Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Virology

Recent advances in computational phylodynamics

Guy Baele[1], Simon Dellicour[1,2], Marc A Suchard[3,4,5], Philippe Lemey[1] and Bram Vrancken[1]

Time-stamped, trait-annotated phylogenetic trees built from virus genome data are increasingly used for outbreak investigation and monitoring ongoing epidemics. This routinely involves reconstructing the spatial and demographic processes from large data sets to help unveil the patterns and drivers of virus spread. Such phylodynamic inferences can however become quite time-consuming as the dimensions of the data increase, which has led to a myriad of approaches that aim to tackle this complexity. To elucidate the current state of the art in the field of phylodynamics, we discuss recent developments in Bayesian inference and accompanying software, highlight methods for improving computational efficiency and relevant visualisation tools. As an alternative to fully Bayesian approaches, we touch upon conditional software pipelines that compromise between statistical coherence and turn-around-time, and we highlight the available software packages. Finally, we outline future directions that may facilitate the large-scale tracking of epidemics in near real time.

Addresses
[1] KU Leuven Department of Microbiology and Immunology, Rega Institute, Laboratory of Evolutionary and Computational Virology, Leuven, Belgium
[2] Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Bruxelles, Belgium
[3] Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA
[4] Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, CA, USA
[5] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Corresponding author: Baele, Guy (guy.baele@kuleuven.be)

Current Opinion in Virology 2018, **31**:24–32

This review comes from a themed issue on **Viral evolution**

Edited by **Gustavo Palacios** and **Jens H Kuhn**

For a complete overview see the Issue and the Editorial

Available online 22nd September 2018

https://doi.org/10.1016/j.coviro.2018.08.009

Introduction

Omics technologies the collective technologies used to explore the roles, relationships and behaviour of large families of cellular molecules such as genes and proteins [1] — have gained a strong foothold in virology and have brought about a considerable intensification of computational research. The wealth of genomic data produced by modern sequencing technologies dramatically increases our capacity to gain insights into the origins, evolution and epidemiology of viruses. Technological innovations have stimulated bioinformatics developments in support of efficiently producing high-quality genomes and deep sequencing data from targeted or metagenomic sequencing experiments, with many tools specifically developed for viral applications. Inference techniques that build on phylogenetic concepts to analyse the resulting genomic data in an evolutionary context have enjoyed a rich and longstanding history of development, and for rapidly evolving pathogens, this has evolved into a specific field called 'phylodynamics'. However, various challenges in analysing viral genomic data have emerged, such as the need to rapidly produce good-quality phylogenetic estimates for large virus data sets and the desire to meld different types of data and models in a phylogenetic framework in order to extract integrated information on epidemiological processes.

Pathogen phylodynamics was originally defined as a framework to study how epidemiological, immunological, and evolutionary processes shape pathogen phylogenies [2,3]. Owing to their rapid rate of evolution, epidemiological and immunological processes can impact viral evolutionary dynamics on short time scales, making their genomes sampled over time a rich resource for outbreak investigation. Developments in the field of pathogen phylodynamics now target various questions regarding the evolution and spread of pathogens. While such questions have been addressed for different viruses with a longstanding circulation history in humans (e.g. HIV [4–6], HCV [7,8] and influenza [9–11]), the ability to rapidly generate genomic data in a 'real-time' fashion now opens up such opportunities for viral outbreaks on smaller time-scales [12••,13]. These scenarios not only require rapid generation, but also rapid analysis, of the genomic data in order to inform outbreak responses in a timely fashion. For particular viruses with a longstanding transmission history in humans, large amounts of sequence data have accumulated in continuously growing databases (e.g. http://www.hiv.lanl.gov/ for HIV and https://www.gisaid.org/ for influenza), of which the analysis requires scaleable phylogenetic methodology.

Whether the interest is to reconstruct evolution on long or short time scales, there is always considerable benefit in calibrating the time scale in calendar time units. Therefore, time-stamped or 'heterochronous' data are the

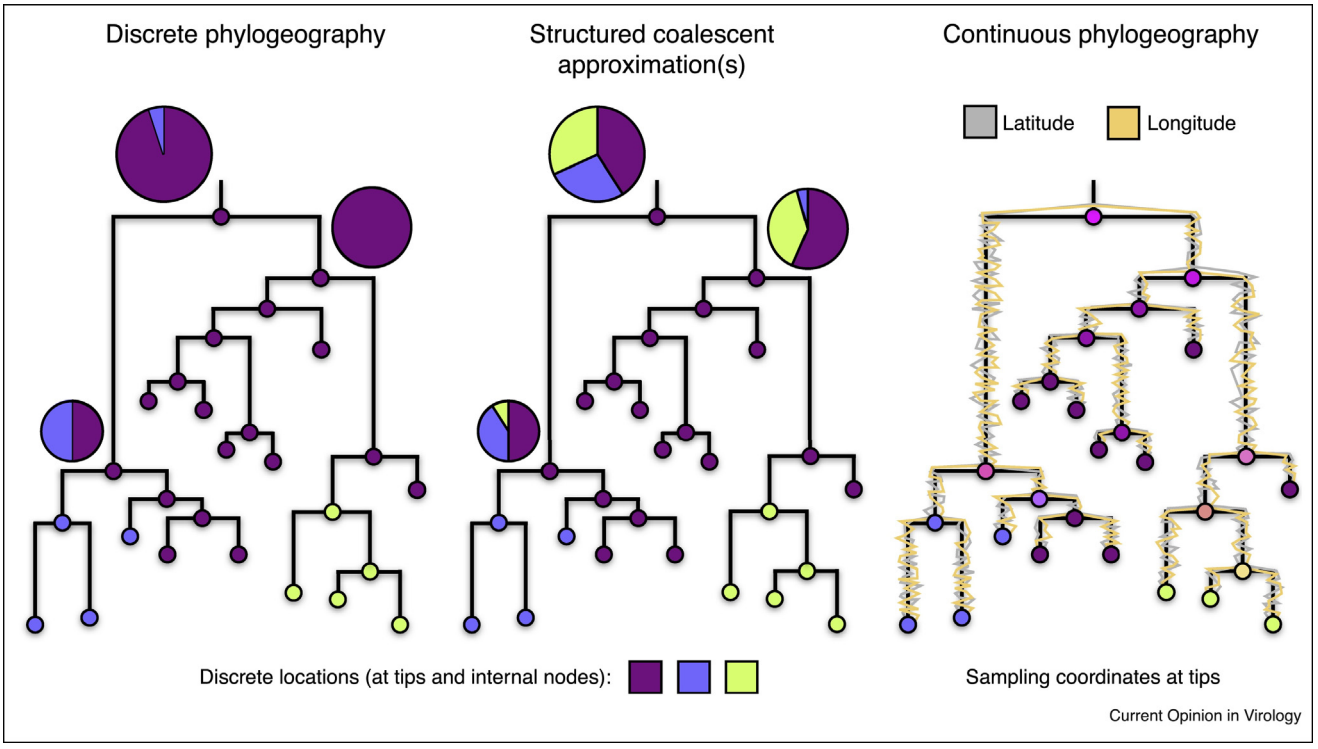
primary ingredients for pathogen phylodynamics because they inform molecular clock models, provide estimates of time-measured trees and date epidemic events [14]. Further, by connecting sampling locations to genetic data, phylodynamic approaches may elucidate the spatial pattern of connectivity underlying viral transmission that may be difficult to assess from surveillance data alone. Phylogeographic models used for this purpose may also find their use in reconstructing the evolution of other viral characteristics, such as host [15] and phenotype [16], and this is rapidly spurring novel developments in Bayesian evolutionary inference methodology. This has resulted for example in the ability to incorporate covariates of different phylodynamic processes, which shifts the focus from reconstructing patterns to testing possible explanatory processes (for an overview, see [17]). These are only few of the novel statistical and computational approaches that have emerged and that allow to extract evolutionary and epidemiological information from pathogen genomes. Here, we review the main inference methods currently in use in the field of pathogen phylodynamics, the software packages in which they are implemented, and tools to visualise the resulting estimates. We first discuss fully Bayesian inference frameworks, which are well known for their statistically coherent handling of sequence and associated trait data, and then consider conditional approaches involving software or pipelines that assume a given topology to obtain divergence time or other phylodynamic estimates.

Bayesian inference of pathogen phylodynamics

Phylogeographic inference

Bayesian phylogeographic inference is a widely used approach to estimate epidemic origins and track geographic spread of pathogens based on discrete or continuous location data [18–20]. The most popular of the discrete phylogeographic approaches (or more generally 'discrete trait analyses'; DTA; Figure 1) consider migration as instantaneous movements (or jumps, [21]) occurring at different frequencies between the set of observed discrete sampling locations. DTA rely on continuous-time Markov chain models, which are also used to model sequence substitution, and can be extended through generalized linear modeling (GLM; [22]) to enable evaluating the relevance and impact of covariates on the dispersal *frequency*. This approach has proven useful in a number of settings, from demonstrating the complex interaction between influenza virus migration and host mobility [22,23] and identifying the drivers of rabies host jumping in American bats [15], to testing the association between virus movements and geographic, administrative

Figure 1



Conceptual representation of the most popular phylogeographic inference approaches. Left: discrete trait analysis for three sampled locations, placing much confidence on particular states on ancestral nodes; middle: structured coalescent approximations showing large uncertainty when reconstructing ancestral nodes; right: continuous phylogeographic inference employs principles from Brownian motion to reconstruct geographic coordinates (that differ from the sampled locations) at ancestral nodes.

and demographic variables during the 2013-2016 West African Ebola virus epidemic [24••].

Despite their popularity, DTA make a number of restrictive assumptions that can be inappropriate when applied to the migration of lineages between geographic locations. DTA potentially under-represent ancestral trait uncertainty and are known to be sensitive to biased sampling of subpopulations [25••] (Figure 1). An alternative approach to discrete phylogeography involves the structured coalescent, an extension of Kingman's coalescent [26] that builds on the standard Wright-Fisher model by specifying a number of discrete subpopulations, allowing location to explicitly affect the coalescent rate. While computational inference under such models remained restricted to only a few populations for a long time, various developments paved the way towards structured coalescent inference with improved statistical properties [27,28]. Recently, De Maio *et al.* [25••] have made inference under the structured coalescent far more efficient when confronted with larger numbers of subpopulations and migration events, although performing inference — including ancestral reconstruction — for high state spaces remains challenging. Their Bayesian structured coalescent approximation (BASTA; available in BEAST 2; [29]) employs a new model-based approach that achieves a close approximation to the structured coalescent, while still integrating over all possible migration histories in a computationally efficient manner. In a study of the validity and robustness of approximations to the structured coalescent, Müller *et al.* [30] present an exact numerical solution to the structured coalescent that — while currently computationally intractable for more than a few lineages and states — led to the development of an improved structured coalescent approximation and an accompanying implementation in BEAST 2 [29]. Müller *et al.* [30] show that, in cases of biased sampling, their method allows for unbiased inferences of migration rates as opposed to previous approximations.

Different from DTA and structured coalescent approaches, continuous diffusion models allow ancestral reconstruction of the spatial history that is not limited to the observed sampling locations (Figure 1). Extensions of the standard Brownian diffusion process that model branch-specific variation in dispersal rates [19] can provide a more realistic description of the spatial diffusion process, which contributed to the popularity of continuous diffusion models (e.g. [31,32]). While the use of such relaxed random walk (RRW) diffusion processes is also very useful for modeling other continuous trait data, they may not appropriately describe the evolutionary process in certain situations. For example, the evolution of influenza antigenic change towards different optima through time or the population evolutionary process towards increased resistance of HIV-1 against neutralization may invalidate the assumption of a zero-mean displacement of Brownian diffusion and its RRW extension. To address this by incorporating directional trends, Gill *et al.* [33] provide a continuous diffusion model that permits permits multiple directional trends on a phylogenetic tree and allows to infer the number of directional trends supported by the data as well as the locations of trend changes.

Complementary to discrete phylogeographic testing, modeling migration as a continuous process allows to test for an impact of external features on the virus dispersal *velocity*. This has recently been achieved by comparing the time elapsed over the branches of the phylogeny with the environmental distances computed between the corresponding edge nodes in a *post hoc* fashion [34••]. The environmental distances are derived from maps that reflect the spatial heterogeneity of the tested feature, and the procedure is repeated over a set of plausible trees to accommodate phylogenetic uncertainty. Similarly, Jacquot *et al.* [35] show how to accommodate variables that lack a spatial component. These approaches have for instance been used to investigate the environmental drivers of rabies virus (RABV) spread in the North-East American raccoon population (Figure 2) [36] and elsewhere [37] and bluetongue virus spread in Europe [35].

Reconstructing viral population dynamics

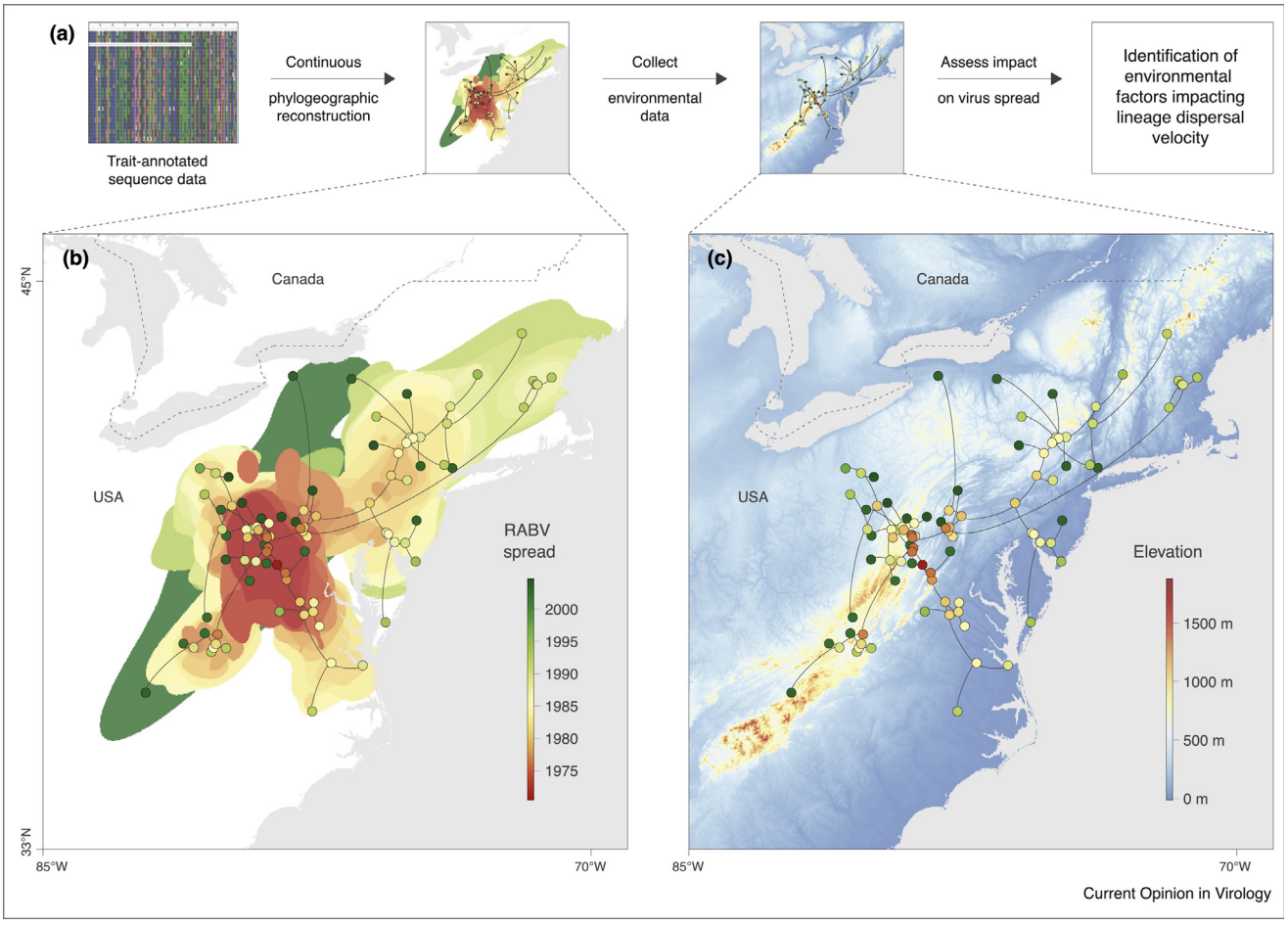
The shape of time-measured phylogenies reflects the dynamic process of pathogen transmission through time. This is the premise of coalescent or birth-death models that are widely used to extract phylodynamic information from viral genetic data in a quantitative manner. A rich history of coalescent development has culminated in flexible models that can accommodate complex dynamics (i.e. the 'sky' models, [39–41]). In several studies, these approaches have allowed interpreting viral spread in a historical setting, pointing at factors that could have contributed to transmission [42,43]. Recent developments seek to more formally examine explanatory variables of viral population dynamics, for example by incorporating covariates into the inference procedure [44,45].

Another promising line of research involves coupling coalescent or birth-death models to complex models of infectious disease dynamics. This builds on earlier work that connects tree generative models to compartmental models (i.e. SI, SIS or SIR models) in order to estimate fundamental quantities like the basic reproductive number (R_0, [46–50]). By adopting novel statistical techniques such as particle filtering (or Sequential Monte Carlo, SMC), prevalence and incidence trajectories [51] or parameters of tailored transmission processes can now be estimated from genetic data [52••].

Software packages

Many phylodynamic inference approaches have been implemented in a Bayesian inference framework,

Figure 2



Workflow to assess the impact of environmental factors on lineage dispersal velocity, with an example of rabies spread in North-East American raccoons. **(a)** Software pipeline to determine environmental impact on lineage dispersal velocity, based on continuous diffusion modeling (either through joint inference or conditional on an existing tree). **(b)** Reconstruction of spatio-temporal diffusion of rabies virus spread in North-East American raccoons, based on a data set of 47 viral sequences [38]: mapped maximum clade credibility (MCC) tree and 95% highest posterior density (HPD) regions obtained from continuous phylogeographic inference; tree nodes and corresponding HPD regions are coloured according to time. **(c)** The MCC tree mapped over environmental data (in this case an elevation map).

allowing to produce estimates with adequately quantified uncertainty for full probabilistic models, which can integrate timed evolutionary histories of sequence evolution under different molecular clock models with trait evolutionary and demographic processes. As indicated earlier, these processes can also be extended to take into account covariates for efficient hypothesis testing. The Bayesian Evolutionary Analysis by Sampling Trees (BEAST) 1 (recently released version 1.10; [53•]) and 2 (currently at version 2.5.0; [29]) software packages have become widely adapted for performing Bayesian phylogenetic and phylodynamic inference from genetic sequence data. Both versions are now developed largely independently, with somewhat different interests such as the modularity and extensibility of the code base through a plug-in package system in BEAST 2 and a focus on data integration in BEAST 1 (we refer to [53•] for more details on the relation of BEAST 1 to BEAST 2).

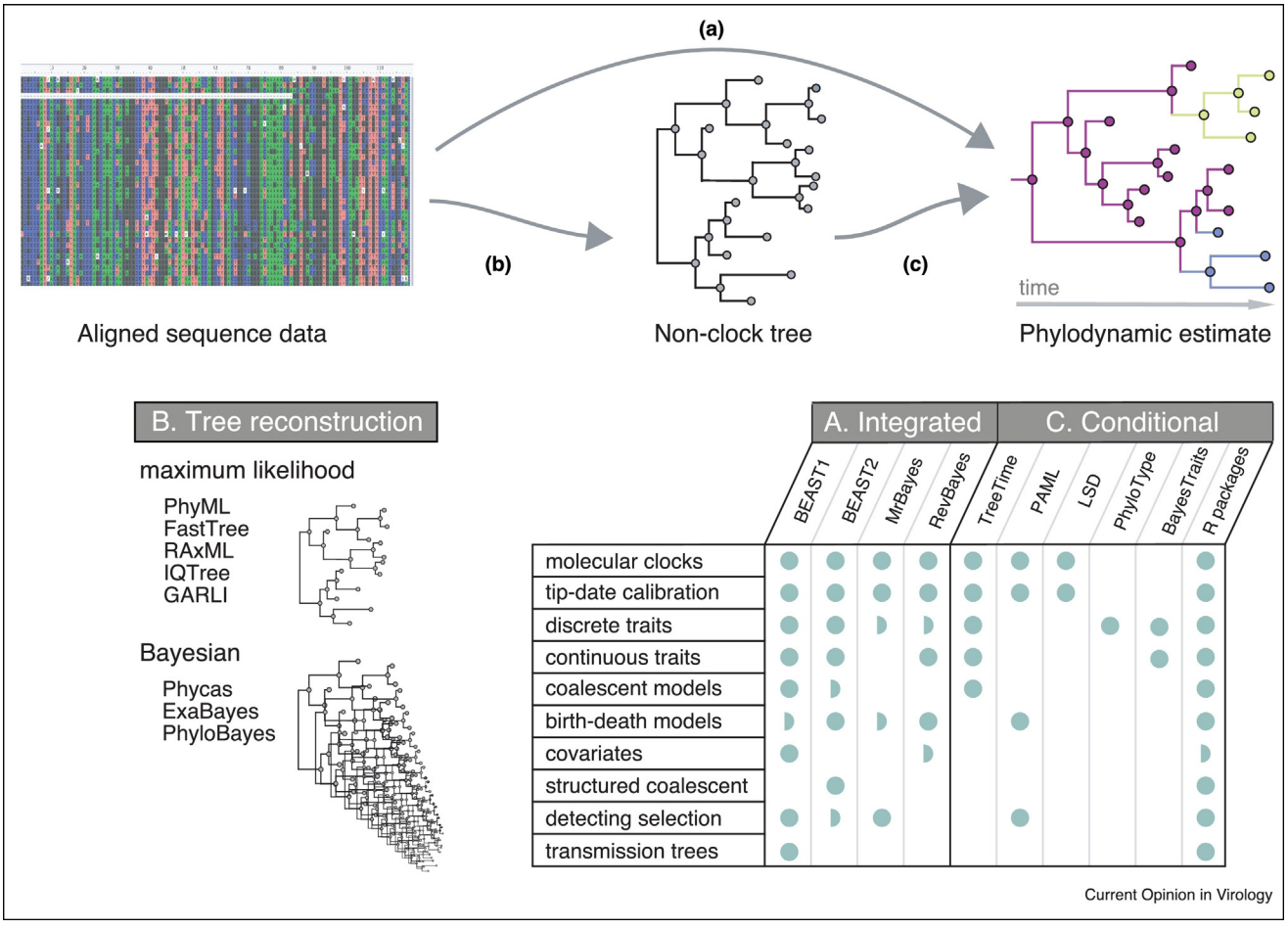
MrBayes [54]) and RevBayes [55] also offer strict and relaxed clock models to estimate divergence times from serially sampled sequences, but present fewer options for reconstructing trait evolutionary processes and coalescent and birth-death models for phylodynamic inference (Figure 3).

Faster methods for large-scale or near real-time phylodynamics

Methodology

Alternatives to integrated Bayesian approaches consist of software pipelines that reduce computation time by combining non-clocklike (fast) maximum-likelihood (ML) or

Figure 3



Schematic overview of phylodynamic estimation procedures and popular software implementations. Integrated approaches **(a)** combine a wide range of models to produce time-stamped phylogenetic trees along with other estimates, such as ancestral trait reconstruction. Conditional approaches first estimate a phylogeny **(b)** that serves as input for software packages that require a fixed phylogenetic tree for phylodynamic inference **(c)**. Maximum likelihood methods produce a point estimate of the topology whereas Bayesian methods estimate a set of evolutionary histories that are all plausible explanations of the sequence relationships. We list different modeling and analysis options for integrated and conditional approaches; half circles indicate limited capacity for the corresponding feature relative to other software packages.

Bayesian tree estimation with divergence time estimation using molecular clocks, ancestral trait reconstruction and estimation of effective population size over time, to name a few (see Figure 3). This entails the initial estimation of one or more unrooted non-clocklike phylogenetic trees from solely the sequence data, which are then treated as known without error in further downstream analyses. By combining bootstrap resampling [56] with ML methods or by taking a Bayesian approach to the tree estimation it is possible to incorporate phylogenetic variability in the *post hoc* analyses. Although quantifying uncertainty is desirable in many hypothesis testing settings as it avoids the information loss and bias associated with the use of a single tree estimate, procedures that produce point estimates may be the only practical way to analyse very large data sets.

Software packages

Popular maximum likelihood and Bayesian software packages geared towards tree inference are listed in Figure 3. A number of these exploit high-performance computational libraries to more efficiently use the available computer hardware (see Box 1). Different conditional approaches take a tree as input to produce a similar estimate, but they may employ different models or estimation procedures. For instance, TreeTime [57], PAML [58] and LSD [59••] can use time-stamped sequences to rescale branch lengths from substitutions per site into calendar time units. Likewise, both PhyloType [60] and BayesTraits [61] may prove useful for reconstructing for example the spatial spread of an epidemic. While these software tools focus on a selected subset of inference options, the modular setup of the *R* software has

Box 1 High-performance computational libraries

The past decades have seen the rise of multi-core processors, both in the CPU and GPU processor markets, enabling massively parallel computations that can be exploited by phylogenetic and phylodynamic software packages. Applications such as IQ-TREE [65] and RaxML/ExaML [66,67] are well known for their use of such architectures to boost performance using a parallel MPI (Message Passing Interface) system. While each of these applications contain their own application-specific software routines to exploit multi-core hardware systems, high-performance computational libraries remove the need for such application-specific code to parallelise computations across multiple processor cores and graphics processing units (GPUs). The BEAGLE (Broad-platform Evolutionary Analysis General Likelihood Evaluator) library is aimed at performing phylogenetic calculations for existing software packages to make more effective use of available multi-core hardware, including GPUs [68]. Since its release, BEAGLE has been adopted by BEAST 1 [53•], BEAST 2 [29], MrBayes [54], PhyML [69] and GARLI [70].

Alternatively, the Phylogenetic Likelihood Library (PLL; [71]) is a highly optimized application programming interface for performing likelihood-based phylogenetic inference, targeting multi-core processors and host processors such as the Intel Xeon Phi using MPI. The PLL has so far been integrated with two phylogenetic software packages: DPPDiv [72] — a Bayesian tool for estimating divergence times on a fixed tree topology — and IQ-TREE [65]. A recent extension of PLL is currently known as LibPLL, a lower-level redesign of PLL that can make use of advanced instruction set extensions to the x86 architecture to improve computational performance [73].

facilitated the development of several dedicated packages capable of (some of) the various post-processing steps indicated in Figure 3. This includes, but is not limited to, *ape* [61], *phytools* [63], *treedater* [64], *rcolgem* [46], *skygrowth* [45] and *phyland* (Frost *et al.*, unpublished). Other useful packages for phylogenetic and phylodynamic inference can be found at https://cran.r-project.org/web/views/Phylogenetics.html and http://www.repidemicsconsortium.org/.

Visualisation tools

Many software tools have been developed to visualise phylogenetic trees, but discussing these is beyond the scope of this review. Instead, we focus on specific applications within the phylodynamics community. Following the initial development of models that accommodate spatial diffusion in discrete and continuous space (e.g. [18,19]), SPREAD (Spatial Phylogenetic Reconstruction of EvolutionAry Dynamics) was developed as a cross-platform application to summarize and visualize Bayesian phylogeographic reconstructions [74] by producing input for Google Earth. Novel web technologies, such as JavaScript and Data-Driven Documents (D3), have enabled the development of browser-based visualization packages. Follow-up work on SPREAD resulted in SpreaD3, which enables modern and interactive web-based visualization of phylogeographic estimates as well as any phylogenetic trait history of interest [75••].

The popularity of browser-based tools is also illustrated by Microreact [76] and Nextstrain [77••]. The former is a web application that allows visualizing data sets consisting of any combination of trees, geographical, temporal and associated (meta)data [76]. Nextstrain aims to visualise epidemics as they unfold in as close to real time as possible by regularly polling public databases for the latest sequences and using fast maximum-likelihood reconstructions [78,66,57]. The framework consists of a database of viral genomes, a bioinformatics pipeline for phylodynamics analysis, and an interactive visualisation platform based on Python and JavaScript [77••]. Using this pipeline, Nextstrain's accompanying website http://www.nextstrain.org/ keeps track of the current state of emerging infectious diseases such as those caused by Zika and dengue viruses.

The computational and methodological advances that made phylodynamic inference possible for increasingly large data sets have also created a need to visualize large phylogenetic trees as well as any annotated information (e.g. virus and patient data). iTol and PhyloGeoTool are particular examples that address this need by supporting interactive navigation of large phylogenies and exploration associated clinical and epidemiological data [79,80].

Future directions

Future phylodynamic model developments will likely see an increasing emphasis on their scalability, as reflected by our discussion of conditional methods and high-performance computational libraries. For computationally demanding approaches, a promising avenue of research involves the development of techniques for updating existing phylogenies/analyses as new data are added. Various software packages allow to perform a likelihood-based insertion of new data on a branch of an unrooted fixed phylogenetic tree, be it sequences (pplacer; [81]) or short reads (EPA, which is integrated into RaxML; [82]), either by resorting to maximum-likelihood or Bayesian estimates or by performing faster approximate methods.

For Bayesian approaches, sequential Monte Carlo (SMC) may provide more formal approach to accommodate novel data into existing analyses in a flexible way [83]. The basic idea is that new sequences are attached to the growing phylogenetic tree following a proposal that relies on genetic similarity. Recently, Dinh *et al.* [84••] published theoretical results on the consistency and stability of such SMC methods for online Bayesian phylogenetic inference, offering important insights for future development. Fourment *et al.* [85] show that poor placement of new sequences within the existing phylogeny can result in poor results, and guided insertion strategies are needed to tackle this problem. Further research into this direction is needed to establish a flexible and easy-to-use online

Bayesian phylodynamic framework that offers clear performance benefits over current inference approaches.

Conflict of interest statement

Nothing declared.

Acknowledgements

GB acknowledges support from the Interne Fondsen KU Leuven/Internal Funds KU Leuven. BV is a post-doctoral research fellow funded by the Fonds Wetenschappelijk Onderzoek (FWO, Belgium). SD is a post-doctoral research fellow funded by the Fonds Wetenschappelijk Onderzoek (FWO, Belgium) and by the Fonds National de la Recherche Scientifique (FNRS, Belgium). PL and MAS acknowledge funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 725422-ReservoirDOCS) and from the Wellcome Trust Collaborative Award, 206298/Z/17/Z. PL acknowledges support by the Special Research Fund, KU Leuven ('Bijzonder Onderzoeksfonds', KU Leuven, OT/14/115), and the Research Foundation — Flanders ('Fonds voor Wetenschappelijk Onderzoek — Vlaanderen', G066215N, G0D5117N and G0B9317N). MAS acknowledges funding from the National Science Foundation under Grant DMS 1264153.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Hasin Y, Seldin M, Lusis A: **Multi-omics approaches to disease**. *Genome Biol* 2017, **18**.

2. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC: **Unifying the epidemiological and evolutionary dynamics of pathogens**. *Science* 2004, **303**:327-332.

3. Volz EM, Koelle K, Bedford T: **Viral phylodynamics**. *PLoS Comput Biol* 2013, **9**:e1002947.

4. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain LV, Liegeois F, Loul S, Ngole EM, Bienvenue Y, Delaporte E, Brookfield JFY, Sharp PM, Shaw GM, Peeters M, Hahn BH: **Chimpanzee reservoirs of pandemic and nonpandemic HIV-1**. *Science* 2006, **313**:523-526.

5. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pépin J, Posada D, Peeters M, Pybus OG, Lemey P: **The early spread and epidemic ignition of HIV-1 in human populations**. *Science* 2014, **346**:56-61.

6. Worobey M, Watts TD, McKay RA, Suchard MA, Granade T, Teuwen DE, Koblin BA, Heneine W, Lemey P, Jaffe HW: **1970s and 'patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America**. *Nature* 2016, **539**:98-101.

7. Iles JC, Raghwani J, Harrison GLA, Pepin J, Djoko CF, Tamoufe U, LeBreton M, Schneider BS, Fair JN, Tshala FM, Kayembe PK, Muyembe JJ, Edidi-Basepeo S, Wolfe ND, Simmonds P, Klenerman P, Pybus OG: **Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa**. *Virology* 2014, **464–465**:233-243.

8. Al-Qahtani AA, Baele G, Khalaf N, Suchard MA, Al-Anazi MR, Abdo AA, Sanai FM, Al-Ashgar HI, Khan MQ, Al-Ahdal MN, Lemey P, Vrancken B: **The epidemic dynamics of hepatitis C virus subtypes 4a and 4d in Saudi Arabia**. *Sci Rep* 2017, **7**:44947.

9. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC: **The genomic and epidemiological dynamics of human influenza A virus**. *Nature* 2008, **453**:615-619.

10. Russell CA, Jones TC, Barr IG, Cox NJ, Garten RJ, Gregory V, Gust ID, Hampson AW, Hay AJ, Hurt AC, de Jong JC, Kelso A, Klimov AI, Kageyama T, Komadina N, Lapedes AS, Lin YP, Mosterin A, Obuchi M, Odagiri T, Osterhaus ADME, Rimmelzwaan GF, Shaw MW, Skepner E, Stohr K, Tashiro M, Fouchier RAM, Smith DJ: **The global circulation of seasonal influenza a (H3N2) viruses**. *Science* 2008, **320**:340-346.

11. Worobey M, Han G-Z, Rambaut A: **A synchronized global sweep of the internal genes of modern avian influenza virus**. *Nature* 2014, **508**:254-257.

12. •• Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, Bore JA, Koundouno R, Dudas G, Mikhail A, Ouédraogo N, Afrough B, Bah A, Baum JH, Becker-Ziaja B, Boettcher J-P, Cabeza-Cabrerizo M, Camino-Sanchez A, Carter LL, Doerrbecker J, Enkirch T, Dorival IGG, Hetzelt N, Hinzmann J, Holm T, Kafetzopoulou LE, Koropogui M, Kosgey A, Kuisma E, Logue CH, Mazzarelli A, Meisel S, Mertens M, Michel J, Ngabo D, Nitzsche K, Pallash E, Patrono LV, Portmann J, Repits JG, Rickett NY, Sachse A, Singethan K, Vitoriano I, Yemanaberhan RL, Zekeng EG, Trina R, Bello A, Sall AA, Faye O, Faye O, Magassouba N, Williams CV, Amburgey V, Winona L, Davis E, Gerlach J, Washington F, Monteil V, Jourdain M, Bererd M, Camara A, Somlare H, Camara A, Gerard M, Bado G, Baillet B, Delaune D, Nebie KY, Diarra A, Savane Y, Pallawo RB: **Real-time, portable genome sequencing for Ebola surveillance**. *Nature* 2016, **530**:228-232.

Seminal work demonstrating the feasibility of near real-time virus genome data generation in resource-challenged settings, a prerequisite for using virus genome data as an information source for guiding control measures in near real-time.

13. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, Burton DR, Lewis-Ximenez LL, de Jesus JG, Giovanetti M, Hill SC, Black A, Bedford T, Carroll MW, Nunes M, Alcantara LC Jr, Sabino EC, Baylis SA, Faria NR, Loose M, Simpson JT, Pybus OG, Andersen KG, Loman NJ: **Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples**. *Nat Protoc* 2017, **12**:1261-1276.

14. Pybus OG, Rambaut A: **Evolutionary analysis of the dynamics of viral infectious disease**. *Nat Rev Genet* 2009, **10**:540-550.

15. Faria NR, Suchard MA, Rambaut A, Streicker DG, Lemey P: **Simultaneously reconstructing viral cross-species transmission history and identifying the underlying constraints**. *Philos Trans R Soc B* 2013, **368**:20120196.

16. Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, Hay AJ, McCauley JW, Russell CA, Smith DJ, Rambaut A: **Integrating influenza antigenic dynamics with molecular evolution**. *eLife* 2014, **3**:e01914.

17. Baele G, Suchard MA, Rambaut A, Lemey P: **Emerging concepts of data integration in pathogen phylodynamics**. *Syst Biol* 2016, **66**:e47-e65.

18. Lemey P, Rambaut A, Drummond AJ, Suchard MA: **Bayesian phylogeography finding its roots**. *PLoS Comput Biol* 2009, **5**:e1000520.

19. Lemey P, Rambaut A, Welch JJ, Suchard MA: **Phylogeography takes a relaxed random walk in continuous space and time**. *Mol Biol Evol* 2010, **27**:1877-1885.

20. Edwards CJ, Suchard MA, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, O'Connell TC, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, Willerslev E, Baryshnikov GF, Rambaut A, Thomas MG, Bradley DG, Shapiro B: **Ancient hybridization and an Irish origin for the modern polar bear matriline**. *Curr Biol* 2011, **21**:1251-1258.

21. Minin VM, Suchard MA: **Fast, accurate and simulation-free stochastic mapping**. *Philos Trans R Soc Lond B Biol Sci* 2008, **363**:3985-3995.

22. Lemey P, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, Russell CA, Smith DJ, Pybus OG, Brockmann D, Suchard MA: **Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2**. *PLoS Pathog* 2014, **10**:e1003932.

23. Nelson MI, Viboud C, Vincent AL, Culhane MR, Detmer SE, Wentworth DE, Rambaut A, Suchard MA, Holmes EC, Lemey P: **Global migration of influenza A viruses in swine**. *Nat Commun* 2015, **6**:6696.

24. •• Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, Park DJ, Ladner JT, Arias A, Asogun D, Bielejec F, Caddy SL, Cotten M, D'Ambrozio J, Dellicour S, Caro AD, Diclaro JD II, Durrafour S, Elmore MJ, Fakoli LS III, Faye O, Gilbert ML, Gevao SM, Gire S, Gladden-Young A, Gnirke A, Goba A, Grant DS, Haagmans BL, Hiscox JA, Jah U, Kargbo B, Kugelman JR, Liu D, Lu J, Malboeuf CM, Mate S, Matthews DA, Matranga CB, Meredith LW, Qu J, Quick J, Pas SD, Phan MVT, Pollakis G, Reusken CB, Sanchez-Lockhart M, Schaffner SF, Schieffelin JS, Sealfon RS, Simon-Loriere E, Smits SL, Stoecker K, Thorne L, Tobin EA, Vandi MA, Watson SJ, West K, Whitmer S, Wiley MR, Winnicki SM, Wohl S, Wölfel R, Yozwiak NL, Andersen KG, Blyden SO, Bolay F, Carroll MW, Dahn B, Diallo B, Formenty P: **Virus genomes reveal factors that spread and sustained the Ebola epidemic**. *Nature* 2017, **544**:309-315.
A detailed phylodynamic analysis providing the 'genetic anatomy' of the West African Ebola virus outbreak.

25. •• De Maio N, Wu C-H, O'Reilly KM, Wilson D: **New routes to phylogeography: a Bayesian structured coalescent approximation**. *PLoS Genet* 2015, **11**:e1005421.
Development and implementation of a structured coalescent approach to discrete phylogeography that allows inferring migration histories that involve more than a few locations.

26. Kingman JFC: **On the genealogy of large populations**. *J Appl Probab* 1982, **19**:27-43.

27. Ewing G, Nicholls G, Rodrigo A: **Using temporally spaced sequences to simultaneously estimate migration rates, mutation rate and population sizes in measurably evolving populations**. *Genetics* 2004, **168**:2407-2420.

28. Vaughan TG, Kühnert D, Popinga A, Welch D, Drummond AJ: **Efficient Bayesian inference under the structured coalescent**. *Bioinformatics* 2014, **30**:2272-2279.

29. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Suchard MA, Rambaut A, Drummond AJ: **BEAST 2: a software platform for Bayesian evolutionary analysis**. *PLoS Comput Biol* 2014, **10**:e1003537.

30. Müller NF, Rasmussen DA, Stadler T: **The structured coalescent and its approximations**. *Mol Biol Evol* 2017, **34**:2970-2981.

31. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, Gray RR, Arinaminpathy N, Stramer SL, Busch MP, Delwart EL: **Unifying the spatial epidemiology and molecular evolution of emerging epidemics**. *Proc Natl Acad Sci USA* 2012, **109**:15066-15071.

32. Trov ao NS, Suchard MA, Baele G, Gilbert M, Lemey P: **Bayesian inference reveals host-specific contributions to the epidemic expansion of influenza A H5N1**. *Mol Biol Evol* 2015, **32**:3264-3275.

33. Gill MS, Tung Ho LS, Baele G, Lemey P, Suchard MA: **A relaxed directional random walk model for phylogenetic trait evolution**. *Syst Biol* 2017, **66**:299-319.

34. •• Dellicour S, Rose R, Faria NR, Lemey P, Pybus OG: **SERAPHIM: studying environmental rasters and phylogenetically informed movements**. *Bioinformatics* 2016, **32**:3204-3206.
Comprehensive analytical framework for assessing the impact of environmental factors on virus dispersal.

35. Jacquot M, Nomikou K, Palmarini M, Mertens P, Biek R: **Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference**. *Philos R Soc B Biol Sci* 2017, **284**.

36. Dellicour S, Rose R, Pybus OG: **Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data**. *BMC Bioinform* 2016, **17**:1-12.

37. Dellicour S, Rose R, Faria NR, Vieira LFP, Bourhy H, Gilbert M, Lemey P, Pybus OG: **Using viral gene sequences to compare and explain the heterogeneous spatial dynamics of virus epidemics**. *Mol Biol Evol* 2017, **34**:2563-2571.

38. Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA: **A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus**. *Proc Natl Acad Sci USA* 2007, **104**:7993-7998.

39. Drummond AJ, Rambaut A, Shapiro B, Pybus OG: **Bayesian coalescent inference of past population dynamics from molecular sequences**. *Mol Biol Evol* 2005, **22**:1185-1192.

40. Minin VM, Bloomquist EW, Suchard MA: **Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics**. *Mol Biol Evol* 2008, **25**:1459-1471.

41. Gill MS, Lemey P, Faria NR, Rambaut A, Shapiro B, Suchard MA: **Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci**. *Mol Biol Evol* 2013, **30**:713-724.

42. Pybus OG, Drummond AJ, Nakano T, Robertson BH, Rambaut A: **The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach**. *Mol Biol Evol* 2003, **20**:381-387.

43. Moratorio G, Costa-Mattioli M, Piovani R, Romero H, Musto H, Cristina J: **Bayesian coalescent inference of hepatitis A virus populations: evolutionary rates and patterns**. *J Gen Virol* 2007, **88(Pt 11)**:3039-3042.

44. Gill MS, Lemey P, Bennett SN, Biek R, Suchard MA: **Understanding past population dynamics: Bayesian coalescent-based modeling with covariates**. *Syst Biol* 2016, **65**:1041-1056.

45. Volz EM, Didelot X: **Modeling the growth and decline of pathogen effective population size provides insight into epidemic dynamics and drivers of antimicrobial resistance**. *Syst Biol* 2018, **67**:719-728 http://dx.doi.org/10.1093/sysbio/syy007.

46. Volz EM, Pond SLK, Ward MJ, Brown AJL, Frost SDW: **Phylodynamics of infectious disease epidemics**. *Genetics* 2009, **183**:1421-1430.

47. Volz EM: **Complex population dynamics and the coalescent under neutrality**. *Genetics* 2012, **190**:187-201.

48. Dearlove B, Wilson DJ: **Coalescent inference for infectious disease: meta-analysis of hepatitis C**. *Philos Trans R Soc Lond B Biol Sci* 2013, **368**:20120314.

49. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ: **Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)**. *Proc Natl Acad Sci USA* 2013, **110**:228-233.

50. Kühnert D, Stadler T, Vaughan TG, Drummond AJ: **Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth-death SIR model**. *J R Soc Interface* 2014, **11**:20131106.

51. Vaughan TG, Leventhal GE, Rasmussen DA, Drummond AJ, Welch D, Stadler T: **Estimating epidemic incidence and prevalence from genomic data**. *bioRxiv* 2018 https://www.biorxiv.org/content/early/2018/05/03/142570.

52. •• Smith RA, Ionides EL, King AA: **Infectious disease dynamics inferred from genetic data via sequential Monte Carlo**. *Mol Biol Evol* 2017, **34**:2065-2084.
Innovative phylodynamic development that adopts Sequential Monte Carlo algorithms for jointly estimating epidemic process parameters and the evolutionary history.

53. • Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A: **Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10**. *Virus Evol* 2018, **4** http://dx.doi.org/10.1093/ve/vey016.
Latest release of the first Bayesian software that is explicitly geared towards the analysis of tip-dated phylogenies, with this latest version focusing on integration different data sources.

54. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space**. *Syst Biol* 2012, **61**:539-542.

55. Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F: **RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language**. *Syst Biol* 2016, **65**:726-736.

56. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap**. *Evolution* 1985, **39**:783-791.

57. Sagulenko P, Puller V, Neher R: **TreeTime: maximum-likelihood phylodynamic analysis**. *Virus Evol* 2018, **4**:vex042.

58. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood**. *Mol Biol Evol* 2007, **24**:1586-1591.

59. •• To TH, Jung M, Lycett S, Gascuel O: **Fast dating using least-squares criteria and algorithms**. *Syst Biol* 2016, **65**:82-97.
Highly scalable and very fast probabilistic methods for converting unrooted phylogenies to time-measured trees using the sampling dates.

60. Chevenet F, Jung M, Peeters M, de Oliveira T, Gascuel O: **Searching for virus phylotypes**. *Bioinformatics* 2013, **29**:561-570.

61. Barker D, Meade A, Pagel M: **Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes**. *Bioinformatics* 2007, **23**:14-20.

63. Revell LJ: **phytools: an R package for phylogenetic comparative biology (and other things)**. *Methods Ecol Evol* 2012, **3**:217-223.

64. Volz EM, Frost SDW: **Scalable relaxed clock phylogenetic dating**. *Virus Evol* 2017, **3**:vex025.

65. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ: **IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies**. *Mol Biol Evol* 2015, **32**:268-274.

66. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies**. *Bioinformatics* 2014, **30**:1312-1313.

67. Kozlov AM, Aberer AJ, Stamatakis A: **ExaML version 3: a tool for phylogenomic analyses on supercomputers**. *Bioinformatics* 2015, **31**:2577-2579.

68. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, Huelsenbeck JP, Ronquist F, Swofford DL, Cummings MP, Rambaut A, Suchard MA: **BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics**. *Syst Biol* 2012, **61**:170-173.

69. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0**. *Syst Biol* 2010, **59**:307-321.

70. Zwickl DJ: *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion (Ph.D. thesis)*. The University of Texas at Austin; 2006.

71. Flouri T, Izquierdo-Carrasco F, Darriba D, Aberer AJ, Nguyen L-T, Minh BQ, von Haeseler A, Stamatakis A: **The phylogenetic likelihood library**. *Syst Biol* 2015, **64**:356-362.

72. Heath TA, Holder MT, Huelsenbeck JP: **A dirichlet process prior for estimating lineage-specific substitution rates**. *Mol Biol Evol* 2012, **29**:939-955.

73. Kobert K, Stamatakis A, Flouri T: **Efficient detection of repeating sites to accelerate phylogenetic likelihood calculations**. *Syst Biol* 2017, **66**:205-217.

74. Bielejec F, Rambaut A, Suchard MA, Lemey P: **SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics**. *Bioinformatics* 2011, **27**:2910-2912.

75. •• Bielejec F, Baele G, Vrancken B, Suchard MA, Rambaut A, Lemey P: **Spread3: Interactive visualization of spatiotemporal history and trait evolutionary processes**. *Mol Biol Evol* 2016, **33**:2167-2169.
A versatile tool for analysing and visualising discrete and continuous trait evolutionary histories on phylogenetic trees.

76. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM: **Microreact: visualizing and sharing data for genomic epidemiology and phylogeography**. *Microb Genom* 2016, **2**:e000093.

77. •• Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA: **Nextstrain: real-time tracking of pathogen evolution**. *bioRxiv* 2018 http://dx.doi.org/10.1093/bioinformatics/bty407.
Real-time tracking tool for virus pathogens, with the goal to aid epidemiological understanding and improve outbreak response.

78. Price MN, Dehal PS, Arkin AP: **FastTree 2 – approximately maximum-likelihood trees for large alignments**. *PLoS ONE* 2010, **5**:e9490.

79. Letunic I, Bork P: **Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees**. *Nucleic Acids Res* 2016, **44(W1)**:W242-W245.

80. Libin P, Eynden EV, Incardona F, Nowé A, Bezenchek A, Group ES, Sönnerborg A, Vandamme A-M, Theys K, Baele G: **PhyloGeoTool: interactively exploring large phylogenies in an epidemiological context**. *Bioinformatics* 2017, **33**:3993-3995.

81. Matsen FA, Kodner RB, Armbrust EV: **pplacer: linear time maximum-likelihood and bayesian phylogenetic placement of sequences onto a fixed reference tree**. *BMC Bioinform* 2010, **11**:538.

82. Berger SA, Stamatakis A: **Aligning short reads to reference alignments and trees**. *Bioinformatics* 2011, **27**:2068-2075.

83. Doucet A, de Freitas N, Gordon N (Eds): *Sequential Monte Carlo Methods in Practice*. Springer; 2001.

84. •• Dinh V, Darling AE, Matsen FA: **Online Bayesian phylogenetic inference: theoretical foundations via sequential Monte Carlo**. *Syst Biol* 2018, **67**:503-517.
First steps towards developing an online Bayesian phylogenetic framework.

85. Fourment M, Claywell BC, Dinh V, McCoy C, Matsen FA IV, Darling AE: **Effective online Bayesian phylogenetics via sequential Monte Carlo with guided proposals**. *Syst Biol* 2018, **67**:490-502.