



Estimating Evolutionary Rates and Timescales from Time-Stamped Data

10

Sebastian Duchêne and David A. Duchêne

Abstract

Methods of molecular dating are playing increasingly valuable roles in evolutionary biology. These methods require independent information to calibrate the molecular clock and obtain meaningful estimates of evolutionary rates and times. One source of such information is the age of the molecular samples, such that the data are said to be time-stamped. In this chapter, we present an outline of current practice and the latest advances in methods for molecular dating using time-stamped data. In addition, there is a broad range of approaches for identifying whether time-stamped data contain sufficient information for estimating evolutionary rates and timescales. We describe a fully Bayesian approach for this purpose and illustrate its performance in analyses of sequence data from H1N1 influenza virus and from *Mycobacterium tuberculosis*. The approaches outlined here provide the foundations for the analysis of time-stamped

data in the era of high-throughput sequencing and high-performance computing.

Keywords

Tip-dating · Measurably evolving populations · Tests of temporal structure · Bayesian inference · Microbial evolution · Ancient DNA

10.1 Introduction

Molecular clock models in phylogenetics are widely used for estimating evolutionary rates and timescales. In addition to information about genetic divergence, molecular clocks often use information about the timing of evolutionary events, also known as a time calibration. Such calibrations provide the raw material for estimating absolute evolutionary rates and times from sequence data. A popular source of calibrating information for molecular clock analyses is the timing of sample collection (Rambaut 2000). Data sets that contain samples collected at different points in time are described as time-stamped or heterochronous. In contrast, isochronous data sets contain samples of similar or identical ages and their evolution is most appropriately represented using an ultrametric time-tree (phylogenetic tree with branch lengths in time units and where ultrametricity means that the distance from the root to each of the tips is the

S. Duchêne

Department of Biochemistry and Molecular Biology,
Bio21 Molecular Sciences and Biotechnology Institute,
University of Melbourne, Melbourne, VIC, Australia

D. A. Duchêne (✉)

Centre for Evolutionary Hologenomics, University of
Copenhagen, Copenhagen, Denmark
e-mail: david.duchene@sund.ku.dk

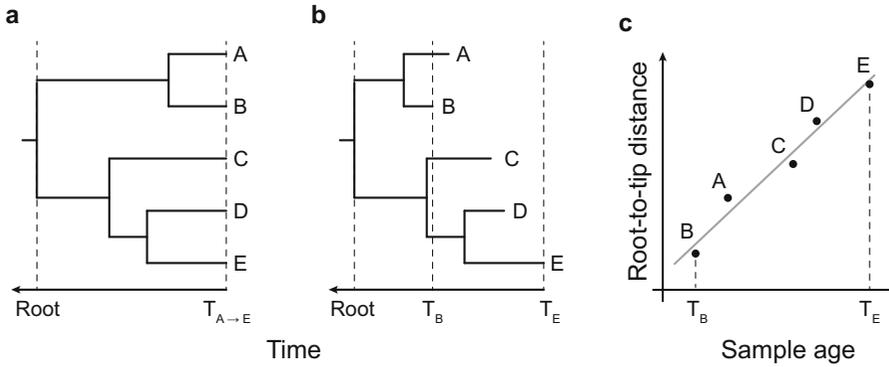


Fig. 10.1 Examples of (a) isochronous (i.e., ultrametric) and (b) heterochronous trees. Data sets with time-stamped sequences are expected to produce heterochronous trees where an appreciable amount of evolutionary change has occurred over the sampling window. The limited sampling window in **a** is insufficient to measure evolutionary change, while that in **b** between T_B and T_E is a candidate for being measurably evolving. (c) The root-to-tip distances plotted as a function of their sampling times.

The evolutionary rate is the slope of the regression line and is intuitively equivalent to the difference in root-to-tip distances between any pair of tips (such as tips A and B) divided by the difference in their sampling times (T_A and T_B). The x -intercept corresponds to the time to the most recent common ancestor and the degree to which the points deviate from the line (R^2) reflects the extent to which the data have departed from strictly clocklike evolution

same). For the sampling times to be useful for calibration of the molecular clock, the period of collection, or sampling window, must have been sufficient for molecular evolution to have left a signature (Fig. 10.1). Samples that have been collected over sufficiently broad periods of time to accumulate evolutionary change are said to come from ‘measurably evolving populations’ (Drummond et al. 2002, 2003).

Many of the principles of phylogenetic inference using molecular clocks in isochronous data also hold for time-stamped data, such as the requirement of using substitution and molecular clock models. The central difference between analyses of these two types of data is methods of fitting a molecular clock and the statistical tests used to confirm that the sampling times span a sufficiently long time, known as assessment of temporal structure (Rieux and Balloux 2016).

Time-stamped data have most frequently been used for the study of evolutionary events involving individuals sampled from a single population or species, as opposed to divergence events among species or higher taxonomic groups. This means that the principles of microevolution and population genetics often play an important role

in analyses of time-stamped data (Arbogast et al. 2002). Combining the methods used in phylogenetics and population genetics largely relies on genealogy-based inference using the principles outlined in coalescent theory (Kingman 1982; Griffiths and Tavaré 1994). By drawing from population genetics theory, analyses of time-stamped data can lead to a range of insights about demography and epidemiology. The power of these approaches is exemplified by the thriving field of phylodynamics (Grenfell et al. 2004).

The growth of efficient, low-cost sequencing has had a substantial impact on the analysis of time-stamped data. Before the major advances in sequencing and computational technologies, studies of pathogen populations using time-stamped data were restricted to RNA viruses (Drummond et al. 2003). This was due to the high rates of evolution of RNA viruses compared with those of other microbes, which allowed for the accumulation of sufficient molecular change over calendar time, even for short genomic regions. Nowadays, high-throughput sequencing allows the extraction of vast amounts of data, including complete genomes, from more slowly evolving microbes. Similarly, novel sequencing

technologies have revolutionized the extraction of target DNA from highly degraded samples, making way for whole-genome analysis of ancient DNA from plants and animals (Millar et al. 2008; Der Sarkissian et al. 2015). As a consequence, methods of analysing these data have also seen appreciable progress, for instance by lifting their former restriction to population-level processes (e.g., Stadler et al. 2013; Grealy et al. 2017) or to rapidly evolving microbes (Biek et al. 2015; Menardo et al. 2019).

Time-stamped data have the advantage that sampling times alone can be used to calibrate the molecular clock, often without the need for other forms of calibration using divergence times or rates. Incorporating sample ages as time calibrations is done in the same way as for node calibrations, either by treating each sample age as a fixed time point or by specifying a probability distribution that accounts for age uncertainty (Rieux and Balloux 2016). In this chapter, we discuss the methods of data collection and analysis of two commonly used types of time-stamped data: those coming from populations of pathogens and those sampled from subfossil material as extracted using ancient DNA techniques.

10.2 Measurably Evolving Populations

10.2.1 Microbial Evolution over Calendar Time

Early phylogenetic analyses of RNA viruses revealed that their substitution rates were sufficiently high that the viruses were able to accumulate an appreciable number of substitutions over weeks or months (Holmes 2009). For example, influenza viruses have been found to evolve at rates of up to 10^{-2} substitutions per site per year; with a genome size of around 13.5 kb, they can accumulate several substitutions per day (Duffy and Holmes 2009). Human immunodeficiency virus (HIV) also undergoes very rapid evolutionary change, with a rate of about 10^{-3} substitutions per site per year (Duchêne et al. 2014a), and can accumulate at least one substitution per week

(assuming a genome size of about 9.5 kb). In a seminal study, Korber et al. (2000) took advantage of the rapid evolution of HIV to calibrate the molecular clock to date its origin in human populations, which revealed that some strains of HIV probably originated in the early twentieth century.

Whole-genome sequencing has revolutionized studies of more slowly evolving microbes, notably bacteria. The evolutionary rates of bacteria are much lower than those of viruses, implying that they would need a much wider sampling window than viruses for their evolutionary rates to be estimated reliably. However, bacteria also have much larger genomes than viruses, such that even with lower rates it is sometimes possible to treat them as measurably evolving. As a case in point, estimates of the evolutionary rate of *Salmonella enterica* range from about 10^{-7} to about 10^{-6} substitutions per site per year (Duchêne et al. 2016b), and hence are at least three orders of magnitude lower than those of some RNA viruses. If only a small portion of its ~5.3 Mb genome is sequenced, for example 10 kb, it would take about 10 years to observe a single substitution. In contrast, when the complete genome is sequenced, up to four substitutions might be observed per month and the samples can be treated as measurably evolving (Zhou et al. 2018). As a result, the growing prevalence of whole-genome sequencing means that many bacteria can now be analysed as measurably evolving populations (Biek et al. 2015).

Although whole-genome sequencing has expanded the range of analyses that are possible in microbes, it has also revealed biological patterns that are not correctly modelled by standard techniques. The most notable problem is homologous recombination, which is very common in some bacterial groups (Yahara et al. 2016). The most obvious limitation of phylogenetic analyses of data sets with substantial recombination is that the whole genome cannot be assumed to follow a single phylogenetic tree topology and that estimates of branch lengths will be incorrect (Hedge and Wilson 2014). While some methods explicitly attempt to model recombination events (Didelot and Wilson 2015;

Vaughan et al. 2017), the most popular approach is to remove recombining regions and to conduct phylogenetic analysis on a ‘core genome’ that includes only sites that have been inherited vertically (Croucher et al. 2014). Failing to account for recombination can give the impression of an erratic molecular clock, and removing such regions can improve the extent to which the data can be treated as measurably evolving (Schultz et al. 2016). It is important to note that downstream analyses based on estimates of evolutionary timescales, such as skyline plots (Pybus et al. 2000), can produce biased inferences when recombinant regions are removed (Lapierre et al. 2016). Accordingly, it is preferable to model recombination explicitly, although this is usually computationally intensive.

10.2.2 Ancient DNA for Temporal Calibration

DNA taken from subfossils of plants and animals usually comes from highly degraded material and requires specialized extraction techniques (Gamba et al. 2016). Until recently, ancient DNA was primarily retrieved from mitochondrial genomes, which are more abundant and have a lower rate of degradation than nuclear genomes (Allentoft et al. 2012). The mitochondrial DNA molecule usually has a highly stable circular structure and has additional protection from decay due to the double membrane of the organelle. In most animals, the rate of evolution of mitochondrial DNA is much higher than that of most nuclear DNA. These characteristics make mitochondrial DNA particularly useful for inferring population-level dynamics over short geological timescales (de Bruyn et al. 2011; Ho and Shapiro 2011). Fast-evolving ancient DNA has been instrumental for inferring population-size fluctuations in a great range of taxa, including the woolly mammoth (Palkopoulou et al. 2013), steppe bison (Shapiro et al. 2004), musk ox (Campos et al. 2010), collared lemming (Brace et al. 2012), and hominids (Posth et al. 2017), among many others (e.g., Lorenzen et al. 2011).

The advent of genome-scale sequencing technologies has greatly facilitated the recovery of ancient DNA data. High-throughput sequencing methods can target highly fragmented DNA molecules, which enables vast amounts of nuclear DNA to be retrieved. This has allowed whole-genome sequences to be recovered from ancient remains (Prüfer et al. 2014). Similarly, it is now commonplace to recover sequence data from materials with trace amounts of the target DNA (Grealy et al. 2017). As a result of novel extraction and sequencing technologies, older samples can now be included in genetic studies.

Some ancient tissue samples used for ancient DNA sequencing have known ages, for instance as documented dates of collection, but others are too old for their ages to be known exactly. Therefore, the ages of samples in time-stamped data sets can have a degree of uncertainty that should not be ignored in phylogenetic analysis. One common source of age uncertainty in ancient DNA samples that are less than around 55,000 years old is that arising from radiocarbon dating (Guilderson et al. 2005). An additional complication is that radiocarbon dates are different from absolute ages, due to fluctuations in atmospheric ^{14}C content. To allow the radiocarbon date estimates to be compared with the timing of other events, such as those of climatic changes, radiocarbon ages need to be converted to calendar time. This conversion can be done by using estimates of atmospheric ^{14}C content in the past, which are becoming increasingly accurate (Stuiver and Reimer 1993; Reimer et al. 2013).

The distribution of uncertainty that emerges from radiocarbon dating can be multimodal, so using a point summary such as the mean or median is a poor description of sample ages (Molak et al. 2015). To solve this problem, some phylogenetics software allow the implementation of parametric distributions to account for uncertainty in sample ages (Shapiro et al. 2011). There are also applications that allow the use of nonparametric distributions to model the uncertainty in radiocarbon dates (Molak et al. 2015). Nonetheless, using the point mean or median estimates of sample ages of time-stamped data strikingly often leads to reasonable estimates

of uncertainty in times and rates (Molak et al. 2013).

An ancient sample can also be dated using indirect methods. The age estimate of the archaeological or stratigraphic location of a sample, or ages estimated from nearby samples, can be used for calibration. However, dating based on the boundaries of stratigraphic layers is often associated with much greater uncertainty than direct estimates. Dates estimated using this method can also be highly inaccurate if the deposit has been reburied or mixed.

10.3 Popular Approaches for Molecular Dating Using Time-Stamped Data

Since the early 2000s, a range of methods have been developed for calibrating the molecular clock using sampling times: root-to-tip regression, likelihood or optimality methods, and Bayesian inference. The intuition behind using sampling times for calibration is that the evolutionary rate is approximately the difference in evolutionary distance between a pair of tips divided by the difference in their sampling times. In the phylogenetic tree in Fig. 10.1b, the rate of evolution can be calculated as the difference in the root-to-tip distance between tips B and E divided by their difference in sampling times ($T_E - T_B$). To obtain a time-tree, the branch lengths of the phylogenetic tree (in units of substitutions per site) can be divided by the evolutionary rate estimate (substitutions per site per year). Clearly, the inclusion of a larger number of time-stamped tips gives more opportunities to calculate the evolutionary rate, thereby improving its accuracy. A fundamental consideration with all methods that use time-stamped data is that the estimates depend on the position of the root, which can be selected or estimated in a number of ways.

10.3.1 Root-to-Tip Regression

One of the earliest molecular clock approaches to time-stamped data was implemented by Korber et al. (2000) to infer the age of the most recent common ancestor of HIV pandemic strains. The data consisted of molecular sequences of the *gag* and *env* genes, with the samples collected over about 10 years. Their method consisted of inferring a phylogenetic tree using maximum likelihood and assuming a constant evolutionary rate (i.e., a strict molecular clock). They conducted a regression of the distance from the root of the tree to each of the tips as a function of their sampling times. The expectation is that samples that are collected later should have undergone more molecular evolutionary change than those closer to the root of the tree.

In such root-to-tip regression, the slope of the line corresponds to the evolutionary rate and the x -intercept is the age of the most recent common ancestor (Fig. 10.1c). The optimal position of the root is that which maximizes clocklike behaviour, which is typically quantified with the R^2 of the regression, although a range of regression statistics can be used. Alternatively, the position of the root can be specified by including an outgroup taxon. The root-to-tip regression method is implemented in the program TempEst (Rambaut et al. 2016).

Benefits of the root-to-tip regression include that it only requires a phylogenetic tree with branch lengths in units of evolutionary distance that can be inferred using different methods (distance-based, maximum-likelihood, or Bayesian approaches), it is computationally very efficient, and it gives a measure of the clocklike behaviour of the data. Empirical studies suggest that it can produce estimates of evolutionary rates that are comparable to those of more sophisticated methods (Duchêne et al. 2016a; Tong et al. 2018). However, the root-to-tip regression has some important limitations. Measuring the root-to-tip distance for every tip means that there is

substantial pseudoreplication because the path from the root to each of the tips will go through the internal branches multiple times and it does not report uncertainty in a meaningful way. In turn, using a p -value to determine the significance of the association of evolutionary distance and time is statistically invalid (Rambaut et al. 2016). Although the phylogenetic tree is used to measure evolutionary distance, the branching order is not taken into account in the regression so that this potentially useful information is discarded. Finally, modelling rate variation among lineages is not straightforward. For these reasons, the root-to-tip regression is mostly used for visual inspection of the data, rather than as a rigorous molecular clock method (see Sect. 10.4).

10.3.2 Optimality Methods

Approaches based on optimizing a function to fit a molecular clock fall in the category of *optimality methods* and include those based on maximum likelihood, least squares, and genetic distance. Rambaut (2000) devised a likelihood function where branch lengths in the tree are treated as the product of evolutionary rates and times. Given a phylogenetic tree and sampling times, it is possible to estimate the evolutionary rate that maximizes this likelihood. This can be performed under the assumption that there is a strict molecular clock, or by allowing rates among branches to be governed by a probability distribution (Seo et al. 2002; Volz and Frost 2017; Sagulenko et al. 2018). Nonparametric methods also optimize a likelihood (or penalized likelihood) function to fit a molecular clock with different degrees of rate variation among lineages (Sanderson 2003; Fourment and Holmes 2014; Chap. 12). There exist several software programs to fit molecular clocks to time-stamped data using likelihood, including TreeDater (Volz and Frost 2017), TreeTime (Sagulenko et al. 2018), TipDate (Rambaut 2000), Physher (Fourment and Holmes 2014), and r8s (Sanderson 2003).

In the program LSD, To et al. (2016) implemented a least-squares dating method that is similar in principle to the Langley–Fitch model

(Langley and Fitch 1974), which assumes a strict molecular clock. The new method differs in that errors in evolutionary rates are assumed to follow a Gaussian, rather than a Poisson, distribution. The objective function depends on the evolutionary rate and the branch lengths. The optimization is conducted via weighted least squares, where the weights are the uncertainty of the Gaussian distribution that governs rates (To et al. 2016). This method assumes a strict molecular clock and aims to minimize evolutionary rate variation among lineages. To obtain uncertainty in the estimates of node times and evolutionary rates, LSD conducts a parametric bootstrap of branches. The position of the root can be optimized in the program, or specified using an outgroup or a particular branch. A useful feature of this method is that it is possible to estimate the ages of samples with unknown collection times.

The optimality methods described here are computationally very efficient, which makes them amenable to very large data sets. For example, LSD has been used to infer the evolutionary rate and timescale of over 1000 strains of influenza within a few minutes on a standard laptop (To et al. 2016). Such computational efficiency is due to the fact that these methods require an estimated phylogenetic tree as an input, instead of inferring the tree directly from the sequence data as is the case with most Bayesian methods. The most obvious limitation is that any uncertainty in LSD estimates typically reflects evolutionary rate variation but not uncertainty in the tree topology or branch lengths. However, these sources of uncertainty can be incorporated using indirect methods, such as repeating the analyses on a set of bootstrap trees.

10.3.3 Bayesian Methods

Most Bayesian molecular clock methods naturally incorporate uncertainty in the estimates of the tree topology, branch lengths, and evolutionary rates via the posterior distribution (see Chap. 6). They can also implement sophisticated models to describe complex patterns of evolutionary rate variation and demographic dynamics. It is

of particular relevance to ancient DNA studies that Bayesian methods allow the researcher to assign a prior distribution for the ages of tips, for example to reflect the uncertainty in ^{14}C dating, and their posterior distribution will be estimated as for other parameters (Shapiro et al. 2011). The most widely used programs that incorporate a full Bayesian model include BEAST 1 (Suchard et al. 2018) and BEAST 2 (Bouckaert et al. 2019), MrBayes (Ronquist et al. 2012b), and RevBayes (Höhna et al. 2016).

In its simplest form, the full Bayesian model consists of a time-tree prior (du Plessis and Stadler 2015) to describe the branching process, a molecular clock model to describe the prior on branch rates, and a substitution model. The phylogenetic likelihood of the sequence data given the tree and the substitution model is calculated by treating branch lengths as the product of times (from the time-tree prior) and rates (from the clock model) (Heath and Moore 2014; Bromham et al. 2018). The position of the root of the tree is informed by the tree prior, instead of being optimized independently as in optimality methods and the root-to-tip regression. The range of clock models that can be used is the same as that for isochronous data, but only some of the available tree priors are valid for heterochronous data.

The most common tree priors posit that branching events are described by either a coalescent or a birth–death process (Drummond and Stadler 2015). Coalescent models are backwards-in-time processes that are conditioned on the ages and number of samples. The rate at which lineages coalesce back in time is determined by a mathematical function of population size over time (Rosenberg and Nordborg 2002). For example, an exponential function can be used to estimate the growth rate of a pathogen population based on the temporal distribution of nodes (Volz et al. 2009). An array of flexible skyline-plot methods can also use the coalescent to infer more complex population dynamics using non-parametric and semiparametric approaches (Ho and Shapiro 2011). Because coalescent models do not explicitly describe the sampling

process, they only require a few modifications to make them applicable to heterochronous data (Rodrigo and Felsenstein 1999; Drummond et al. 2002).

Birth–death models are forwards-in-time processes and they have an expectation of the number of samples and of their ages. The simplest model is known as the Yule process and it assumes constant diversification and no extinction, or death, of lineages (Yule 1924). The result of the Yule process is always an isochronous time-tree, so it cannot be used for analyses of heterochronous data. A birth–death process with explicit sampling assumes that lineages can go extinct and can be sampled with some probability (Stadler 2010), and hence can be applied to heterochronous data. A key consideration relating to the birth–death model is that the sampling parameter should reflect the process under which the data were sampled; the constant birth–death assumes constant sampling effort over time and lineages, whereas the birth–death skyline allows the user to specify periods of time with variable sampling effort (Stadler et al. 2013). There also exist multiple birth–death models that allow certain lineages to be sampled with a higher probability (Stadler and Bonhoeffer 2013). Recent studies have suggested that the choice of sampling scheme can have a considerable effect on the birth–death tree prior, producing time priors for internal nodes that are more informative than those under the coalescent (Boskova et al. 2018). As with all Bayesian analyses, it is important to choose a prior distribution that is reasonable for the data at hand.

Some Bayesian methods do not implement a full Bayesian model. Instead of relying on sequence data, they assume an estimate of the phylogenetic tree with branch lengths (Thorne et al. 1998; Yang 2007; Didelot et al. 2018). These approaches are usually more computationally efficient than those that use the full Bayesian model. However, they currently have a limited range of tree priors available and their computational efficiency comes at the expense of ignoring phylogenetic uncertainty.

10.4 Verifying Temporal Structure

Estimating evolutionary rates and times using time-stamped data requires sufficient molecular evolution between sampling times (Duchêne et al. 2015b; Murray et al. 2015). If this requisite is met, the data are said to have temporal structure. If the molecular data have evolved too slowly relative to the timeframe covered by the samples, then they might not have temporal structure and can produce spurious inferences of evolutionary rates and times (Rambaut 2000). Failing a test of temporal structure generally means that either a more rapidly evolving molecular marker must be sampled, or the sampling window must be widened by the inclusion of new samples from times outside the existing window. Below we outline the methods that have been proposed to test whether time-stamped data have temporal structure.

10.4.1 Root-to-Tip Regression

A fast and popular approach to test temporal structure is to employ a root-to-tip regression under the assumption that the data follow a molecular clock, as described above. The test only requires estimation of the root-to-tip distances, which are the summed lengths of branches from the root of the tree to each tip. This method tests for a linear relationship between the molecular substitutions accumulated and the ages of the samples (Fitch et al. 1991; Fig. 10.1c). The slope must be positive because it is a crude estimate of the evolutionary rate, and a high R^2 coefficient of determination indicates clocklike evolution (Korber et al. 2000).

The root-to-tip regression has several known shortcomings. In many time-structured data sets, the samples come from only a small number of time points; this means that the results could be based on only a small number of data points, leading to low statistical power. In addition, many data sets violate the assumption of the molecular clock, such that a poor root-to-tip regression can lead to falsely taking the data as

lacking temporal structure (Firth et al. 2010; Duchêne et al. 2020). More critically, the root-to-tip measurements used in this method are not statistically independent, as explained in Sect. 10.3.1. Nonetheless, root-to-tip regression is extremely fast and it is commonly used as an exploratory diagnostic of the reliability of rate estimates (Duchêne et al. 2016a; Rambaut et al. 2016; Tong et al. 2018).

10.4.2 Date-Randomization Test

A more robust test of temporal structure known as the date-randomization test involves permuting the dates of samples (Ramsden et al. 2008). The goal of permuting the sample ages is to create data sets where the association between sample age and molecular evolution is broken. A large collection of data sets with randomized tips can be taken to represent a null distribution of rate estimates. Temporal structure is said to be lacking if the rate estimates obtained with the correct sampling times resemble those estimated from the date-randomized replicates (Fig. 10.2a).

The date-randomization test can be used to evaluate temporal structure using Bayesian and optimality methods, and two test criteria have been proposed. The first criterion (CR1) assesses whether the mean rate estimated from the empirical data falls within the 95% credible interval of the rate estimates from the date-randomized replicates (Fig. 10.2a). The second criterion (CR2) assesses whether the 95% credible interval of the rate estimates with correct sampling times overlaps with the range of those from the date-randomized replicates (Duffy and Holmes 2009; Ramsden et al. 2009). CR2 provides a more conservative assessment and is recommended, with minimal chances of failing to reject a data set with no temporal structure (Duchêne et al. 2015b). However, this criterion also brings a moderate chance of incorrectly rejecting data sets as lacking structure, equivalent to a high Type II error rate. An implementation of this test in LSD, or any optimality method, is computationally less expensive and it is feasible to conduct a large number of

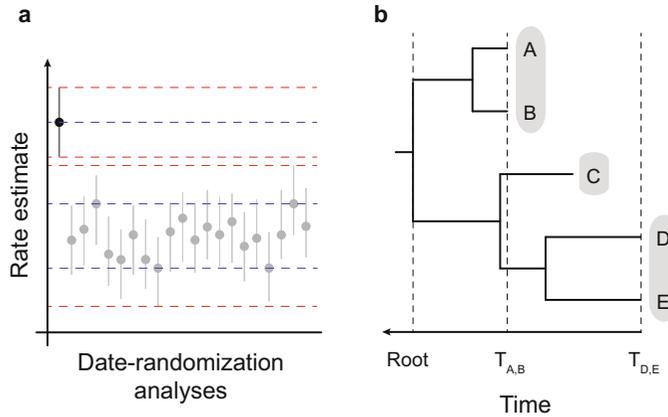


Fig. 10.2 (a) Example of results from a date-randomization test indicating strong temporal structure. Evolutionary rate estimates correspond to the estimate with the correct sampling times (black) and those from 20 date-randomized replicates (grey). Solid circles are the mean rate estimates and the error bars are the 95% credible intervals. The blue dashed lines denote the mean rate values with the correct sampling times and the range in mean rates coming from the randomizations, as used in the CR1 method of testing temporal structure. Similarly, the red dashed lines denote the 95% credible intervals from

the data with correct sampling times and the randomizations, and can be used as a stringent criterion for assessing temporal structure CR2. In CR2, the data are considered to have strong temporal structure if the credible interval for the estimate with correct sampling times does not overlap with those from any of the date-randomized replicates. The tree in **b** presents an example of phylogenetic and temporal clustering, where samples A and B have similar sampling times to each other, as do samples D and E. In this case, a cluster-based date-randomization test is more appropriate

randomizations. In this case, one can compute where the estimate with correct sampling times falls with respect to those from the randomizations, providing the equivalent of a frequentist p -value (Duchêne et al. 2018). Although a large number of randomizations is desirable, several studies have used 20 with reasonable results (e.g., Kerr et al. 2012; Duchêne et al. 2015b).

A critical consideration when performing the date-randomization test comes about when data have a nonuniform temporal sampling. In many time-structured data sets, dates are grouped in such a way that close relatives have similar sampling ages, a pattern known as phylogenetic and temporal clustering (Fig. 10.2b).

In cases of nonuniform temporal sampling, the temporal and phylogenetic information is confounded and this can lead to severe overestimation of molecular rates. A possible reason for the poor rate estimates is that such data sets provide few independent instances of comparison

between molecular and temporal data, and therefore less information about molecular rates (Murray et al. 2015). Interestingly, data sets that yield highly phylogenetically imbalanced trees (those that look pectinate or comb-like) also tend to yield overestimates of molecular rates (Duchêne et al. 2015a), which might in part be explained by the common confounding of temporal and phylogenetic data observed in imbalanced trees.

Most tests of temporal structure fail to reject data sets when the temporal and phylogenetic information are confounded. This means that data will be falsely identified as containing temporal structure. One solution to this problem is to use the clustered date-randomization test (Duchêne et al. 2015b), in which sampling times are permuted among samples but not among those that share the same age. Such a clustered approach to date randomization leads to a reliable test of temporal structure (Duchêne et al. 2015b; Murray et al. 2015).

10.4.3 Bayesian Test of Model Fit

The statistical fit of models with different sample dates can provide an alternative test of temporal structure. For example, treating a heterochronous data set as isochronous is expected to lead to a poorer statistical fit than if samples are assigned their true dates. In Bayesian molecular dating, testing for temporal structure using model fit is done by estimating the marginal likelihood of two different models: one using the empirical sampling times, and one where all the samples are assigned the most recent date (Baele et al. 2012; Murray et al. 2015). If the data contain temporal structure, the marginal likelihood for the model with the original sample dates is expected to be superior. Approximate methods for computing marginal likelihoods are often regarded as computationally expensive and sometimes unreliable, but some are likely to be sufficient (e.g., path sampling, stepping-stone sampling; Xie et al. 2011; Baele et al. 2013). Marginal likelihoods can be readily estimated using popular software such as BEAST.

Analyses of empirical data have shown that this method can be misleading if a poor marginal-likelihood estimator is used, with a tendency to support the presence of temporal structure even in analyses that yield incorrect estimates of evolutionary rates and times (Murray et al. 2015). However, recent work has demonstrated that a highly accurate estimator, generalized stepping-stone sampling (Fan et al. 2011; Baele et al. 2016), can effectively detect temporal structure in simulations and empirical data (Duchêne et al. 2020).

10.4.4 Comparing the Prior and Posterior to Assess Temporal Structure

The broad uptake of the date-randomization test is largely due to the possibility of implementing it in popular Bayesian frameworks, such as BEAST. However, the interpretation of its result is not strictly Bayesian, instead bearing some

resemblance to frequentist methods; the goal is to test a hypothesis (whether the data have temporal structure or not) with some confidence level (similar to p -value testing using a significance value, α). In contrast, a fully Bayesian approach should assess statistical support for including sampling times (Baele et al. 2012; Murray et al. 2015; Duchêne et al. 2020) or assessing the extent to which the sequence data and sampling times are informative about the inferences. The former method has been previously assessed (see Sect. 10.4.3), but the latter has received limited attention.

In general, sequence data are considered informative if the posterior distribution is considerably different from the prior (with the notable exception of internal-node calibrations; Heath and Moore 2014). The expectation is that with informative sequence data, the posterior should be more precise and closer to the true value than the prior, a behaviour also known as statistical consistency. However, even sequence data with very low information content can drive, and sometimes mislead, estimates of some parameters in the full phylogenetic model. As a case in point, Möller et al. (2018) found that uninformative sequence data can produce precise, but incorrect, estimates of tree length and of the evolutionary rate. This probably occurs because sequence data that are uninformative for estimating evolutionary rates and timescales can still contain sufficient information to resolve the topology. Under these circumstances, a limited set of trees will be sampled and lead to a posterior that is much more precise than the prior. Other parameters, including the age of the root node, do not appear to suffer from this problem.

Here we describe an approach that involves comparing the prior and posterior distributions of different parameters to assess information content in molecular sequence data and their association with their sampling times. Our method of assessing temporal structure consists of quantifying information content in the posterior relative to that of the prior for the age of the root node. A simple measure is to take the 95% quantile width divided by the mean, an analogue to the coefficient of variation and referred to here

as CV. We calculate this for the prior, CV_{prior} , and for the posterior, $CV_{\text{posterior}}$, and take the ratio $CV_{\text{ratio}} = CV_{\text{prior}} / CV_{\text{posterior}}$. A CV_{ratio} of 1 means that the prior and posterior are equally informative, whereas a CV_{ratio} of more than 1 means that the posterior is more informative than the prior.

We expect that data with temporal structure should have a higher CV_{ratio} than those with no temporal structure. However, this can depend on the parameter in question and its corresponding prior. For example, if the evolutionary rate has a very broad prior, even sequence data with no temporal structure can produce a posterior that is much more informative than the prior, with a potentially large but misleading CV_{ratio} . This is expected because the rate will be a function of the number of variable sites and the prior on the age of the root node. In contrast, the age of the root node will require data with strong temporal structure to obtain an informative posterior and high CV_{ratio} .

To determine the behaviour of this approach, we simulated the evolution of DNA sequences using parameters inferred for H1N1 influenza virus, which typically has clocklike behaviour and strong temporal structure (Hedge et al. 2013). We used the HKY+ Γ substitution model, a strict molecular clock with an evolutionary rate of 3.66×10^{-3} substitutions per site per year, and an exponential coalescent process for the branching times. One hundred data sets were simulated to have temporal structure, with sampling times that span 7 months and which match those of some data sets of the 2009 influenza pandemic in North America (Hedge et al. 2013), while another 100 were generated on ultrametric trees and with no temporal structure. All data sets contained 50 samples, sequence lengths of 13,156 nt, and about 350 variable sites to match typical genome data sets from influenza virus.

We analysed the data in BEAST 2.5 (Bouckaert et al. 2014, 2019) using a substitution model and tree prior that matched those used to generate the data, and a Markov chain Monte Carlo simulation with 5×10^7 steps, sampling every 5000 steps. For the data with temporal structure, we used the correct sampling times for calibration, but, for the data with no temporal

structure, we set sampling times from a typical influenza outbreak (Hedge et al. 2013). We used a relaxed-clock model with a lognormal distribution. This model has good performance even for data that follow a strict clock (Drummond et al. 2006), and it can accommodate apparent rate variation among lineages that might arise when specifying sampling times for the data with no temporal structure. The priors were all proper, such that each integrates to 1, and were selected according to previous analyses of these data (Duchêne et al. 2019). Ideally, one could compare the prior selected for each parameter with its marginal posterior distribution. However, such user-specified priors often differ from the marginal prior, particularly those for ages of nodes (including that of the root node) which can interact with the topology and other priors (Duchêne et al. 2014b). To obtain the marginal prior one can run the analyses without sequence data (equivalent to selecting the option ‘sample from prior’ in BEAST 2).

The simulations demonstrate that analyses of data with temporal structure result in a posterior that is much more informative than the prior (Fig. 10.3a, b), with a CV_{ratio} between 3 and 11 for the evolutionary rate and between 4 and 13 for the age of the root node. In 97 out of 100 simulations the posterior 95% credible interval of the evolutionary rate included the value used to generate the data. Analyses of the data with no temporal structure yielded rate estimates that never included the true evolutionary rate, but the posterior for this parameter was nonetheless always more informative than the prior (Fig. 10.3c), with a CV_{ratio} between 1 and 12. The CV_{ratio} values of the evolutionary rate are similar for both sets of simulations, despite those with no temporal structure always yielding incorrect rate estimates. This illustrates the point that comparing the prior and posterior of the rate can provide a misleading assessment of temporal structure (Fig. 10.4a). This probably occurs because sequence data are informative about the topology and the total amount of sequence divergence even in the absence of temporal structure.

The CV_{ratio} of the age of the root node is a more useful diagnostic to assess temporal structure than that of the evolutionary rate (e.g.,

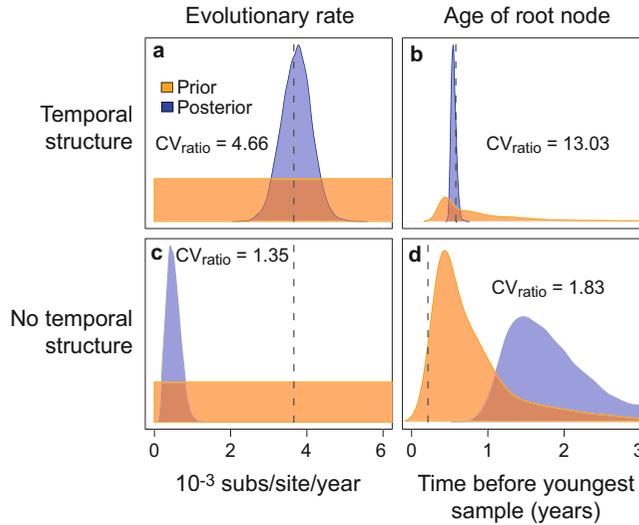


Fig. 10.3 Prior and posterior densities of the mean evolutionary rate and the age of the root node for a simulated data set with temporal structure (a, b) and without temporal structure (c, d). CV_{ratio} is a measure of information content. For the prior and the posterior, we calculate the 95% interval width divided by the mean, and the ratio of

this quantity of the prior and the posterior is the CV_{ratio} . A value of 1 indicates that the prior and posterior are similarly informative, with higher values suggesting a more informative posterior. The dashed line corresponds to the ‘true’ value used to generate the data

Fig. 10.3b, d). Its value ranged between 4.8 and 14 for the simulated data with temporal structure and between 1 and 2.5 for those with no temporal

structure (Fig. 10.4). According to these results, a posterior for the age of the root node that is about fivefold more informative than the prior

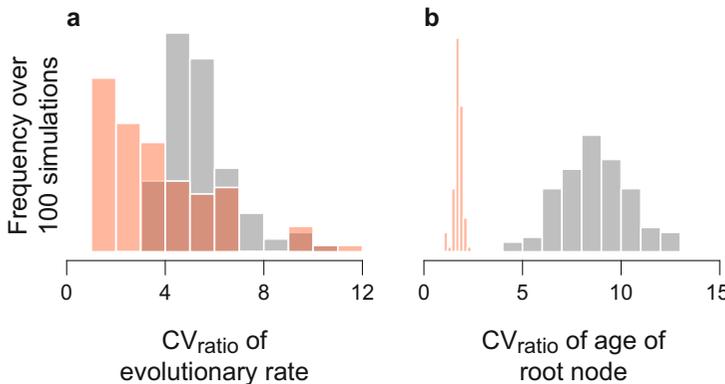


Fig. 10.4 Histograms of the CV_{ratio} for the evolutionary rate and age of the root node for 100 simulations with temporal structure (grey) and without temporal structure (red). Higher values indicate a more informative posterior distribution relative to the prior. (a) CV_{ratio} of the evolutionary rate is similar whether the data have temporal structure or not, but, in the data with no temporal structure, this parameter was never estimated correctly, meaning that

this statistic is misleading for assessing temporal structure. (b) In contrast, CV_{ratio} of the age of the root node is much higher for the simulated data with temporal structure. The distribution of CV_{ratio} with no temporal structure (red) is lower and does not overlap with that from the simulated data with temporal structure (grey). As such, CV_{ratio} of the age of the root node is an informative statistic for assessing temporal structure

($CV_{\text{ratio}} > 5$) can be used as evidence of temporal structure.

Although this test of temporal structure appears to be effective, it requires careful consideration of the priors, especially those on node times as imposed by the tree prior. Here we have used a coalescent tree prior that is conditioned on the sampling times. In contrast, birth–death tree priors can provide information about sampling times and, if provided, they can be treated as part of the data to inform diversification parameters and the age of the root node (Boskova et al. 2018). For this reason, it is important to sample from the prior to determine whether it is reasonable (Heled and Drummond 2012). An obvious problem is when the prior on the age of the root node is over-informative, such that the posterior is very similar regardless of whether the data have strong temporal structure or not. In such circumstances, inferences of evolutionary rates and times are driven by the tree prior and temporal structure might only have a small impact. Although this is sometimes desirable, for example when internal-node calibrations are used in combination with sampling times, it should be explicitly acknowledged when interpreting the estimates.

10.5 Heterochronous Data Analysis in Practice

To provide an illustration of how temporal structure can be evaluated using Bayesian methods, we present here an analysis of two empirical data sets. Although there has been extensive use and validation of the date-randomization test and the root-to-tip regression, less attention has been given to comparing prior and posterior distributions to assess temporal structure. We analysed two previously published data sets of 2009 H1N1 influenza virus (Hedge et al. 2013) and of an outbreak of the bacterium *Mycobacterium tuberculosis* in the Swiss city of Bern (Kühnert et al. 2018) to show how the results from the simulations in Sect. 10.4.4 can be applied to empirical data. The influenza data set consists of 100 whole genomes collected between

February and August 2009 in North America, while the *M. tuberculosis* data set consists of 68 samples collected in Bern over a 10-year period. Our analyses are similar to those described in Sect. 10.4.4, with the same tree prior, substitution model, and Markov chain Monte Carlo settings.

The evolutionary rate estimates from both data sets were similar to those of the original studies, at 0.22 SNPs per genome per year for *M. tuberculosis*, and 3.66×10^{-3} substitutions per site per year for H1N1 influenza, although slightly lower for *M. tuberculosis*, reported at about 0.5 SNPs per genome per year by Kühnert et al. (2018). The estimate of the age of the root node of influenza is around the start of 2009, which is consistent with the expected origin of the 2009 influenza outbreak in the Northern Hemisphere (Fig. 10.5). According to the simulations in Sect. 10.4.4, a CV_{ratio} for the age of the root node of at least 5 would indicate evidence for temporal structure. As such, there appears to be strong evidence of temporal structure for the influenza data set, with a CV_{ratio} of 8.91, whereas that for the *M. tuberculosis* data is only 1.32. The low CV_{ratio} of the *M. tuberculosis* data is consistent with a low R^2 (0.05) from a root-to-tip regression in the original study (Kühnert et al. 2018). Comparing prior and posterior distributions of the age of the root node appears to be effective for analyses of empirical data. It has the key benefits of an intuitive interpretation and ease of use.

10.6 Conclusions and Future Directions

Calibrating the molecular clock using heterochronous data has been valuable for estimating evolutionary rates and timescales in rapidly evolving organisms and in ancient DNA studies. There has been dramatic progress since the proposal of the early root-to-tip regression and strict-clock methods (Korber et al. 2000; Seo et al. 2002), towards incorporating more sophisticated models of rate variation (Ho and Duchêne 2014; Bromham et al. 2018), modelling

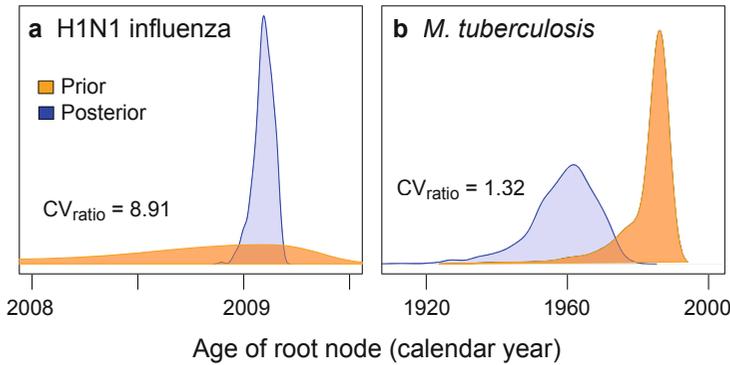


Fig. 10.5 Posterior density of the age of the root node for two empirical data sets of (a) H1N1 influenza virus and (b) *Mycobacterium tuberculosis*. CV_{ratio} of the root height is reported in each case, and it is much higher for the H1N1 influenza data set and within the range of values from the

simulated data with temporal structure, indicating that it has strong temporal structure. In contrast, the *M. tuberculosis* data set has a much lower CV_{ratio} that lies within the range of values from the simulated data with no temporal structure

uncertainty in sampling times (Shapiro et al. 2011; Molak et al. 2013), and handling very large data sets (To et al. 2016).

With most current methods it is important to verify temporal structure to avoid misleading inferences. Several methods to do this have been described in this chapter, but regardless of the choice of method for assessing temporal structure, the results should be carefully considered before any of the data are discarded. For example, many bacterial data sets have strong temporal structure, but it is often obscured by recombination (Schultz et al. 2016), so correctly accounting for recombination is an important development (Vaughan et al. 2017; Didelot et al. 2018). In some viruses, notably Hepatitis B virus, even data sets that include samples from about 500 years ago still show little temporal structure, a pattern that has been attributed to mutational saturation (Patterson Ross et al. 2018). Accordingly, developing more realistic substitution and molecular clock models is likely to improve the resulting inferences. In cases when the best available methods still detect no temporal structure in the data, it might be necessary to resort to adding calibrating information via internal-node calibration or previous rate estimates, to widen the sampling window, or to sequence more informative genomic regions.

Bayesian approaches have been particularly popular because they allow simultaneous estimation of a multitude of parameters of interest, such as migration rates or epidemiological spread (Lemey et al. 2009; Kühnert et al. 2011), and because they can combine different sources of information for calibration (Ronquist et al. 2012a; Zhang et al. 2015). Recent developments, mostly in the Bayesian framework, include models that allow ancient samples to be placed as direct ancestors to modern samples (Gavryushkina et al. 2014), and those that can treat fossil taxa as tips in the phylogenetic tree instead of using them indirectly for internal-node calibrations (Heath et al. 2014). The flexibility of many Bayesian software programs, such as BEAST 2 and RevBayes (Höhna et al. 2016; Bouckaert et al. 2019), presents a key opportunity to develop more realistic approaches for including heterochronous data in complex evolutionary scenarios.

References

- Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway RN, Bunce M (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc R Soc B* 279:4724–4733

- Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu Rev Ecol Syst* 33:707–740
- Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* 29:2157–2167
- Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P (2013) Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* 30:239–243
- Baele G, Lemey P, Suchard MA (2016) Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. *Syst Biol* 65:250–264
- Biek R, Pybus OG, Lloyd-Smith JO, Didelot X (2015) Measurably evolving pathogens in the genomic era. *Trends Ecol Evol* 30:306–313
- Boskova V, Stadler T, Magnus C (2018) The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. *Virus Evol* 4: vex044
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput Biol* 10: e1003537
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Poppinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu C-H, Xie D, Zhang C, Stadler T, Drummond AJ (2019) BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLOS Comput Biol* 15:e1006650
- Brace S, Palkopoulou E, Dalén L, Lister AM, Miller R, Otte M, Germonpré M, Blockley SPE, Stewart JR, Barnes I (2012) Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proc Natl Acad Sci USA* 109:20532–20536
- Bromham L, Duchêne S, Hua X, Ritchie AM, Duchêne DA, Ho SYW (2018) Bayesian molecular dating: opening up the black box. *Biol Rev* 93:1165–1191
- Campos PF, Willerslev E, Sher A, Orlando L, Axelsson E, Tikhonov A, Aaris-Sørensen K, Greenwood AD, Kahlke R-D, Kosintsev P, Krakhmalnaya T, Kuznetsova T, Lemey P, MacPhee R, Norris CA, Shepherd K, Suchard MA, Zazula GD, Shapiro B, Gilbert MTP (2010) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proc Natl Acad Sci USA* 107:5675–5680
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR (2014) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15
- de Bruyn M, Hoelzel AR, Carvalho GR, Hofreiter M (2011) Faunal histories from Holocene ancient DNA. *Trends Ecol Evol* 26:405–413
- Der Sarkissian C, Allentoft ME, Ávila-Arcos MC, Barnett R, Campos PF, Cappellini E, Ermini L, Fernández R, da Fonseca R, Ginolhac A, Hansen AJ, Jónsson H, Korneliussen T, Margaryan A, Martin MD, Moreno-Mayar JV, Raghavan M, Rasmussen M, Velasco MS, Schroeder H, Schubert M, Seguin-Orlando A, Wales N, Gilbert MTP, Willerslev E, Orlando L (2015) Ancient genomics. *Philos Trans R Soc B* 370:20130387
- Didelot X, Wilson DJ (2015) ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLOS Comput Biol* 11:e1004041
- Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ (2018) Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res* 46: e134
- Drummond AJ, Stadler T (2015) Evolutionary trees. In: Drummond AJ, Bouckaert R (eds) *Bayesian evolutionary analysis with BEAST*. Cambridge University Press, Cambridge, UK, pp 21–43
- Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161:1307–1320
- Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends Ecol Evol* 18:481–488
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLOS Biol* 4:e88
- Duchêne S, Holmes EC, Ho SYW (2014a) Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. *Proc R Soc B* 281:20140732
- Duchêne S, Lanfear R, Ho SYW (2014b) The impact of calibration and clock-model choice on molecular estimates of divergence times. *Mol Phylogenet Evol* 78:277–289
- Duchêne D, Duchêne S, Ho SYW (2015a) Tree imbalance causes a bias in phylogenetic estimation of evolutionary timescales using heterochronous sequences. *Mol Ecol Resour* 15:785–794
- Duchêne S, Duchêne D, Holmes EC, Ho SYW (2015b) The performance of the date-randomization test in phylogenetic analyses of time-structured virus data. *Mol Biol Evol* 32:1895–1906
- Duchêne S, Geoghegan JL, Holmes EC, Ho SYW (2016a) Estimating evolutionary rates using time-structured data: a general comparison of phylogenetic methods. *Bioinformatics* 32:3375–3379
- Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC (2016b) Genome-scale rates of evolutionary change in bacteria. *Microb Genom* 2:e000094
- Duchêne S, Duchêne DA, Geoghegan JL, Dyson ZA, Hawkey J, Holt KE (2018) Inferring demographic

- parameters in bacterial genomic data using Bayesian and hybrid phylogenetic methods. *BMC Evol Biol* 18:95
- Duchêne S, Bouckaert R, Duchêne DA, Stadler T, Drummond AJ (2019) Phylogenetic model adequacy using posterior predictive simulations. *Syst Biol* 68:358–364
- Duchêne S, Stadler T, Ho SYW, Duchêne DA, Dhanasekaran V, Baele G (2020) Bayesian evaluation of temporal signal in measurably evolving populations. *Mol Biol Evol* 37:3363–3379
- Duffy S, Holmes EC (2009) Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *J Gen Virol* 90:1539–1547
- du Plessis L, Stadler T (2015) Getting to the root of epidemic spread with phylodynamic analysis of genomic data. *Trends Microbiol* 23:383–386
- Fan Y, Wu R, Chen M-H, Kuo L, Lewis PO (2011) Choosing among partition models in Bayesian phylogenetics. *Mol Biol Evol* 28:523–532
- 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
- Fitch WM, Leiter JME, Li X, Palese P (1991) Positive Darwinian evolution in human influenza A viruses. *Proc Natl Acad Sci USA* 88:4270–4274
- Fourment M, Holmes EC (2014) Novel non-parametric models to estimate evolutionary rates and divergence times from heterochronous sequence data. *BMC Evol Biol* 14:163
- Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA, Al-Rasheid KAS, Bradley DG, Orlando L (2016) Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. *Mol Ecol Resour* 16:459–469
- Gavryushkina A, Welch D, Stadler T, Drummond AJ (2014) Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLOS Comput Biol* 10:e1003919
- Grealy A, Phillips M, Miller G, Gilbert MTP, Rouillard J-M, Lambert D, Bunce M, Haile J (2017) Eggshell palaeogenomics: Palaeognath evolutionary history revealed through ancient nuclear and mitochondrial DNA from Madagascan elephant bird (*Aepyornis* sp.) eggshell. *Mol Phylogenet Evol* 109:151–163
- Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* 303:327–332
- Griffiths RC, Tavaré S (1994) Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc B* 344:403–410
- Guilderson TP, Reimer PJ, Brown TA (2005) The boon and bane of radiocarbon dating. *Science* 307:362–364
- Heath TA, Moore BR (2014) Bayesian inference of species divergence times. In: Chen M-H, Kuo L, Lewis PO (eds) *Bayesian phylogenetics: methods, algorithms, and applications*. CRC Press, Boca Raton, FL, pp 277–318
- Heath TA, Huelsenbeck JP, Stadler T (2014) The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc Natl Acad Sci USA* 111:E2957–E2966
- Hedge J, Wilson DJ (2014) Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. *mBio* 5:e02158–e02114
- Hedge J, Lycett SJ, Rambaut A (2013) Real-time characterization of the molecular epidemiology of an influenza pandemic. *Biol Lett* 9:20130331
- Heled J, Drummond AJ (2012) Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst Biol* 61:138–149
- Ho SYW, Duchêne S (2014) Molecular-clock methods for estimating evolutionary rates and time scales. *Mol Ecol* 23:5947–5975
- Ho SYW, Shapiro B (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Mol Ecol Resour* 11:423–434
- Höhna S, Landis MJ, Heath TA, Boussau B, Lartillot N, Moore BR, Huelsenbeck JP, Ronquist F (2016) RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol* 65:726–736
- Holmes EC (2009) *The evolution and emergence of RNA viruses*. Oxford University Press, Oxford, UK
- Kerr PJ, Ghedin E, DePasse JV, Fitch A, Cattadori IM, Hudson PJ, Tschärke DC, Read AF, Holmes EC (2012) Evolutionary history and attenuation of myxoma virus on two continents. *PLOS Pathog* 8:e1002950
- Kingman JFC (1982) The coalescent. *Stoch Proc Appl* 13:235–248
- Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T (2000) Timing the ancestor of the HIV-1 pandemic strains. *Science* 288:1789–1796
- Kühnert D, Wu C-H, Drummond AJ (2011) Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infect Genet Evol* 11:1825–1841
- Kühnert D, Coscolla M, Brites D, Stucki D, Metcalfe J, Fenner L, Gagneux S, Stadler T (2018) Tuberculosis outbreak investigation using phylodynamic analysis. *Epidemics* 25:47–53
- Langley CH, Fitch WM (1974) An examination of the constancy of the rate of molecular evolution. *J Mol Evol* 3:161–177
- Lapierre M, Blin C, Lambert A, Achaz G, Rocha EPC (2016) The impact of selection, gene conversion, and biased sampling on the assessment of microbial demography. *Mol Biol Evol* 33:1711–1725
- Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. *PLOS Comput Biol* 5:e1000520
- Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, Ugan A, Borregaard MK,

- Gilbert MTP, Nielsen R, Ho SYW, Goebel T, Graf KE, Byers D, Stenderup JT, Rasmussen M, Campos PF, Leonard JA, Koepfli K-P, Froese D, Zazula G, Stafford TW, Aaris-Sørensen K, Batra P, Haywood AM, Singarayer JS, Valdes PJ, Boeskorov G, Burns JA, Davydov SP, Haile J, Jenkins DL, Kosintsev P, Kuznetsova T, Lai X, Martin LD, McDonald HG, Mol D, Meldgaard M, Munch K, Stephan E, Sablin M, Sommer RS, Sipko T, Scott E, Suchard MA, Tikhonov A, Willerslev R, Wayne RK, Cooper A, Hofreiter M, Sher A, Shapiro B, Rahbek C, Willerslev E (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* 479:359–364
- Menardo F, Duchêne S, Brites D, Gagneux S (2019) The molecular clock of *Mycobacterium tuberculosis*. *PLOS Pathog* 15:e1008067
- Millar CD, Huynen L, Subramanian S, Mohandesan E, Lambert DM (2008) New developments in ancient genomics. *Trends Ecol Evol* 23:386–393
- Molak M, Lorenzen ED, Shapiro B, Ho SYW (2013) Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. *Mol Biol Evol* 30:253–262
- Molak M, Suchard MA, Ho SYW, Beilman DW, Shapiro B (2015) Empirical calibrated radiocarbon sampler: a tool for incorporating radiocarbon-date and calibration error into Bayesian phylogenetic analyses of ancient DNA. *Mol Ecol Resour* 15:81–86
- Möller S, du Plessis L, Stadler T (2018) Impact of the tree prior on estimating clock rates during epidemic outbreaks. *Proc Natl Acad Sci USA* 115:4200–4205
- Murray GGR, Wang F, Harrison EM, Paterson GK, Mather AE, Harris SR, Holmes MA, Rambaut A, Welch JJ (2015) The effect of genetic structure on molecular dating and tests for temporal signal. *Methods Ecol Evol* 7:80–89
- Palkopoulou E, Dalén L, Lister AM, Vartanyan S, Sablin M, Sher A, Edmark VN, Brandström MD, Germonpré M, Barnes I, Thomas JA (2013) Holarctic genetic structure and range dynamics in the woolly mammoth. *Proc R Soc B* 280:20131910
- Patterson Ross Z, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, Poinar D, Douglas MW, Eden J-S, Holmes EC (2018) The paradox of HBV evolution as revealed from a 16th century mummy. *PLOS Pathog* 14:e1006750
- Posth C, Wißing C, Kitagawa K, Pagani L, van Holstein L, Racimo F, Wehrberger K, Conard NJ, Kind CJ, Bocherens H, Krause J (2017) Deeply divergent archaic mitochondrial genome provides lower time boundary for African gene flow into Neanderthals. *Nat Commun* 8:16046
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Mojrjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PLF, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49
- Pybus OG, Rambaut A, Harvey PH (2000) An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155:1429–1437
- Rambaut A (2000) Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 16:395–399
- Rambaut A, Lam TT, Carvalho LM, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* 2:vew007
- Ramsden C, Melo FL, Figueiredo LM, Holmes EC, Zanotto PMA, VGDN Consortium (2008) High rates of molecular evolution in hantaviruses. *Mol Biol Evol* 25:1488–1492
- Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol Biol Evol* 26:143–153
- Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Ramsey CB, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Haffidason H, Hajdas I, Hatté C, Heaton TJ, Hoffmann DL, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney CSM, van der Plicht J (2013) IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55:1869–1887
- Rieux A, Balloux F (2016) Inferences from tip-calibrated phylogenies: a review and a practical guide. *Mol Ecol* 25:1911–1924
- Rodrigo AG, Felsenstein J (1999) Coalescent approaches to HIV population genetics. In: Crandall KA (ed) *The evolution of HIV*. Johns Hopkins University Press, Baltimore, MD, pp 233–272
- Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP (2012a) A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst Biol* 61:973–999
- Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012b) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542
- Rosenberg NA, Nordborg M (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet* 3:380–390
- Sagulenko P, Puller V, Neher RA (2018) TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol* 4:vex042

- Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302
- Schultz MB, Duy PT, Nhu TDH, Wick RR, Ingle DJ, Hawkey J, Edwards DJ, Kenyon JJ, Nguyen PHL, Campbell JI, Thwaites G, Nguyen TKN, Hall RM, Fournier-Level A, Baker S, Holt KE (2016) Repeated local emergence of carbapenem-resistant *Acinetobacter baumannii* in a single hospital ward. *Microb Genom* 2:e000050
- Seo TK, Thorne JL, Hasegawa M, Kishino H (2002) A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics* 18:115–123
- Shapiro B, Drummond AJ, Rambaut A, Wilson MC, Matheus PE, Sher AV, Pybus OG, Gilbert MTP, Barnes I, Binladen J, Willerslev E, Hansen AJ, Baryshnikov GF, Burns JA, Davydov S, Driver JC, Froese DG, Harington CR, Keddie G, Kosintsev P, Kunz ML, Martin LD, Stephenson RO, Storer J, Tedford R, Zimov S, Cooper A (2004) Rise and fall of the Beringian steppe bison. *Science* 306:1561–1565
- Shapiro B, Ho SYW, Drummond AJ, Suchard MA, Pybus OG, Rambaut A (2011) A Bayesian phylogenetic method to estimate unknown sequence ages. *Mol Biol Evol* 28:879–887
- Stadler T (2010) Sampling-through-time in birth-death trees. *J Theor Biol* 167:696–404
- Stadler T, Bonhoeffer S (2013) Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos Trans R Soc B* 368:20120198
- Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ (2013) Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci USA* 110:228–233
- Stuiver M, Reimer PJ (1993) Extended ^{14}C data base and revised CALIB 3.0 ^{14}C age calibration program. *Radiocarbon* 35:215–230
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016
- Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15:1647–1657
- To T-H, Jung M, Lycett S, Gascuel O (2016) Fast dating using least-squares criteria and algorithms. *Syst Biol* 65:82–97
- Tong KJ, Duchêne DA, Duchêne S, Geoghegan JL, Ho SYW (2018) A comparison of methods for estimating substitution rates from ancient DNA sequence data. *BMC Evol Biol* 18:70
- Vaughan TG, Welch D, Drummond AJ, Biggs PJ, George T, French NP (2017) Inferring ancestral recombination graphs from bacterial genomic data. *Genetics* 205:857–870
- Volz EM, Frost SDW (2017) Scalable relaxed clock phylogenetic dating. *Virus Evol* 3:vex025
- Volz EM, Kosakovsky Pond SL, Ward MJ, Brown AJL, Frost SDW (2009) Phylodynamics of infectious disease epidemics. *Genetics* 183:1421–1430
- Xie W, Lewis PO, Fan Y, Kuo L, Chen MH (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol* 60:150–160
- Yahara K, Didelot X, Jolley KA, Kobayashi I, Maiden MCJ, Sheppard SK, Falush D (2016) The landscape of realized homologous recombination in pathogenic bacteria. *Mol Biol Evol* 33:456–471
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yule GU (1924) A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S. *Philos Trans R Soc B* 213:21–87
- Zhang C, Stadler T, Klopstein S, Heath TA, Ronquist F (2015) Total-evidence dating under the fossilized birth–death process. *Syst Biol* 65:228–249
- Zhou Z, Lundstrøm I, Tran-Dien A, Duchêne S, Alikhan N-F, Sergeant MJ, Langridge G, Fotakis AK, Nair S, Stenøien HK (2018) Pan-genome analysis of ancient and modern *Salmonella enterica* demonstrates genomic stability of the invasive Para C lineage for millennia. *Curr Biol* 28:2420–2428