



HAL
open science

Confidence intervals in molecular dating by maximum likelihood

Emmanuel Paradis, Santiago Claramunt, Joseph Brown, Klaus Schliep

► **To cite this version:**

Emmanuel Paradis, Santiago Claramunt, Joseph Brown, Klaus Schliep. Confidence intervals in molecular dating by maximum likelihood. *Molecular Phylogenetics and Evolution*, 2023, 178, pp.107652. 10.1016/j.ympev.2022.107652 . ird-04892561

HAL Id: ird-04892561

<https://ird.hal.science/ird-04892561v1>

Submitted on 16 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Confidence intervals in molecular dating by maximum likelihood

Emmanuel Paradis^{a,*}, Santiago Claramunt^b, Joseph Brown^b, Klaus Schliep^c

^a*ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France*

^b*Department of Natural History, Royal Ontario Museum, Toronto, ON, 5S2C6 Canada*

^c*Institute of Computational Biotechnology, Technology University Graz, Austria*

Abstract

Molecular dating has been widely used to infer the times of past evolutionary events using molecular sequences. This paper describes three bootstrap methods to infer confidence intervals under a penalized likelihood framework. The basic idea is to use data pseudoreplicates to infer uncertainty in the branch lengths of a phylogeny reconstructed with molecular sequences. The three specific bootstrap methods are nonparametric (direct tree bootstrapping), semiparametric (rate smoothing), and parametric (Poisson simulation). Our extensive simulation study showed that the three methods perform generally well under a simple strict clock model of molecular evolution; however, the results were less positive with data simulated using an uncorrelated or a correlated relaxed clock model. Several factors impacted, possibly in interaction, the performance of the confidence intervals. Increasing the number of calibration points had a positive effect, as well as increasing the sequence length or the number of sequences although both latter effects depended on the model of evolution. A case study is presented with a molecular phylogeny of the Felidae (Mammalia: Carnivora). A comparison was made with a Bayesian analysis: the results were very close in terms of confidence intervals and there was no marked tendency for an approach to produce younger or older bounds compared to the other.

Keywords: bootstrap, Felidae, penalized likelihood, phylogenetics

*Corresponding author

Email address: Emmanuel.Paradis@ird.fr (Emmanuel Paradis)

1. Introduction

2 The inference of divergence dates using molecular sequence data has many applications such
as estimating the dates of major evolutionary radiations more than one billion years ago (Parsons
4 et al., 2021) or quantifying the dynamics of viral epidemics over several months (?). Pioneer works
on the molecular clock by Zuckerkandl and Pauling (1962, 1965b,a) opened up the development
6 of numerous methods on these issues (Rutschmann, 2006; Bromham, 2019; Guindon, 2020). A
crucial step forward was achieved by the integration of molecular clock models with phylogenetic
8 tree inference (Felsenstein, 1984, 1987) making possible to account for the variation in rates of
molecular evolution both among lineages and among genomic regions (Zuckerkandl, 1987; Ho
10 et al., 2011).

Assessing variation in rates of molecular evolution in DNA sequences involves two distinct
12 levels. First, the rate of substitution (the replacement over a long time of one base by another) can
vary with the type of substitution (e.g., TS/TV ratio) or the position of the base along the sequence.
14 This level of variation can be taken into account during phylogenetic inference. The fundamental
model of phylogenetic inference assumes time-reversibility so a reconstructed tree is unrooted
16 and its branch lengths are expressed in expected numbers of substitutions (Felsenstein, 1981).
Therefore, variation in substitution rates among lineages cannot be generally inferred (unless in
18 some special cases; see below). Second, molecular dating takes into account variation among
lineages in order to estimate dates of the nodes of a (assumed rooted) phylogeny. The simplest
20 model (commonly called ‘strict clock’ model) assumes that the substitution rate is the same in all
lineages. The models that relax this assumption are called under the collective name of ‘relaxed
22 clock’ models. The present paper is concerned with this second level of variation.

Incorporating heterogeneity in molecular evolutionary rates when analyzing large data sets,
24 possibly covering large numbers of evolutionary lineages and of genomic regions, can be done in
a statistical framework (Sanderson, 2002; Drummond et al., 2006). A critical aspect of statistical
26 inference is to assess the confidence or uncertainty in parameter estimates. Bayesian methods treat
this problem by integrating over a set of prior parameter distributions in order to infer their pos-
terior distributions given the data (e.g., Zhu et al., 2015; Bromham et al., 2018). With frequentist
28 methods, confidence intervals (CIs) of the parameter can be derived with different methods based
30 on assumptions about the distribution of the estimators or on the shape of the likelihood function

(Casella and Berger, 2002). However, such methods may not be reliable if their basic assumptions
32 are not met. The bootstrap provides a general approach to compute CIs of an estimate with few
assumptions (Efron, 1981). An attractive feature of the bootstrap is that it can be used in complex
34 situations such as phylogenetic inference (Felsenstein, 1985; Efron et al., 1996).

Some efforts have been made to increase the realism of relaxed clock models, for instance
36 the universal pacemaker model (?), or specialized models to analyze bacterial genomes (??). Fur-
thermore, recent works have emphasized the importance of calibration points in molecular dating
38 (??); this issue is further discussed below. In a recent review, ? list seventeen computer programs
aimed at performing molecular dating: two of them are based on a distance-based approach, six
40 on a likelihood approach, and nine on a Bayesian approach.

Baldwin and Sanderson (1998) used nonparametric bootstrap data matrices to infer confidence
42 intervals of the inferred dates using a strict clock model (which was not rejected from their data).
Uncertainty in tree topology was also assessed in their study. This approach was refined and inte-
44 grated into the program r8s (Sanderson, 2003), although this has been removed from its later ver-
sions (see r8s's manual dated December 2004; [https://sourceforge.net/projects/
46 r8s/](https://sourceforge.net/projects/r8s/); accessed 2022-04-20). This approach has been used extensively in the literature (Burbrink
and Pyron, 2008; Sauquet, 2013, and references therein). Britton et al. (2002) developed the mean
48 path length (MPL) method that uses a rooted tree with branch lengths estimated with no assump-
tion of a strict molecular clock (see next section) to estimate the divergence dates with the variance
50 of the lengths of the paths from nodes to tips. Under the assumption of Poisson-distributed sub-
stitutions, standard-errors of these estimated dates can be calculated. Several approaches based
52 on least squares also proposed to use bootstrap resampling to assess uncertainty in the estimated
dates (Xia and Yang, 2011; To et al., 2016). Tao et al. (2020) developed an analytical method
54 to compute CIs based on the RelTime method (Tamura et al., 2012), a method that shares some
similarities with the MPL method. Using simulated and real data, they showed that their method
56 gives CIs similar to the credible intervals from a Bayesian analysis.

In this paper, we further elaborate on the nonparametric bootstrap introduced in the previous
58 paragraph, and present semiparametric, and parametric bootstrap methods to infer CIs in molecular
dates inferred by maximum likelihood (Sanderson, 2002; Paradis, 2013). These methods, which
60 use pseudoreplication together with likelihood maximization, are likely to be attractive in big
phylogenetic problems where a Bayesian analysis is unfeasible (e.g., Harvey et al., 2020). An

62 extensive simulation study was run to assess several statistical properties of these three methods.
A case study is presented on the molecular phylogeny of the Felidae (Mammalia, Carnivora).

64 **2. Methods**

Throughout this paper, it is assumed that the data are made of n aligned sequences with s sites
66 with one or more additional sequences representing an outgroup (i.e., the n sequences constitute a
monophyletic group with respect to the outgroup) also aligned with the n sequences, so the initial
68 data set is made of n' sequences (with $n' > n$). In absence of an outgroup, the tree can be rooted
by other means that may be less optimal (e.g., To et al., 2016) which we do not consider here.
70 Besides, we adhere to the terminology of calling a ‘phylogram’ a phylogenetic tree with branch
lengths in numbers of substitutions, a ‘chronogram’ a tree with lengths in units of time, and a
72 ‘cladogram’ a tree with no branch lengths.

2.1. Maximum likelihood framework

74 The aligned sequences are first analyzed with phylogenetic tree reconstruction by maximum
likelihood (ML) yielding an unrooted binary phylogram with $2n' - 3$ branches. After rooting the
76 tree with the outgroup and removing the latter from the tree, we obtain a rooted binary phylogram
with n tips and $2n - 2$ branches. The branches are now time-oriented from the root towards the
78 tips. We denote each branch length as l_i ($i = 1, \dots, 2n - 2$). ML phylogenetic inference usually
assumes a Markovian, time-reversible model of molecular evolution so that multiple substitutions
80 are taken into account, and the estimated branch length \hat{l}_i is interpreted as the mean number of
substitutions per site between the two nodes along branch i (Felsenstein, 2004). We denote the
82 substitution rate along branch i as r_i . This rate is expressed in substitutions per site per unit of
time. We further denote the time separating the two nodes along branch i as t_i , and write:

$$l_i = r_i \times t_i. \tag{1}$$

84 We make the following assumptions on the possible values of these three quantities: $t_i > 0$, $r_i \geq 0$,
and therefore $l_i \geq 0$.

86 We now aim to find the dates of the nodes: we denote these dates as d_j ($j = 1, \dots, n - 1$). If
we assume that the sequences are contemporary, we can define constraints among the t_i 's and the

88 dates in relation to the tree topology. For instance, if two terminal branches, say x and y , define a
cherry in the tree (i.e., they are connected to the same node), then $t_x = t_y = d_k$ where k is the node
90 ancestor of the branches x and y . Even with such constraints, there are still too many parameters for
a straightforward likelihood approach (the number of cherries in a tree depends on its topology and
92 varies between one and $\lfloor n/2 \rfloor$). Additional constraints on how the substitution rates vary among
the branches of the tree make possible to use a penalized likelihood (PL) approach (Sanderson,
94 2002). The PL function, Ψ , is:

$$\Psi = \ln L - \lambda \Phi, \quad (2)$$

where L is the likelihood function of a model of the branch lengths based on eq. (1), λ is a smooth-
96 ing parameter, and Φ is a penalty function constraining rate variation. Ψ is defined in terms of its
parameters (i.e., the dates d_j 's and the rates r_i 's) and needs to be maximized with respect to these
98 parameters to find their estimates, \hat{d}_j and \hat{r}_i . An important feature of molecular dating methods
is the way substitution rates vary across the tree. In this paper, we consider three main models:
100 the strict clock model where the rate is constant, the correlated relaxed clock model where rates in
contiguous branches are similar, and uncorrelated relaxed clock model where rates vary without
102 correlation. We emphasize here that there are some small differences in the exact definitions and
implementations of these models depending on the software and whether they are used for simu-
104 lation or inference (see Sect. ?? for details on the models considered in this study). In particular,
we point out that the 'relaxed' model implemented in `ape` (Paradis and Schliep, 2019) is actually
106 an uncorrelated relaxed clock model (the term 'relaxed' is actually often used in the literature to
qualify all clock models outside of the strict clock model).

108 If no calibration point is available, it is possible to fix the age of the root of the tree to a fixed
value (e.g., one) so that the estimated dates are in relative units of time. If some absolute dates can
110 be used as calibration points (e.g., from fossils) then these dates can be calculated in absolute time
units.

112 The penalized framework is one way to impose restrictions on the distribution of substitution
rates. Another way is to define a discrete distribution of these rates throughout the tree but without
114 assigning a specific value to each branch. The likelihood function is calculated with the contri-
bution of each branch integrated with respect to the possible values of rates weighted by their
116 frequencies (e.g., the strict clock model is a special case with a single category; see Lepage et al.,

2007). This approach makes use of a full likelihood function instead of a penalized likelihood.
118 Both approaches can be seen as parts of a general framework of molecular dating with likelihood
methods.

120 2.2. *Uncertainty assessment*

In the ML framework, uncertainty in parameter estimation is commonly assessed by comput-
122 ing the second derivatives of the likelihood function: this amounts to assessing the reliance to be
placed in the estimates in different samples from the same statistical population (Fisher, 1922).
124 This approach assumes a parametric formulation of the data distribution; however, it is not di-
rectly applicable to a PL approach because of the penalty term. On the other hand, the bootstrap
126 makes no parametric assumption about the distribution of the data (Efron and Tibshirani, 1993;
Hesterberg, 2015). The resampling procedure of the bootstrap aims to provide alternative samples
128 from the distribution under investigation using the original data as an approximation for this dis-
tribution. In the context of phylogenetic inference, bootstrap samples are obtained by resampling
130 the sites of the aligned sequences (Felsenstein, 1985). The idea used in this paper is to obtain a
set of estimates of the branch lengths, assess their variation, and use this variation to quantify the
132 uncertainty in the estimated dates (\hat{d}_j). Three bootstrap methods are presented in the followings
(Fig. 1).

134 2.2.1. *Nonparametric bootstrap*

The first method uses the bootstrap trees directly into the penalized likelihood dating proce-
136 dure. Bootstrap samples are first generated by resampling the sites of the sequence alignment with
replacement; each chronogram is then estimated using the same time calibration constraints. Note
138 that the usual phylogenetic bootstrap performs topology search for each bootstrap sample and is
used to assess uncertainty in the inferred bipartitions of the unrooted tree (Czech et al., 2017). By
140 contrast, in the present approach, the tree topology is fixed to the one found by the ML tree search
performed on the original sequence data. Therefore, the bootstrap samples will be used to infer
142 a set of bootstrap trees all with that same topology, but different branch lengths, and then input
in the chronogram estimation procedure. The variation of the estimated dates from the bootstrap
144 trees will give a measure of their uncertainty.

2.2.2. Semiparametric (smoothed) bootstrap

146 Because resampling the sequence alignment is done with replacement, the same site can be
drawn several times in the same bootstrap sample. If two sequences differ by only a few sites,
148 then the configurations of the bootstrap samples is limited and they will vary discretely which
is likely to produce biases for summary statistics based on them. Smoothing the distribution
150 of bootstrap samples has been proposed as a solution to this problem (Silverman and Young,
1987; De Angelis and Young, 1992; ?). Kernel density estimation (Wand and Jones, 1995) can
152 be used to fit a non-parametric distribution to the bootstrap samples of branch lengths for each
branch, then randomly sample this distribution to obtain a set of branch lengths subsequently
154 used in the penalized likelihood procedure. Two constraints are imposed on this distribution:
it must be unimodal, and its density must be equal to zero at $l = 0$. Because of the discrete
156 nature of sequences and the variable sites therein, it is common that bootstrap branch lengths vary
also discretely, so that a non-parametric density estimation might be multimodal (Fig. 2). The
158 bandwidth, b , of the density estimation should not be too large to avoid inferring positive densities
for unrealistic values of l (see below), and not too small to avoid a multimodal density curve. We
160 used an initial bandwidth using Silverman's (1986) 'rule of thumb' which is:

$$b = \min \left(\sigma(l_i), \frac{Q_3(l_i) - Q_1(l_i)}{1.34} \right) \times 0.9 \times B^{-0.2},$$

where $\sigma(l_i)$ is the standard-deviation calculated with the bootstrap sample of the lengths of branch
162 l_i , $Q_1(l_i)$ and $Q_3(l_i)$ are its first and third quartiles, respectively, and B is the number of bootstrap
replicates. The density is then estimated by local Gaussian smoothing. If the resulting density is
164 multimodal, then b is increased by $b/10$ until the density is unimodal.

In order to avoid sampling negative or zero branch lengths, the density was defined for $l \geq$
166 10^{-8} . Because we used a local Gaussian smoothing, it could result in density being positive for
 $l \leq 0$. If this happened, the density function was modified to be zero for $l < 10^{-8}$, and rescaled
168 so that the total density area given by its integral is equal to one. The final density function was
transformed into a cumulative density function (CDF) which was then used to generate random
170 samples of branch lengths by first generating random variates from a uniform distribution on $[0, 1]$
and then finding the values of l from the CDF.

172 *2.2.3. Parametric bootstrap: Poisson simulation*

Under stochastic models of molecular evolution, the number of substitutions along a branch
174 of the tree is expected to follow a Poisson distribution. Since the branch length estimated by ML \hat{l}_i
is the expected number of substitutions along branch i , it can be taken as an estimate of the mean
176 of this Poisson distribution. A sample of branch lengths can thus be generated by simulating from
this distribution and scaling by the sequence length. These samples of branch lengths are then used
178 to estimate a sample of chronograms. The information from these chronograms is summarized in
the same way as for the two other methods.

180 *2.3. Implementation*

We implemented the methods presented in this paper in R (R Core Team, 2024). The ML tree
182 estimation and tree bootstrapping use functions in `phangorn` (Schliep, 2011) which makes possi-
ble to perform phylogenetic bootstrapping with a fixed tree topology. The branch length smoothing
184 and simulations were implemented with functions in the `stats` package. Some additional code has
been written for the present study and is available on GitHub (see Data accessibility) or upon
186 request from the authors.

2.4. Simulation study

188 The questions addressed in the present simulation study were: (1) What are the coverages (the
propability that the true date is within the inferred CI) of the bootstrap methods proposed here?
190 and (2) How robust are these bootstrap methods to model misspecification?

All simulations shared the same basic structure: a tree was simulated under a birth–death
192 process with a fixed value of n using Stadler’s (2011) reverse-time algorithm setting speciation
rate depending on n (as explained below) and extinction rate equal to zero, as coded in `ape`. The
194 branch lengths were rescaled so that the age of the root was equal to 50. An outgroup with a single
tip was attached to this tree (so $n' = n + 1$). Three different models of substitution rate variation
196 among lineages were used as implemented in NELSI (Ho et al., 2015): a strict clock model (with
parameters $\text{rate} = 0.01$ and $\text{noise} = 10^{-5}$), Kishino’s model of autocorrelated rates ($\text{rate} = 10^{-5}$,
198 $\text{standard-deviation} = 0.3$), and a model of uncorrelated, Γ -distributed rates ($\text{shape} = 0.5$, $\text{rate} = 1$).
The justifications for the choice of these parameters are given in the Supplementary information.
200 The tree output by NELSI was used to simulate a set of n' sequences with s sites under a Jukes–
Cantor model (Jukes and Cantor, 1969) using the function `simSeq` in `phangorn`. Before running

202 the dating procedure, κ calibration points were selected from the branching times of the birth–
death tree; except for the root which was always included, the nodes were chosen randomly. The
204 dates of these calibration points were assumed to be known without error. These simulations were
potentially highly stochastic because of the combination of three random processes (birth–death,
206 substitution rate variation, and sequence evolution), so that several adjustments were made. First,
the speciation rate was adjusted with respect to n to avoid very short branches for large values of
208 n : it was set to 0.1, 0.05, and 0.025 for $n = 20, 50, \text{ and } 100$, respectively. Second, the branch
lengths of the tree returned by NELSI were rescaled so the shortest terminal branch had a length
210 of at least 0.001, so that it was unlikely that two sequences were identical. And third, the rate of
the Jukes–Cantor model in phangorn was adjusted so that no more than 50% of the sequence sites
212 were polymorphic. It was assumed that the tree topology was known, the ML tree estimated by
phangorn was rooted by removing the outgroup, and then input in the molecular dating procedure.

214 In the first run, the parameters were $n = \{20, 50, 100\}$; $s = \{10^3, 10^4\}$; and $\kappa = \{1, 5, 10\}$. The
three models described in the previous paragraph (strict clock, correlated relaxed clock, and uncor-
216 related relaxed clock) were used to simulate rates with NELSI. These simulations were replicated
100 times for each combination of n , s , and κ , and each model. The model was assumed to be
218 known, so the model used for the simulation was also used for the chronogram estimation. For
each simulated data set and each bootstrap method, 100 bootstrap replicates were used. The 50%
220 and 95% CIs of each estimated date were calculated and stored with the branching times of the
simulated birth–death tree, as well as the running times of each bootstrap analysis. This run was
222 duplicated independently on a cluster in Canada and on the Occigen supercomputer in Montpellier.

In the second run, two models were considered: strict clock and correlated relaxed clock. The
224 values $n = 20$ and $s = 10^3$ were chosen, while $\kappa = \{1, 5, 10\}$. This run is similar to the previous one
except that the model was assumed to be misspecified: the dating procedure used the correlated
226 model if the data were simulated with the strict clock model, and vice versa. The simulations were
replicated 100 times for each combination of the two models with the different values of κ .

228 2.5. Case study

The family Felidae (Mammalia: Carnivora) includes around 40 living species, all with a
230 strictly carnivorous diet, and many fossil forms (Werdelin et al., 2010). In relation to their preda-
tory lifestyle, they have acquired highly specialized adaptations (e.g., long canines, high running

232 speed, night vision, social hunting behaviour). A dated phylogeny is crucial for the evolutionary
study of these adaptations. The data came from a study by Johnson et al. (2006) on the phylo-
234 genetic relationships among the species of Felidae. A search on GenBank under this reference
returned 993 sequences covering 45 taxa, including seven species of Carnivora not belonging to
236 Felidae and which served as outgroup in this study. All sequences were downloaded: they covered
23 genes although not all were sequenced for the 45 taxa: the number of sequences for each gene
238 varied between 38 and 45. Each gene was aligned separately with MUSCLE (Edgar, 2004), and
the alignments were then combined in a single matrix. This matrix was analyzed with **phangorn**
240 using a GTR+ Γ +I model of molecular evolution. A topological search using stochastic tree rear-
rangements (Nguyen et al., 2015) was performed and all parameters were estimated by maximizing
242 the likelihood. The final topology was used for molecular dating after dropping the outgroup. A
single calibration point was used assuming an age of the root (i.e., the most recent common ances-
244 tor of all living Felidae) between 16 Ma and 20 Ma (Benton and Donoghue, 2007; Werdelin et al.,
2010). Three molecular dating models were fitted: the strict clock model, the correlated relaxed
246 clock model, and the uncorrelated relaxed clock model. Their respective fit was assessed with
the penalized information criterion (Φ IC) which is computed with Φ IC = AIC + $\lambda\delta_{\{\Phi\}}$, where
248 AIC is the Akaike criterion information (AIC = $-2\ln L + 2k$, where k is the number of estimated
parameters) and $\delta_{\{\Phi\}}$ is the singular value of Φ (Paradis, 2013). For comparison, the data were
250 also analyzed with BEAST (Drummond et al., 2006), and with a model of rooted phylogenetic
inference (Felsenstein, 2004, p. 266); see Supplementary information.

252 **3. Results**

3.1. Simulation study

254 With data simulated from a strict clock model, the coverage of dates was mostly high and var-
ied around 95% (Fig. 3a). Longer sequences did not increase the performance unless the number
256 of sequences was large too. Overall, increasing the number of calibrations improved the coverage
of the CIs except for $n = 20$. For all values of n , s , and κ , the parametric bootstrap performed less
258 well than the two other methods, and there was a tendency for the semiparametric bootstrap to per-
form better than the nonparametric one. Increasing s improved the performance of the parametric
260 bootstrap. With data simulated with the correlated relaxed clock model, the performances were

less satisfying; however, in this case, they improved with more sequences and also with more calibration points (Fig. 3b). A similar result was obtained with data simulated from the uncorrelated relaxed clock model (Fig. 3c). For these two models, we observed the same difference between the three bootstrap methods noted with the strict clock model.

The observed widths of the 95% CIs were related to the previous results: the smaller the proportions covering the true value, the narrower the CIs (Fig. 4). However, there were some important exceptions to this general pattern. For data simulated from the strict clock model, the CI widths decreased when the number of sequences increased (Fig. 4a). Particularly, the parametric bootstrap showed narrower CIs when s was increased while the coverages were improved (Figs. 3a and 4a). On the other hand, this pattern was not observed for the two other models (Fig. 4b,c). For all models, the parametric bootstrap yielded narrower CIs than the two other methods. In some cases (e.g., strict clock model with $n = 20$), increasing the number of calibration points resulted in narrower CIs, but this was not a general result.

The running times varied greatly and were mostly explained by the model and the number of sequences (Fig. 5): the timings for the strict clock model varied between 0.1 sec and 1 min 24 sec whatever the other settings, and were less than 10 sec on average. The parametric bootstrap was the fastest of the three bootstrap methods. Increasing the number of calibration points also decreased running times. Clearly, the number of free parameters is an important factor in the variation of these running times. The semiparametric bootstrap was the slowest of the three methods: most of the analyses took between 20 min and 1 hr for the two relaxed clock models with $n = 100$.

When data were analyzed with a misspecified model, the observed patterns depended on the models: when data were simulated under a strict clock model and analyzed with a correlated relaxed clock model the CIs were narrower and with smaller coverage values than the reverse situation (Fig. 6). Increasing the number of calibration points yielded narrower CIs when the data were simulated from a correlated relaxed clock model, but not with the strict clock model (Fig. 6c,d).

3.2. Case study

The alignment obtained with MUSCLE had a total of 619,785 bp ($n' = 45$, $s = 13,773$) and included 50,011 alignment gaps which were very clustered: 833 sites (6% out of 13,773) had eight gaps or more. A majority of these gaps (29,808; 60%) were inserted when combining the 23

gene-level alignments as a consequence of missing sequences in some species. The phylogenetic
292 reconstruction gave a log-likelihood = $-49,269.78$. There was evidence for a substantial propor-
tion of invariant sites ($\hat{I} = 0.236$), and for inter-site variation (shape parameter of the Γ distribution:
294 $\hat{\alpha} = 0.909$). The phylogeny was overall consistent with the tree inferred by Johnson et al. (2006)
with only a small number of differences. Although there is room to improve the GTR+ Γ +I model
296 used here, for instance with partitioned models, we limited the present phylogenetic analysis since
it is aimed to illustrate the methods introduced in this paper.

298 The fitted models showed that the strict clock model was the best one ($\Phi IC = 77.6, 224, \text{ and } 227$, for the strict clock, autocorrelated relaxed clock, and uncorrelated relaxed clock models, re-
spectively). The estimated substitution rate was $\hat{r} = 4.27 \times 10^{-4}$ substitution.site⁻¹.Ma⁻¹. The
300 chronogram and its CIs inferred by the three above methods were very similar (Fig. 7). The 95%
CIs were slightly wider for the semiparametric method (mean: 2.53 Ma) than for the nonpara-
metric method (2.48 Ma), whereas the parametric bootstrap resulted in the narrowest intervals
302 (mean: 2.24 Ma). A similar pattern was observed for the 50% CIs (0.96 Ma, 0.88 Ma, 0.83 Ma,
respectively).
304

306 The analysis with BEAST gave results which are very close to those described in the previous
paragraph: the lower and upper bounds of the intervals inferred from both approaches were highly
308 correlated (Pearson correlation coefficient = 0.77 and 0.74, respectively) and the agreement was
particularly strong for the older nodes. There was no marked tendency for a method to produce
310 younger or older bounds compared to the other (see Supplementary information and Figs. ??-??
for details).

312 **4. Discussion**

It is still debated what is the most appropriate approach for molecular dating and assessment of
314 uncertainty around the estimated dates as illustrated by the various approaches and reviews cited
in this paper. Our simulation results give a mixed picture of the success and usefulness of our
316 proposed bootstrap methods: good statistical properties were observed for the simple case of a
strict molecular clock whereas the results were less positive in more complex settings, although
318 some factors had a positive impact on the quality of the CIs such as longer sequences as we discuss
in the following paragraphs.

320 Several points emerge from our simulation results. The number of calibration points was

an important factor, and increasing this number resulted in better CIs in almost all situations.
322 Several recent studies based on simulations reached a similar conclusion (???), with some studies
showing the greatest importance for the calibration points close to the root of the tree (Duchêne
324 et al., 2014; ?; ?). Furthermore, ? showed that “strategically placed” calibration points are likely
to compensate biases due to using simplistic assumptions in the substitution model. We decided
326 to always include the root node date as a calibration point in order to simplify the comparisons of
our simulation results.

328 Another result is the importance of the model of substitution rate evolution along the tree. Not
surprisingly, using a strict clock model to simulate and analyze the data gave good results with
330 respect to all measures we considered in this study. To et al. (2016) observed a similar result
with a least squares dating method (further discussed below). Surely, a critical feature in our
332 study is that estimation with this model implies the estimation of a single rate thus simplifying
the optimization of the likelihood function (this is confirmed by the short running times with the
334 strict clock model). The two other models require to calculate a penalty function which needs to
estimate a rate for each branch of the tree. Generally when one fits a statistical model, increasing
336 the number of parameters while keeping the same data is likely to increase the CIs of the parameter
estimates because less information (or degrees of freedom) is used to estimate each parameter.
338 Increasing the number of calibration points partially compensated for this increased imprecision,
while increasing the sequence length did not (a similar result that was also observed by Duchêne
340 et al., 2014).

Furthermore, it is likely that the parameter values used to simulate the data were influential on
342 our results. We selected these values in order to simulate trees that are substantially non-clock like
(see Supplementary information) which may lead to difficulties in estimating correctly the dates.
344 As an interesting comparison, Tao et al. (2020) obtained good coverage intervals when simulating
data with Kishino’s uncorrelated model; however, they simulated data using $\nu = 0.1$ while we
346 observed that this parameter value led to trees that are close to clock-like trees. Instead we used
 $\nu = 0.3$ in our simulations which resulted in much more non-ultrametric trees.

348 In agreement with our expectations, sequence length had an effect on our simulation results.
Generally in statistical analyses, sample size is expected to increase the precision of bootstrap
350 methods because larger samples lead to better estimates of data distributions (Hesterberg, 2015).
We observed that the longer sequences generally resulted in better coverage of the CIs, particularly

352 when a single calibration point was used. Sequence length had an effect on the widths of the CIs
but, as discussed above on the number of calibration points, this depended on several factors: with
354 the strict clock model CIs were narrower with longer sequences but this was not observed with
 $n = 100$ sequences. A similar pattern was found for the correlated relaxed clock model except with
356 the parametric bootstrap which always gave wider CIs when sequence length was increased. With
the uncorrelated relaxed clock model, the CIs were generally narrower with longer sequences.
358 Nevertheless, the improvement due to increased sequence length was visibly less important than
the improvement due to increased number of calibration points.

360 A result from our simulations that needs further attention is the fact that the correlated and
uncorrelated relaxed clock models gave narrower CIs than the strict clock model. This difference
362 tended to be less when increasing the number of calibration points, and also for the semiparamet-
ric bootstrap compared to the nonparametric one, particularly for the uncorrelated relaxed clock
364 model. We looked at some simulated trees and observed that with the uncorrelated relaxed clock
model there was a tendency to generate branch lengths with a very skewed distribution, so that
366 one or two branches were very long. When bootstrapping the data, these long branches were con-
sistently represented in the bootstrap trees thus leading to narrow CIs. Smoothing the distribution
368 of branch lengths partially compensated for this effect.

The running times of computing the CIs with $n = 20$ did not exceed fifteen minutes. These
370 times varied a lot and depended on the model: they were ten times longer for the correlated
relaxed clock model compared to the strict clock model, and twice longer for the uncorrelated
372 relaxed clock model compared to the correlated relaxed clock model. Interestingly, increasing the
sequence length resulted in shorter running times which is likely explained by the fact that longer
374 sequences lead to better estimates of branch lengths (\hat{l}_i 's) and, consequently, to easier estimation
of dates during the PL optimization (i.e., the PL function is “less flat” with more accurate \hat{l}_i 's).
376 Increasing the number of calibration points also decreased running times which may be explained
by the lesser number of dates to estimate.

378 The simulation results showed that the nonparametric and semiparametric methods gave very
close results, while the parametric method gave the narrowest CIs with both simulated and real
380 data. This is expected given that the last method generates pseudo-samples based only on the
stochasticity of the model of substitution and ignores the additional error due to sampling se-
382 quences of finite length. The two other methods capture both sources of error, resulting in slightly

wider and more accurate CIs with real data. With simulated data, the parametric bootstrap resulted
384 in CIs that did not cover well the real dates, so that these narrow CIs were not well paid-off. On
the other hand, we note that the simulations with the strict clock model showed that the para-
386 metric bootstrap yielded narrower and more precise CIs when the sequence length was increased,
therefore suggesting that this method behaves correctly at least in some situations (e.g., long se-
388 quences). Clearly, a more extensive study is required to investigate how the present approach
behaves in a wider variety of situations.

390 The application with the Felidae data gave results broadly consistent with previous works,
although differing on several points. Most estimated dates are a bit more ancient than those previ-
392 ously published although the fossil record of Felidae is still poorly known (Werdelin et al., 2010).
For instance, Tseng et al. (2014) described a skull closely related to *Panthera* that they dated back
394 to 7 Ma, whereas it was previously estimated that the big cats belonging to this genus diverged
about 4 Ma (Werdelin et al., 2010). With all three methods, the CIs were wider for the older nodes
396 than with the recent ones. Further analyses not shown here found that much narrower CIs were
obtained if the age of the root is fixed, especially for the Pantherinae (*Panthera* + *Neofelis*), al-
398 though the dates were not themselves substantially modified. Concerning the BEAST analysis, it
would be interesting to explore the impact of tree priors on the estimated dates and could explain
400 the observed discrepancies with the PL analysis, especially for the shallower nodes.

It is not clear from the literature how similar Bayesian posterior distributions and frequentist
402 CIs are (or should be) perceived by phylogeneticists. Some studies have focused on point esti-
mates of dates and showed that chronograms reconstructed by Bayesian and likelihood methods
404 with real data from these two approaches were very close (Lorén et al., 2014). Furthermore, ?
showed that the dates inferred with RelTime and with Bayesian inference are generally in very
406 good agreement, although they apparently did not compare confidence and credibility intervals.
Conceptually, intervals from Bayesian and from frequentist methods should not be confused. First,
408 Bayesian intervals depend on both the prior distributions of the parameters and the data, whereas
CIs depend, for a given fitted model, only on the data. Second, these intervals are obtained by
410 MCMC integration for Bayesian methods, while model fitting optimization is used for frequentist
methods. And third, Bayesian methods usually jointly estimate the tree and the dates (although the
412 cladogram topology can be fixed) whereas frequentist methods compute CIs on a fixed cladogram.
In this respect, it is interesting to note the recent contribution by Bickel (2022) who proposed to

414 include clade uncertainty into the frequentist CIs of estimated dates. A simple practical approach
could be to report together (e.g., in a table) the CIs of the dates with the support values of the
416 nodes.

An attractive feature of the approach presented in this paper is that it can be run in reason-
418 able times: the complete analysis with the Felidae data took less than one hour, which includes
sequence alignment, phylogeny inference with *phangorn*, and molecular dating with calculations
420 of CIs with the three bootstrap methods. The running times may vary a lot, however, depending
on the data size and the models considered. The least-squares dating (LSD) method is another
422 very fast dating method for very large phylogenies ($n > 1000$) with dated tips (To et al., 2016).
This method assumes that the branch lengths of the phylogram follow a normal distribution so that
424 least squares can be used to estimate the dates of the nodes. To et al. (2016) found their method
to be robust to violation of the strict molecular clock as long as the rates are uncorrelated. This is
426 likely a consequence of the normality assumption: if the rates are correlated then the errors due to
the strict clock assumption would accumulate along the paths from root to tips resulting in biased
428 estimates. Besides, it is likely that the LSD method does not perform well with small phylogenies
because the errors would not cancel out on the paths from root to tips if these paths are too short
430 (even in the case of uncorrelated rates). Furthermore, it will be interesting to extend the present
PL framework to the situation of dated (i.e., non-contemporaneous) tips which would require to
432 handle the date constraints in different a way than done in the present study. This clearly requires
further study.

434 The simulation results presented above demonstrate that the CIs inferred by the present ap-
proach have generally satisfying statistical properties. Clearly, more simulations and works are
436 needed to investigate these properties. We suggest that more attention needs to be paid on simu-
lation settings in order to improve our understanding of the present methods and to compare them
438 with previous contributions and other simulation studies from the literature. The present study left
aside the issue of bias: in theory, bootstrap replicates make possible to estimate the expected value
440 of the estimator biases; however, bias estimates can have high variability (Hesterberg, 2015), or
require computationally intensive methods such as double bootstrapping, although approximate
442 methods exist (DiCiccio and Efron, 1996; Burbrink and Pyron, 2008).

CRedit authorship contribution statement

444 **Emmanuel Paradis:** Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing – Original Draft, Writing – Review & Editing, Visualization. **Santiago**
446 **Claramunt:** Conceptualization, Methodology, Software, Writing – Review & Editing. **Joseph**
Brown: Methodology, Software, Validation, Resources, Writing – Review & Editing. **Klaus**
448 **Schliep:** Methodology, Software, Writing – Review & Editing.

Declaration of Competing Interest

450 The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful to three anonymous reviewers and the Associate Editor for their constructive
454 comments on a previous version of this article. The calculations used for data analyzes benefited from the ISEM computing cluster platform and the Occigen Supercomputer under the Grand
456 Équipement National de Calcul Intensif (GENCI). This is publication ISEM 2022-255.

Funding

458 SC acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery Grant RGPIN-2018-06747.

Data accessibility

All data analyzed in this paper are deposited in GenBank. Additional code is available on
462 <https://github.com/josephwb/chronos> included for the simulations and the case study (folder `examples/`).

Supplementary information

Supplementary figures ??–?? are available online.

466 **References**

- Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword
468 alliance (Compositae). *Proc. Natl. Acad. Sci. USA* 95, 9402–9406. doi:10.1073/pnas.95.
16.9402.
- 470 Benton, M.J., Donoghue, P.C.J., 2007. Paleontological evidence to date the tree of life. *Mol. Biol.*
Evol. 24, 26–53. doi:10.1093/molbev/msl150. erratum vol. 24, pp. 889–891.
- 472 Bickel, D.R., 2022. Propagating clade and model uncertainty to confidence intervals of divergence
times and branch lengths. *Mol. Phyl. Evol.* 167. doi:10.1016/j.ympev.2021.107357.
- 474 Britton, T., Oxelman, B., Vinnersten, A., Bremer, K., 2002. Phylogenetic dating with con-
fidence intervals using mean path lengths. *Mol. Phyl. Evol.* 24, 58–65. doi:10.1016/
476 S1055-7903(02)00268-3.
- Bromham, L., 2019. Six impossible things before breakfast: assumptions, models, and belief in
478 molecular dating. *Trends Ecol. Evol.* 34, 474–486. doi:10.1016/j.tree.2019.01.017.
- Bromham, L., Duchêne, S., Hua, X., Ritchie, A.M., Duchêne, D.A., Ho, S.Y.W., 2018. Bayesian
480 molecular dating: opening up the black box. *Biol. Rev.* 93, 1165–1191. doi:10.1111/brv.
12390.
- 482 Burbrink, F.T., Pyron, R.A., 2008. The taming of the skew: estimating proper confidence intervals
for divergence dates. *Syst. Biol.* 57, 317–328. doi:10.1080/10635150802040605.
- 484 Casella, G., Berger, R.L., 2002. *Statistical inference* (second edition). Duxbury, Pacific Grove,
CA.
- 486 Czech, L., Huerta-Cepas, J., Stamatakis, A., 2017. A critical review on the use of support values
in tree viewers and bioinformatics toolkits. *Mol. Biol. Evol.* 34, 1535–1542. doi:10.1093/
488 molbev/msx055.
- De Angelis, D., Young, G.A., 1992. Smoothing the bootstrap. *Int. Statist. Rev.* 60, 45–56. doi:10.
490 2307/1403500.
- DiCiccio, T.J., Efron, B., 1996. Bootstrap confidence intervals (with discussion). *Statist. Sci.* 11,
492 189–212. doi:10.1214/ss/1032280214.
- Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and
494 dating with confidence. *PLoS Biol.* 4, 699–710. doi:10.1371/journal.pbio.0040088.
- Duchêne, S., Lanfear, R., Ho, S.Y.W., 2014. The impact of calibration and clock-model choice

496 on molecular estimates of divergence times. *Mol. Phyl. Evol.* 78, 277–289. doi:10.1016/j.
ympev.2014.05.032.

498 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high through-
put. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.

500 Efron, B., 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other
methods. *Biometrika* 68, 589–599. doi:10.1093/biomet/68.3.589.

502 Efron, B., Halloran, E., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees.
Proc. Natl. Acad. Sci. USA 93, 13429–13434. doi:10.1073/pnas.93.23.13429.

504 Efron, B., Tibshirani, R.J., 1993. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca
Raton, FL.

506 Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach.
J. Mol. Evol. 17, 368–376. doi:10.1007/BF01734359.

508 Felsenstein, J., 1984. Distance methods for inferring phylogenies: a justification. *Evolution* 38,
16–24. doi:10.2307/2408542.

510 Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolu-
tion* 39, 783–791. doi:10.2307/2408678.

512 Felsenstein, J., 1987. Estimation of hominoid phylogeny from a DNA hybridization data set. *J.
Mol. Evol.* 26, 123–131. doi:10.1007/BF02111286.

514 Felsenstein, J., 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA.

Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc.*
516 *Lond. A* 222, 309–368. doi:10.1098/rsta.1922.0009.

Guindon, S., 2020. Rates and rocks: strengths and weaknesses of molecular dating methods.
518 *Frontiers Genet.* 11, 526. doi:10.3389/fgene.2020.00526.

Harvey, M.G., Bravo, G.A., Claramunt, S., Cuervo, A.M., Derryberry, G.E., Battilana, J., See-
520 holzer, G.F., McKay, J.S., O’Meara, B.C., Faircloth, B.C., Edwards, S.V., Pérez-Emán, J.,
Moyle, R.G., Sheldon, F.H., Aleixo, A., Smith, B.T., Chesser, R.T., Silveira, L.F., Cracraft,
522 J., Brumfield, R.T., Derryberry, E.P., 2020. The evolution of a tropical biodiversity hotspot.
Science 370, 1343–1348. doi:10.1126/science.aaz6970.

524 Hesterberg, T.C., 2015. What teachers should know about the bootstrap: resampling in the un-
dergraduate statistics curriculum. *Am. Stat.* 69, 371–386. doi:10.1080/00031305.2015.
526 1089789.

- Ho, S.Y.W., Duchêne, S., Duchêne, D., 2015. Simulating and detecting autocorrelation of
528 molecular evolutionary rates among lineages. *Mol. Ecol. Res.* 15, 688–696. doi:10.1111/
1755-0998.12320.
- 530 Ho, S.Y.W., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G., Cooper, A.,
2011. Time-dependent rates of molecular evolution. *Mol. Ecol.* 20, 3087–3101. doi:10.
532 1111/j.1365-294X.2011.05178.x.
- Johnson, W.E., Eizirik, E., Pecon-Slattery, J., Murphy, W.J., Antunes, A., Teeling, E., O'Brien,
534 S.J., 2006. The Late Miocene radiation of modern Felidae: a genetic assessment. *Science* 311,
73–77. doi:10.1126/science.1122277.
- 536 Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules, in: Munro, H.N. (Ed.),
Mammalian Protein Metabolism. Academic Press, New York, pp. 21–132. doi:10.1016/
538 B978-1-4832-3211-9.50009-7.
- Lepage, T., Bryant, D., Philippe, H., Lartillot, N., 2007. A general comparison of relaxed molec-
540 ular clock models. *Mol. Biol. Evol.* 24, 2669–2680. doi:10.1093/molbev/msm193.
- Lorén, J.G., Farfán, M., Fusté, M.C., 2014. Molecular phylogenetics and temporal diversification
542 in the genus *Aeromonas* based on the sequences of five housekeeping genes. *PLoS ONE* 9,
e88805. doi:10.1371/journal.pone.0088805.
- 544 Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective
stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–
546 274. doi:10.1093/molbev/msu300.
- Paradis, E., 2013. Molecular dating of phylogenies by likelihood methods: a comparison of
548 models and a new information criterion. *Mol. Phyl. Evol.* 67, 436–444. doi:10.1016/j.
ympev.2013.02.008.
- 550 Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
analyses in R. *Bioinformatics* 35, 526–528. doi:10.1093/bioinformatics/bty633.
- 552 Parsons, C., Stüeken, E.E., Rosen, C.J., Mateos, K., Anderson, R.E., 2021. Radiation of nitrogen-
metabolizing enzymes across the tree of life tracks environmental transitions in Earth history.
554 *Geobiology* 19, 18–34. doi:10.1111/gbi.12419.
- R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation
556 for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org>.
- Rutschmann, F., 2006. Molecular dating of phylogenetic trees: A brief review of current methods

558 that estimate divergence times. *Divers. Distrib.* 12, 35–48. doi:10.1111/j.1366-9516.
2006.00210.x.

560 Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence
times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109. doi:10.1093/
562 oxfordjournals.molbev.a003974.

Sanderson, M.J., 2003. r8s: inferring absolute rates of molecular evolution and divergence
564 times in the absence of a molecular clock. *Bioinformatics* 19, 301–302. doi:10.1093/
bioinformatics/19.2.301.

566 Sauquet, H., 2013. A practical guide to molecular dating. *Comptes Rendus Palevol* 12, 355–367.
doi:10.1016/j.crpv.2013.07.003.

568 Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592–593. doi:10.
1093/bioinformatics/btq706.

570 Silverman, B.W., 1986. *Density Estimation*. Chapman and Hall, London.

Silverman, B.W., Young, G.A., 1987. The bootstrap: to smooth or not to smooth? *Biometrika* 74,
572 469–479. doi:10.1093/biomet/74.3.469.

Stadler, T., 2011. Simulating trees with a fixed number of extant species. *Syst. Biol.* 60, 676–684.
574 doi:10.1093/sysbio/syr029.

Tamura, K., Battistuzzi, F.U., Billing-Ross, P., Murillo, O., Filipowski, A., Kumar, S., 2012. Es-
576 timating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci. USA* 109,
19333–19338. doi:10.1073/pnas.1213199109.

578 Tao, Q.Q., Tamura, K., Mello, B., Kumar, S., 2020. Reliable confidence intervals for RelTime
estimates of evolutionary divergence times. *Mol. Biol. Evol.* 37, 280–290. doi:10.1093/
580 molbev/msz236.

To, T.H., Jung, M., Lycett, S., Gascuel, O., 2016. Fast dating using least-squares criteria and
582 algorithms. *Syst. Biol.* 65, 82–97. doi:10.1093/sysbio/syv068.

Tseng, Z.J., Wang, X., Slater, G.J., Takeuchi, G.T., Li, Q., Liu, J., Xie, G., 2014. Himalayan
584 fossils of the oldest known pantherine establish ancient origin of big cats. *Proc. R. Soc. Lond.*
B 281, 20132686. doi:10.1098/rspb.2013.2686.

586 Wand, M.P., Jones, M.C., 1995. *Kernel smoothing*. Chapman & Hall/CRC, Boca Raton, USA.

Werdelin, L., Yamaguchi, N., Johnson, W.E., O'Brien, S.J., 2010. Phylogeny and evolution of cats
588 (Felidae), in: Macdonald, D.W., Loveridge, A.J. (Eds.), *The Biology and Conservation of Wild*

Felids. Oxford University Press, Oxford, pp. 59–82.

590 Xia, X., Yang, Q., 2011. A distance-based least-square method for dating speciation events. *Mol. Phyl. Evol.* 59, 342–353. doi:10.1016/j.ympev.2011.01.017.

592 Zhu, T.Q., Dos Reis, M., Yang, Z.H., 2015. Characterization of the uncertainty of divergence time estimation under relaxed molecular clock models using multiple loci. *Syst. Biol.* 64, 267–280. doi:10.1093/sysbio/syu109.

Zuckerkandl, E., 1987. On the molecular evolutionary clock. *J. Mol. Evol.* 26, 34–46. doi:10.1007/BF02111280.

596 Zuckerkandl, E., Pauling, L., 1962. Molecular disease, evolution, and genic heterogeneity, in: Kasha, M., Pullman, B.. (Eds.), *Horizons in Biochemistry*. New York, pp. 189–225.

Zuckerkandl, E., Pauling, L., 1965a. Evolutionary divergence and convergence in proteins, in: 600 Bryson, V., Vogel, H.J. (Eds.), *Evolving Genes and Proteins*. New York, pp. 97–165.

Zuckerkandl, E., Pauling, L., 1965b. Molecules as documents of evolutionary history. *J. Theor. Biol.* 8, 357–366. doi:10.1016/0022-5193(65)90083-4.

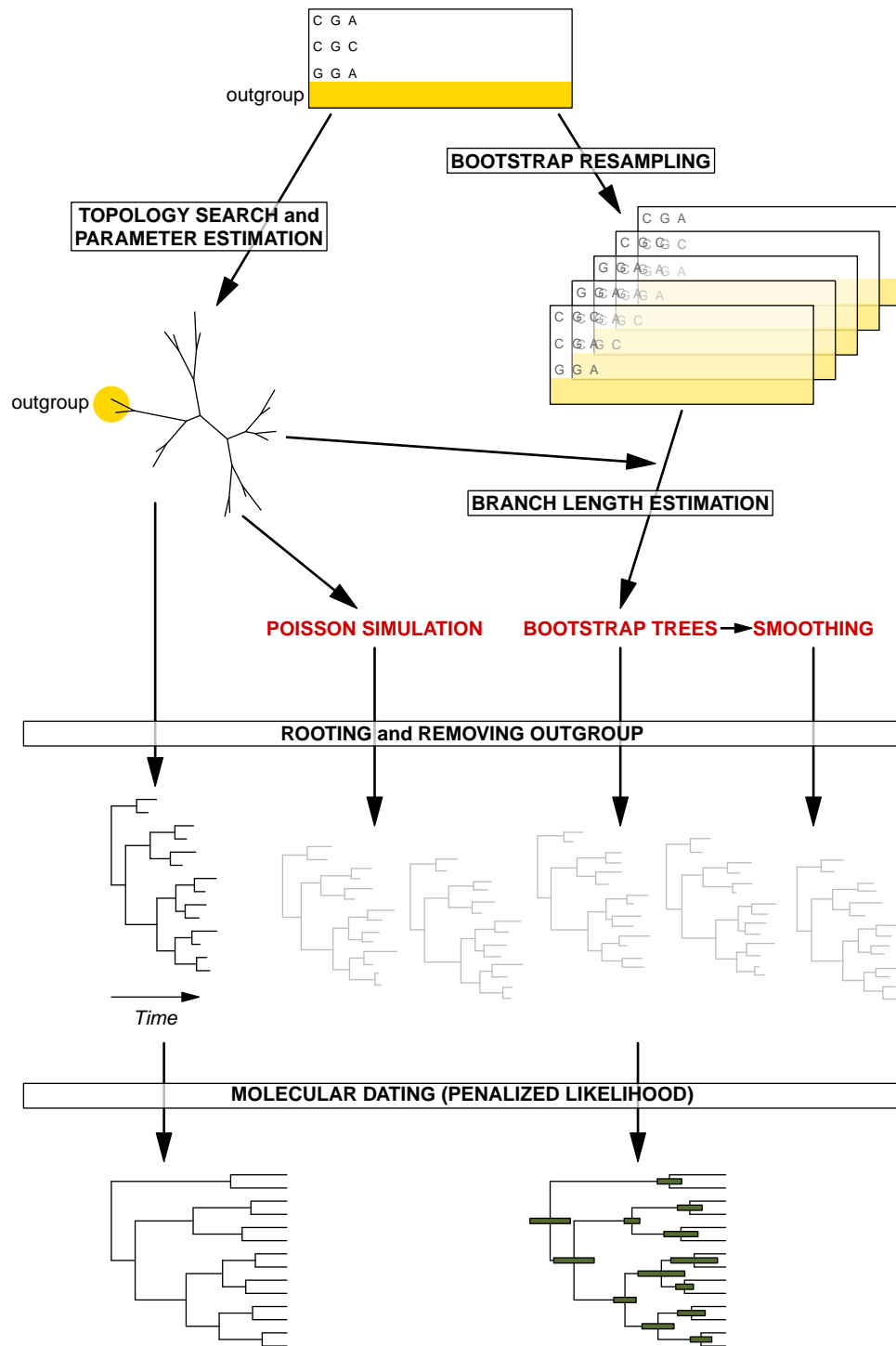


Figure 1: The framework described in this paper. The leftmost workflow is the standard procedure of phylogenetic inference and downstream molecular dating. The three approaches shown in red are described in the text: nonparametric bootstrap (bootstrap trees), semiparametric bootstrap (smoothing), and parametric bootstrap (Poisson simulation).

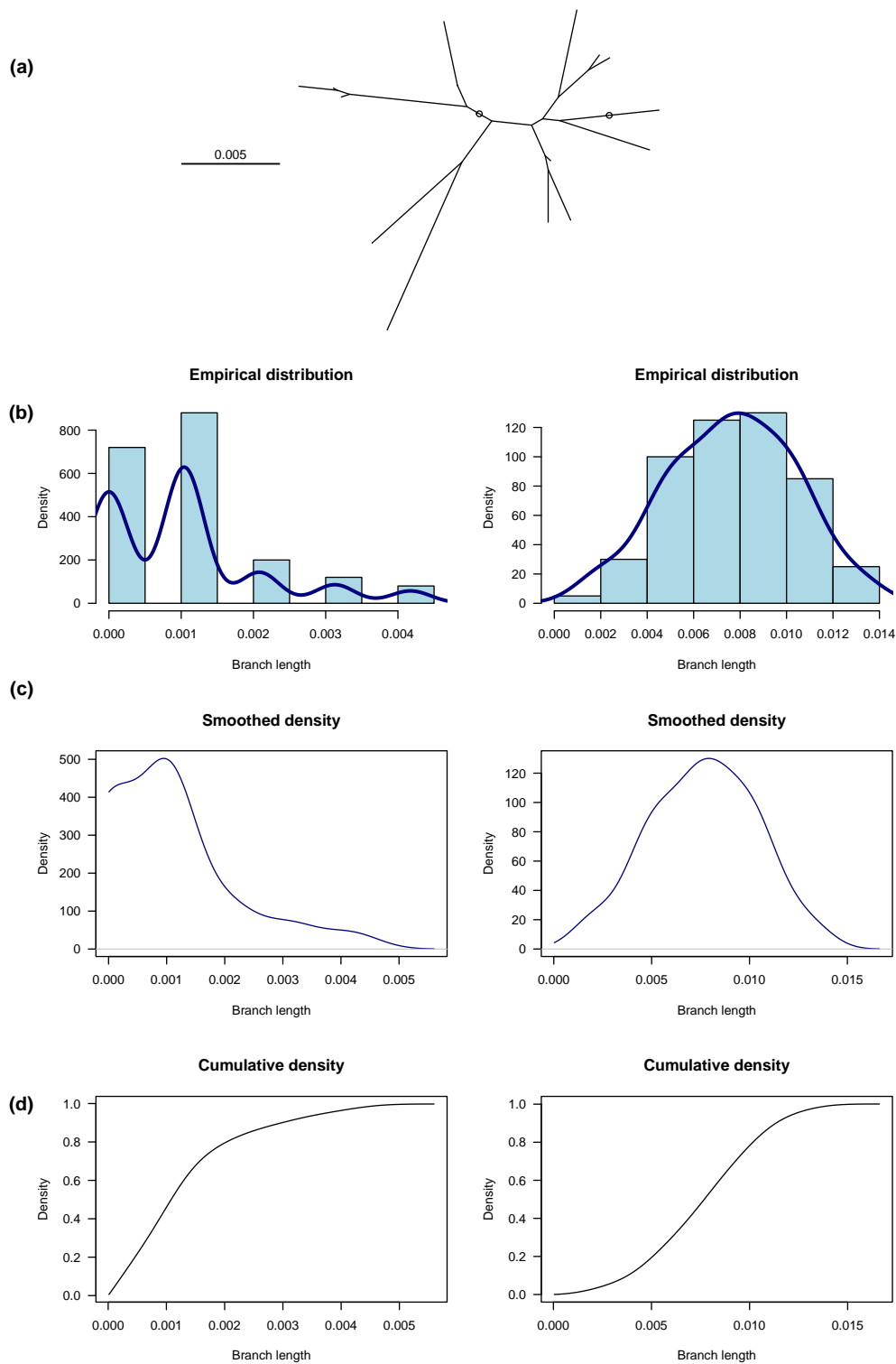


Figure 2: (a) An unrooted tree with two branches shown with circles and analyzed after bootstrapping on the left and right panels, respectively. (b) The raw distribution of branch lengths (histogram) with a local Gaussian density (curve). (c) The smoothed density estimated by the procedure described in the text. (d) The final cumulative density used to generate random samples of branch lengths.

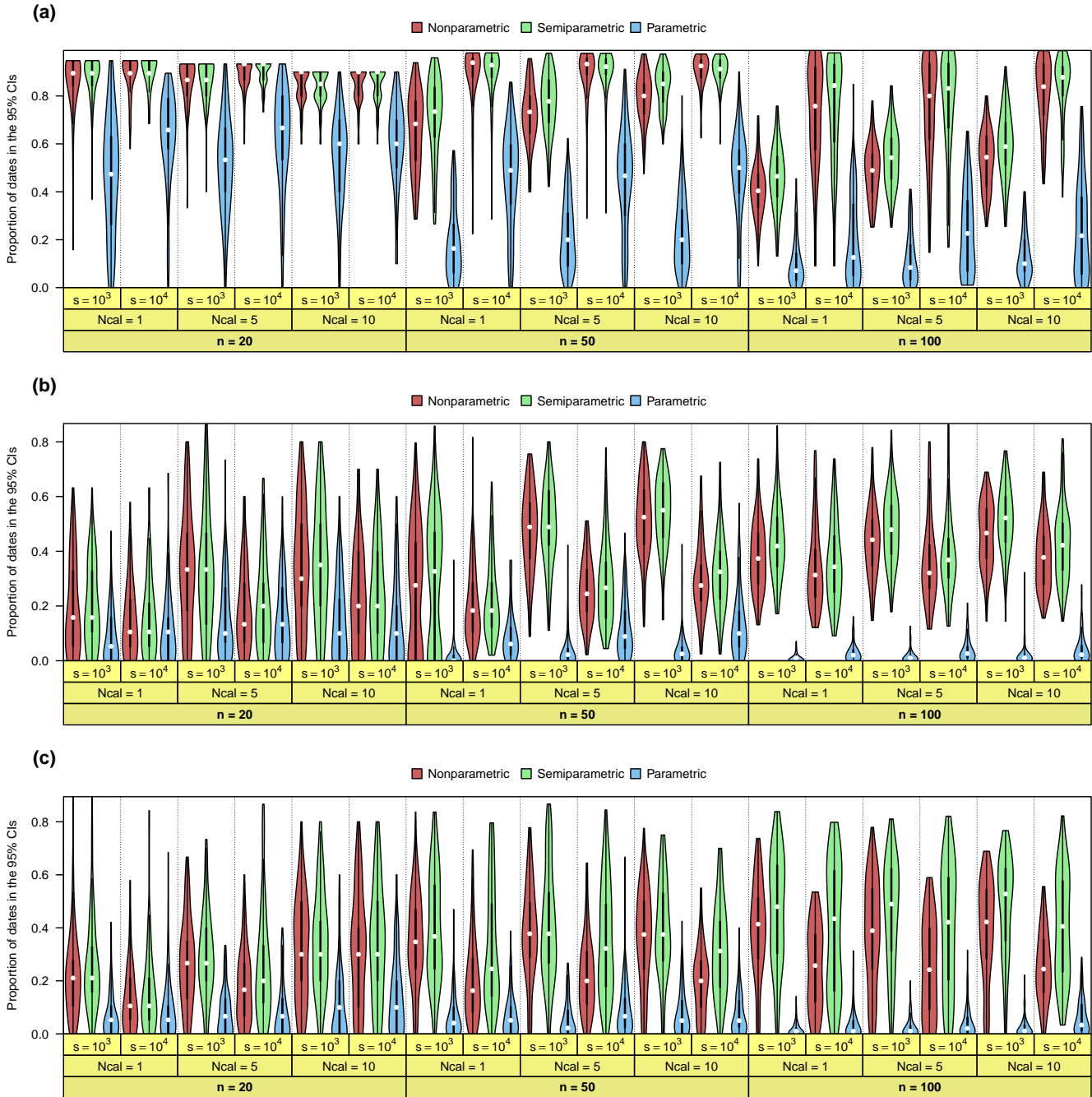


Figure 3: Proportion of true dates within the inferred 95% confidence intervals. (a) Strict clock model. (b) Correlated relaxed clock model. (c) Uncorrelated relaxed clock model. See main text for the implementations of these models during simulations and for model fitting. s : sequence length; $Ncal$: number of calibration points; n : number of sequences.

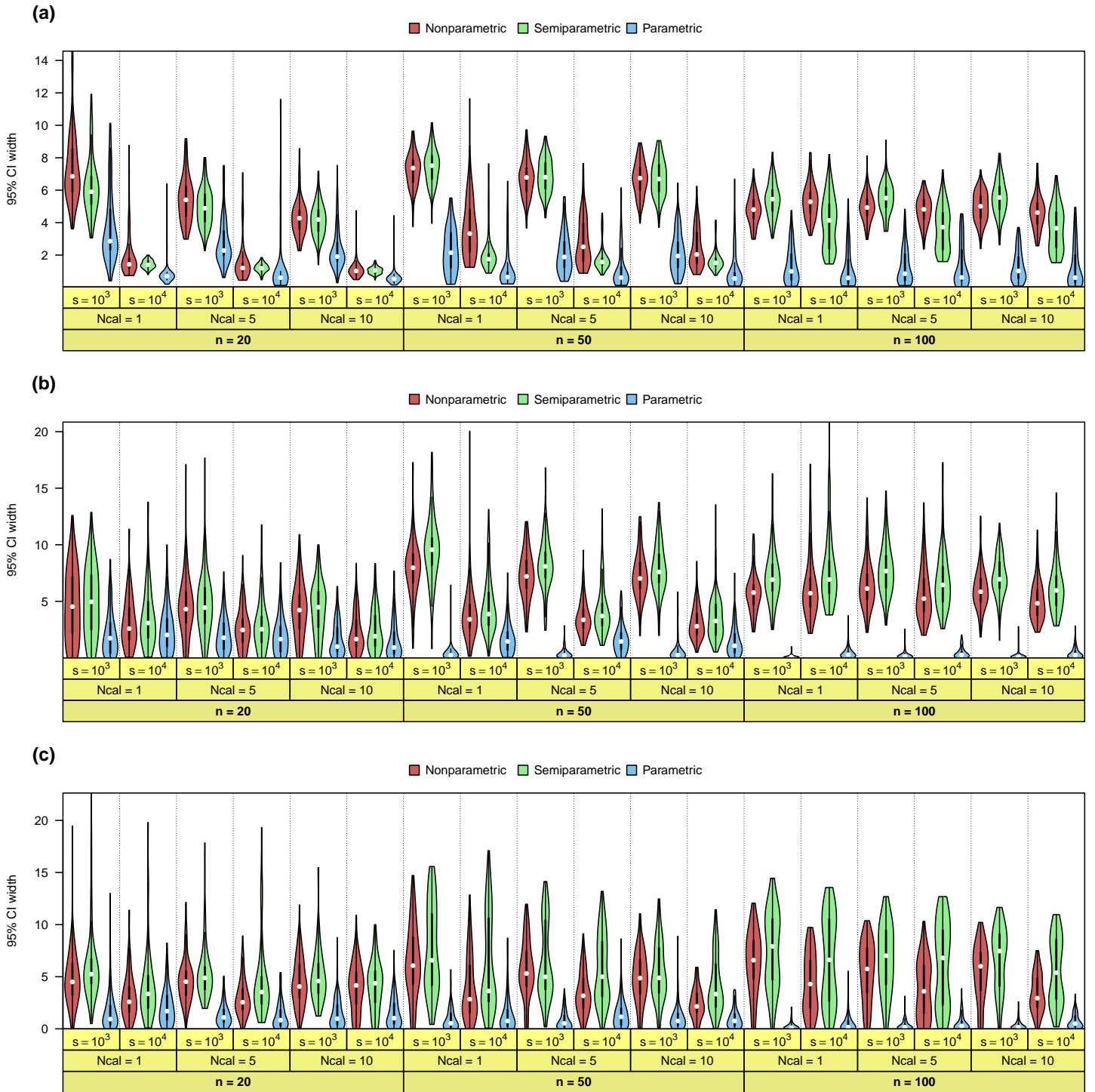


Figure 4: Mean width of the inferred 95% confidence intervals. (a) Strict clock model. (b) Correlated relaxed clock model. (c) Uncorrelated relaxed clock model. See main text for the implementations of these models during simulations and for model fitting. s : sequence length; N_{cal} : number of calibration points; n : number of sequences.

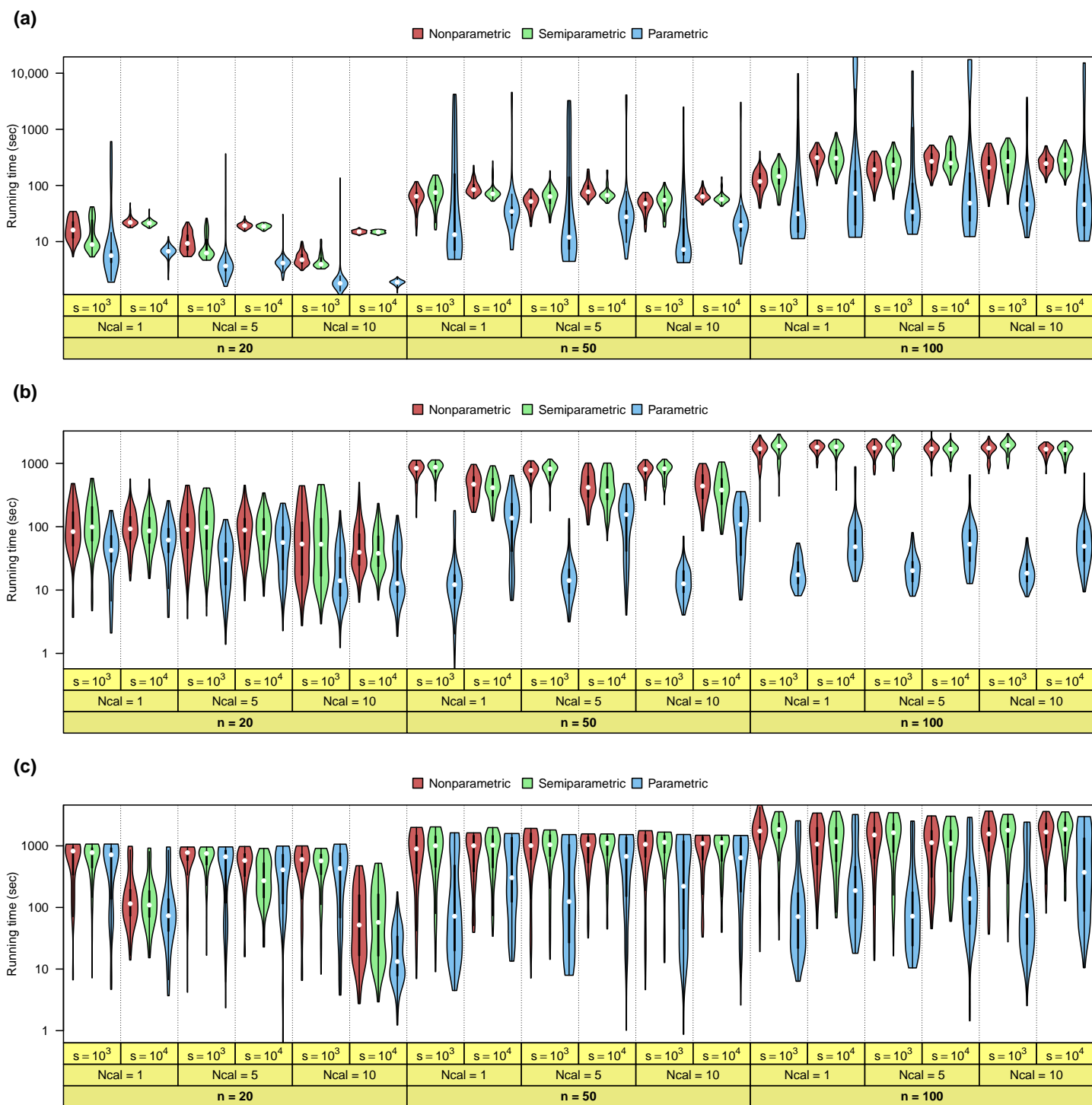


Figure 5: Running times. (a) Strict clock model. (b) Correlated relaxed clock model. (c) Uncorrelated relaxed clock model. See main text for the implementations of these models during simulations and for model fitting. s : sequence length; N_{cal} : number of calibration points; n : number of sequences.

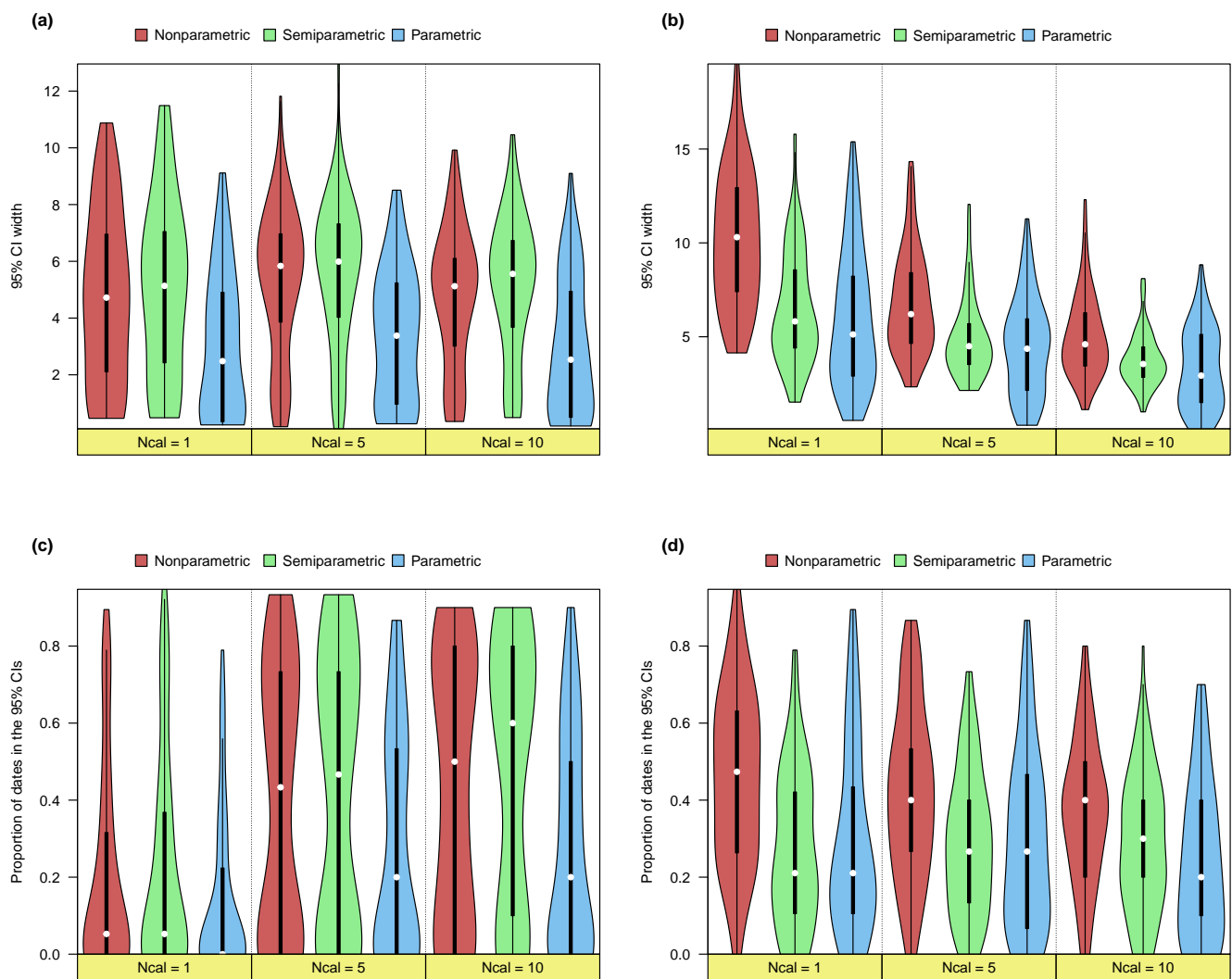


Figure 6: Simulation with misspecified models. (a) and (c) Data simulated with a strict clock model and analyzed with a correlated relaxed clock model. (b) and (d) Data simulated with a correlated relaxed clock model and analyzed with a strict clock model. Ncal: number of calibration points.

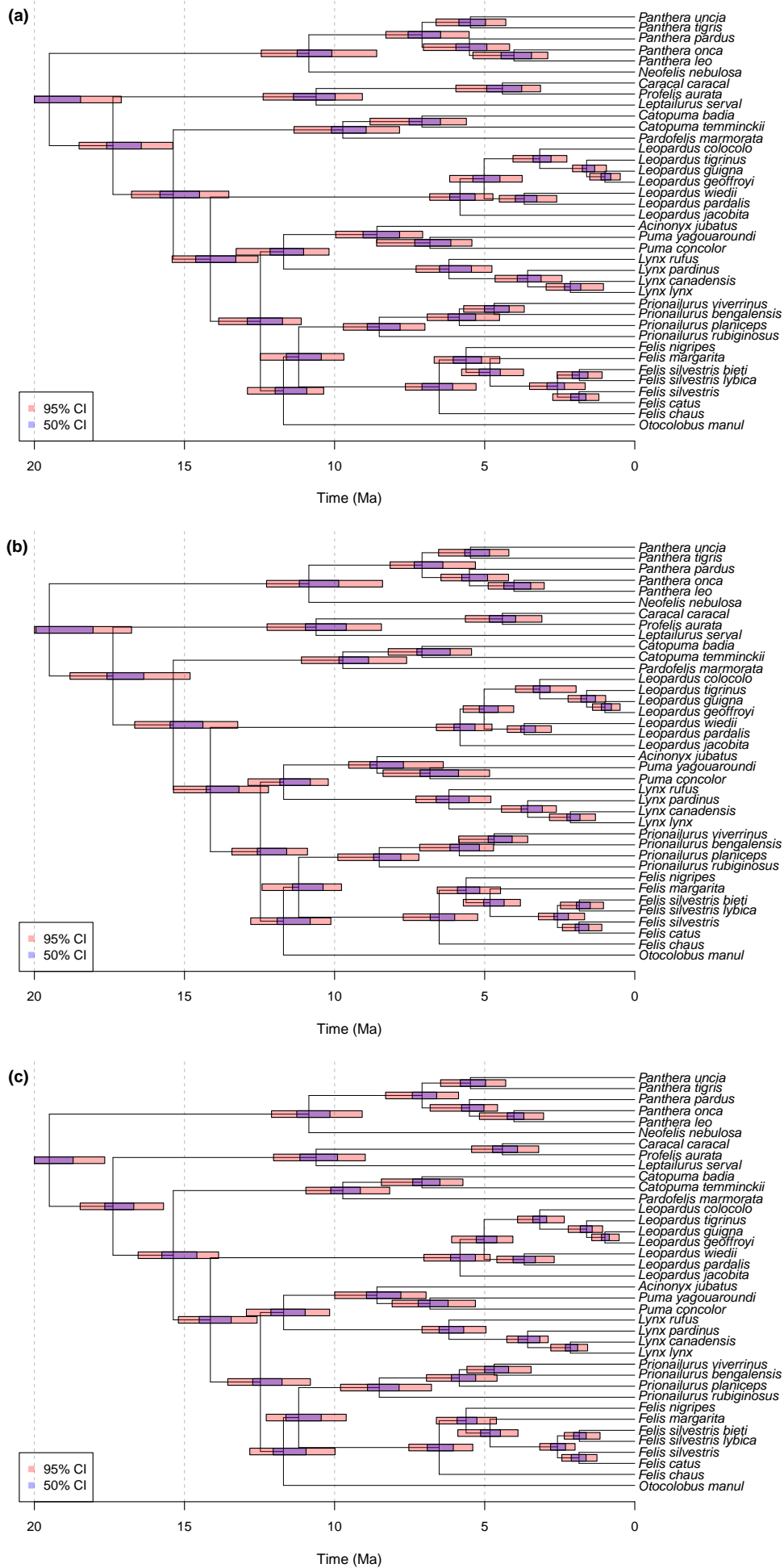


Figure 7: Chronogram with 50% and 95% confidence intervals of the Felidae. (a) Semiparametric, (b) nonparametric, and (c) parametric bootstrap.