Collaboration) (2009) Pandemic potential of a novel strain of influenza A (H1N1): early findings. *Science* 324: 1557-1561. doi: 10.1126/science.1176062 ***Initial report on the potential transmission of the 2009 swine flu virus.***
7. Smith GJD, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CM, Raghwani J, Bhatt S, Peiris JS, Guan Y, Rambaut A. (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459: 1122-1125. doi: 10.1038/nature08182 ***Investigation into the origins of the 2009 H1N1 swine flu virus.***
8. Supporting letter from Director, Public Health England Reference Microbiology Services (held on file). ***Confirming the use of phylodynamic methods implemented in BEAST to inform analyses of 2009 swine flu pandemic in the UK, and 2012-13 outbreak of MERS coronavirus.***
9. Supporting letter from Chief Expert on AIDS, Chinese Centre for Disease Control and Prevention (held on file). ***Confirming the use of BEAST in analyses and control of Chinese HIV epidemic.***
10. Supporting letter from Senior Investigator, AIDS Research Centre, National Institute of Infectious Diseases, Tokyo, Japan (held on file).
11. Supporting letter from Director of the National Reference Laboratory for Arboviruses, Brazilian Ministry of Health, and the Head of the Centre for Technological Innovation, Brazilian Ministry of Health (held on file).
12. Supporting letter from Head of Virology Department, Animal Health and Veterinary Laboratories Agency, UK (held on file).