



HAL
open science

Choosing the best ancestral character state reconstruction method.

Manuela Royer-Carenzi, Pierre Pontarotti, Gilles Didier

► **To cite this version:**

Manuela Royer-Carenzi, Pierre Pontarotti, Gilles Didier. Choosing the best ancestral character state reconstruction method.. *Mathematical Biosciences*, 2013, 242 (1), pp.95-109. 10.1016/j.mbs.2012.12.003 . hal-00830797

HAL Id: hal-00830797

<https://hal.science/hal-00830797v1>

Submitted on 17 Feb 2022

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Choosing the Best Ancestral Character State Reconstruction Method

Manuela Royer-Carenzi¹
Pierre Pontarotti¹
Gilles Didier²

¹ LATP, UMR CNRS 7352 FR 3098 IFR 48, Evolution Biologique et Modélisation,
Aix-Marseille University, 13331 Marseille Cedex 3 - FRANCE

² IML, UMR CNRS 6206, Aix-Marseille University,
13288 Marseille Cedex 9 - FRANCE

February 16, 2022

Abstract

Despite its intrinsic difficulty, ancestral character states reconstruction is an essential tool for testing evolutionary hypothesis. Two major classes of approaches of this question can be distinguished, depending upon they are parsimony- or likelihood-based. We focus here on the second class of methods, more specifically on approaches based on continuous-time Markov modeling of character evolution. Among them, we consider the most likely ancestor reconstruction, the posterior probability reconstruction, the likelihood ratio method, and the Bayesian approach.

We discuss and compare the above-mentioned methods over several phylogenetic trees, adding the most parsimonious reconstruction performance in confrontation. Under the assumption that the character evolves following a continuous-time Markov process, we compute and compare the expectations of success of each method for a broad range of evolution parameter values. Moreover, we show how the knowledge of the evolution model parameters yields to compute upper bounds of reconstruction performances, which are provided as references.

The results of all the reconstruction methods are quite close one to another and the expectations of success are not so far from their theoretical upper bounds. But the performances ranking heavily depends on the topology of the studied tree, on the ancestral node to infer as well as on the parameter values. Consequently, we propose a protocol providing for each parameters value the best method in terms of expectation of success, with regard to the phylogenetic tree and the ancestral node to infer.

Keywords: Maximum likelihood; Ancestral state reconstruction; Continuous Markov process, Comparative methods

1 Introduction

Ancestral reconstruction may concern either sequences (Yang et al., 1995; Koshi and Goldstein, 1996; Zhang and Nei, 1997; Krishnan et al., 2004) or spe-

cific ecological, phenotypic, or biogeographic traits, called characters (Felsenstein, 1985). Even though this question suffers from theoretical inherent limitations (Mossel, 2003), the reconstruction of ancestral states is hard to avoid when one wants to test evolutionary theories about selection process (Messier and Stewart, 1997; Bishop et al., 2000), neutral evolution (Langlais and Fitch, 1974; Templeton, 1996), homoplasy (Ree and Donoghue, 1998), functional divergence *etc.* The trait to be reconstructed can be more or less complex, discrete or continuous. In the present work, we consider the most basic case: binary characters, that is where a trait takes only two different values, typically the presence or the absence of a given feature. We focus on the reconstruction problem, which can be introduced as follows. The whole phylogenetic history of a set of organisms is assumed known and is represented as a tree with specified branch lengths. The character states of contemporary organisms (tips of the tree) are also given. The reconstruction process, delineated in Figure 1, aims to assign, in the most relevant way, the character states of the ancestor organisms (the internal nodes of the tree).

Different points of view can be used to evaluate the relevance of a particular ancestral reconstruction. Among them, one can distinguish two major classes. In the first one, the idea is to explain the contemporary states in the simplest way, that is generally with the fewer number of changes of state between an ancestor and its child. Such approaches are said parsimonious (Fitch, 1971; Swoford and Maddison, 1992; Collins et al., 1994; Maddison and Maddison, 2009) and are still widely used to reconstruct ancestral states. The other main point of view is to model the character evolution as a stochastic process, generally as a continuous time Markov model, and to consider the likelihoods of the possible ancestral character states computed from this model, in order to perform the reconstruction (Felsenstein, 1981; Koshi and Goldstein, 1996; Schluter et al., 1996; Pagel, 1999; Huelsenbeck and Ronquist, 2001; Nielsen, 2002). One of the main advantages of likelihood approaches is that they yield to take into account divergence times (branch lengths) while parsimonious methods consider each evolution step in the same way. Another concern of parsimonious approaches is that they are based either on parameters (generalized parsimony) or on strong assumptions, often controversial, like irreversibility for Dollo parsimony. The choice of parsimony parameters remains difficult (Ree and Donoghue, 1998; Didier, 2011), while maximum likelihood estimation or Bayesian approaches are natural answers to the same question for stochastic methods. Several authors discuss the advantage of likelihood approaches with regard to parsimonious ones (Schluter et al., 1996; Mooers and Schluter, 1999; Pagel, 1999; Nielsen, 2002; Huelsenbeck et al., 2003). We argue that the points of view originating these two classes of approaches differ in such an amount (time influence !) that the relevance of one or another can only be tested with regard to real biological data sets. For instance, Clark et al. (2008) study Pacific *Cyrtandra* and conclude that in this particular case, likelihood-based methods offer more congruent results than parsimonious ones.

In this study, we focus on several likelihood-based reconstruction methods which are all based on the continuous-time Markov modeling of character evolution. The *most likely ancestor reconstruction* returns the states corresponding to the ancestral history maximizing the total likelihood of the phylogenetic tree (Yang et al., 1995). The *posterior probability reconstruction* as described in Yang et al. (1995) computes for each ancestral organisms the posterior proba-

bilities of all possible states and associates the state having the greatest posterior probability to this organism. The *likelihood ratio reconstruction*, introduced in Pagel (1999), computes the likelihoods associated to each possible state of an ancestral node to determine the most likely. Finally the *Bayesian approach* considers weighted sums of posterior probabilities computed from the possible model parameters values, with weights given by prior distributions (Huelsenbeck and Ronquist, 2001). Here we rather use an empirical Bayesian method since some of the prior distributions parameters are estimated from the data (Maritz and Lwin, 1989; Pagel et al., 2004). These likelihood-based methods are confronted to the most parsimonious reconstruction with equal gain/loss costs (Fitch parsimony criterion (Fitch, 1971)).

In Maddison (1995), author proposes a method to compute the probability of correct reconstruction by parsimony, under the assumption that the evolution of the binary character follows a discrete-time Markov process. We apply the same general scheme to likelihood-based reconstruction methods, but here under the assumption that character evolution is driven by a continuous-time Markov chain, which turns out to be the same assumption underlying the likelihood approaches of reconstruction. It allows us to compute the expectations of success of each method and thus effectively compare their respective performances.

The rest of the paper is organized as follows. In the first Section, we formally present the ancestral reconstruction problem and the evolutionary models used in likelihood approaches. We briefly introduce the likelihood-based approaches of character state reconstruction above-mentioned, detailing in particular the strong relation between the likelihood ratio and the posterior probability reconstruction. In the second section, we start by showing how to compute the expectation of success of any reconstruction method, either likelihood-based or not. We also provide upper bounds of these reconstruction expectations of success. Next, we discuss the general influence of the parameters driving the evolution process on the reconstruction performances and compare the expectations of success obtained by the methods. Finally we propose a practical protocol designed to select the most relevant reconstruction method with regard to a given phylogenetic tree, with known contemporary states. The protocol is illustrated over a biological dataset from Webster and Purvis (2002).

2 Materials and methods

2.1 Formal presentation of the ancestral reconstruction problem

The notation \mathcal{T} designs both a phylogenetic tree (including branch length information) and its set of nodes. The (complete) evolutionary history of a binary character over \mathcal{T} associates to each node a state belonging to $\{0, 1\}$, says 1 for present and 0 for absent. Formally, it is an element \mathbf{x} of $\{0, 1\}^{\mathcal{T}}$. The state of a node α is then noted \mathbf{x}_α . We put \mathcal{A} and \mathcal{C} for the set of the internal nodes (ancestors) and for the set of tips (contemporary organisms) of \mathcal{T} respectively. In the ancestral reconstruction problem, the states of nodes in \mathcal{C} are known and stored in an element \mathbf{c} of $\{0, 1\}^{\mathcal{C}}$, while the states of nodes \mathcal{A} are unknown. The ancestral state reconstruction problem consists in associating binary states to all internal nodes, in other words, to find, in a way somehow optimal, an element

\mathbf{a} in $\{0, 1\}^A$, which altogether with the known state configuration $\mathbf{c} \in \{0, 1\}^C$ of the tips, gives the whole evolutionary history of the binary character over \mathcal{T} .

2.2 Ancestral reconstruction methods

2.2.1 Evolutionary model: parameters and notations

A basic presentation of continuous time Markov models, including formula of likelihood computation, is given in Appendix A. We just recall here that for a binary character, such a model \mathcal{M} can be defined with two parameters π_1 and λ , where π_1 is the probability of state 1 in the stationary distribution of \mathcal{M} and λ can be thought as a scale factor speeding up evolutionary time. More intuitively, π_1 reflects the propensity of staying in state 1 as well as evolving from 0 to 1. Informally, λ is a positive real number which indicates the strength of the relation, also modulated by branch lengths, between the state of a node and the state of its direct ancestor: the smaller is λ , the unlikelier becomes a change between a node and its child, while when λ is very high, evolution essentially turns out to draw a state following the stationary distribution (influence of direct ancestors tends to become negligible). From a given model \mathcal{M} , one associates to any time t , the transition matrix $\mathbb{P}_{\mathcal{M}}(t)$ in which entry $(a, b) \in \{0, 1\}^2$ contains the probability to evolve from state a to state b in time t under model \mathcal{M} (see Appendix A).

The probability of a complete configuration of states $\mathbf{x} \in \{0, 1\}^{\mathcal{T}}$ under the model \mathcal{M} is noted:

$$\mathbb{P}_{\mathcal{T}, \mathcal{M}}(X_{\mathcal{T}} = \mathbf{x})$$

where $X_{\mathcal{S}}$ denotes the states variables vector of the nodes of a subset $\mathcal{S} \subset \mathcal{T}$.

If \mathcal{I} and \mathcal{J} are two non-intersecting subsets of nodes, \mathbf{i} and \mathbf{j} two states configurations belonging respectively to $\{0, 1\}^{\mathcal{I}}$ and $\{0, 1\}^{\mathcal{J}}$, the probability of the configuration obtained by merging \mathbf{i} and \mathbf{j} is noted:

$$\mathbb{P}_{\mathcal{T}, \mathcal{M}}(X_{\mathcal{I}} = \mathbf{i}, X_{\mathcal{J}} = \mathbf{j}).$$

If \mathcal{H} is a subset of \mathcal{T} and \mathbf{h} a configuration of $\{0, 1\}^{\mathcal{H}}$, the probability of the partial configuration \mathbf{h} , known only over \mathcal{H} , is given by:

$$\mathbb{P}_{\mathcal{T}, \mathcal{M}}(X_{\mathcal{H}} = \mathbf{h}) = \sum_{\mathbf{r} \in \{0, 1\}^{\mathcal{T} \setminus \mathcal{H}}} \mathbb{P}_{\mathcal{T}, \mathcal{M}}(X_{\mathcal{H}} = \mathbf{h}, X_{\mathcal{T} \setminus \mathcal{H}} = \mathbf{r}).$$

2.2.2 Most likely ancestor reconstruction

The most likely ancestor reconstruction associates to a tree \mathcal{T} and a configuration of known states $\mathbf{c} \in \{0, 1\}^C$ of tips, the configuration of ancestor states which makes the probability of the corresponding whole evolutionary history as great as possible. The evolution model parameters are not given and we have to estimate them from the data available, that are the tips states. This is done by numerically computing the model $\widehat{\mathcal{M}}$ which maximizes the likelihood of the tips states. Model $\widehat{\mathcal{M}}$ is used to compute the most likely ancestor reconstruction which is the configuration $\mathbf{m}_{\widehat{\mathcal{M}}} \in \{0, 1\}^A$ defined by:

$$\mathbf{m}_{\widehat{\mathcal{M}}} = \arg \max_{\mathbf{a} \in \{0, 1\}^A} \mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}}(X_{\mathcal{A}} = \mathbf{a}, X_{\mathcal{C}} = \mathbf{c}).$$

We also denote by $\mathbf{m}_{\widehat{\mathcal{M}}}(\alpha)$ or simply $\text{MLA}_{\widehat{\mathcal{M}}}(\alpha)$ the state of node α in the most likely reconstruction $\widehat{\mathbf{m}}_{\widehat{\mathcal{M}}}$.

Remark that the algorithm computing the most likely ancestor does not compute the likelihoods of all possible ancestors but performs in a very same way as the algorithm computing the most parsimonious reconstruction. In particular its time of computation is linear with the number of nodes of the tree.

2.2.3 Posterior probability reconstruction

Let α be a node of \mathcal{T} of unknown state. The posterior probability of state 1 for node α is the probability of observing state 1 on α , conditioned to what is known on the tree (the configuration \mathbf{c} of contemporary states). Again the computation of this conditional probability needs an evolution model which has to be estimated by maximum likelihood like in the case of most likely ancestor reconstruction. This estimated model is still noted $\widehat{\mathcal{M}}$ and is used to compute the posterior probability:

$$\begin{aligned} \text{PP}_{\widehat{\mathcal{M}}}(\alpha) &= \mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}}(X_\alpha = 1 \mid X_{\mathcal{C}} = \mathbf{c}) \\ &= \frac{\mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}}(X_\alpha = 1, X_{\mathcal{C}} = \mathbf{c})}{\mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}}(X_{\mathcal{C}} = \mathbf{c})}. \end{aligned}$$

The ancestral reconstruction derived from posterior probability basically associates to each unknown node α , the state 1 if

$$\text{PP}_{\widehat{\mathcal{M}}}(\alpha) = \mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}}(X_\alpha = 1 \mid X_{\mathcal{C}} = \mathbf{c}) > \mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}}(X_\alpha = 0 \mid X_{\mathcal{C}} = \mathbf{c})$$

and the state 0 otherwise. Since these two conditional probabilities are complementary, the previous inequality is equivalent to $\text{PP}_{\widehat{\mathcal{M}}}(\alpha) > 0.5$. We will see later that if we replace the estimated model $\widehat{\mathcal{M}}$ by the real underlying evolution model in the preceding inequality, the state that we guess for node α is actually the one which minimizes the error expectation.

The computation time of the posterior probability for a given node is proportional with the total number of nodes of the tree. Since it has to be performed once for each unknown node, the total time of computation of the reconstruction is quadratic with respect to the size of the tree.

2.2.4 Likelihood ratio reconstruction

The likelihood ratio reconstruction was introduced in Pagel (1999). To reconstruct the state of an unknown node α , this method starts by computing the maximum likelihood estimations of the Markov model $\widehat{\mathcal{M}}_1$ and $\widehat{\mathcal{M}}_0$, obtained from \mathcal{T} , with the known states (that of tips) and by setting the state of α respectively to 1 and 0. Next, it computes the likelihoods corresponding to the two alternative reconstructions for node α , under the two corresponding models, and finally considers their ratio:

$$\text{LR}_{\widehat{\mathcal{M}}_1, \widehat{\mathcal{M}}_0}(\alpha) = \frac{\mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}_1}(X_\alpha = 1, X_{\mathcal{C}} = \mathbf{c})}{\mathbb{P}_{\mathcal{T}, \widehat{\mathcal{M}}_0}(X_\alpha = 0, X_{\mathcal{C}} = \mathbf{c})}.$$

The likelihood ratio $\text{LR}_{\widehat{\mathcal{M}}_1, \widehat{\mathcal{M}}_0}(\alpha)$ indicates which hypothesis of reconstruction to favor for node α . As in the case of posterior probability, one needs a threshold to

perform effectively the reconstruction. A natural choice is to reconstruct state 1 for node α if its likelihood ratio is greater than one and state 0 otherwise.

The computation time of this method is way greater than that of the preceding methods because we have to estimate the parameters of model twice per nodes. Since this estimation is done numerically, it implies the computation of a great number of likelihoods over the tree.

2.2.5 Likelihood ratio and posterior probability

The likelihood ratio is actually very close to the posterior probability. To explain this point, we recall that to compute the likelihood ratio, models $\widehat{\mathcal{M}}_1$ and $\widehat{\mathcal{M}}_0$ are estimated from the tips and by setting the state of α respectively to 1 and 0. The greater is the tree, the closer are these estimations one to another and to the estimation $\widehat{\mathcal{M}}$ obtained from the only states of the tips (*i.e.* the influence of setting the state of α becomes negligible). It comes that for a sufficient size of tree, we have:

$$\text{LR}_{\widehat{\mathcal{M}}_1, \widehat{\mathcal{M}}_0}(\alpha) \approx \frac{\mathbb{P}_{\tau, \widehat{\mathcal{M}}}(X_\alpha = 1, X_C = \mathbf{c})}{\mathbb{P}_{\tau, \widehat{\mathcal{M}}}(X_\alpha = 0, X_C = \mathbf{c})}.$$

A basic calculus using the expression of the posterior probability gives:

$$\text{PP}_{\widehat{\mathcal{M}}}(\alpha) \approx \frac{\text{LR}_{\widehat{\mathcal{M}}_1, \widehat{\mathcal{M}}_0}(\alpha)}{1 + \text{LR}_{\widehat{\mathcal{M}}_1, \widehat{\mathcal{M}}_0}(\alpha)}.$$

In particular the corresponding reconstruction methods tends to becomes equivalent since we expect $\text{PP}_{\widehat{\mathcal{M}}}(\alpha) > \frac{1}{2}$ if $\text{LR}_{\widehat{\mathcal{M}}_1, \widehat{\mathcal{M}}_0}(\alpha) > 1$ and reciprocally.

2.2.6 Bayesian reconstruction methods

Bayesian approaches aim to incorporate uncertainty in the reconstruction process. This uncertainty is formalized by a prior probability distribution μ over the model parameters. This prior is used to compute the corresponding Bayesian posterior probability relative to each unknown node α and defined in the following way:

$$\text{B}(\alpha) = \int_{\widetilde{\mathcal{M}}} \mathbb{P}_{\tau, \widetilde{\mathcal{M}}}(X_\alpha = 1 | X_C = \mathbf{c}) \mu(\widetilde{\mathcal{M}})$$

Basically the idea is to sum the posterior probabilities obtained over all the possible parameters of the binary Markov model of evolution, weighted by the their prior probabilities. The choice of prior distributions for the parameters λ and π_1 will be discussed further. We get a Bayesian posterior probability which can be used in the same way as the posterior probability obtained from the maximum likelihood estimation of the model to perform an ancestral reconstruction. Bayesian reconstruction approach also provides a confidence interval for this Bayesian posterior probability, revealing if its value is significantly greater (resp. smaller) than 0,5 in order to reconstruct state 1 (resp. 0) or if it is not relevant enough.

Finally, remark that the Bayesian reconstruction approach can be extended in order to take into account uncertainty not only on the model of evolution but also on the phylogenetic tree supporting the character evolution (it then needs a prior probability on trees) (Ronquist, 2004). This last point is not discussed further here because the phylogenetic tree is assumed to be known.

2.2.7 Choice of prior distributions in Bayesian reconstructions

We have to choose prior distributions for the parameters λ and π_1 . We chose prior distributions as suggested in Maritz and Lwin (1989) and in Pagel et al. (2004). As prior for parameter λ , we use an exponential distribution with parameter estimated by maximum likelihood from the tips states, and we use an uniform prior for parameter π_1 . Estimating, or determining from biological consideration, the parameter of the exponential prior distribution can hardly be avoided since the evolution speed of a character depends on its own nature.

The Bayesian posterior probability is generally computed by stochastic simulations methods like Markov Chain Monte Carlo techniques. Here we evaluate an approximation of the Bayesian posterior probability by sampling the space of parameters as follows. For a given number n , we split the set of possible values of λ (that is non-negative real numbers) into n parts of probability $\frac{1}{n}$ with regard to the exponential distribution considered (in all simulations presented here the Bayesian probability is computed by splitting the domain of the scale parameter into $n = 10$ parts). Each part is then represented by the parameter value corresponding to its median. Our approximation for the Bayesian posterior probability is then computed as the mean of the posterior probabilities obtained for all pairs of parameters (λ', π'_1) where λ' is the median of a part and π'_1 taken in $\{0.1, 0.2, \dots, 0.9\}$. We choose n in such a way that increasing the number of parts in which we split the domain of parameters λ or π_1 for the approximation of the Bayesian posterior probability, does not change the expectations of success.

3 Assessing performance

3.1 Reconstruction accuracy

In Maddison (1995), author provides exact calculations of the probabilities of success for a parsimonious reconstruction under the assumption that the character evolves following a symmetric discrete time Markov model. In order to evaluate performances of likelihood-based reconstruction approaches, we proceed in a same way but we assume here that the character follows a continuous-time Markov model.

3.2 Conditional expectation of success

Let \mathcal{M} be the unknown model used for generating the data. We consider a particular evolutionary history, where we denote by \mathbf{c} the set of contemporary states and by \mathbf{a} the set of ancestral states that are obtained. A particular reconstruction method m infers ancestral states \mathbf{r}^c from the set of contemporary states \mathbf{c} . It succeeds in reconstructing node α if $\mathbf{a}_\alpha = \mathbf{r}_\alpha^c$. But under the evolution model \mathcal{M} and conditioned to the contemporary states, node α has either state \mathbf{a}_α with probability

$$\mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = \mathbf{a}_\alpha \mid X_{\mathcal{C}} = \mathbf{c}),$$

or state $\mathbf{b}_\alpha = 1 - \mathbf{a}_\alpha$ with the complementary probability.

Consequently, given the contemporary states \mathbf{c} , a reconstruction method predicting $\mathbf{r}_\alpha^{\mathbf{c}}$ for node α succeeds with expectation

$$\sum_{\mathbf{u}_\alpha \in \{0,1\}} \delta_{\mathbf{r}_\alpha^{\mathbf{c}}, \mathbf{u}_\alpha} \mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = \mathbf{u}_\alpha \mid X_{\mathcal{C}} = \mathbf{c})$$

where δ is the Kronecker delta:

$$\delta_{\mathbf{r}_\alpha^{\mathbf{c}}, \mathbf{u}_\alpha} = \begin{cases} 1, & \text{if } \mathbf{r}_\alpha^{\mathbf{c}} = \mathbf{u}_\alpha \\ 0, & \text{if } \mathbf{r}_\alpha^{\mathbf{c}} \neq \mathbf{u}_\alpha. \end{cases}$$

This is nothing but $\mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = \mathbf{r}_\alpha^{\mathbf{c}} \mid X_{\mathcal{C}} = \mathbf{c})$.

3.3 Expectation of success

We denote by $S_{m, \alpha}$ the binary random variable indicating whether the reconstruction method m succeeds in reconstructing node α . The expectation of success of m is nothing but that of $S_{m, \alpha}$

$$\mathbb{E}(S_{m, \alpha}) = \sum_{\mathbf{c} \in \{0,1\}^{\mathcal{C}}} \mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = \mathbf{r}_\alpha^{\mathbf{c}} \mid X_{\mathcal{C}} = \mathbf{c}) \mathbb{P}_{\tau, \mathcal{M}}(X_{\mathcal{C}} = \mathbf{c}) \quad (1)$$

Consequently, for a given continuous time Markov model of evolution \mathcal{M} of parameters π_1 and λ , computing the value

$$p_{\mathcal{M}, \alpha, \mathbf{c}} = \mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = 1 \mid X_{\mathcal{C}} = \mathbf{c}) \quad (2)$$

for any set of contemporary states \mathbf{c} provides the expectation of success of any reconstruction method at a chosen node α .

Finally, if one wants to select a reconstruction method from its expected number of correct reconstructed states over the whole phylogenetic tree, one has to sum $\mathbb{E}(S_{m, \alpha})$ over all the ancestral nodes α .

3.4 Upper bound for expectation of success - Optimal reconstruction

As yet pointed out, notably in Mossel (2003), ancestral reconstruction suffers from limits inherent to the stochastic character of evolution. Here we are somehow able to explicit this limit.

From Equation (1), we derive that

$$\mathbb{E}(S_{m, \alpha}) \leq \sum_{\mathbf{c} \in \{0,1\}^{\mathcal{C}}} \max_{\mathbf{u}_\alpha \in \{0,1\}} \mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = \mathbf{u}_\alpha \mid X_{\mathcal{C}} = \mathbf{c}) \mathbb{P}_{\tau, \mathcal{M}}(X_{\mathcal{C}} = \mathbf{c}).$$

This implies that the expectation of success of any reconstruction method m is smaller than the expectation of success of the reconstruction $\text{OP}_{\mathcal{M}}$ inferring \mathbf{p} at the ancestral nodes in the following way :

$$\mathbf{p}_\alpha = \begin{cases} 1, & \text{if } \mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = 1 \mid X_{\mathcal{C}} = \mathbf{c}) > \mathbb{P}_{\tau, \mathcal{M}}(X_\alpha = 0 \mid X_{\mathcal{C}} = \mathbf{c}) \\ 0, & \text{otherwise.} \end{cases}$$

The reconstruction method $OP_{\mathcal{M}}$ will be referred to as the \mathcal{M} -optimal reconstruction since no approach can give better results in terms of expectation of success. Note that the only difference with the posterior probability reconstruction lies in the model used to compute the conditional probability. Namely, this model is the one generating the data for the optimal reconstruction while being its maximum likelihood estimation from observed states in the case of the posterior probability reconstruction. Naturally in the real life, we never have access to the model underlying the biological evolution. Nevertheless, in the particular case of our study, we do know this model and the optimal reconstructions allows us to present optimal upper bounds of expectations success.

Remark that under the assumption that the evolution model parameters are not fixed but follows given distributions, a similar argument shows that the method achieving the best performances is the Bayesian reconstruction with these distributions as priors.

3.5 Saturation

We are interested here in defining what is a relevant range of values for parameters π_1 and λ . Namely, without a priori knowledge, π_1 may describe the interval $[0, 1]$, whereas λ is constrained by an evolution scale neither too high ($\lambda \leq \lambda_{\max}$) to keep a phylogenetic signal strong enough almost on the shortest branch, nor too low ($\lambda \geq \lambda_{\min}$) to observe some evolution events, almost on the longest branch. Thus, we fix a threshold ε such that the state of a child must depend on the state of its direct ancestor:

$$|p_{00}(t) - p_{10}(t)| \geq \varepsilon \Leftrightarrow \lambda \leq \frac{-\ln(\varepsilon)}{t} \quad (3)$$

and such that a child and its direct ancestor do not almost surely have the same state:

$$p_{00}(t) - p_{10}(t) \leq 1 - \varepsilon \Leftrightarrow \lambda \geq \frac{-\ln(1 - \varepsilon)}{t}. \quad (4)$$

This provides thresholds $\lambda_{\min} = \frac{-\ln(1 - \varepsilon)}{t_{\max}}$ and $\lambda_{\max} = \frac{-\ln(\varepsilon)}{t_{\min}}$ that have to be computed for each tree. Thus success results show that saturation is reached as soon as λ is greater than a threshold λ_{\max} .

4 Results

4.1 Data

We considered several trees (with branch lengths information) as support for comparing the efficiency of reconstruction methods for binary characters: the (sampled) tree-of-life restricted to eukaryote organisms (Letunic and Bork, 2007) (Fig. 2a) and two subtrees extracted from Pleistocene planktic Foraminifera phylogeny (Webster and Purvis, 2002) (Fig. 2b and 2c).

Branch lengths are normalized in order to get an average time from root to tips equal to 100.

4.2 Parameters used for computing the expectations of success

Applying formulae (3) and (4) with $\varepsilon = 0.01$, we get that suitable values for λ must range 0.0001 to 1. We then consider λ in the logarithmic scale $\{0.0001, 0.001, 0.01, 0.1, 1\}$. As results for $\lambda = 0.1$ and $\lambda = 1$ are very close because saturation is nearly reached at $\lambda = 0.1$, results for $\lambda = 0.1$ will not be shown.

For each tree we compare the accuracy of methods to reconstruct the state of all the ancestors (including the root of the tree). Thus for any possible tips configuration, we perform a parsimonious, an optimal, a most likely ancestor, a posterior probability, a likelihood ratio and a Bayesian reconstruction for every node α (when needed, the parameters of the model used in reconstruction are estimated knowing only the states of the tips). We also compute the likelihood the probability $p_{\mathcal{M},\alpha,\mathbf{c}}$ for this internal node to be in state 1, given each tips configuration, considering combinations of parameters π_1 and λ with π_1 taken among $\{0.1, 0.2, \dots, 0.9\}$ and λ among $\{0.0001, 0.001, 0.01, 0.1, 1\}$. Then we deduce the conditional probability of success associated to configuration \mathbf{c} (either $p_{\mathcal{M},\alpha,\mathbf{c}}$ or $1 - p_{\mathcal{M},\alpha,\mathbf{c}}$). Finally, by summing over all the possible configurations \mathbf{c} , we obtain for every node the expectation of success associated to each reconstruction method.

The results are displayed in Figure 3, where we plot the expectation of success of each method *vs* the parameter π_1 for all the values of λ considered. Actually, for every tree, we display the results only for the root and for the internal nodes labeled A and B in every tree (Fig. 2).

4.3 Parameters influence

Everything stated here stands for all the reconstruction methods, and for all the studied trees. As expected, we observe in Figure 4, that all the reconstruction approaches perform better as λ is smaller. Indeed, this corresponds to a model where character states evolve slowly, and consequently reconstruction is quite easy.

The parameter π_1 also influences the expectations of success of all methods since, for a fixed λ , they are better with π_1 close to 0 (or symmetrically close to 1) than with π_1 about 0.5. Moreover, this influence of π_1 grows with λ . Notably for $\lambda = 0.0001$, the difference between the expectations of success of reconstructions obtained from simulations using $\pi_1 = 0.1$ and simulations using $\pi_1 = 0.5$ is less than 0.005 while this difference is about 0.4 for $\lambda = 0.1$. We recall that however small is the difference between expectations of success, it is significant.

4.4 Saturation

For greater values of parameter λ , the ancestor/child relation turns out to be weaker. Hence, the classifications become less reliable and the influence of the parameter π_1 over the reconstruction results increases. Indeed as parameter λ increases, the expectations of success tend to decrease linearly with the parameter π_1 (for $\pi_1 \geq \frac{1}{2}$), and symmetrically. When λ is greater than 1, the optimal expectation of success follows the so-called curve of no-discrimination.

Essentially this draws the curve we would obtain by inferring the state 1 if $\pi_1 > \frac{1}{2}$ and the state 0 otherwise, without taking into account the state of the other nodes. For such values of λ , the phylogenetic signal is no longer strong enough to be relevant with regard to the reconstruction problem.

4.5 Comparison of \mathcal{M} -optimal, most likely ancestor, posterior probability, likelihood ratio, Bayesian and most parsimonious reconstructions

As stated above, the \mathcal{M} -optimal method provides upper bounds for expectations of success when evolution is driven by \mathcal{M} . Despite the fact that there are significant difference between the expectations of success achieved by the methods, it is worth noting that no method shows really bad results. Indeed, all the expectations of success are quite close to the optimal one, especially when λ is small.

The reconstruction methods based on likelihood generally obtain better performances comparing with the parsimonious one, except for some of the nodes as showed in Figure 3, when λ is small enough.

We observe that when we reconstruct the root of every tree, the curves corresponding to the expectations of success of the posterior probability and the likelihood ratio reconstructions are not distinguishable. A possible explanation could be that time from the common ancestor is great enough to make parameters estimations $\widehat{\mathcal{M}}_0$, $\widehat{\mathcal{M}}_1$ very close to $\widehat{\mathcal{M}}$, and consequently the reconstructions coincide.

As expected, the \mathcal{M} -optimal reconstruction has the best performances, when evolution is driven by model \mathcal{M} of parameters (λ, π_1) . Since the real parameters (λ, π_1) are unknown, the \mathcal{M} -optimal reconstruction can not be achieved. Then one has to choose between the other reconstruction methods. But the expectations of success depend on the topology of the tree, the node to infer, the parameters values and hence no rule can be guessed.

5 Discussion

5.1 None reconstruction method can be the best

Under the assumption that the character evolution is driven by a Markov model \mathcal{M} , we have seen that the best method is the so-called \mathcal{M} -optimal method $OP_{\mathcal{M}}$. It clearly implies that finding a “real” reconstruction method (i.e. working only from the knowledge of the contemporary states) which gives better performances than any other method for any model \mathcal{M} is hopeless. Though it certainly has bad performances if the evolution follows a model different from \mathcal{M} , $OP_{\mathcal{M}}$ outperforms any other methods if \mathcal{M} actually drives evolution. This is the reason for which we propose the following protocol as an help to choose the most suitable reconstruction method with regard to the phylogenetic tree considered, the internal node of which we want the state to be reconstructed as well as the evolutionary assumptions we make.

5.2 A Protocol to make the best choice

5.2.1 Description

Suppose now that we study a phylogenetic tree with character states known for several species (generally the contemporary species). All the reconstruction methods infer a state for the ancestral node α , but they do not necessarily coincide. Indeed, in the examples previously studied (common ancestor and internal nodes A, B), the states inferred by the likelihood-based reconstruction methods do not coincide in at least 10% configurations (see Table 1). Actually, if we included the most parsimonious reconstruction in this count, the number of conflicting reconstructions would be much greater. Indeed, the most parsimonious reconstruction, lying on the minimization of a discrete criterion, often reconstructs either state 0 or 1 with the same cost. In this case, reconstruction is considered failed, and the most parsimonious reconstruction does not coincide with the other reconstruction methods.

Thus, if all the reconstruction methods return the same inferred for node α , then any reconstruction method can be chosen. Otherwise the configuration provides conflicting reconstructed states and one has to select the best reconstruction method in terms of conditional expectation of success. We recall that expectations of success are computed with fixed values of parameters λ and π_1 , defining the model \mathcal{M} . Indeed, given the configuration \mathbf{c} for contemporary states, the conditional expectation of success of a reconstruction method m is merely $p_{\mathcal{M},\alpha,\mathbf{c}}$ defined in Equation (2), if it predicts state 1 for the node α , or $1 - p_{\mathcal{M},\alpha,\mathbf{c}}$ if it predicts state 0. In other words, the conditional expectation of success of any reconstruction method at fixed parameters (λ, π_1) is merely the probability to get the reconstructed state at the ancestral node under model (λ, π_1) , given configuration \mathbf{c} . Hence, all the reconstruction methods are split into two categories : the accurate ones, forecasting the state associated to $\max(p_{\mathcal{M},\alpha,\mathbf{c}}, 1 - p_{\mathcal{M},\alpha,\mathbf{c}})$, and the other ones, forecasting the state associated to $\min(p_{\mathcal{M},\alpha,\mathbf{c}}, 1 - p_{\mathcal{M},\alpha,\mathbf{c}})$.

If the real parameters values (λ, π_1) were known, one would infer state associated to $\max(p_{\mathcal{M},\alpha,\mathbf{c}}, 1 - p_{\mathcal{M},\alpha,\mathbf{c}})$. Else we explore all the possible evolutionary models with a given step and identify for every parameters value the most accurate state to infer for node α .

5.2.2 Illustration

We take as example the (sub-)tree-of-life (Fig. 2a), and consider one of the 128 configurations where likely ancestor, posterior probability, likelihood ratio, Bayesian and most parsimonious reconstruction methods do not return a same state for the root and for internal nodes A, B (Fig. 5). We underline that though the inferred state does not depend on the values of the real underlying parameters (λ, π_1) , the conditional expectation of success does so. Hence, to make the best choice, we compare the conditional expectations of success of most likely ancestor, posterior probability, likelihood ratio, Bayesian and most parsimonious reconstruction methods with respect to parameters (λ, π_1) . Since we have already computed $\lambda_{\min} = 0.0001$ and $\lambda_{\max} = 1$ for this tree, we consider parameter λ with values in interval $[0.0001, 1]$. A particular grid is chosen to sweep all the range, becoming more detailed for small values.

Figure 6 illustrates the results. Thus when predicting a state for the root (Fig. 6a), the most parsimonious reconstruction is inappropriate because it is unable to choose between the states ; the likelihood ratio, the posterior probability and the most likely ancestor reconstructions infer state 1, with a great expectation of success when parameter λ is not too small, associated with a parameter π_1 greater than 0.5 or when parameter λ is very small, associated with π_1 smaller than 0.5 ; whereas the Bayesian reconstruction infers state 0, which corresponds to a reliable prediction for the other pairs of parameters (λ, π_1) . But methods are neither split in the same way when reconstructing node A, nor node B. Indeed, when predicting a state for the internal node A (Fig. 6b), the posterior probability and the most likely ancestor reconstructions infer state 1, with a great expectation of success when parameters λ and π_1 are both very small or both great enough; whereas the most parsimonious, the Bayesian and the likelihood ratio reconstructions infer state 0, which corresponds to a reliable prediction for the other pairs of parameters (λ, π_1) . Finally, when predicting a state for the internal node B (Fig. 6c), only the most likely ancestor reconstruction infers state 1, with a great expectation of success when parameters λ and π_1 are both very small or both great enough; whereas the other reconstruction methods infer state 0, which corresponds to a reliable prediction for the other pairs of parameters (λ, π_1) .

As stated above, it is impossible to define a priori the most appropriate reconstruction method. Indeed, the accuracy of a reconstruction method depends on the topology of the studied tree, on the ancestral node to infer as well as on the real parameter values. Assuming that the evolutionary speed λ and the parameter π_1 do not vary upon the phylogenetic tree, choosing the same reconstruction method for inferring a state to whole the nodes may be misleading. Hence when reconstructing any internal node, we suggest to lead the protocol for this node, and then deduce the suitable inferred state, according to evolution assumptions.

5.3 A biological example

Here we apply the protocol to the complete data given by (Webster and Purvis, 2002) : 20 morphospecies of Pleistocene planktic Foraminifera. In Webster and Purvis (2002), several continuous characters are studied : area, ellipse major axis and ellipse minor axis. For these characters, the states of the tips as well as the states of 13 ancestral nodes labeled from A to M (Fig. 7) are given. The states for the 13 ancestors were directly measured on fossils closed to the phylogenetic positions.

We apply the protocol to the continuous character area, transformed into a discrete binary character by a 2-means clustering, coding by state 1 the greatest area ($> 0.1 mm^2$) and by 0 the smallest. Thus the tips configuration is given and, as in Webster and Purvis (2002), we aim to reconstruct the states of the 13 ancestral nodes labeled from A to M. For every internal node, we ran the protocol, we compared the state inferred by most likely ancestor, posterior probability, likelihood ratio, Bayesian and most parsimonious reconstructions with the fossil state (given by Webster and Purvis (2002)). Then we identified the good reconstruction methods, pointed out in Figure 8 by their color in the pie. A filled pie means that none of these reconstruction methods can be successful.

It appears that internal nodes A, B, E, F, G, I, J, K, L are well reconstructed by all the reconstruction methods ; the most parsimonious reconstruction is the only method failing in reconstructing nodes C, D, H ; whereas no reconstruction method was able to predict node M. Hence in this biological example, only the most parsimonious reconstruction method is inappropriate, and any likelihood-based reconstruction method would be accurate.

In addition, the probability to predict the state given by the fossil record at any ancestral node A to L depends on the parameters values λ and π_1 . In figure 9, the colored symbols define the panel of values corresponding to a high probability (greater than 0.5) for every ancestral node. Hence every node provides a convenient range of values for parameters λ and π_1 . Assuming that parameters λ and π_1 do not vary over the tree, this remark suggests that the real parameters generating the character evolution would lie in the common range of values, identified in Figure 10 by the symbols that are not lightened. Besides, under such models (λ, π_1) , the probability of observing state 1 at node M is more likely than state 0, but state 0 is not excluded, especially if the evolutionary speed is great. Indeed, under a great evolutionary speed λ and an intermediate parameter π_1 , such that model (λ, π_1) keeps convenient with the reconstruction of ancestral states A to L, the probability of observing state 1 at node M rounds 0.5, and then nearly equals the probability of observing state 0. Thus in Webster and Purvis (2002), state 0 is associated to the ancestral node M whereas state 1 is associated to all its direct descendants (see Fig. 8), which is consistent with the fact that the area character seems to be led by a great evolutionary speed and a moderate tendency to the loss of the character (great sizes are likely to be lost).

6 Conclusion

We presented a tour of the most used likelihood-based reconstruction approaches, and we compared them one against each other and against the most parsimonious reconstruction method.

The performances, in term of expectation of success, of each method were evaluated over phylogenetic trees where the evolution of a binary character is supported by a Markov model. Moreover all the methods are systematically compared to the theoretical upper bounds of expectations of success that a reconstruction process can achieve with regard to the random character of evolution.

Since the expectation of success of any reconstruction method depends on the phylogenetic tree and on the node to infer, our general conclusion is that no reconstruction method can be designed a priori as the best. Worst, a single method may not be suitable for whole the ancestral nodes. Hence, before inferring a state to an ancestral node, one has to lead a study, as described in our protocol, which is a guideline to choose the method suitable under the parameters (λ, π_1) likely to have generated the evolution.

References

- Bishop, J., Dean, A., and Mitchell-Odds, T. (2000). Rapid evolution of plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proceedings of the National Academy of Science USA*, 97:5322–5327.
- Clark, J., Ree, R., Alfaro, M., King, M., Wagner, W., and Roalson, E. (2008). A Comparative Study in Ancestral Range Reconstruction Methods: Retracing the Uncertain Histories of Insular Lineages. *Syst. Biol.*, 57(5):693–707.
- Collins, T., Wimberger, P., and Naylor, G. (1994). Compositional bias, character-state bias, and character-state reconstruction using parsimony. *Syst. Biol.*, 43:482–496.
- Didier, G. (2011). Parametric Maximum Parsimonious Reconstruction on Trees. *Bull Math Biol*, 73:1447–1502.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17:368–376.
- Felsenstein, J. (1985). Phylogenies and the comparative methods. *American Naturalist*, 125:1–15.
- Fitch, W. (1971). Towards defining the course of evolution: Minimum change for a specific tree topology. *Syst. Zool.*, 20:406–416.
- Huelsenbeck, J., Nielsen, R., and Bollback, J. (2003). Stochastic mapping of morphological characters. *Syst. Biol.*, 52:131–158.
- Huelsenbeck, J. and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogeny. *Phys. Rev. E*, 8:754–755.
- Koshi, J. and Goldstein, R. (1996). Probabilistic reconstruction of ancestral protein sequences. *J. Mol. Evol.*, 42:313–320.
- Krishnan, N., Seligmann, H., Stewart, C.-B., Jason de Koning, A., and Pollock, D. (2004). Ancestral Sequence Reconstruction in Primate Mitochondrial DNA: Compositional Bias and Effect on Functional Inference. *Mol. Biol. Evol.*, 21(10).
- Langlais, C. and Fitch, W. (1974). An estimation of the consistency of the rate of molecular evolution. *Journal of Molecular Evolution*, 3:161–177.
- Letunic, I. and Bork, P. (2007). Interactive tree of life (itol): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1):127–128.
- Maddison, W. (1995). Calculating the probability distributions of ancestral states reconstructed by parsimony on phylogenetic trees. *Syst. Biol.*, 44(4):474–481.
- Maddison, W. and Maddison, D. (2009). Mesquite: a modular system for evolutionary analysis. version 2.71. Technical report, Software available from, <http://mesquiteproject.org>.
- Maritz, J. and Lwin, T. (1989). *Empirical Bayes Methods*. Chapman and Hall, New York. Second Edition.

- Messier, W. and Stewart, C. (1997). Episodic adaptative evolution of primate isozymes. *Nature*, 385:151–154.
- Mooers, A. and Schluter, D. (1999). Reconstructing Ancestor States with Maximum Likelihood: Support for One- and Two-Rate Models. *Syst. Biol.*, 48(3):623–633.
- Mossel, E. (2003). On the impossibility of reconstructing ancestral data and phylogenies. *J. Comput. Biol.*, 10(5):669–676.
- Nielsen, R. (2002). Mapping mutations on phylogenies. *Syst. Biol.*, 51:729–739.
- Pagel, M. (1999). The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Syst. Biol.*, 48:612–622.
- Pagel, M., Meade, A., and Barker, D. (2004). Bayesian Estimation of Ancestral character States on Phylogenies. *Syst. Biol.*, 53(5):673–684.
- Ree, R. and Donoghue, M. (1998). Step matrices and the interpretation of homoplasy. *Syst. Biol.*, 47(4):582–588.
- Ronquist, F. (2004). Bayesian inference of character evolution. In *TRENDS in Ecology and Evolution*, volume 19 of *Phylogenetics Series*, pages 475–481. Elsevier.
- Schluter, D., Price, T., Mooers, A., and Ludwig, D. (1996). Likelihood of ancestor states in adaptative radiation. *Evolution*, 51(6):1699–1711.
- Swofford, D. and Maddison, W. (1992). Parsimony, character-state reconstructions, and evolutionary inferences. In *R. L. Mayden (Ed.)*, Systematics, historical ecology, and North American freshwater fishes, pages 187–223. Stanford, Stanford University Press.
- Templeton, A. (1996). Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the cytochrome oxidase ii gene in the hominoid primates. *Genetics*, 144:1263–1270.
- Webster, A. and Purvis, A. (2002). Testing the accuracy for reconstructing ancestral states of continuous characters. *Proc. R. Soc. Lond. B*, 269:143–149.
- Yang, Z., Kumar, S., and Nei, M. (1995). A New Method of Inference of Ancestral Nucleotide and Amino Acid Sequences. *Genetics*, 141:1641–1650.
- Zhang, J. and Nei, M. (1997). Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. *J. Mol. Evol.*, 44(Suppl. 1):S139–S146.

A Details on ML reconstruction methods and notations

A.1 Binary continuous time Markov models parameters

We refer to Ronquist (2004) for a more complete survey of topic. Basically, a binary continuous time Markov model is defined from two instantaneous transition rates: the gain rate q_{01} and the loss rate q_{10} . As it will be recalled below, the transitions probabilities associated to \mathcal{M} are computed from its infinitesimal generator $Q_{\mathcal{M}}$:

$$Q_{\mathcal{M}} = \begin{pmatrix} -q_{01} & q_{01} \\ q_{10} & -q_{10} \end{pmatrix}.$$

Note that each row in $Q_{\mathcal{M}}$ sums to zero.

An equivalent and more intuitive way of defining such a model \mathcal{M} is to write its infinitesimal generator $Q_{\mathcal{M}}$ as:

$$Q_{\mathcal{M}} = \lambda \begin{pmatrix} -\pi_1 & \pi_1 \\ 1 - \pi_1 & -(1 - \pi_1) \end{pmatrix},$$

where $\lambda = q_{01} + q_{10}$ and $\pi_1 = \frac{q_{01}}{q_{01} + q_{10}}$. This parametrization is quite more explicit. Indeed $(1 - \pi_1, \pi_1)$, also denoted by (π_0, π_1) is the stationary distribution of \mathcal{M} while λ is a scale factor representing the relation between a child and its direct ancestor, also modulated by the branch lengths (this relation being stronger as λ is smaller – see below).

A.2 Transition probabilities

The probabilities of transition in a time t are then given by the entries of the matrix $P_{\mathcal{M}}(t)$: entry $P_{\mathcal{M}}(t)_{ab}$ is the transition probability from a to b in time t . The matrix $P_{\mathcal{M}}(t)$ is defined from the infinitesimal generator of \mathcal{M} as:

$$P_{\mathcal{M}}(t) = e^{tQ_{\mathcal{M}}} = \begin{pmatrix} 1 - \pi_1(1 - e^{-\lambda t}) & \pi_1(1 - e^{-\lambda t}) \\ \pi_0(1 - e^{-\lambda t}) & 1 - \pi_0(1 - e^{-\lambda t}) \end{pmatrix}.$$

In particular, if $\lambda = 0$ then $P_{\mathcal{M}}(t)$ is the identity matrix (no change can occur).

When $\lambda \rightarrow +\infty$, matrix $P_{\mathcal{M}}(t)$ tends to $\begin{pmatrix} \pi_0 & \pi_1 \\ \pi_0 & \pi_1 \end{pmatrix}$ which means that the state of a child does not depend on the state of its ancestor and is just drawn following the stationary distribution.

A.3 Evolution history likelihood

Let \mathcal{T} be a given phylogenetic tree \mathcal{T} (topology and branch lengths). A binary character-state evolution history over \mathcal{T} is a configuration \mathbf{x} associating a state $\mathbf{x}_n \in \{0, 1\}$ to each node n of \mathcal{T} . The random vector $X_{\mathcal{T}}$ describes the state of the evolutionary history on the tree. The probability/likelihood of a configuration \mathbf{x} of \mathcal{T} under a binary continuous Markov model \mathcal{M} is denoted

$\mathbb{P}_{\mathcal{T},\mathcal{M}}(X_{\mathcal{T}} = \mathbf{x})$ and is then computed as:

$$\mathbb{P}_{\mathcal{T},\mathcal{M}}(X_{\mathcal{T}} = \mathbf{x}) = \pi_{\mathbf{x}_r} \prod_{n \in \mathcal{T}} \prod_{m \in \mathcal{D}(n)} \mathbb{P}_{\mathcal{M}}(t_{nm})_{\mathbf{x}_n \mathbf{x}_m}$$

where r is the root of \mathcal{T} , $\mathcal{D}(n)$ designs the set of child nodes of n in \mathcal{T} and t_{nm} is the branch length connecting n to m .

Tables and Figures

TABLE 1 - Percentage of configurations with conflicting reconstructions for the common ancestor, and internal nodes A and B within trees 2a, 2b and 2c.

Tree	root	internal node A	internal node B	configurations number
Tree-of-life 2a	10.9%	17.2%	25.0%	128
Subtree 2b	16.0%	10.9%	16.8%	512
Subtree 2c	17.6%	18.0%	25.0%	512

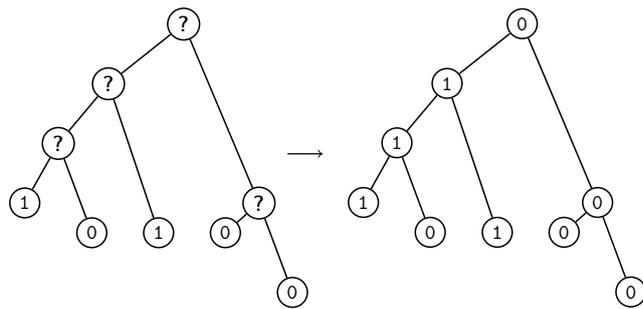
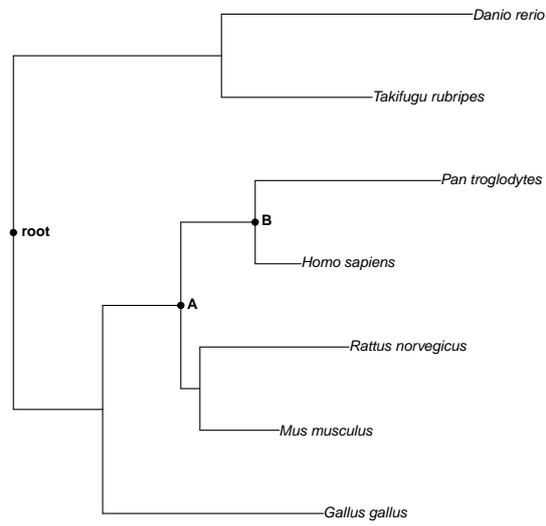
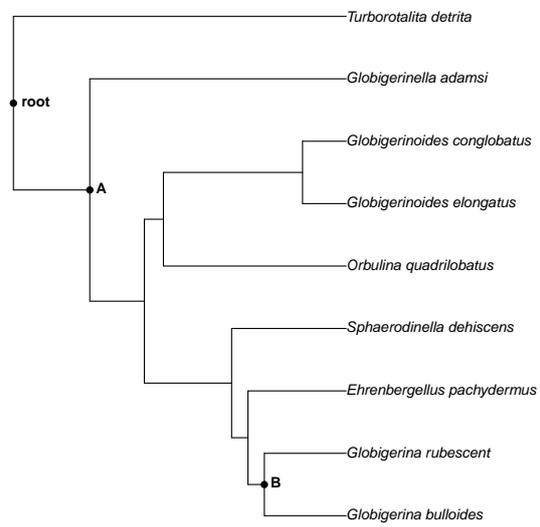


Figure 1: This diagram illustrates the principle of reconstruction: unknown character states are inferred from the known ones using an evolution model that rules the character state evolution.

a)



b)



c)

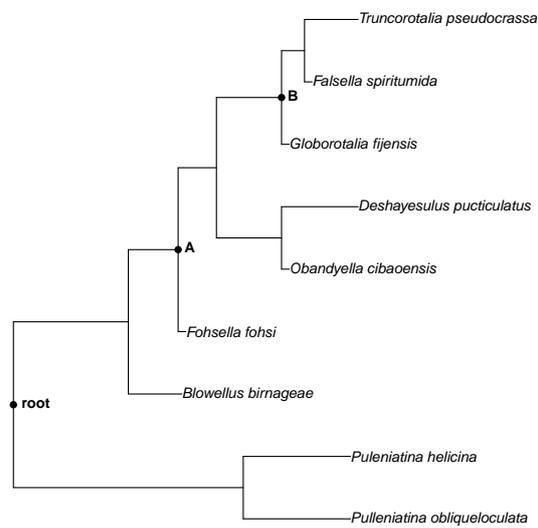
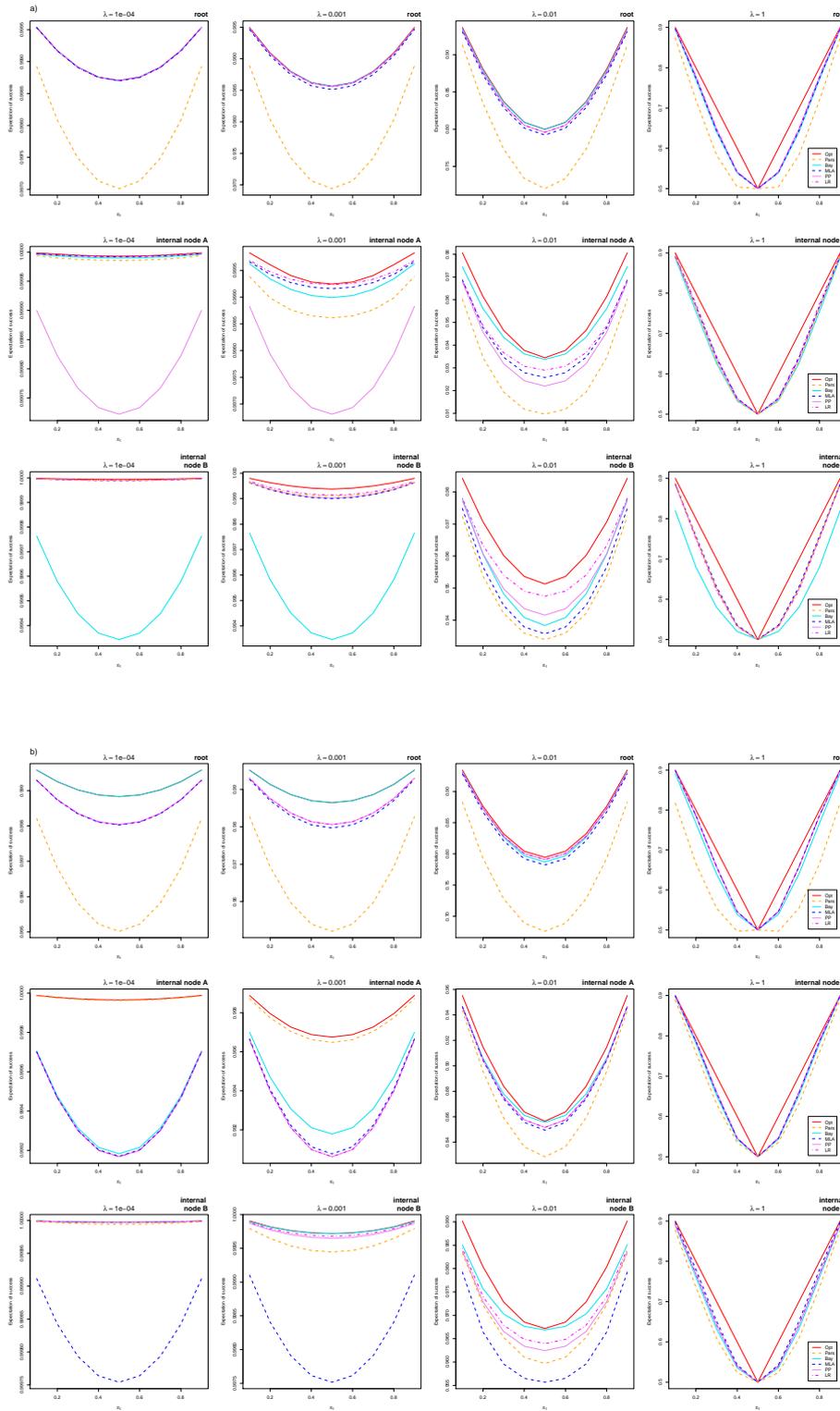


Figure 2: The trees used for supporting the evolution of binary character states in order to assess performance of reconstruction methods: (a) the tree-of-life restricted to eukaryote organisms (Letunic and Bork, 2007); (b) and (c) two subtrees extracted from Pleistocene planktic Foraminifera phylogeny (Webster and Purvis, 2002). In every tree, two internal nodes are labeled A and B, to be referred to in further studies.



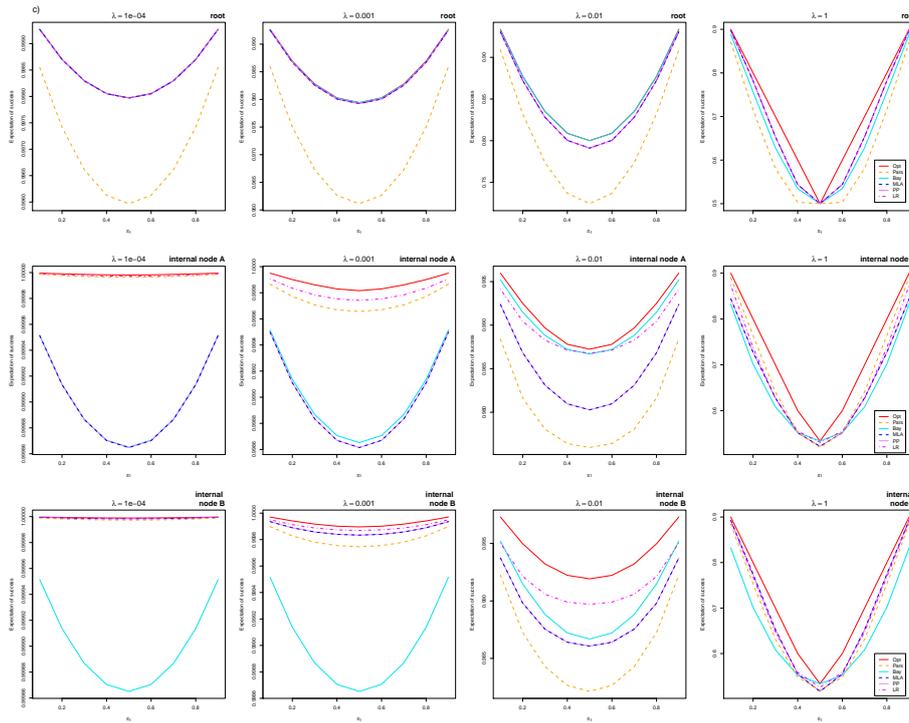


Figure 3: Expectations of success for \mathcal{M} -optimal, most likely ancestor, posterior probability, likelihood ratio, Bayesian and most parsimonious reconstructions. Each plot displays the expectations of success *versus* parameter π_1 , for a fixed value of parameter λ among $\{0.0001, 0.001, 0.01, 1\}$. Scale of plots changes with parameter λ . Plots are ordered by line for each node. From top to bottom: root, internal node A and internal node B. Trees are displayed in separate figures : Fig. 3a) for tree 2a), Fig. 3b) for tree 2b) and Fig. 3c) for tree 2c).

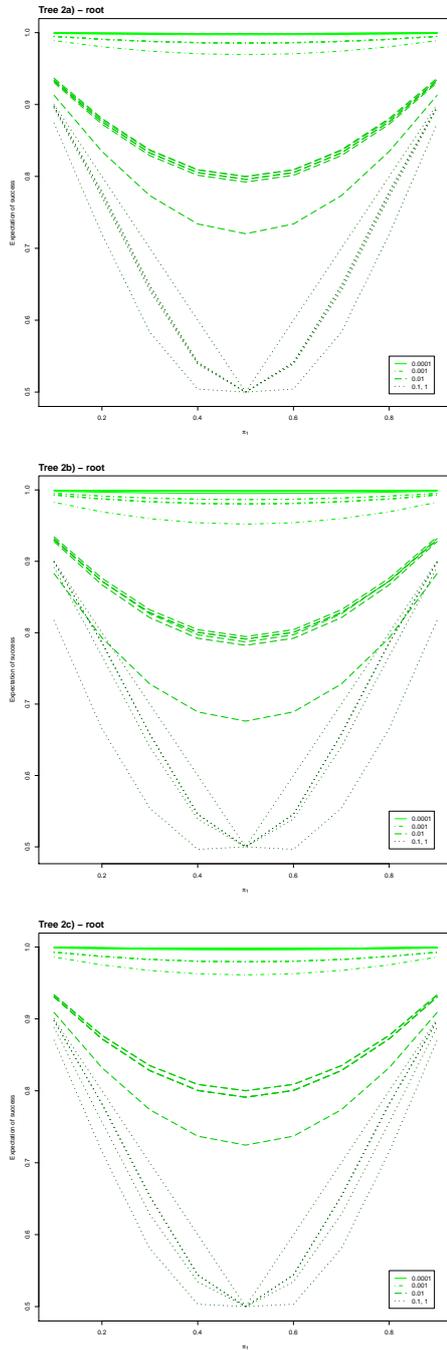


Figure 4: Expectations for success when predicting the state of the common ancestor with \mathcal{M} -optimal, most parsimonious, Bayesian, most likely ancestor, likelihood ratio and posterior probability reconstruction methods. Each plot displays the expectations of success *versus* parameter π_1 , with parameter λ varying among $\{0.0001, 0.001, 0.01, 0.1, 1\}$. The expectations of success for $\lambda = 0.1$ and $\lambda = 1$ are not distinguishable, so only curves for $\lambda = 1$ are plotted. Plots are ordered from top to bottom referring to the trees 2a), 2b) and 2c).

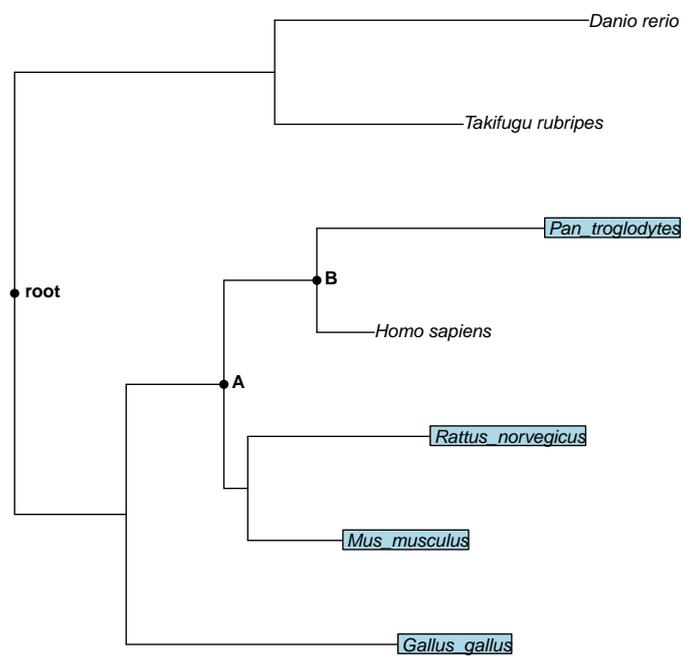
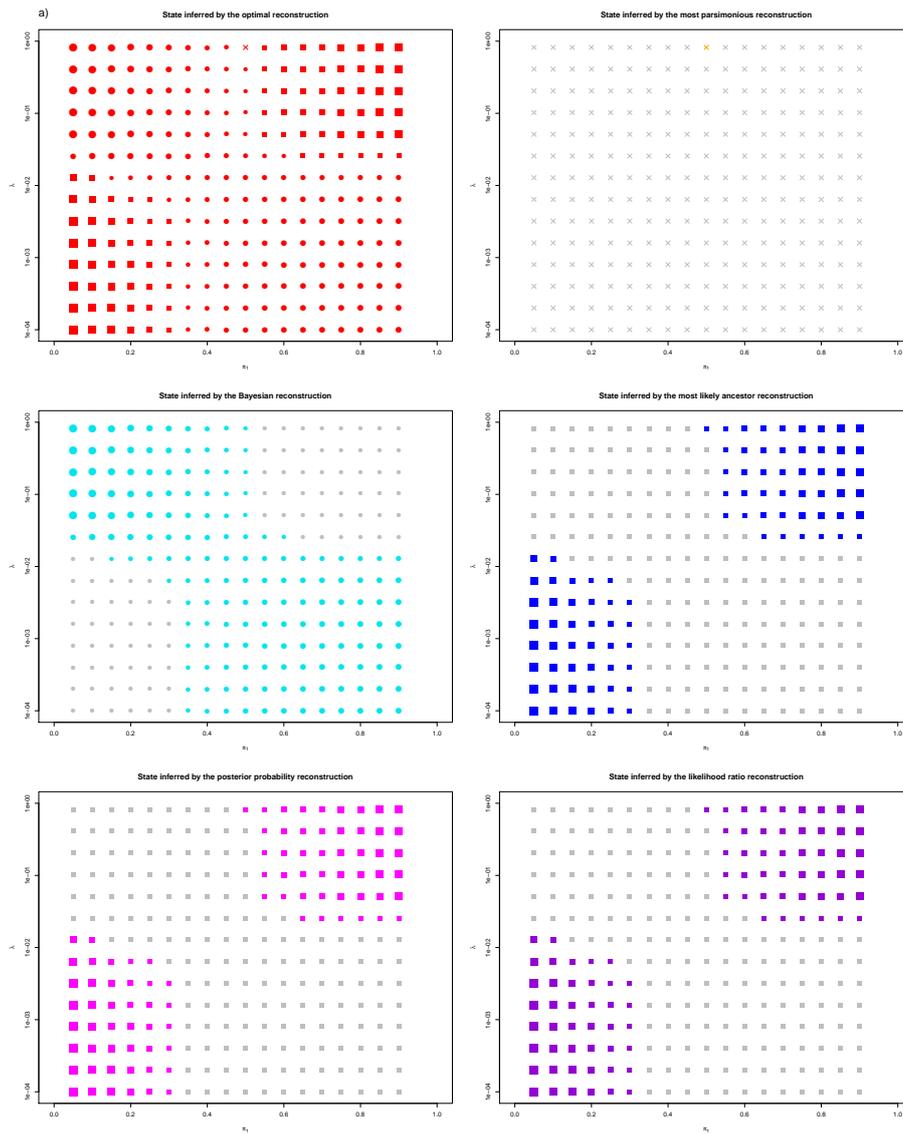
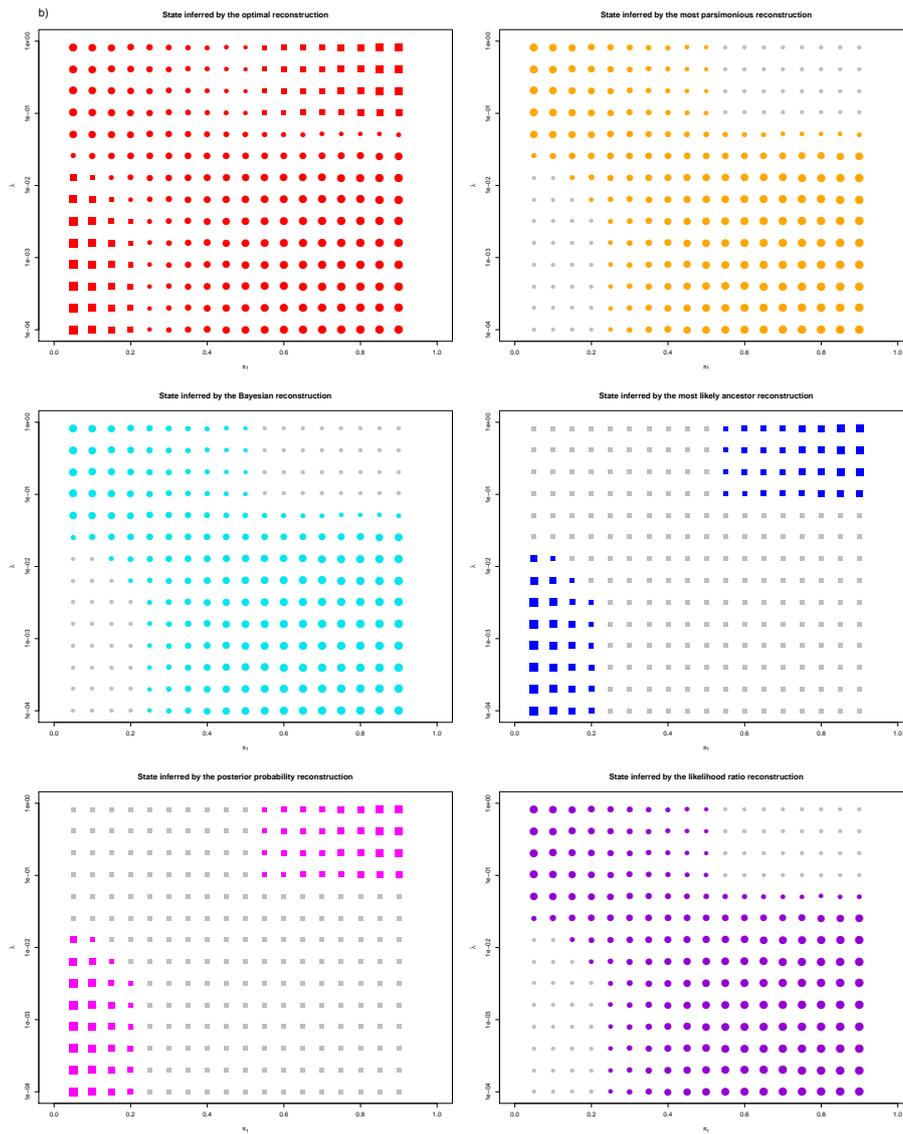


Figure 5: A particular tips configuration of tree 2a), used for illustrating the protocol. The states of the tips are either 1 for tips with a rectangular frame printed around their label and a blue background, or 0 for the other tips. The root and the internal nodes A and B will be reconstructed.





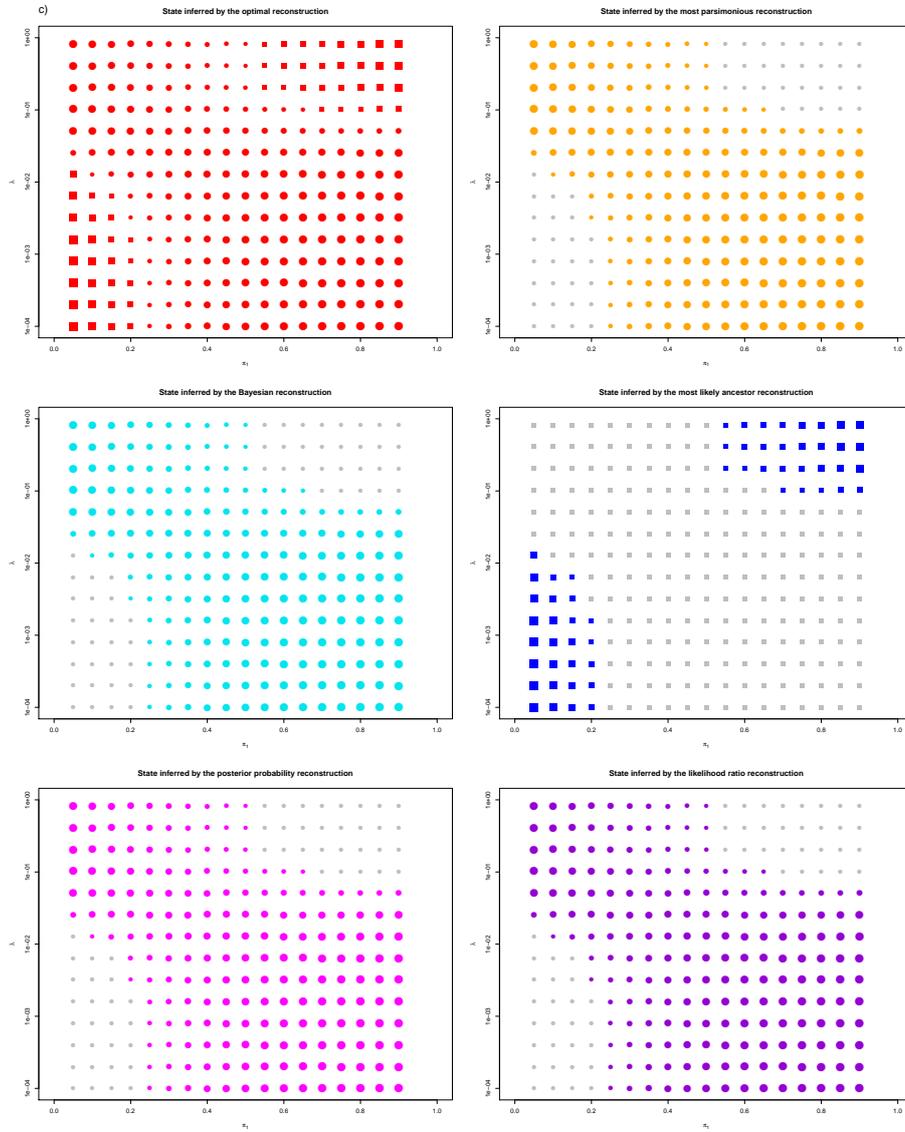


Figure 6: From the particular tips configuration given in Fig. 5, the root and the internal nodes A and B are inferred by \mathcal{M} -optimal, most parsimonious, Bayesian, most likely ancestor, likelihood ratio and posterior probability reconstructions. The inferred state does not depend on the parameters (λ, π_1) ; 0 is represented with a solid circle and 1 by a solid square. But the symbol size varies according to λ and π_1 values, because it represents the associated conditional expectation of success. When the conditional expectation of success is greater than 0.5, the reconstruction method is considered as accurate and the symbol is colored, else it remains grey. Note that for the \mathcal{M} -optimal reconstruction, symbols are always colored, whatever parameters λ and π_1 . Results are displayed with a log-scale on y -axis for the root in Fig. 6a), for internal node A in 6b) and for internal node B in 6c).

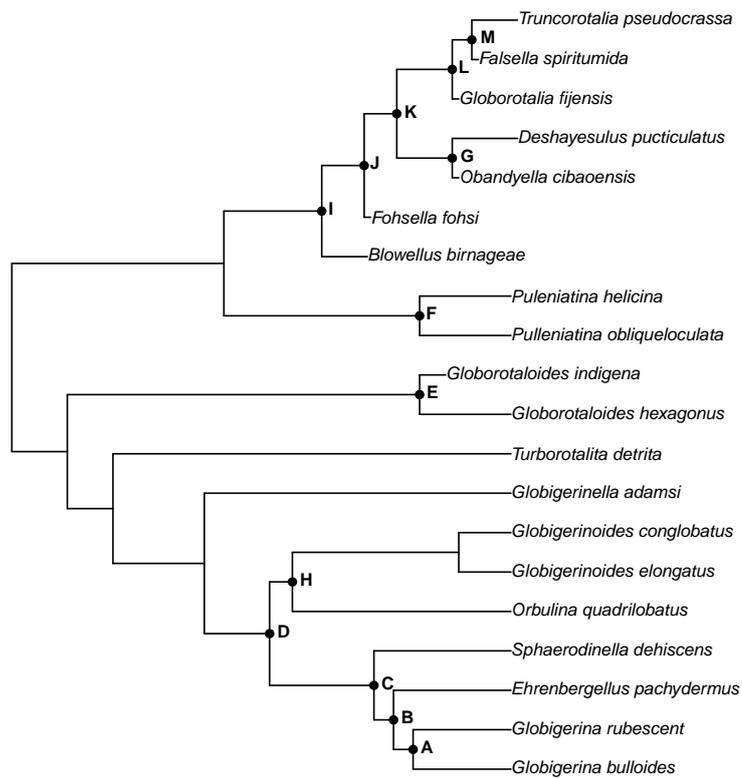


Figure 7: The tree used for supporting the protocol : Pleistocene planktic Foraminifera phylogeny (Webster and Purvis, 2002), with 13 ancestral nodes labeled from A to M.

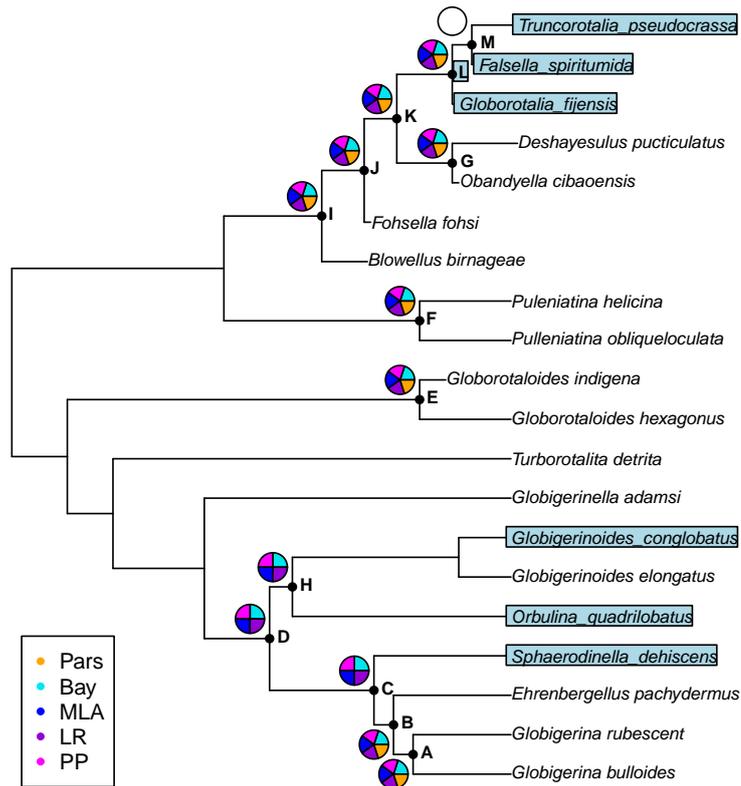


Figure 8: The tree used for supporting the protocol with the area states given for the contemporary species and also for the 13 ancestors. The states of the tips are either 1 for tips with a rectangular frame printed around their label and a blue background, or 0 for the other tips. For every ancestral state, a pie is drawn and filled with the colors associated to the accurate reconstruction methods, in terms of predicting the theoretical state given in Webster and Purvis (2002).

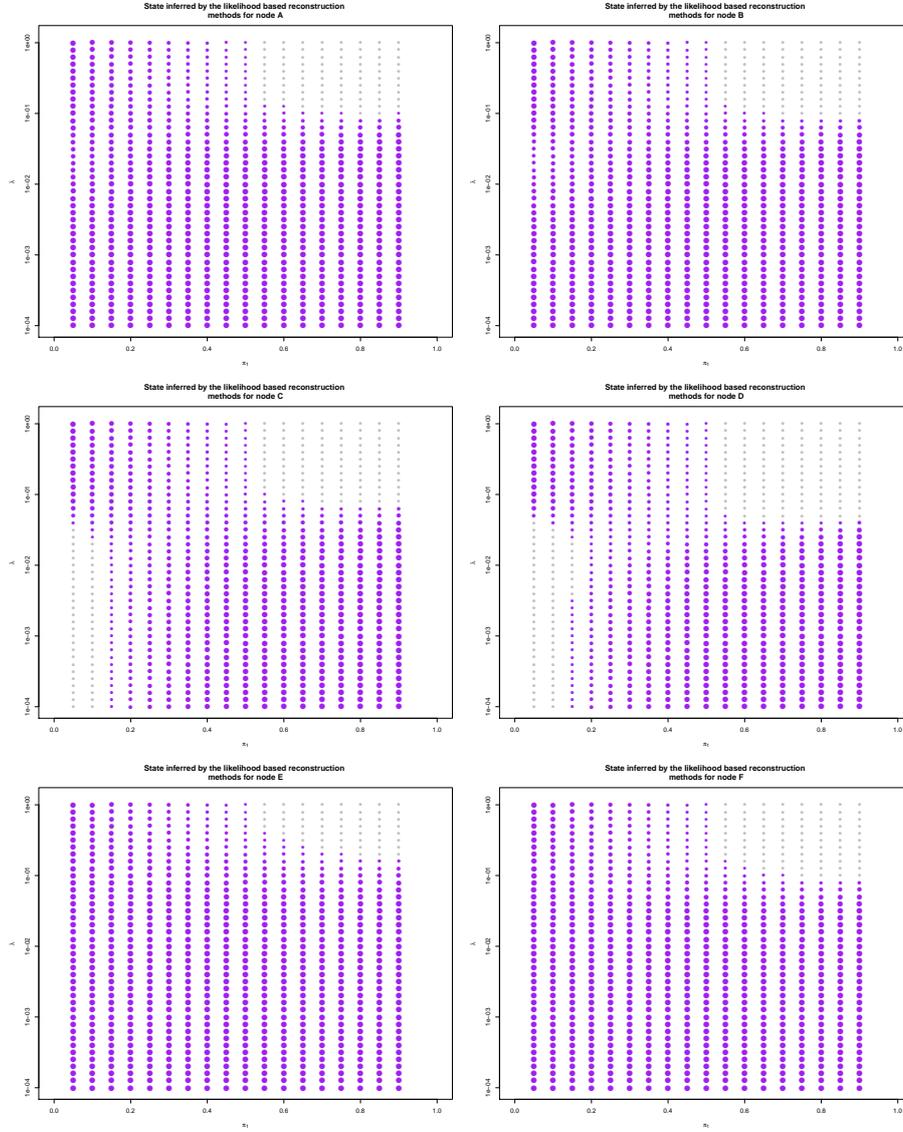


Figure 9: Probability of observing state 0 (solid circle) or 1 (solid square) under models (λ, π_1) , with a log-scale on y -axis. The symbol size varies according to λ and π_1 values, because it represents the associated conditional expectation of success. When the conditional expectation of success is greater than 0.5, the reconstruction method is considered as accurate and the symbol is colored, else it remains grey. Since all likelihood-based methods coincide when reconstructing nodes A to L, only the common graphic is given for every node, instead of one by method.

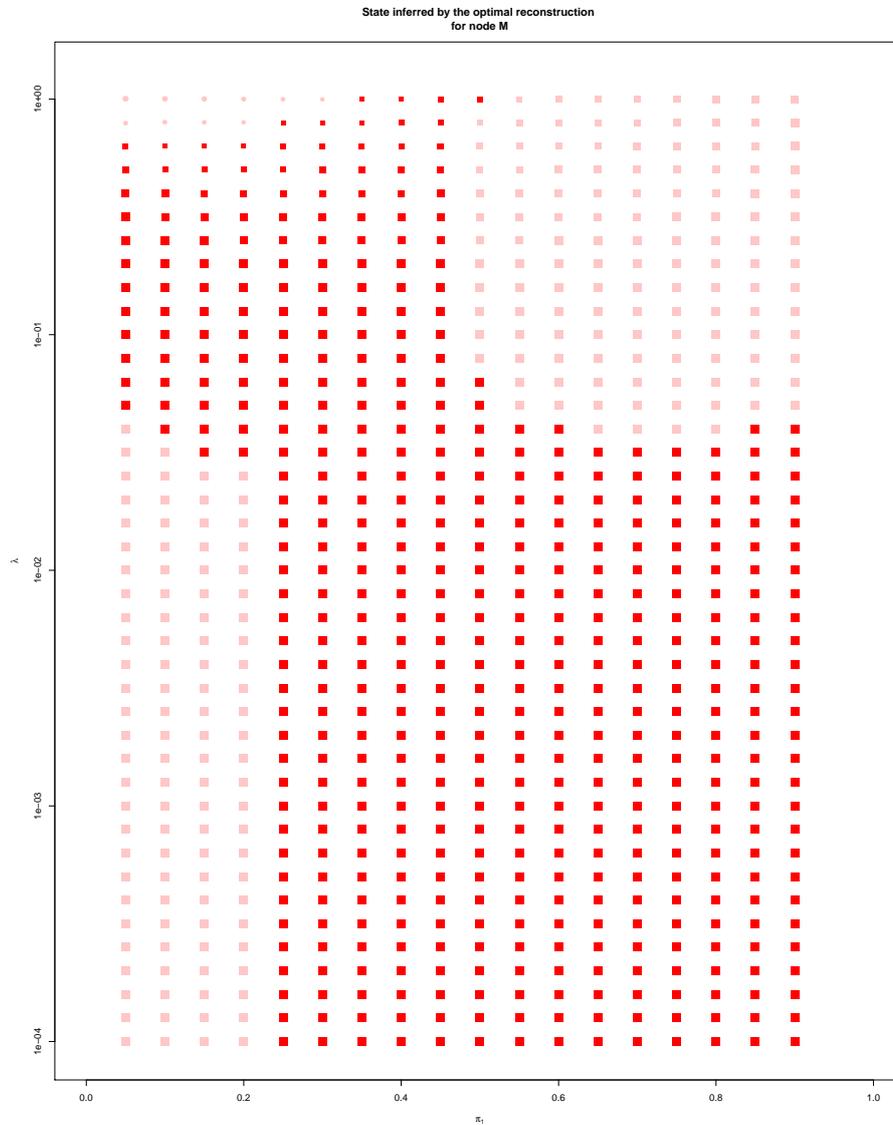


Figure 10: Probability of observing state 0 (solid circle) or 1 (solid square) under models (λ, π_1) , with a log-scale on y -axis. The symbol size varies according to λ and π_1 values, because it represents the associated conditional expectation of success for the optimal reconstruction method. By construction of the optimal reconstruction method, its conditional expectation of success is always greater than 0.5. We lightened the symbols associated to models (λ, π_1) not convenient with accurate reconstructions of nodes A to L.