Additive uncorrelated relaxed clock models for the dating of genomic epidemiology phylogenies

Xavier Didelot[1,*] and Erik M Volz[2]

[1] School of Life Sciences and Department of Statistics, University of Warwick, United Kingdom

[2] Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, United Kingdom

[*] Corresponding author. Tel: 0044 (0)2476 572827. Email: xavier.didelot@gmail.com

INTRODUCTION

Dated phylogenies have branch lengths measured in unit of time instead of evolutionary distance as in standard phylogenies. The leaves of a dated phylogeny are aligned on the time axis with their isolation dates (which is usually known) and the internal nodes are aligned with the time when the corresponding common ancestors existed (which is usually unknown but can be estimated). Dated phylogenies represent a very useful and popular tool for genomic epidemiology, for example to study population dynamics (1), transmission (2), pathogen population structure (3) or host population structure (4). Dated phylogenies can be built directly from the genetic data for example using BEAST (5, 6, 7, 8), although this can sometimes be slow especially when working with large datasets. Consequently, an alternative two-step approach has emerged over the past few years, which is based on first building a standard phylogeny, and second estimating the date of each node in this phylogeny. The first step (standard phylogenetics) can be performed for example using RAxML (9), PhyML (10), FastTree (11) or IQ-TREE (12). The second step (phylogeny dating) can be performed for example using LSD (13), node.dating (14), treedater (15), TreeTime (16) or BactDating (17). In this article for ease of presentation we focus on this phylogeny dating approach, although our findings on the choice of clock models are relevant to the integrated approach too.

An important consideration when building a dated phylogeny is the choice of the clock model, which represents the way in which mutations accumulate during the evolution of the population (18, 19, 20, 21). In the phylogeny dating approach, the clock model represents the stochastic relationship between the known evolutionary length x_i and the unknown length of time l_i for each branch i of the phylogeny. To keep this relationship simple, let us consider that on each branch each site mutates at most once, which is equivalent to assuming an infinite sites model (22). This assumption is usually acceptable when working with genomic epidemiological data, where the overall time scale is relatively small. It would be straightforward to extend our results to a finite sites model (23), but we prefer to focus on the infinite sites model to keep the presentation as simple as possible. In this case, x_i represents exactly the number of mutations that happened on each branch. The simplest model is called the strict clock model and assumes a constant rate μ of mutation on all the branches (24). Therefore, a branch of duration l_i will contain a number of mutations x_i which is Poisson distributed with parameter μl_i.

The strict clock model (24) has just a single parameter μ and this simplicity is attractive, but it is often too simple because of variations in the mutation rate from one lineage to another. A number of alternatives have been proposed, with the most popular being the uncorrelated relaxed clock model (25). Under this model, each branch has its own mutation rate m_i, and these per-branch rates are independent of one another. In current implementations of the uncorrelated relaxed clock model, the rates m_i are drawn independently and identically from a well defined rate distribution, for example a Lognormal distribution (25), an Exponential distribution (25, 13), a Normal distribution (16), or a Gamma distribution (15, 17). However, we found that the use of the same distribution for all per-branch rates of the uncorrelated relaxed clock model is inconsistent with the intuitive biological expectation of additivity between branches of the phylogeny. For example, if we consider two branches i and j of the tree with durations l_i and l_j respectively, then the distribution of $x_i + x_j$ is not the same as the distribution for a branch of length $l_i + l_j$. The currently used models are therefore not robust to adding or removing genomes in the phylogeny, since the way these genomes find common ancestors with the remaining genomes will cause some branches to be split or merged. This non-additivity issue of the frequently used relaxed clock models becomes clear when we consider splitting or merging branches of the tree, but it is of course important even if the user has no intention to add or remove genomes, since it means that the dating results are not robust to the selection of genomes used for analysis.

Using an additive model is likely to be especially important for applications of dating in genomic epidemiology where many branches of short duration are considered, due to very large sample sizes and epidemic processes of interest sometimes occurring in a matter of days (26, 27). It is also very relevant to real-time studies of pathogen outbreaks (28, 29), where new cases are continuously added onto the phylogeny over time, splitting ancestral branches. Here we propose alternative robust uncorrelated relaxed clock models which solve this issue and therefore have better statistical and biological properties compared to the current models. We consider both the case where the number of mutations on a branch is discrete or continuous. We illustrate the difference between our models and previous models using simulations, and show that previous models can lead to misleading conclusions on both simulated and real genomic epidemiology datasets.

MATERIALS AND METHODS

Additivity of the strict clock model

We start with the simple strict clock (SC) model (24) in order to set notations and define the additivity property in this context. Under the SC model, we have that each branch mutates as a Poisson process with rate μ. The number of mutations x_i on a branch of length l_i is therefore:

$$x_i \sim \text{Poisson}(l_i\mu) \tag{1}$$

Note that we use lower case symbols for both random variables and their realisations which is a frequently used abuse of notation in the field (and also more generally when greek symbols are used). Let us now consider two branches of lengths l_1 and l_2. Under the SC model, the distribution of the convolution $x_1 + x_2$, ie the sum of the number of mutations on both branches, is the same as the distribution of the number x of mutations on a branch of length $l = l_1 + l_2$, because:

$$x_1 \sim \text{Poisson}(l_1\mu) \text{ and } x_2 \sim \text{Poisson}(l_2\mu) \implies x_1 + x_2 \sim \text{Poisson}((l_1 + l_2)\mu) \tag{2}$$

We call this property the additivity of the SC model, and note that it is a consequence of the infinite divisibility of the Poisson distribution.

Non-additivity of previous uncorrelated relaxed clock models

The uncorrelated relaxed clock (RC) model was first proposed by Drummond et al (25). In this model, each branch has its own mutation rate m_i. A convenient choice for the distribution of the m_i rates is a Gamma(k, θ) distribution, since this is the conjugate of the Poisson distribution of x_i given l_i. As previously noted (15) this choice leads to:

$$x_i \sim \text{NegBin}\left(k, \frac{\theta l_i}{1 + \theta l_i}\right) \tag{3}$$

More generally, let μ and σ^2 denote the mean and variance of the distribution of per-branch rates m_i. In the case of the Gamma(k, θ) distribution, this is achieved by setting $k = \frac{\mu^2}{\sigma^2}$ and $\theta = \frac{\sigma^2}{\mu}$. Using the laws of of total expectation and variance of x_i we can show that:

$$\mathbf{E}(x_i) = \mathbf{E}(\mathbf{E}(x_i|m_i l_i)) = \mathbf{E}(m_i l_i) = \mu l_i \tag{4}$$

$$\mathbf{V}(x_i) = \mathbf{E}(\mathbf{V}(x_i|m_i l_i)) + \mathbf{V}(\mathbf{E}(x_i|m_i l_i)) = \mathbf{E}(m_i l_i) + \mathbf{V}(m_i l_i) = \mu l_i + \sigma^2 l_i^2 \tag{5}$$

We note that the expectation is the same as in the SC model, whereas the variance is increased by an additive factor $\sigma^2 l_i^2$. The fact that the variance is increased makes sense since RC is a relaxation of the SC model. However, the variance is increased by a factor that is not proportional to the branch length l_i, and this implies that the model does not have the additivity property. In particular, we find that the variance of the number of mutations x on a branch of length $l = l_1 + l_2$ is greater than the variance of $x_1 + x_2$ where x_1 and x_2 are numbers of mutations on branches of lengths l_1 and l_2 respectively:

$$\mathbf{V}(x) = \mu l + \sigma^2 l^2 > \mathbf{V}(x_1 + x_2) = \mu(l_1 + l_2) + \sigma^2(l_1^2 + l_2^2) \tag{6}$$

Since the variances of x and $x_1 + x_2$ are not the same, their distributions are clearly not identical and so the RC is not additive like the SC model. This is true for the RC model in Equation 3 which is based on the same Gamma distribution for all per-branch rates, but the calculation above was not based on any particular distribution, so that it also applies to any other RC model based on any other identical distribution for the per-branch rates. The fact that the RC model does not have the additivity property is problematic both from a statistical and biological point of view.

Additive uncorrelated relaxed clock model

In order to obtain the additivity property in a relaxed clock model, we propose an alternative model which we call the additive uncorrelated relaxed clock (ARC) model. This model has parameters μ and ω such that a branch of duration l_i has mutation rate m_i with expectation $\mathbf{E}(m_i) = \mu$ and variance $\mathbf{V}(m_i) = \mu\omega/l_i$. Using the laws of total expectation and variance as previously, we find that:

$$\mathbf{E}(x_i) = \mathbf{E}(\mathbf{E}(x_i|m_i l_i)) = \mathbf{E}(m_i l_i) = \mu l_i \tag{7}$$

$$\mathbf{V}(x_i) = \mathbf{E}(\mathbf{V}(x_i|m_i l_i)) + \mathbf{V}(\mathbf{E}(x_i|m_i l_i)) = \mathbf{E}(m_i l_i) + \mathbf{V}(m_i l_i) = \mu l_i(1 + \omega) \tag{8}$$

The expected number of mutations under the ARC model is therefore the same as in the SC model and RC model. The variance is increased relative to the SC model by a multiplicative factor $1 + \omega$. The values of the expectation and variance on the number of mutations are therefore compatible with the desired additivity property of the proposed model. However, this is a necessary but not sufficient condition. For the model to be additive, we need the distributions to be additive, not just their expectations and variances. We can obtain this full additivity property using a Gamma distribution for the mutation rate m_i of a branch of length l_i as follows:

$$m_i \sim \text{Gamma}\left(\frac{\mu l_i}{\omega}, \frac{\omega}{l_i}\right) \tag{9}$$

Since the Gamma distribution is the conjugate prior to the Poisson($m_i l_i$) distribution of x_i given m_i we get:

$$x_i \sim \text{NegBin}\left(\frac{\mu l_i}{\omega}, \frac{\omega}{1+\omega}\right) \tag{10}$$

This ARC model clearly verifies the additivity property, since the sum of two Negative-Binomial random variables with the same second parameter is also a Negative-Binomial random variable. Specifically:

$$x_1 \sim \text{NegBin}\left(\frac{\mu l_1}{\omega}, \frac{\omega}{1+\omega}\right) \text{ and } x_2 \sim \text{NegBin}\left(\frac{\mu l_2}{\omega}, \frac{\omega}{1+\omega}\right) \implies x_1+x_2 \sim \text{NegBin}\left(\frac{\mu(l_1+l_2)}{\omega}, \frac{\omega}{1+\omega}\right) \tag{11}$$

Like the Poisson distribution used in the SC model (Equation 1), the Negative-Binomial distribution used here (ie with a constant second parameter $\frac{\omega}{1+\omega}$) is infinitely divisible for its first parameter which is proportional to the branch length.

Continuous relaxed clock models

In this section we consider models in which the branch lengths x_i of the phylogenetic tree, measured in units of substitutions, are continuous. This is useful because most standard phylogenetic software return trees where branch lengths are continuous, in order to accommodate uncertainties in ancestral sequence reconstructions (30) and to account for non-uniform mutation models which give different weights to different types of mutations (31). Gamma distributions are a natural choice for this as previously noted (17). For example, in the case of a continuous strict clock (CSC) model with rate μ, instead of the discrete Poisson distribution from Equation 1 we can use the Gamma distribution with the same expectation and variance, namely:

$$x_i \sim \text{Gamma}(\mu l_i, 1) \tag{12}$$

This CSC model satisfies the additivity property, since:

$$x_1 \sim \text{Gamma}(\mu l_1, 1) \text{ and } x_2 \sim \text{Gamma}(\mu l_2, 1) \implies x_1 + x_2 \sim \text{Gamma}(\mu(l_1+l_2), 1) \tag{13}$$

A continuous uncorrelated relaxed clock (CRC) model was recently proposed (17) based on the assumption that each branch has its own mutation rate m_i with mean μ and variance σ^2, as in the discrete RC model. Specifically, x_i was proposed to be Gamma distributed as follows:

$$x_i \sim \text{Gamma}\left(\frac{\mu^2 l_i}{\mu + \sigma^2 l_i}, 1 + \frac{\sigma^2 l_i}{\mu}\right) \tag{14}$$

This choice is analogous to the discrete RC model (25) previously mentioned, and suffers from the same issue of non-additivity. In particular, we can use the laws of total expectation and variance of

x_i to get $\mathbf{E}(x_i) = \mu l_i$ and $\mathbf{V}(x_i) = \mu l_i + \sigma^2 l_i^2$ exactly as in the discrete case (cf Equations 4 and 5). If $\sigma^2 = 0$ this model reduces to the CSC model (Equation 12) which is additive, but otherwise this model does not have the additive property. This is true for the CRC model in Equation 14 but also for any other CRC model based that assumes that the per-branch rates are identically and independently distributed.

We can remedy this issue in a similar way as we did for the discrete case, and define a continuous additive relaxed clock (CARC) model. We consider the model with parameters μ and ω such that a branch of duration l_i has mutation rate m_i with the same expectation and variance as in the discrete case, ie $\mathbf{E}(m_i) = \mu$ and variance $\mathbf{V}(m_i) = \mu\omega/l_i$. By application of the laws of total expectation and variance, we get the same expectation and variance for x_i as in the discrete case, cf Equations 7 and 8. These formula for the expectation and variance of x_i are necessary for the additivity of the model, but as noted in the discrete case they are not sufficient since we also need the distributions themselves to be additive. To obtain this property we define the CARC using the following Gamma distribution:

$$x_i \sim \text{Gamma}\left(\frac{\mu l_i}{1+\omega}, 1+\omega\right) \tag{15}$$

If $\omega = 0$, this model reduces to the CSC model (Equation 12). The CARC model has the additivity property since the sum of two Gamma distributed random variables with the same scale parameter is also Gamma distributed with the same scale. Specifically:

$$x_1 \sim \text{Gamma}\left(\frac{\mu l_1}{1+\omega}, 1+\omega\right) \text{ and } x_2 \sim \text{Gamma}\left(\frac{\mu l_2}{1+\omega}, 1+\omega\right) \implies x_1 + x_2 \sim \text{Gamma}\left(\frac{\mu(l_1+l_2)}{1+\omega}, 1+\omega\right) \tag{16}$$

Note that there is a difference in the way we derived this continuous model (CARC, Equation 15) compared to the discrete model (ARC, Equation 10): in the latter we selected a distribution on m_i to deduce the distribution of x_i whereas in the former we selected a distribution of x_i directly, without worrying about the distribution of m_i. There is however no difference in practice between these two approaches: in the discrete case the distribution of m_i was selected to get the distribution of x_i we wanted (ie with the additivity property) which is not statistically more principled than directly specifying the distribution of x_i.

Model properties and implementation

We have shown above that the existing strict clock models for both discrete (SC, Equation 1) and continuous (CSC, Equation 12) cases satisfy the desired additivity property, whereas the existing uncorrelated relaxed clock models (RC and CRC) do not. However, we have defined two new relaxed clock models for the discrete (ARC, Equation 10) and continuous cases (CARC, Equation 15) that have the additivity property. The SC model is a simple Poisson process on the branches of the phylogeny, whereas the ARC model corresponds to a negative binomial process (32, 33). The CSC and CARC models both correspond to a gamma process, and these three processes are all Lévy processes, which means that they have stationary and independent increments (34). Lévy processes generate infinitely divisible random variables, which implies the additive property that we sought, since a branch may be divided into any number of parts when samples are added into a phylogenetic tree, and this division

should not affect the distribution of the number of mutations on that branch. The ARC model in Equation 10 can therefore be obtained by considering that branches are made of L infinitesimal units, each of which as an associated number of substitutions distributed as NegBin $\left(\frac{\mu l_i}{Ls}, \frac{\omega}{1+\omega}\right)$. The sum of these L random variables corresponds to the number of substitutions on the whole branch, which is distributed as in Equation 10 using the Negative-Binomial summation rule (Equation 11). Likewise, the CARC model in Equation 15 can be derived using Gamma $\left(\frac{\mu l_i}{L(1+\omega)}, 1+\omega\right)$ for the distribution of substitution of each infinitesimal unit and using the Gamma summation rule (Equation 16). One of the earliest proposed models for a relaxed molecular clock (35) was based on a compound Poisson process which is another type of Lévy process and therefore satisfies the additive property, but this model has not been used in a phylogenetic framework. More generally, Lévy processes are natural to describe biological phenomena in time, and have been proposed several times recently to model evolutionary jumps (36, 37, 38), which is similar to the relaxation of the molecular clock we want to model in this study.

The treedater software can perform dating of the nodes of a phylogeny using maximum likelihood (15). treedater previously implemented the SC and RC models, and we have extended it so that it can now also use the new ARC model (Equation 10). The BactDating software can perform dating of the nodes of a phylogeny using Bayesian inference (17). BactDating previously implemented the SC, RC, CSC and CRC models, and we have extended it so that it can now also use the new ARC and CARC models (Equation 10 and 15). Futhermore, BactDating can simulate datasets based on all six clock models described above.

RESULTS

Comparison of models

Model	Full name	Relaxed	Additive	Continuous	Equation	Reference
SC	Strict Clock	N	Y	N	1	(24)
RC	Relaxed Clock	Y	N	N	3	(25)
ARC	Additive Relaxed Clock	Y	Y	N	10	This study
CSC	Continuous Strict Clock	N	Y	Y	12	(17)
CRC	Continuous Relaxed Clock	Y	N	Y	14	(17)
CARC	Continuous Additive Relaxed Clock	Y	Y	Y	15	This study

Table 1: Summary of the six clock models under study and their properties.

The six clock models considered in this article and their properties are summarised in Table 1. We compared the discrete distributions of number of substitutions implied by the strict SC model, the relaxed RC model and the new ARC model, varying both the duration of the branches considered and the level of relaxation in the RC and ARC models (Figure 1). Increasing the variance of the per-branch rates in the RC model (parameter σ^2) and the ARC model (parameter ω) made the distributions of substitutions increasingly diffuse relative to the SC model, as expected. There are however marked differences in behaviour between the RC and ARC models: in the RC model the distribution mode for longer branches quickly shifts to small values as relaxation is increased, whereas this is not the case in the ARC model. Conversely, for short branches even a high σ^2 in the RC model does not imply much relaxation, whereas a high ω in the ARC model has a much clearer effect for small branches.

We performed a similar comparison of the continuous distributions of number of substitutions implied by the strict CSC model, the relaxed CRC model and the new CARC model (Figure 2). We note that these results are very similar to the discrete case for all six models considered, ie SC vs CSC, RC vs CRC and ARC vs CARC (compare Figures 1 and 2). This indicates that the Gamma distributions used in the three continuous models are good continuous equivalent to the Poisson and NegBin distributions used in the three discrete models. In particular, comparison between CRC and ARC shows very similar features to the ones described above between RC and ARC concerning the effect on short vs long branches. A useful property in the continuous case (and not in the discrete case) is that the CSC model is a special case of both the CRC model (by setting $\sigma^2 = 0$) and the CARC (by setting $\omega = 0$). This property is useful for model selection, since it means that CSC is embedded within CRC and CARC.

Application to multiple simulated datasets

We simulated 100 datasets, each of which consisted of 100 genomes sampled at regular intervals between 2010 and 2020. The dated phylogeny was simulated from the coalescent model (39) with an expected coalescent time for any two lineages equal to $\alpha = N_e g = 5$ years. The CARC model was used to simulate mutations along the branches of this dated phylogeny, with a mean rate of $\mu = 5$ mutations per year, and a dispersion parameter varying between $\omega = 0$ (in which case the model reduced to the strict CSC) and $\omega = 10$. The resulting unrooted phylogenies were used as input trees in both BactDating (17) and treedater (15). In BactDating, separate MCMC runs were performed assuming either the old CRC model or the new CARC model. Each MCMC was run for 10^5 iterations which took approximately 10 minutes on a single node of a standard desktop computer. Good convergence and mixing properties of the MCMC results were found using both the Gelman-Rubin diagnostic (40, 41) and effective sample size test implemented in CODA (42).

We compared the fit of these two models by computing the deviance information criterion (DIC) of both models (43). We found that the CARC had significantly better fit (ie smaller DIC) for all simulations with $\omega > 1$, which is as expected since the data was simulated from the CARC model. This model comparison was more ambiguous when $\omega < 1$, which again is as expected since when ω is close to zero both the CARC and CRC models reduce to the CSC model. Figure 3 shows the difference between real and estimated time to the most recent common ancestor (TMRCA) and the estimated mean mutation rate μ for both models, as well as the estimates of the parameter ω for the CARC model. The 95% credibility intervals of both the TMRCA and μ almost always include the correct values of zero and five respectively, but the intervals are slightly larger in the CRC model for μ, and much larger for the TMRCA. This indicates that even using the CRC does not result in biased estimated, but that more precise estimates can be obtained using the CARC model, especially for dating nodes. The estimates of ω under the CARC model follows the true values of ω used in the simulation, which is as expected when the same model is used for simulation and inference but also shows that there is significant statistical power, even in these relatively small datasets, to correctly infer the level of relaxation of the molecular clock.

We applied treedater (15) to the same datasets using the ARC model and computed parametric bootstrap values for the TMRCA, mean mutation rate μ and relaxation parameter ω (Figure S1). The inferred values of ω followed the correct values used in the simulations, which is as expected since the ARC model used for inference is a discrete version of the CARC used for simulation. The TMRCA and μ were correctly inferred with no evidence of bias, but the 95% confidence intervals estimated using parametric bootstrapping were wider than the Bayesian credible intervals in BactDating, which is certainly the result of inherent differences between these two statistical approaches rather than

differences between the continuous and discrete models.

Application to real datasets

Use DIC (43) to compare fit of models.

DISCUSSION

todo

References

[1] Ho SYW, Shapiro B. Skyline-plot methods for estimating demographic history from nucleotide sequences. Mol Ecol Resour. 2011;11(3):423–434.

[2] Didelot X, Fraser C, Gardy J, Colijn C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. Mol Biol Evol. 2017;34:997–1007.

[3] Volz EM, Wiuf C, Grad YH, Frost SDW, Dennis AM, Didelot X. Identification of hidden population structure in time-scaled phylogenies. bioRxiv. 2019;p. 704528.

[4] Volz EM, Koelle K, Bedford T. Viral Phylodynamics. PLoS Comput Biol. 2013 mar;9(3):e1002947.

[5] Drummond AJ, Rambaut A. BEAST : Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.

[6] Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29(8):1969–1973.

[7] Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014 apr;10(4):e1003537.

[8] 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(1):vey016.

[9] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313.

[10] 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 may;59(3):307–21.

[11] Price MN, Dehal PS, Arkin AP. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One. 2010 jan;5(3):e9490.

[12] Nguyen LT, 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(1):268–274.

[13] To TH, Jung M, Lycett S, Gascuel O. Fast dating using least-squares criteria and algorithms. Syst Biol. 2016;65(1):82–97.

[14] Jones BR, Poon AFY. Node.dating: Dating ancestors in phylogenetic trees in R. Bioinformatics. 2017;33(6):932–934.

[15] Volz EM, Frost SDW. Scalable relaxed clock phylogenetic dating. Virus Evol. 2017;3(2):vex025.

[16] Sagulenko P, Puller V, Neher RA. TreeTime: Maximum likelihood phylodynamic analysis. Virus Evol. 2018;4:vex042.

[17] Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. 2018;46:e134.

[18] Kumar S. Molecular clocks: four decades of evolution. Nat Rev Genet. 2005;6:654–662.

[19] Lepage T, Bryant D, Philippe H, Lartillot N. A general comparison of relaxed molecular clock models. Mol Biol Evol. 2007 dec;24(12):2669–80.

[20] Drummond AJ, Suchard MA. Bayesian random local clocks, or one rate to rule them all. BMC Biol. 2010;8(1):114.

[21] Lartillot N, Phillips MJ, Ronquist F. A mixed relaxed clock model. Philos Trans R Soc B Biol Sci. 2016;371:20150132.

[22] Kimura M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics. 1969;61(4):893–903.

[23] Jukes TH, Cantor CR. Evolution of protein molecules. In: Mamm. protein Metab. Academic Press; 1969. p. 21–132.

[24] Zuckerkandl E, Pauling L. Molecular Disease, Evolution, and Genic Heterogeneity. In: Horizons Biochem. Academic Press Inc; 1962. p. 189–222.

[25] Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006 may;4(5):e88.

[26] Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. Nature. 2015;524(7563):97–101.

[27] Faria NR, Quick J, Claro IM, Thézé J, De Jesus JG, Giovanetti M, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546(7658):406–410.

[28] Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228–232.

[29] Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: Real-time tracking of pathogen evolution. Bioinformatics. 2018;34(23):4121–4123.

[30] Yang Z, Rannala B. Molecular phylogenetics: principles and practice. Nat Rev Genet. 2012 may;13(5):303–14.

[31] Liò P, Goldman N. Models of molecular evolution and phylogeny. Genome Res. 1998;8(12):1233–1244.

[32] Barndorff-Nielsen O, Yeo GF. Negative Binomial Processes. J Appl Probab. 1969;6(3):633–647.

[33] Kozubowski T, Podgorski K. Distributional properties of the negative binomial Lévy process. Probab Math Stat. 2009;29(Fasc. 1):43–71.

[34] Applebaum D. Lévy Processes - From Probability to Finance and Quantum Groups. Not AMS. 2004;51(11):1336–1347.

[35] Takahata N. On the overdispersed molecular clock. Genetics. 1987;116(1):169–179.

[36] Jourdain B, Méléard S, Woyczynski WA. Lévy flights in evolutionary ecology. J Math Biol. 2012;65(4):677–707.

[37] Landis MJ, Schraiber JG, Liang M. Phylogenetic analysis using Lévy processes: Finding jumps in the evolution of continuous traits. Syst Biol. 2013;62(2):193–204.

[38] Duchen P, Leuenberger C, Szilágyi SM, Harmon L, Eastman J, Schweizer M, et al. Inference of Evolutionary Jumps in Large Phylogenies using Levy Processes. Syst Biol. 2017;66(6):950–963.

[39] Kingman JFC. The coalescent. Stoch Process their Appl. 1982 sep;13(3):235–248.

[40] Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. Stat Sci. 1992;7(4):457–511.

[41] Brooks SPB, Gelman AG. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1998;7(4):434–455.

[42] Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006;6(March):7–11.

[43] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Ser B (Statistical Methodol. 2002;64(4):583–639.

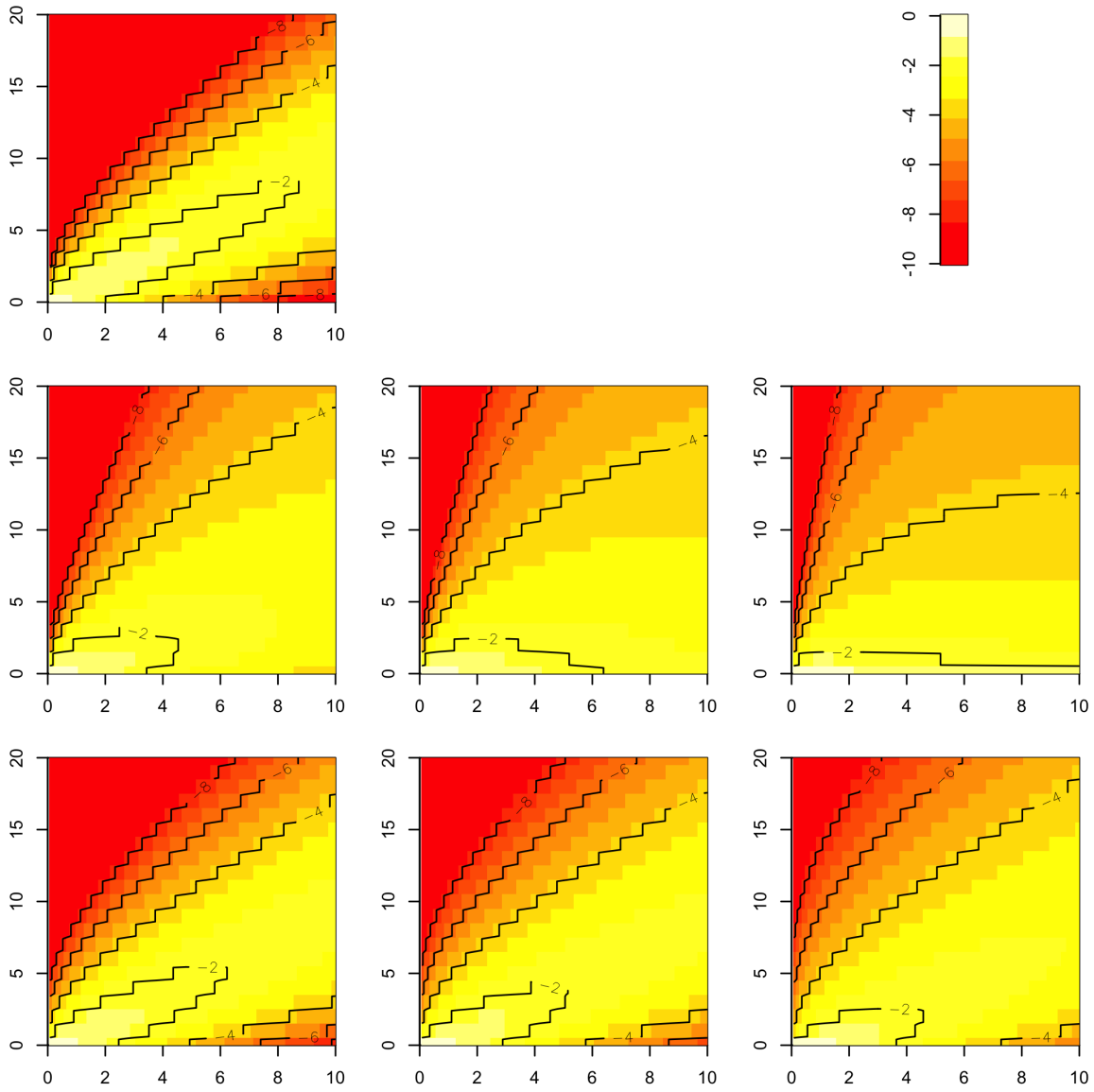

Figure 1: **Comparison of discrete clock models.** The top-left plot shows the SC model, with $\mu = 1$. The second row shows the RC model, with $\mu = 1$ and $\sigma^2 = 0.5$, 1 and 2 respectively from left to right. The third row shows the ARC model, with $\mu = 1$ and $\omega = 0.5$, 1 and 2 respectively from left to right. In each plot, the x-axis shows values of l_i, the y-axis shows values of x_i and the color represents the value of the log of $p(x_i|l_i)$ as per the legend.

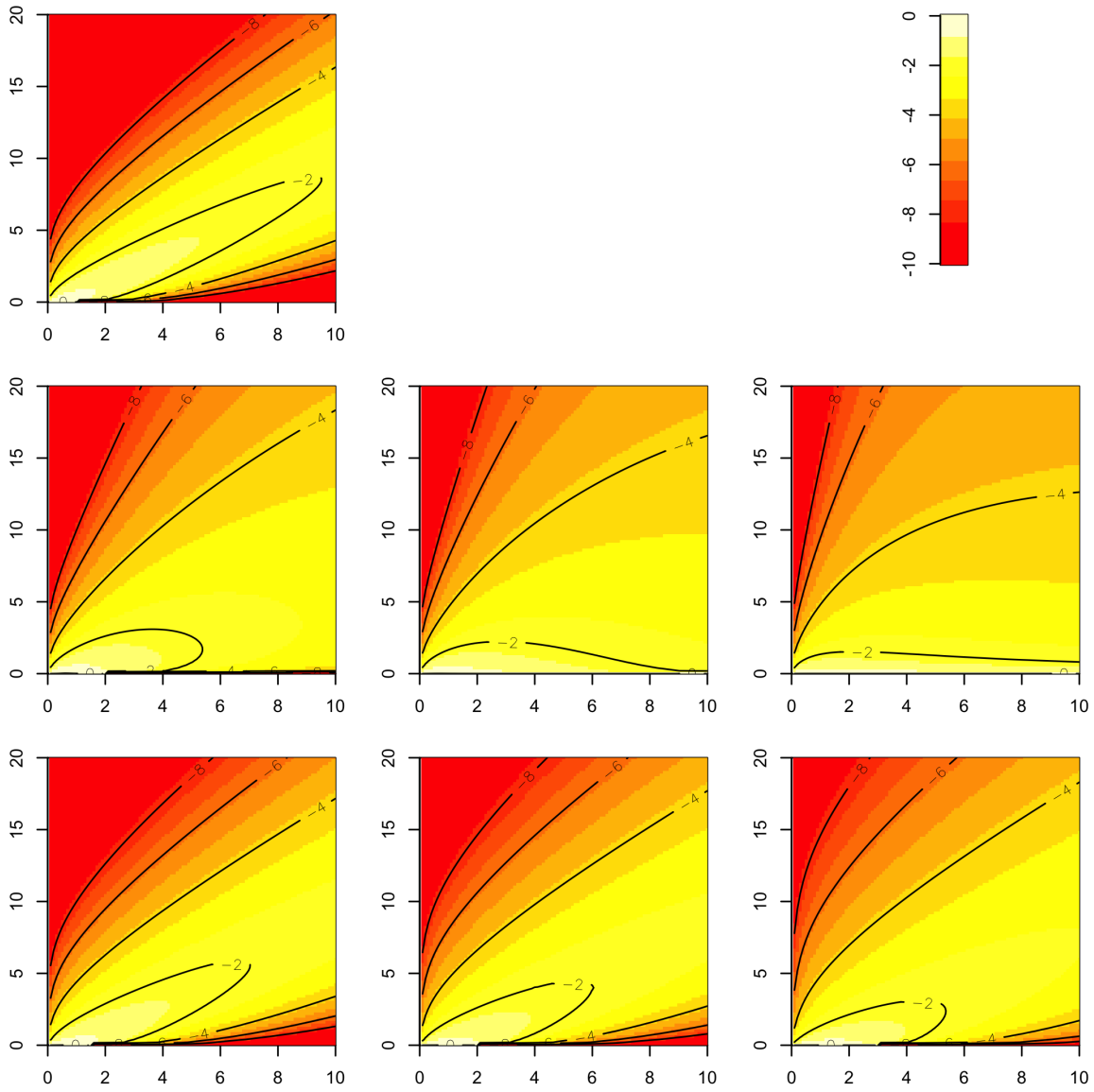

Figure 2: **Comparison of continuous clock models.** The top-left plot shows the CSC model, with $\mu = 1$. The second row shows the CRC model, with $\mu = 1$ and $\sigma^2 = 0.5$, 1 and 2 respectively from left to right. The third row shows the CARC model, with $\mu = 1$ and $\omega = 0.5$, 1 and 2 respectively from left to right. In each plot, the x-axis shows values of l_i, the y-axis shows values of x_i and the color represents the value of the log of $p(x_i|l_i)$ as per the legend.

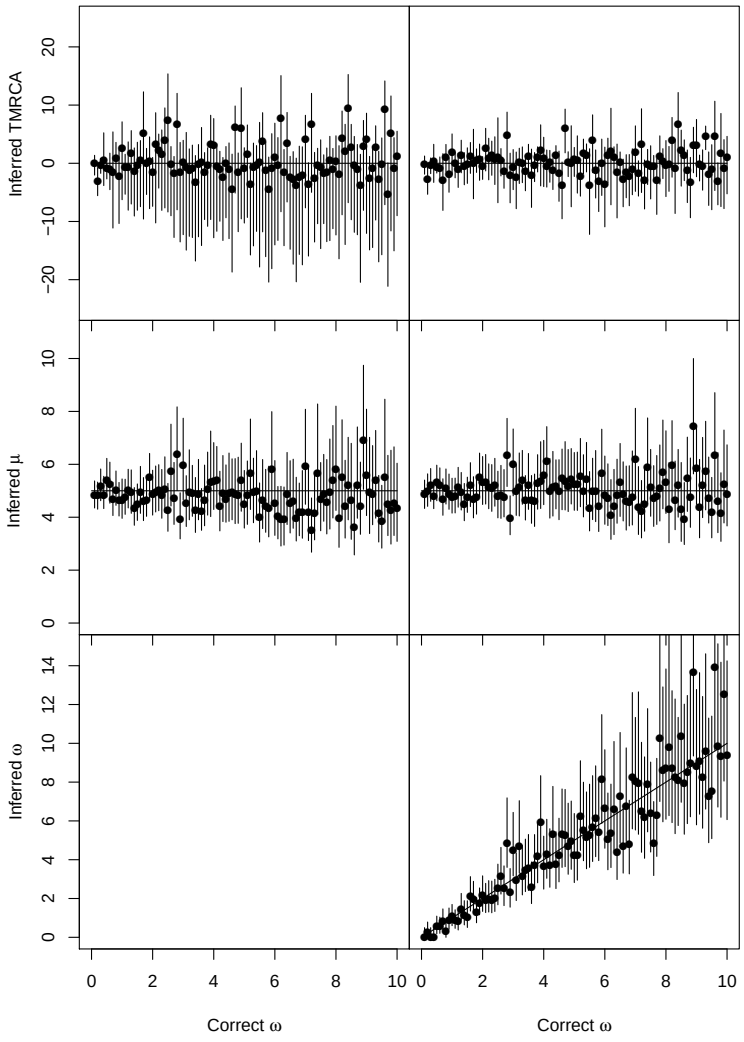

Figure 3: **Application of BactDating to 100 simulated datasets.** On the left inference used the CRC model and on the right the CARC model. The top row shows inferred values of the TMRCA (relative to the correct value), the middle row shows inferred values of the mean mutation rate μ, and the bottom row shows inferred values of the relaxation parameter ω for the CARC model. In each plot, the x-axis represents the value of ω used in the simulations (varied between 0 and 10) and the y-axis represents the inferred values, with a dot for the posterior mean and a bar for the 95% credibility interval.

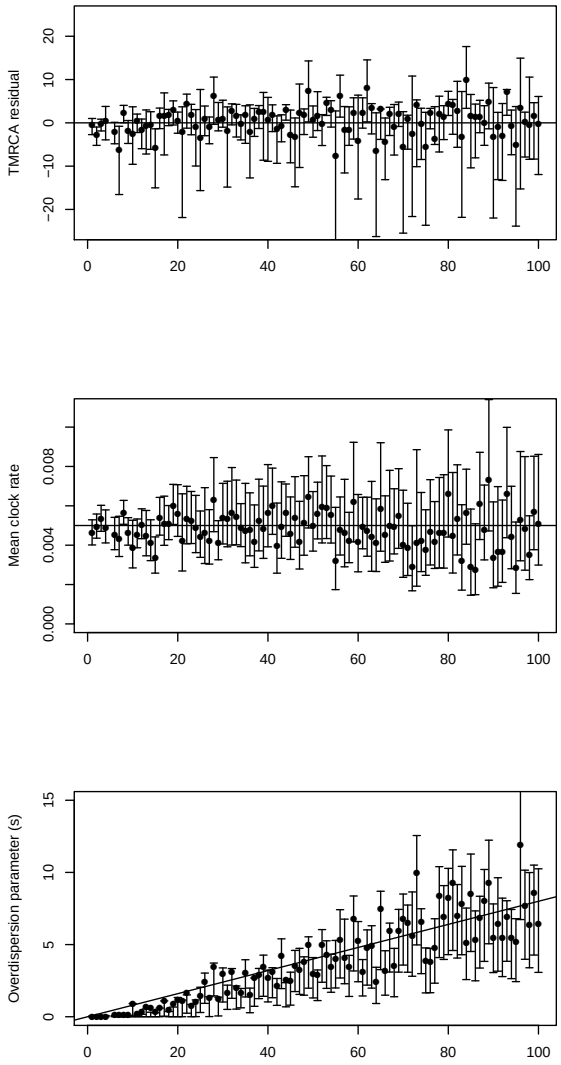

Figure S1: **Application of treedater to 100 simulated datasets.** The top, middle and bottom rows show inferred values of the TMRCA, mean mutation rate μ and relaxation parameter ω, respectively.