Time-dependent-asymmetric-linear-parsimonious ancestral state reconstruction

Gilles Didier

Time-dependent-asymmetric-linear-parsimonious ancestral state reconstruction

Gilles Didier

Aix-Marseille Université, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, FRANCE

gilles.didier@univ-amu.fr

November 25, 2016

Abstract

The time-dependent-asymmetric-linear parsimony is an ancestral state reconstruction method which extends the standard linear parsimony (a.k.a. Wagner parsimony) approach by taking into account both branch lengths and asymmetric evolutionary costs for reconstructing quantitative characters (asymmetric costs amount to assuming an evolutionary trend toward the direction with the lowest cost).

A formal study of the influence of the asymmetry parameter shows that the time-dependent-asymmetric-linear parsimony infers states which are all taken among the known states, except for some degenerate cases corresponding to special values of the asymmetry parameter. This remarkable property holds in particular for the Wagner parsimony.

This study leads to a polynomial algorithm which determines, and provides a compact representation of, the parametric reconstruction of a phylogenetic tree, that is for all the unknown nodes, the set of all the possible reconstructed states associated to the asymmetry parameters leading to them. The time-dependent-asymmetric-linear parsimony is finally illustrated with the parametric reconstruction of the body size of cetaceans.

Keywords : maximum parsimony, ancestral state reconstruction, quantitative characters, parametric methods

1 Introduction

Testing hypotheses about evolutionary mechanisms like environment influence, homoplasy etc. calls for information not only about the extant organisms but also about the ancestral ones, which is mostly inaccessible, with a few exceptions when related fossils can be found. More generally, how to study the evolution since we cannot observe the process ongoing over a significant time scale, but only see its result? A common way to cope with this issue is to infer the unknown, and essentially unknowable, information about the ancestral organisms from that observed on the extant taxa, by using the phylogenetic relationships between these last ones. Such an inference is sometimes called *character mapping* or more often, *ancestral state reconstruction* [12, 7, 2]. Note that the problem of determining the phylogenetic relationships between extant taxa is generally treated independently of, and prior to, the ancestral reconstruction. In short, ancestral state reconstruction approaches generally assume that both the phylogeny and the character states of extant taxa are given and aim to infer the ancestral states from these data.

Ancestral reconstruction is a challenging and important question in evolutionary biology. As such, it motivated the development of several approaches [5, 9, 8, 6, 13]. They all rely on two general points of view which are quite different. Namely, ancestral reconstruction methods are based either on the parsimony principle or on stochastic models of character evolution. Note that reconstruction approaches differs not only in the principle underlying them, but also in the nature of the states that they aim to reconstruct. One does not use the same methods for reconstructing continuous/quantitative features (e.g. size, weight, cranial volume...) or discrete characters which are here the characters taking only a finite number of values like the number of fingers, the presence or absence of a given feature etc. In [12], authors differentiate discrete characters according to the way in which their possible values can follow one another during evolution. For instance, *linear ordered characters* have ordered values and are such that evolving from a value to another requires to pass through all the intermediate values.

The present work focuses on the parsimonious method for reconstructing continuous character in which the cost of an evolution is the absolute difference between its ending and starting states. This method is referred to as the Wagner or the linear parsimony [4, 11, 1]. Following [1], we consider an asymmetric version of this method, allowing the cost of an increasing evolution to be different from that of a decreasing evolution of the same amount. Moreover, we allow the cost of a character variation during an interval of time, to depend on the length of this interval (in practice, lengths of intervals of times are branch lengths). The reconstruction method such obtained is called *time-dependent-asymmetric-linear parsimony* (*TDALP*).

Our starting point is a detailed study of the influence of asymmetry parameter on the reconstructed states. This study is based on the characterization of the function associating a state x and an asymmetry parameter γ, with the smallest cost, under γ, of a reconstruction forced to infer x at the root of the tree [1]. The properties of this function allow us to prove that, whatever the asymmetry parameter, there always exists a reconstruction in which all the inferred states are taken among the known states (generally those of the tips, but we make no assumption on this point). Reconstructions containing states which are not among the known ones are degenerates cases which only occurs for a finite number of special values of the asymmetry parameter. This strong property, which is much stronger than saying that the reconstructed states lies in the range of the known ones, makes the approach totally relevant to deal with discrete characters, in particular with linear ordered ones. To our knowledge, the TDALP is the only reconstruction approach which can be applied to both discrete and continuous characters. For instance, reconstruction methods based on stochastic models handling continuous character use Brownian motion to model their evolution, and are very different to those dealing with discrete characters which use Markov models for the same purpose.

The results obtained about the influence of the asymmetry parameter are next used for designing an algorithm which determines, for all unknown nodes of the tree, the different states reconstructed by the TDALP according to the asymmetry parameter. For an unknown node, the set of all the possible reconstructed states with the corresponding asymmetry parameters, is called its *parametric reconstruction*. The algorithmic complexity of the algorithm is polynomial (quartic with the size of tree) both in time and memory space. The TDALP can be applied to a wide variety of biological datasets for reconstructing characters of any nature whatsoever (see above).

I developed a software implementing the approach presented here. This software takes as inputs a phylogenetic tree with or without branch lengths, and a file containing the known character states (in standard ".csv" table format) and outputs the parametric reconstruction of all the unknown nodes in several formats, notably in graphical ones as displayed in Figures 2 and 3. Its source code, written in C language, is freely available at `https://github.com/gilles-didier/TDALP`.

The rest of the paper is organized as follows. We introduce the notations, a formal presentation of the ancestral reconstruction problem and the first definitions in Section 2. The detailed study of the influence of the asymmetry parameter is carried out in Section 3. It leads to the definition of the parametric reconstruction of a character at a node, which is formally defined at the beginning of Section 4. This section continues with the presentation of an algorithm computing the parametric reconstruction of all the unknown nodes of a given tree and the study of its complexity. We conclude with Section 5 by illustrating our approach with the reconstruction of the body size of cetaceans.

2 Definitions and notations

The cardinal of a finite set $\mathcal{S}$ is noted $|\mathcal{S}|$.

Let T be a rooted tree which is not required to be binary. As it should lead to no confusion, we still write T for its set of nodes. For all nodes $n \in T$, we put

- $\mathcal{C}_n$ for the set of child nodes of n,
- τ_n for the length of the branch ending at n,
- a_n for the direct ancestor of n,
- T_n for the subtree of T rooted at n.

Before introducing the reconstruction problem, let us start by considering a subset $\mathcal{K}$ of of nodes of T, a map ϑ from $\mathcal{K}$ to the set of real numbers $\mathbb{R}$. The map ϑ will be referred to as the *initial function* and the nodes of $\mathcal{K}$ are said *known*. In the standard ancestral state reconstruction problem, $\mathcal{K}$ is exactly the set of tips of T but we make here no assumption on $\mathcal{K}$. In plain English, any node of the tree can be known or not. For all nodes n of T, we put $\mathcal{K}_n$ for the subset of known nodes of the subtree T_n, i.e. $\mathcal{K}_n = \mathcal{K} \cap T_n$. The values of $\{\vartheta(k) \mid k \in \mathcal{K}\}$ are the *known states* of T. For all nodes n, we put $\vartheta(\mathcal{K}_n)$ for the set $\{\vartheta(k) \mid k \in \mathcal{K}_n\}$.

For technical reasons, we will consider two special nodes o and p not belonging to T, for which, by convention, we set $\vartheta(o) = -\infty$ and $\vartheta(p) = +\infty$.

The *ϑ-assignments* of T are the maps ξ from T to the set of real numbers which extend ϑ (i.e. such that $\xi(n) = \vartheta(n)$ for all nodes $n \in \mathcal{K}$). The reconstruction problem of T with regard to ϑ consists in finding the most relevant (in some sense) ϑ-assignment of T.

2.1 Parsimonious reconstruction

In the parsimony framework, the relevance of an assignment is expressed in terms of cost. More specifically, in the time-dependent-asymmetric-linear parsimony (*TDALP*) case, an ancestor/child transition from value x at node n to value y at its child m is associated with the cost:

$$\Delta_{\gamma,\lambda}(x, y, \tau_m) = \begin{cases} \gamma\phi(\tau_m)(y - x) & \text{if } x < y, \\ \lambda\phi(\tau_m)(x - y) & \text{if } x \geq y, \end{cases}$$

where λ and γ are two positive real numbers and ϕ is a function from $\mathbb{R}_{>0}$ to $\mathbb{R}_{>0}$. In what follows, we make the assumption that $\tau_n > 0$ for all nodes $n \in T$. In other words, we forbid null branch lengths (but we allow polytomies). Remark that the generalized linear parsimony as defined in [1], which generalizes the Wagner parsimony [4], corresponds to the case where ϕ is constant with $\phi(\tau) = 1$ for all τ. In an evolutionary context, it makes sense to choose a decreasing function ϕ (roughly speaking, evolving takes times), but any positive function can be used. In Section 5, we set $\phi(\tau) = \frac{1}{\tau}$ for reconstructing the body size of cetaceans. The special case where $\gamma = \lambda = 1$ and

$\phi(\tau) = 1$ for all τ, corresponds to the Wagner parsimony. Remark that steadiness is always cost-free, i.e. $\Delta_{\gamma,\lambda}(x, x, \tau) = 0$ whatever x, τ, γ, λ and ϕ.

The cost $\Delta_{\gamma,\lambda}(\xi)$ of an assignment ξ of T is then the sum of all the costs of its ancestor/child transitions:

$$\Delta_{\gamma,\lambda}(\xi) = \sum_{n \in T} \sum_{m \in \mathcal{C}_n} \Delta_{\gamma,\lambda}(\xi(n), \xi(m), \tau_m)$$

Finding the most time-dependent-asymmetric-linear parsimonious reconstruction on T with regard to ϑ consists in determining a ϑ-assignment ξ of T with a minimal cost $\Delta_{\gamma,\lambda}(\xi)$.

Let us first remark that multiplying both λ and γ by a positive constant factor does not change the relative order of the assignments costs. Since $\lambda > 0$, we can divide both parameters by λ without changing which assignments are the most parsimonious. From now on, we assume without loss of generality that $\lambda = 1$. The cost of an assignment only depends on the parameter γ, and is called *γ-cost*:

$$\Delta_\gamma(x, y, \tau_m) = \begin{cases} \gamma\phi(\tau_m)(y - x) & \text{if } x < y, \\ \phi(\tau_m)(x - y) & \text{if } x \geq y. \end{cases} \tag{1}$$

Below, γ will be referred to as the *asymmetry parameter*. Intuitively, reconstructing with $\gamma < 1$ (resp. with $\gamma = 1$, with $\gamma > 1$) corresponds to the assumption that the character evolves with a positive trend (resp. without trend, with a negative trend).

A *γ-parsimonious reconstruction* of T with regard to an initial function ϑ is an ϑ-assignment with a minimum γ-cost.

2.2 Cost and stem cost functions

Let us borrow and adapt some definitions of [1]. For all nodes $n \in T$, the *(subtree) cost function* f_n maps a pair $(\gamma, x) \in \mathbb{R}_{>0} \times \mathbb{R}$ to the smallest γ-cost which can be achieved by an assignment ξ of the subtree T_n such that $\xi(n) = x$. An assignment ξ of T_n satisfying $\xi(n) = x$ will be said *x-rooted*. By construction, $f_n(\gamma, x)$ is thus the γ-cost of a x-rooted γ-parsimonious assignment of T_n. By convention, if there exists no x-rooted assignment of T_n (typically when $n \in \mathcal{K}$ with $\vartheta(n) \neq x$) then we set $f_n(\gamma, x) = +\infty$.

Claim 1. *For all $\gamma > 0$ and all real values x, we have that*

1. *if n is an unknown leaf then $f_n(\gamma, x) = 0$;*
2. *if n is a known leaf then*

$$f_n(\gamma, x) = \begin{cases} 0 & \text{if } x = \vartheta(n), \\ +\infty & \text{otherwise;} \end{cases}$$

3. *if n is an unknown internal node then*

$$f_n(\gamma, x) = \sum_{m \in \mathcal{C}_n} \inf_{y \in \mathbb{R}} (\Delta_\gamma(x, y, \tau_m) + f_m(\gamma, y));$$

4. *if n is a known internal node then*

$$f_n(\gamma, x) = \begin{cases} \sum_{m \in \mathcal{C}_n} \inf_{y \in \mathbb{R}} (\Delta_\gamma(x, y, \tau_m) + f_m(\gamma, y)) & \text{if } x = \vartheta(n), \\ +\infty & \text{otherwise.} \end{cases}$$

Items 3 and 4 of Claim 1 lead us to introduce an additional notation. For all non-root nodes m of T, we define the *stem cost function* $\widehat{f}_m$ as

$$\widehat{f}_m(\gamma, x) = \inf_{y \in \mathbb{R}} (\Delta_\gamma(x, y, \tau_m) + f_m(\gamma, y)),$$

which can be understood as the smallest γ-cost which can be achieved by an assignment of the tree only made of the direct ancestor of m and of the subtree T_m which associates the direct ancestor of m, with the state x. Equation of Item 3 in Claim 1 becomes

$$f_n(\gamma, x) = \sum_{m \in \mathcal{C}_n} \widehat{f}_m(\gamma, x) \tag{2}$$

For all non-root nodes m of T, the tree made of the direct ancestor of m and of the subtree T_m will be referred to as the *stem-subtree* of m and noted $\widehat{T}_m$.

Proving properties of f_n and $\widehat{f}_n$ shall follow a same general induction scheme which will be referred to as the *standard induction scheme*. It comes from Lemma 1 of [1] and stands in the three following steps:

Step 1: establish the property for the base cases of $\widehat{f}_n$ and f_n;

Step 2: prove that if the property holds for f_n then the same property holds for $\widehat{f}_n$;

Step 3: prove that if the property holds for all $\widehat{f}_m$ where m is a child of n then it holds for f_n.

Steps 1 and 3 are generally plain. Both maps f_n and $\widehat{f}_n$ have very simple forms for the base cases (i.e. the leaves). In most cases, Step 3 follows straightforwardly from Equation 2. The main point of the proofs actually stands in Step 2.

3 Parametric analysis

We shall study the influence of the parameter γ on the states inferred in the γ parsimonious reconstruction. To this end, we start by studying the nature of cost and stem cost functions.

Theorem 1. *Let n be an unknown node of T. The maps f_n and $\widehat{f}_n$ are piecewise-linear and continuous and, for all $\gamma > 0$, the maps $x \to f_n(\gamma, x)$ and $x \to \widehat{f}_n(\gamma, x)$ are both convex.*

More precisely, if all the nodes of T_n are unknown then $f_n(\gamma, x) = 0$ for all $\gamma > 0$ and all x. Otherwise, there exist:

- *an integer u_n and a strictly increasing positive real sequence $(\Gamma_i^n)_{1 \leq i \leq u_n}$,*
- *an integer sequence $(v_i^n)_{0 \leq i \leq u_n}$ and for, all $0 \leq i \leq u_n$, a sequence $(e_{i,j}^n)_{1 \leq j \leq v_i^n}$ of known nodes of T_n (i.e. of $\mathcal{K}_n$) verifying $\vartheta(e_{i,j}^n) < \vartheta(e_{i,j'}^n)$ iff $j < j'$; by convention we set $e_{i,0}^n = o$ and $e_{i,v_i^n+1}^n = p$ (we recall that $\vartheta(o) = -\infty$ and $\vartheta(p) = +\infty$),*
- *two non-negative real bi-sequences $(A_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$ and $(B_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$,*
- *two real bi-sequences $(C_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$ and $(D_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$*

such that, by setting $\Gamma_0^n = 0$ and $\Gamma_{u_n+1}^n = +\infty$ and for all $0 \leq i \leq u_n$, all $0 \leq j \leq v_i^n$, all $\gamma \in (\Gamma_i^n, \Gamma_{i+1}^n]$ and all $x \in (\vartheta(e_{i,j}^n), \vartheta(e_{i,j+1}^n)]$, we have

$$f_n(\gamma, x) = -A_{i,j}^n \gamma x + B_{i,j}^n x + C_{i,j}^n \gamma + D_{i,j}^n,$$

all the coefficients being such that f_n is continuous. Moreover, the sequence $(-A_{i,j}^n \gamma + B_{i,j}^n)_{0 \leq j \leq v_i^n}$ (i.e. the x-coefficients of f_n) is increasing with $A_{i,0}^n = B_{i,v_i^n}^n \leq \sum_{m \in \mathcal{C}_n} \phi(\tau_m)$ and $A_{i,v_i^n}^n = B_{i,0}^n = 0$.

In the same way, if all the nodes of T_n are unknown then $\widehat{f}_n(\gamma, x) = 0$ for all $\gamma > 0$ and all x. Otherwise there exist $\widehat{u}_n$, $(\widehat{\Gamma}_i^n)_{1 \leq i \leq \widehat{u}_n}$, $(\widehat{v}_i^n)_{0 \leq i \leq u_n}$, $(\widehat{e}_{i,j}^n)_{0 \leq i \leq \widehat{u}_n, 1 \leq j \leq \widehat{v}_i^n}$, $(\widehat{A}_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq \widehat{v}_i^n}$, $(\widehat{B}_{i,j}^n)_{0 \leq i \leq \widehat{u}_n, 0 \leq j \leq \widehat{v}_i^n}$, $(\widehat{C}_{i,j}^n)_{0 \leq i \leq \widehat{u}_n, 0 \leq j \leq \widehat{v}_i^n}$ and $(\widehat{D}_{i,j}^n)_{0 \leq i \leq \widehat{u}_n, 0 \leq j \leq \widehat{v}_i^n}$ verifying the same properties as their f_n-counterparts except that we have $\widehat{A}_{i,0}^n =$

$\widehat{B}^n_{i,\widehat{v}^n_i} \leq \phi(\tau_n)$ *and* $\widehat{A}^n_{i,\widehat{v}^n_i} = \widehat{B}^n_{i,0} = 0$*, and such that, for all* $0 \leq i \leq \widehat{u}_n$*, all* $0 \leq j \leq \widehat{v}^n_i$*, all* $\gamma \in (\widehat{\Gamma}^n_i, \widehat{\Gamma}^n_{i+1}]$ *and all* $x \in (\vartheta(\widehat{e}^n_{i,j}), \vartheta(\widehat{e}^n_{i,j+1})]$*, we have*

$$\widehat{f}_n(\gamma, x) = -\widehat{A}^n_{i,j}\gamma x + \widehat{B}^n_{i,j}x + \widehat{C}^n_{i,j}\gamma + \widehat{D}^n_{i,j}.$$

Proof. We follow the standard induction scheme and start with Step 1. The base cases of f_n, i.e. when n is a leaf, are given by Items 1 and 2 of Claim 1. Let us check the base cases of $\widehat{f}_n$. If n is an unknown leaf then we have $\widehat{f}_n(\gamma, x) = 0$ for all pairs $\gamma > 0$ and all x. If n is a known leaf then we have:

$$\widehat{f}_n(\gamma, x) = \begin{cases} \gamma\phi(\tau_n)(\vartheta(n) - x) & \text{if } x < \vartheta(n), \\ \phi(\tau_n)(x - \vartheta(n)) & \text{otherwise,} \end{cases}$$

for all $\gamma > 0$ and all x. In all cases, $\widehat{f}_n$ is piecewise linear and continuous. Moreover, we remark that, for x small enough (resp. large enough), the x-coefficients of f_n and $\widehat{f}_n$ are either 0 or $-\gamma\phi(\tau_n)$ (resp. either 0 or $\phi(\tau_n)$), according to whether n is unknown or not.

Let us proceed to Step 2 and assume that n is an internal node and that the theorem holds for f_n. If n is known then we have:

$$\widehat{f}_n(\gamma, x) = \begin{cases} \gamma\phi(\tau_n)(\vartheta(n) - x) + f_n(\gamma, \vartheta(n)) & \text{if } x < \vartheta(n), \\ \phi(\tau_n)(x - \vartheta(n)) + f_n(\gamma, \vartheta(n)) & \text{otherwise.} \end{cases}$$

Since f_n is piecewise linear and continuous, the map $\widehat{f}_n$ is well piecewise linear and continuous with $x \to \widehat{f}_n(\gamma, x)$ convex. The same remark as just above about the x-coefficient of $\widehat{f}_n$ holds again.

Let now assume that n is unknown internal node. We still assume that the theorem holds for f_n and we assume in addition that T_n contains at least one known node (the other case being straightforward). Let us consider an index $0 \leq i < u_n$ and a real number $\gamma \in [\Gamma^n_i, \Gamma^n_{i+1})$. With the induction hypothesis, the map $y \to f_n(\gamma, y)$ is convex, continuous and piecewise linear. So is the map defined for any real value x and all $y \geq x$ by

$$g_x(y) = \Delta_\gamma(x, y, \tau_n) + f_n(\gamma, y)$$

For all real values x and $y \geq x$, we have that $g_x(y) = \gamma\phi(\tau_n)(y - x) + f_n(\gamma, y)$, which, under the notations of the theorem, can be written as:

$$g_x(y) = (-(A^n_{i,j} - \phi(\tau_n))\gamma + B^n_{i,j})y - \gamma\phi(\tau_n)x + C^n_{i,j}\gamma - D^n_{i,j}.$$

From the induction hypothesis, the sequence $(-(A^n_{i,j} - \phi(\tau_n))\gamma + B^n_{i,j})_j$ increases with j. Let j^+_γ be the smallest index such that $-(A^n_{i,j^+_\gamma} - \phi(\tau_n))\gamma + B^n_{i,j^+_\gamma} \geq 0$ (i.e. that corresponding to the first interval on which g_x does not decrease with $y \geq x$). Since, from our induction hypothesis, we have that $A^n_{k_n, v^n_i} = 0$, such an integer always exists and we get that

$$\inf_{y \geq x} g_x(y) = \begin{cases} g_x(\vartheta(e^n_{i,j^+_\gamma})) & \text{if } x < \vartheta(e^n_{i,j^+_\gamma}), \\ g_x(x) & \text{otherwise.} \end{cases}$$

Let us remark that

$$g_x(\vartheta(e^n_{i,j^+_\gamma})) \leq g_x(x) = f_n(\gamma, x) \quad \text{for all } x < \vartheta(e^n_{i,j^+_\gamma}). \tag{3}$$

Symmetrically, for all real values x and $y \leq x$, we have that $g_x(y) = \phi(\tau_n)(x - y) + f_n(\gamma, y)$, i.e.

$$g_x(y) = (-A^n_{i,j}\gamma + B^n_{i,j} - \phi(\tau_n))y + \phi(\tau_n)x + C^n_{i,j}\gamma - D^n_{i,j}.$$

Let us define j^-_γ as the greatest integer smaller or equal to $v^n_i + 1$ and such that $-A^n_{i,j^-_\gamma - 1}\gamma + B^n_{i,j^-_\gamma - 1} - \phi(\tau_n) \leq 0$ (i.e. that corresponding to the last interval on which

g_x does decrease with $y \leq x$). Since, from our induction hypothesis, we have that $B^n_{k_n,0} = 0$, such an integer always exists. We have that

$$\inf_{y \leq x} g_x(y) = \begin{cases} g_x(x) & \text{if } x < \vartheta(e^n_{i,j^-_\gamma}), \\ g_x(\vartheta(e^n_{i,j^-_\gamma})) & \text{otherwise.} \end{cases}$$

Let us remark that

$$g_x(\vartheta(e^n_{i,j^+_\gamma})) \leq g_x(x) = f_n(\gamma, x) \quad \text{for all } x \geq \vartheta(e^n_{i,j^-_\gamma}). \tag{4}$$

Since both γ and $\phi(\tau_n)$ are positive, we have that

$$-(A^n_{i,j} - \phi(\tau_n))\gamma + B^n_{i,j} \geq -A^n_{i,j}\gamma + B^n_{i,j} - \phi(\tau_n).$$

It follows that $j^+_\gamma \leq j^-_\gamma$ therefore $\vartheta(e^n_{i,j^+_\gamma}) \leq \vartheta(e^n_{i,j^-_\gamma})$.

Moreover, since

$$\begin{aligned} \widehat{f}_n(\gamma, x) &= \inf_y(\Delta_\gamma(x, y, \tau_n) + f_n(\gamma, y)) \\ &= \min\{\inf_{y \leq x} g_x(y), \inf_{y \geq x} g_x(y)\}, \end{aligned}$$

the inequalities 3 and 4 imply that

$$\widehat{f}_n(\gamma, x) = \begin{cases} \gamma\phi(\tau_n)(\vartheta(e^n_{i,j^+_\gamma}) - x) + f_n(\gamma, \vartheta(e^n_{i,j^+_\gamma})) & \text{if } x < \vartheta(e^n_{i,j^+_\gamma}), \\ f_n(\gamma, x) & \text{if } \vartheta(e^n_{i,j^+_\gamma}) \leq x < \vartheta(e^n_{i,j^-_\gamma}), \\ \phi(\tau_n)(x - \vartheta(e^n_{i,j^-_\gamma})) + f_n(\gamma, \vartheta(e^n_{i,j^-_\gamma})) & \text{if } x \geq \vartheta(e^n_{i,j^-_\gamma}). \end{cases}$$

It straightforwardly follows that, for all γ, the map $x \to \widehat{f}_n(\gamma, x)$ is piecewise linear and continuous. From the induction assumption, $x \to f_n(\gamma, x)$ is convex, in particular between $\vartheta(e^n_{i,j^+_\gamma})$ and $\vartheta(e^n_{i,j^-_\gamma})$. Still from the induction assumption, we have that $A^n_{i,0} = B^n_{i,v^n_i}$ and $A^n_{i,v^n_i} = B^n_{i,0} = 0$. Two possibilities arise:

- If $A^n_{i,0} = B^n_{i,v^n_i} \leq \phi(\tau_n)$ then both $j^+_\gamma = 0$ and $j^-_\gamma = v^n_i + 1$. We have $\widehat{f}_n(\gamma, x) = f_n(\gamma, x)$. From the induction hypothesis, the function $x \to \widehat{f}_n(\gamma, x)$ is well convex and such that $\widehat{A}^n_{i,0} = \widehat{B}^n_{i,\widehat{v}^n_i} \leq \phi(\tau_n)$ and $\widehat{A}^n_{i,\widehat{v}^n_i} = \widehat{B}^n_{i,0} = 0$.
- If $A^n_{i,0} = B^n_{i,v^n_i} > \phi(\tau_n)$, then we have both $j^+_\gamma > 0$ and $j^-_\gamma \leq v^n_i$. The definition of j^+_γ and of j^-_γ ensures that $-A^n_{i,j^+_\gamma+1}\gamma + B^n_{i,j^+_\gamma+1} \geq -\gamma\phi(\tau_n)$ and $-A^n_{i,j^-_\gamma-1}\gamma + B^n_{i,j^-_\gamma-1} \leq \phi(\tau_n)$, which implies that the map $x \to \widehat{f}_n(\gamma, x)$ is still convex, in this case with $\widehat{A}^n_{i,0} = \widehat{B}^n_{i,\widehat{v}^n_i} = \phi(\tau_n)$ and $\widehat{A}^n_{i,\widehat{v}^n_i} = \widehat{B}^n_{i,0} = 0$.

Let us show that $\widehat{f}_n$ is piecewise linear with regard to γ. To do so, we put $\Phi_1, \ldots, \Phi_p$ for the elements of

$$\left(\left\{\frac{B^n_{i,j}}{A^n_{i,j} - \phi(\tau_n)}\middle| 0 \leq j \leq v^n_i\right\} \bigcup \left\{\frac{B^n_{i,j} - \phi(\tau_n)}{A^n_{i,j}}\middle| 0 \leq j \leq v^n_i\right\}\right) \bigcap (\Gamma^n_i, \Gamma^n_{i+1}),$$

indexed in increasing order. By construction, the indexes j^+_γ and j^-_γ are both constant over all the sub-intervals $[\Gamma^n_i, \Phi_1)$, $[\Phi_1, \Phi_2)$, $\ldots$, $[\Phi_{p-1}, \Phi_p)$, $[\Phi_p, \Gamma^n_{i+1})$. Since from our induction hypothesis $\gamma \to f_n(\gamma, x)$ is linear over $(\Gamma^n_i, \Gamma^n_{i+1})$ and for all values x, the map $\gamma \to \widehat{f}_n(\gamma, x)$ is linear over all the sub-intervals above. It follows that the map $\widehat{f}_n$ is piecewise linear and its continuity with regard to γ is straightforward to verify at all bounds Φ_k for $1 \leq k \leq p$.

Step 3 is the last one remaining. Let n be an unknown internal node and let us assume that the theorem holds for all children m of n. In particular, the maps $\widehat{f}_m$ are piecewise linear, continuous and convex with regard to x. As sum of these maps (Equation 2), the map f_n satisfies itself these properties. Moreover, since for all children m of n, $\widehat{A}^m_{i,0} = \widehat{B}^m_{i,\widehat{v}^m_i} \leq \phi(\tau_m)$ and $\widehat{A}^m_{i,\widehat{v}^m_i} = \widehat{B}^m_{i,0} = 0$, we have that $A^n_{i,0} = B^n_{i,v^n_i} \leq \sum_{m \in \mathcal{C}_n} \phi(\tau_m)$ and $A^n_{i,v^n_i} = B^n_{i,0} = 0$. □

Figure 1 shows an example of cost function and its graphical representation.

For all functions f_n (resp. $\widehat{f}_n$) and all $0 < i < u_n$ (resp. all $0 < \ell < \widehat{u}_n$), the bounds Γ_i^n (resp. $\widehat{\Gamma}_i^n$) will be referred to as *asymmetry-bounds* of f_n (resp. of $\widehat{f}_n$) and, for all $0 < j < v_i^n$ (resp. all $0 < j < \widehat{v}_i^n$), the bounds $\vartheta(e_{i,j}^n)$ (resp. $\vartheta(\widehat{e}_{i,j}^n)$) will be referred to as *state-bounds* of f_n (resp. of $\widehat{f}_n$).

Remark 1. *In the generalized parsimony scheme of [1] (i.e. when $\phi(\tau) = 1$ for all τ), both $(A_{i,j}^n)_{i,j}$ and $(B_{i,j}^n)_{i,j}$ are integer bi-sequences and the sequence $(\Gamma_i^n)_i$ is rational.*

Since $\phi(\tau_k)$ is assumed positive for all nodes $k \in T$, the following remark is straightforward to prove by induction.

Remark 2. *If n is an unknown node such that $\mathcal{K}_n \neq \emptyset$, there exists no pair $(i,j) \in \{0, \dots u_n\} \times \{0, \dots, v_i^n\}$ such that $A_{i,j}^n = B_{i,j}^n = 0$.*

The proof of the following corollary is essentially contained in that of Theorem 1.

Corollary 1. *Let γ be a positive real number, n a node of T, m a child of n and, under the notations of Theorem 1, i be the greatest index such that $\gamma \leq \widehat{\Gamma}_{i+1}^m$. If a γ-parsimonious reconstruction assigns the value x to n then it assigns to all unknown children m of n,*

1. *a value between $\vartheta(\widehat{e}_{i,1}^m)$ and $\vartheta(\widehat{e}_{i,2}^m)$ (both included) if $x \leq \vartheta(\widehat{e}_{i,2}^m)$ and $-(\widehat{A}_{i,1}^m - \phi(\tau_m))\gamma - \widehat{B}_{i,1}^m = 0$,*
2. *$\vartheta(\widehat{e}_{i,1}^m)$ if $x \leq \vartheta(\widehat{e}_{i,1}^m)$ and $-(\widehat{A}_{i,1}^m - \phi(\tau_m))\gamma + \widehat{B}_{i,1}^m \neq 0$,*
3. *a value between $\vartheta(\widehat{e}_{i,\widehat{v}_i^m-1}^m)$ and $\vartheta(\widehat{e}_{i,\widehat{v}_i^m}^m)$ (both included) if $x \geq \vartheta(\widehat{e}_{i,\widehat{v}_i^m-1}^m)$ and $-\widehat{A}_{i,\widehat{v}_i^m-1}^m\gamma + \widehat{B}_{i,\widehat{v}_i^m-1}^m + \phi(\tau_m) = 0$,*
4. *$\vartheta(\widehat{e}_{i,\widehat{v}_i^m}^m)$ if $x \geq \vartheta(\widehat{e}_{i,\widehat{v}_i^m}^m)$ and $-\widehat{A}_{i,\widehat{v}_i^m-1}^m\gamma + \widehat{B}_{i,\widehat{v}_i^m-1}^m - \phi(\tau_m) \neq 0$.*
5. *x otherwise.*

Cases 1 (resp. Case 3) occurs if and only if $\gamma = \frac{\widehat{B}_{i,1}^m}{\widehat{A}_{i,1}^m - \phi(\tau_m)} = \widehat{\Gamma}_{i+1}^m$ (resp. $\gamma = \frac{\widehat{B}_{i,\widehat{v}_i^m-1}^m - \phi(\tau_m)}{\widehat{A}_{i,\widehat{v}_i^m-1}^m} = \widehat{\Gamma}_i^m$). In particular, if we have both $-(\widehat{A}_{i,1}^m - \phi(\tau_m))\gamma + \widehat{B}_{i,1}^m \neq 0$ and $-\widehat{A}_{i,\widehat{v}_i^m-1}^m\gamma + \widehat{B}_{i,\widehat{v}_i^m-1}^m - \phi(\tau_m) \neq 0$, which is always true if $\gamma \in (\widehat{\Gamma}_i^m, \widehat{\Gamma}_{i+1}^m)$, then if a γ-parsimonious reconstruction assigns the value x to n then it assigns to all unknown children m of n,

- *$\vartheta(\widehat{e}_{i,1}^m)$ if $x < \vartheta(\widehat{e}_{i,1}^m)$,*
- *x if $\vartheta(\widehat{e}_{i,1}^m) \leq x < \vartheta(\widehat{e}_{i,\widehat{v}_i^m}^m)$,*
- *$\vartheta(\widehat{e}_{i,\widehat{v}_i^m}^m)$ if $x \geq \vartheta(\widehat{e}_{i,\widehat{v}_i^m}^m)$,*

(this point was already stated in [1]).

Theorem 2. *For all asymmetry parameters γ, there exists a γ-parsimonious reconstruction in which all the reconstructed states belong to $\vartheta(\mathcal{K})$. Moreover, if a reconstruction is both γ- and γ'-parsimonious for two positive real numbers $\gamma \neq \gamma'$ then its inferred states all belong to $\vartheta(\mathcal{K})$.*

Proof. Let us start by proving that for all $\gamma > 0$, there exists a γ-parsimonious reconstruction which assigns a known value to the root. This is plain if the root r is known. If r is unknown then, under the notations of Theorem 1 and for all $\gamma > 0$, the map $x \to f_r(\gamma, x)$ is piecewise linear, continuous and convex with

$$f_r(\gamma, x) = -A_{i_\gamma,j}^r \gamma x + B_{i_\gamma,j}^r x + C_{i_\gamma,j}^r \gamma + D_{i_\gamma,j}^r,$$

for all $0 \leq j \leq v^r_{i_\gamma}$ and all $x \in [\vartheta(e^r_{i_\gamma,j}), \vartheta(e^r_{i_\gamma,j+1}))$, where i_γ is such that $\gamma \in [\Gamma^r_{i_\gamma}, \Gamma^r_{i_\gamma+1})$.

Let j_γ be the smallest index such that $-A^r_{i_\gamma,j_\gamma}\gamma + B^r_{i_\gamma,j_\gamma} \geq 0$, which always exists (Theorem 1). Since the sequence $(-A^r_{i_\gamma,j}\gamma + B^r_{i_\gamma,j})_j$ is increasing, we have that

$$\arg\min_{x \in \mathbb{R}} f_r(\gamma, x) = \vartheta(e^r_{i_\gamma,j_\gamma}).$$

Moreover, there exists a real $z \neq \vartheta(e^r_{i_\gamma,j_\gamma})$ such that $\arg\min_{x\in\mathbb{R}} f_r(\gamma, x) = z$ if and only if $-A^r_{i_\gamma,j_\gamma}\gamma + B^r_{i_\gamma,j_\gamma} = 0$. From Remark 2, this implies that if a reconstruction is both γ- and γ'-parsimonious for two positive real numbers $\gamma < \gamma'$ then it assigns $\vartheta(e^r_{i_\gamma,j_\gamma})$ to the root.

By induction and from Corollary 1, if a γ-parsimonious reconstruction assigns a known value to a node n of T then it assigns known values to all nodes of the subtree rooted at n, except for a finite number of particular values of the asymmetry parameter γ. □

In plain English, whatever the function ϕ and the parameter γ, the TDALP reconstructs all the unknown states with values taken among the known states, except for some degenerate cases. In particular, this holds for the Wagner parsimony.

4 Parametric reconstruction

4.1 Definition

Let us start with a remark which follows essentially from the proof of Theorem 2

Remark 3. *Let r be the root of T and $(\dot\Gamma^r_i)_{0\leq i\leq w_r+1}$ be the elements of*

$$\left(\bigcup_{1\leq i\leq u_r}\left(\left\{\frac{B^r_{i,j}}{A^r_{i,j}} \mid 0 \leq j \leq v^r_i\right\} \cap (\Gamma^r_i, \Gamma^r_{i+1})\right)\right) \bigcup \{\Gamma^r_i \mid 0 \leq i \leq u_r + 1\},$$

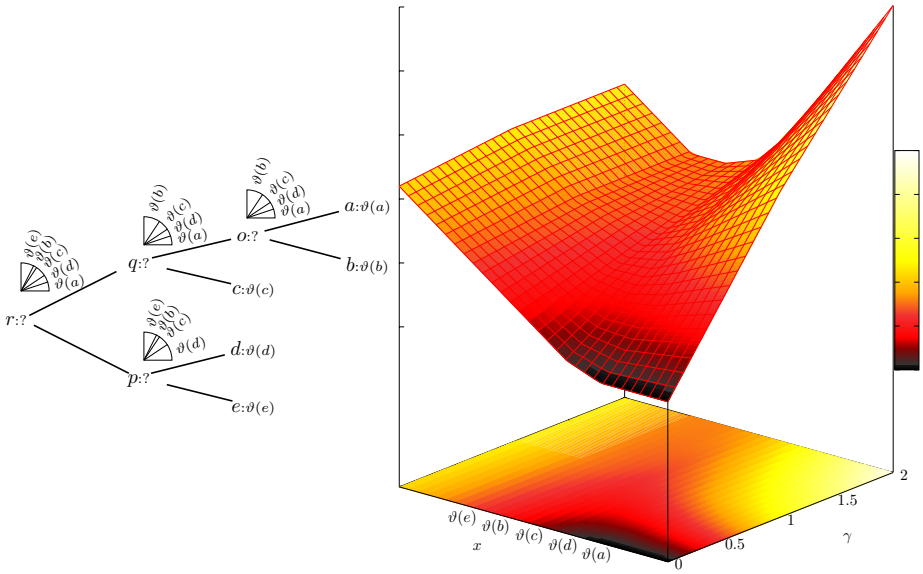
indexed in increasing order (we have that $\dot\Gamma^r_0 = 0$ and $\dot\Gamma^r_{w_n+1} = \infty$). There exists a sequence $(\rho^r_i)_{0\leq i\leq w_r}$ of increasing values in $\vartheta(\mathcal{K})$ such that for all $0 \leq i \leq w_r$ and all $\dot\Gamma^r_i \leq \gamma < \dot\Gamma^r_{i+1}$, a γ-parsimonious reconstruction associates ρ^r_i to r. If $\dot\Gamma^r_i < \gamma < \dot\Gamma^r_{i+1}$ then all γ-parsimonious reconstructions associate ρ^r_i to r.

From Remark 3 and by induction with Corollary 1, we get that, for all unknown node n of T, there exists a triple $(w_n, (\dot\Gamma^n_i)_{0\leq i\leq w_n+1}, (\rho^n_i)_{0\leq i\leq w_n})$ such that $\dot\Gamma^n_0 = 0$, $\dot\Gamma^n_{w_n+1} = \infty$, and, for all $0 \leq i \leq w_n$ and all $\dot\Gamma^n_i \leq \gamma < \dot\Gamma^n_{i+1}$, a γ-parsimonious reconstruction associates ρ^n_i to n. The triple $(w_n, (\dot\Gamma^n_i)_{0\leq i\leq w_n+1}, (\rho^n_i)_{0\leq i\leq w_n})$ will be referred to as the *parametric reconstruction* of n.

A graphical, and useful, way to represent a parametric reconstruction is to cut a quarter pie at each slope $\dot\Gamma^n_i$ for $1 \leq i \leq w_n$ (all these slopes are in the positive quadrant). Each slice of the quarter pie is then associated to a reconstructed value for n and its size reflects the proportion of asymmetry parameters leading to it (Figures 1-top-left, 2 and 3).

4.2 Algorithm

Theorems 1 and 2, Corollary 1 and Remark 3 suggest the approach sketched in Algorithms 1 and 2, for computing the parametric time-dependent-asymmetric-linear reconstruction of a tree T with regard to a known function ϑ. Algorithm 1 presents two functions which compute the cost and stem cost functions. These functions are



$$f_r(\gamma, x)$$

0		1	$+\infty$
$\vartheta(a)$	$\dfrac{-2\gamma x + \gamma[-\vartheta(a) + \vartheta(b) + \vartheta(c) + \vartheta(e)]}{(1-3\gamma)x + \gamma[\vartheta(b) + \vartheta(c) + \vartheta(e)] - \vartheta(a)}$	$-2\gamma x + \gamma[\vartheta(b) + \vartheta(e)] - \vartheta(a) + \vartheta(c)$	
$\vartheta(d)$	$\dfrac{(2-3\gamma)x + \gamma[\vartheta(b) + \vartheta(c) + \vartheta(e)] - \vartheta(a) - \vartheta(d)}{(2-\gamma)x + \gamma[\vartheta(b) - \vartheta(c) + \vartheta(e)] - \vartheta(a) - \vartheta(d)}$	$(1-2\gamma)x + \gamma[\vartheta(b) + \vartheta(e)] - \vartheta(a) + \vartheta(c) - \vartheta(d)$	
$\vartheta(c)$		$\vartheta(b)$ $\dfrac{(3-2\gamma)x + \gamma[\vartheta(b) + \vartheta(e)] - \vartheta(a) - \vartheta(c) - \vartheta(d)}{(2-\gamma)x + \gamma\vartheta(e) - \vartheta(a) + \vartheta(b) - \vartheta(c) - \vartheta(d)}$	
$\vartheta(e)$	$\dfrac{}{2x + \gamma[\vartheta(b) - \vartheta(c)] - \vartheta(a) - \vartheta(d)}$	$2x - \vartheta(a) + \vartheta(b) - \vartheta(c) - \vartheta(d)$	

Figure 1: Top: A tree in which only leaves are known with $\vartheta(a) < \vartheta(d) < \vartheta(c) < \vartheta(b) < \vartheta(e)$ and where the pie representations of the reconstructed values according to the the slope γ are above the unknown nodes (Left) Graphic representation of the subtree cost function of its root (Right). Bottom: The subtree cost function of its root. The subtree cost function is that of the generalized parsimony scheme of [1] (i.e. with $\phi(\tau) = 1$).

used by the function `main` of Algorithm 2 for computing the parametric reconstruction $(w_n, (\Gamma_i^n)_{0\leq i\leq w_n+1}, (\rho_i^n)_{0\leq i\leq w_n})$ of all nodes n of T.

The complexity of the computation of the parametric reconstruction depends on the size of the piecewise linear functions $(f_n)_{n\in T}$ and $(\widehat{f}_n)_{n\in T}$, i.e. the number of intervals required to define them. Though the number of state-bounds is, by construction, smaller than $|\mathcal{K}|+2$, bounding the number of asymmetry-bounds is not that straightforward.

4.3 Bounding u_n

From now on, we assume without loss of generality that all the bounds of the piecewise linear functions that we will consider, are necessary, in the sense that, under the notations of Theorem 1 and for all nodes n of T, the cost functions f_n are such that:

- for all $0 \leq i \leq u_n$ and all $0 \leq j < v_i^n$, we have that $(A_{i,j}^n, B_{i,j}^n, C_{i,j}^n, D_{i,j}^n) \neq (A_{i,j+1}^n, B_{i,j+1}^n, C_{i,j+1}^n, D_{i,j+1}^n)$;
- for all $0 \leq i < u_n$, there exists $0 \leq j \leq v_i^n$ and $0 \leq j' \leq v_{i+1}^n$ such that $(\vartheta(e_{i,j}^n), \vartheta(e_{i,j+1}^n)) \cap (\vartheta(e_{i+1,j'}^n), \vartheta(e_{i+1,j'+1}^n)) \neq \emptyset$ and $(A_{i,j}^n, B_{i,j}^n, C_{i,j}^n, D_{i,j}^n) \neq (A_{i+1,j'}^n, B_{i+1,j'}^n, C_{i+1,j'}^n, D_{i+1,j'}^n)$;

and that the same holds for all stem cost functions $\widehat{f}_n$. For all $0 \leq i \leq u_n$, all $\gamma \in (\Gamma_i^n, \Gamma_{i+1}^n)$ and all $0 \leq j < v_i^n$, we will say that $\vartheta(e_{i,j}^n)$ is a state-bound wrt γ.

For bounding the complexity of Algorithm 2, we shall bound the number of intervals in which one has to split the domain of γ, in order to express the map f_n. By examining the proof of Theorem 1 and Algorithm 1, we first remark that the only stage in which the number of asymmetry-bounds increases, is that computing $\widehat{f}_n$ from f_n for the unknown internal nodes n. The main point of this section is thus to evaluate the maximal increase between u_n and $\widehat{u}_n$. Let us start with a technical lemma.

Lemma 1. *Let n be an unknown node of T. For all $x \in \mathbb{R}$, the x-coefficient of $f_n(\gamma, x)$ (resp. of $\widehat{f}_n(\gamma, x)$) decreases with γ.*

Proof. We prove the lemma by using again the standard scheme. All the base cases are plain, and so is Step 1.

Let us skip to Step 2 and assume that, for all children m of n, the x-coefficient of $f_m(\gamma, x)$ decreases with γ. From the proof of Theorem 1 and for all $\gamma > 0$, if i is such that $\Gamma_{i-1}^n < \gamma \leq \Gamma_i^n$, there exists an index $0 \leq j_\gamma^+ \leq v_i^n$ such that, by setting $p_\gamma^+ = e_{i,j_\gamma^+}^n$, the x-coefficient of $\widehat{f}_m(\gamma, x)$ is equal to $-\gamma\phi(\tau_n)$ for $x < \vartheta(p_\gamma^+)$ and greater than $-\gamma\phi(\tau_n)$ otherwise. Symmetrically, there exists an index $0 \leq j_\gamma^- \leq v_i^n$ such that, by setting $p_\gamma^- = e_{i,j_\gamma^-}^n$, the x-coefficient of $\widehat{f}_m(\gamma, x)$ is equal to $\phi(\tau_n)$ for $x > \vartheta(p_\gamma^-)$ and smaller than $\phi(\tau_n)$ otherwise. The definitions of these nodes altogether with the induction hypothesis implies that the values $\vartheta(p_\gamma^+)$ and $\vartheta(p_\gamma^-)$ both increase with γ.

Let now consider two positive real numbers $\gamma' \leq \gamma''$. We then have $\vartheta(p_{\gamma'}^+) \leq \vartheta(p_{\gamma'}^-)$, $\vartheta(p_{\gamma''}^+) \leq \vartheta(p_{\gamma''}^-)$, $\vartheta(p_{\gamma'}^+) \leq \vartheta(p_{\gamma''}^+)$ and $\vartheta(p_{\gamma'}^-) \leq \vartheta(p_{\gamma''}^-)$. The first two inequalities come from construction (proof of Theorem 1) and the two last ones from the induction hypothesis. This leaves only two cases to investigate:

1. $\vartheta(p_{\gamma'}^+) \leq \vartheta(p_{\gamma'}^-) \leq \vartheta(p_{\gamma''}^+) \leq \vartheta(p_{\gamma''}^-)$ and
2. $\vartheta(p_{\gamma'}^+) \leq \vartheta(p_{\gamma''}^+) \leq \vartheta(p_{\gamma'}^-) \leq \vartheta(p_{\gamma''}^-)$.

With regard to the value of x, five possibilities arise in Case 1:

- if $x \leq \vartheta(p_{\gamma'}^+)$ then the x-coefficients of $\widehat{f}_m(\gamma', x)$ and $\widehat{f}_m(\gamma'', x)$ are equal to $-\gamma'\phi(\tau_n)$ and $-\gamma''\phi(\tau_n)$ respectively;

Function `compute_cost_and_stem_cost`

input : a node n of T

`/* Compute recursively the cost and stem cost functions of all the descendants of` n `and its cost function */`

if $\mathcal{C}_n \neq \emptyset$ **then**

forall $m \in \mathcal{C}_n$ **do**

$f_m \leftarrow$ `compute_cost_and_stem_cost`(m);

$\widehat{f}_m \leftarrow$ `cost_to_stem`(m);

if $n \in \mathcal{K}$ **then** $f_n(\gamma, x) \leftarrow \begin{cases} \sum_{m \in \mathcal{C}_n} \widehat{f}_m(\gamma, \vartheta(n)) & \text{if } x = \vartheta(n) \\ +\infty & \text{otherwise} \end{cases}$;

else $f_n \leftarrow \sum_{m \in \mathcal{C}_n} \widehat{f}_m$;

else

if $n \in \mathcal{K}$ **then** $f_n(\gamma, x) \leftarrow \begin{cases} 0 & \text{if } x = \vartheta(n) \\ +\infty & \text{otherwise} \end{cases}$;

else $f_n(\gamma, x) \leftarrow 0$;

Function `cost_to_stem`

input : a node n of T

`/* Compute the stem cost function of` n `from its cost function */`

if $n \in \mathcal{K}$ **then**

$\widehat{f}_n(\gamma, x) \leftarrow \begin{cases} \gamma\phi(\tau_n)(\vartheta(n) - x) + f_n(\gamma, \vartheta(n)) & \text{if } x < \vartheta(n) \\ \phi(\tau_n)(x - \vartheta(n)) + f_n(\gamma, \vartheta(n)) & \text{if } x \geq \vartheta(n) \end{cases}$;

else

$\widehat{u}_n \leftarrow -1$; $\widehat{\Gamma}^n_0 \leftarrow \Gamma^n_0$;

for $i \leftarrow 0$ **to** u_n **do**

$j_+ \leftarrow \min\{j \mid \frac{B^n_{i,j}}{A^n_{i,j} - \phi(\tau_n)} > \Gamma^n_i\}$; $M_+ \leftarrow \max\{j \mid \frac{B^n_{i,j}}{A^n_{i,j} - \phi(\tau_n)} < \Gamma^n_{i+1}\}$;

$j_- \leftarrow \min\{j \mid \frac{B^n_{i,j} - \phi(\tau_n)}{A^n_{i,j}} > \Gamma^n_i\}$; $M_- \leftarrow \max\{j \mid \frac{B^n_{i,j} - \phi(\tau_n)}{A^n_{i,j}} < \Gamma^n_{i+1}\}$;

while $j_- \leq M_-$ **or** $j_+ \leq M_+$ **do**

$\widehat{u}_n \leftarrow \widehat{u}_n + 1$;

for $\widehat{\Gamma}^n_{\widehat{u}_n} < \gamma < \widehat{\Gamma}^n_{\widehat{u}_n+1}$,

$\widehat{f}_n(\gamma, x) \leftarrow$

$\begin{cases} \gamma\phi(\tau_n)(\vartheta(e^n_{i,j_+}) - x) + f_n(\gamma, \vartheta(e^n_{i,j_+})) \text{ if } x < \vartheta(e^n_{i,j_-}) \\ f_n(\gamma, x) \text{ if } \vartheta(e^n_{i,j_+}) \leq x < \vartheta(e^n_{i,j_+}) \\ \phi(\tau_n)(x - \vartheta(e^n_{i,j_-})) + f_n(\gamma, \vartheta(e^n_{i,j_-})) \text{ if } x \geq \vartheta(e^n_{i,j_-}) \end{cases}$;

if $j_- \leq M_-$ **or** $j_+ \leq M_+$ **then**

$\widehat{\Gamma}^n_{\widehat{u}_n+1} \leftarrow \min\{\frac{B^n_{i,j_+}}{A^n_{i,j_+} - \phi(\tau_n)}, \frac{B^n_{i,j_-} - \phi(\tau_n)}{A^n_{i,j_-}}\}$;

if $\frac{B^n_{i,j_+}}{A^n_{i,j_+} - \phi(\tau_n)} \leq \widehat{\Gamma}^n_{\widehat{u}_n}$ **then** $j_+ \leftarrow j_+ + 1$;

if $\frac{B^n_{i,j_-} - \phi(\tau_n)}{A^n_{i,j_-}} \leq \widehat{\Gamma}^n_{\widehat{u}_n}$ **then** $j_- \leftarrow j_- + 1$;

else $\widehat{\Gamma}^n_{\widehat{u}_n+1} \leftarrow \Gamma^n_{i+1}$;

Algorithm 1: Computation of f_n and $\widehat{f}_n$ for all nodes n of T (under the notations of Theorem 1).

Function `compute_reconstruction`

input : a node n of T

`/* Compute the parametric reconstruction of` n `from its cost function */`

if $n \in \mathcal{K}$ **then**

$\dot{\Gamma}_0^n \leftarrow 0$; $\dot{\Gamma}_1^n \leftarrow \infty$;

$\rho_0^n \leftarrow \vartheta(r)$; $w_n \leftarrow 1$;

else

$w_n \leftarrow -1$; $\dot{\Gamma}_{w_n}^n \leftarrow \Gamma_0^n$;

for $i \leftarrow 0$ **to** u_n **do**

for $j = \min\{k \mid B_{i,k}^n \geq \Gamma_i^n A_{i,k}^n\}$ **to** $\max\{k \mid B_{i,k-1}^n < \Gamma_{i+1}^n A_{i,k-1}^n\}$ **do**

$w_n \leftarrow w_n + 1$;

$\rho_{w_n}^n \leftarrow e_{i,j}^n$;

if $j < \max\{k \mid B_{i,k-1}^n < \Gamma_{i+1}^n A_{i,k-1}^n\}$ **then** $\dot{\Gamma}_{w_n+1}^n \leftarrow \frac{B_{i,j+1}^n}{A_{i,j+1}^n}$;

else $\dot{\Gamma}_{w_n+1}^n \leftarrow \Gamma_{i+1}^n$;

Function `propagate_reconstruction`

input : a node n of T

`/* Compute the parametric reconstruction of` n `from that of its direct ancestor and, recursively, that of all its descendants */`

if $n \in \mathcal{K}$ **then**

$\dot{\Gamma}_0^n \leftarrow 0$; $\dot{\Gamma}_1^n \leftarrow \infty$;

$\rho_0^r \leftarrow \vartheta(n)$; $w_n \leftarrow 1$;

else

$p \leftarrow$ direct ancestor of n;

$i \leftarrow 0$; $j \leftarrow 0$;

$w_n \leftarrow -1$; $\dot{\Gamma}_0^n \leftarrow 0$;

while $i \leq \widehat{u}_n$ **or** $j \leq w_p$ **do**

$w_n \leftarrow w_n + 1$;

if $\rho_j^p < e_{i,1}^n$ **then** $\rho_{w_n}^n \leftarrow e_{i,1}^n$;

if $\rho_j^p \geq e_{i,1}^n$ **and** $\rho_j^p < e_{i,\widehat{v}_i^n}^n$ **then** $\rho_{w_n}^n \leftarrow \rho_j^p$;

if $\rho_j^p \geq e_{i,\widehat{v}_i^n}^n$ **then** $\rho_{w_n}^n \leftarrow e_{i,\widehat{v}_i^n}^n$;

$\dot{\Gamma}_{w_n+1}^n \leftarrow \min\{\widehat{\Gamma}_{i+1}^n, \dot{\Gamma}_{j+1}^p\}$;

if $\widehat{\Gamma}_{i+1}^n \leq \dot{\Gamma}_{w_n+1}^n$ **then** $i \leftarrow i + 1$;

if $\dot{\Gamma}_{j+1}^p \leq \dot{\Gamma}_{w_n+1}^n$ **then** $j \leftarrow j + 1$;

forall $m \in \mathcal{C}_n$ **do** `propagate_reconstruction` (m);

Function `main`

input : a tree T and a known function ϑ

$r \leftarrow$ root of T;

`compute_cost_and_stem_cost` (r);

`compute_reconstruction` (r);

forall $m \in \mathcal{C}_r$ **do** `propagate_reconstruction` (m);

Algorithm 2: Parametric reconstruction of a tree T.

- if $\vartheta(p^+_{\gamma'}) \leq x \leq \vartheta(p^-_{\gamma'})$ then the x-coefficient of $\widehat{f}_m(\gamma', x)$ is equal to that of $f_m(\gamma', x)$, thus greater than $-\gamma'\phi(\tau_n)$, and the x-coefficient of $\widehat{f}_m(\gamma'', x)$ is equal to $-\gamma''\phi(\tau_n)$;
- if $\vartheta(p^-_{\gamma'}) \leq x \leq \vartheta(p^+_{\gamma''})$ then the x-coefficients of $\widehat{f}_m(\gamma', x)$ and $\widehat{f}_m(\gamma'', x)$ are $\phi(\tau_n)$ and $-\gamma''\phi(\tau_n)$ respectively;
- if $\vartheta(p^+_{\gamma'}) \leq x \leq \vartheta(p^-_{\gamma''})$ then the x-coefficient of $\widehat{f}_m(\gamma', x)$ is equal to $\phi(\tau_n)$ and that of $\widehat{f}_m(\gamma'', x)$ is equal to that of $f_m(\gamma'', x)$, thus smaller than $\phi(\tau_n)$;
- if $x \geq \vartheta(p^-_{\gamma''})$ then the x-coefficients of $\widehat{f}_m(\gamma', x)$ and $\widehat{f}_m(\gamma'', x)$ are both equal to $\phi(\tau_n)$.

Similarly, five possibilities arise in Case 2:

- if $x \leq \vartheta(p^+_{\gamma'})$ then the x-coefficients of $\widehat{f}_m(\gamma', x)$ and $\widehat{f}_m(\gamma'', x)$ are equal to $-\gamma'\phi(\tau_n)$ and $-\gamma''\phi(\tau_n)$ respectively;
- if $\vartheta(p^+_{\gamma'}) \leq x \leq \vartheta(p^+_{\gamma''})$ then the x-coefficient of $\widehat{f}_m(\gamma', x)$ is equal to that of $f_m(\gamma', x)$, thus greater than $-\gamma'\phi(\tau_n)$, and the x-coefficient of $\widehat{f}_m(\gamma'', x)$ is equal to $-\gamma''\phi(\tau_n)$;
- if $\vartheta(p^+_{\gamma''}) \leq x \leq \vartheta(p^-_{\gamma'})$ then the x-coefficients of $\widehat{f}_m(\gamma', x)$ and $\widehat{f}_m(\gamma'', x)$ are equal to that of $f_m(\gamma', x)$ and $f_m(\gamma'', x)$, respectively;
- if $\vartheta(p^-_{\gamma'}) \leq x \leq \vartheta(p^-_{\gamma''})$ then the x-coefficient of $\widehat{f}_m(\gamma', x)$ is $\phi(\tau_n)$ and that of $\widehat{f}_m(\gamma'', x)$ is equal to that of $f_m(\gamma'', x)$, thus smaller than $\phi(\tau_n)$;
- if $x \geq \vartheta(p^-_{\gamma''})$ then the x-coefficients of $\widehat{f}_m(\gamma', x)$ and $\widehat{f}_m(\gamma'', x)$ are both equal to $\phi(\tau_n)$.

In all the situations cover by Cases 1 and 2, the x-coefficient of $\widehat{f}_m(\gamma', x)$ is always greater than that of $\widehat{f}_m(\gamma'', x)$, either directly or from the induction hypothesis. The x-coefficient of $\widehat{f}_m(\gamma, x)$ does decrease with γ.

Step 3 follows straightforwardly from Equation 2 and yields to conclude that, for all reals x, the x-coefficient of $f_n(\gamma, x)$ decreases with γ. □

Lemma 2. *Let n be a non-root unknown internal node n of T with $\widehat{u}_n > u_n$, ℓ be an index in $\{1, \ldots, \widehat{u}_n\}$ such that $\widehat{\Gamma}^n_\ell \notin \{\Gamma^n_k \mid 1 \leq k \leq u_n\}$ and i be such that $\Gamma^n_i < \widehat{\Gamma}^n_\ell < \Gamma^n_{i+1}$. There exists an index $0 \leq j \leq v^n_i$ such that at least one of the following assertions holds:*

1. $\widehat{\Gamma}^n_\ell = \dfrac{B^n_{i,j}}{A^n_{i,j} - \phi(\tau_n)}$ *and* $\vartheta(\widehat{e}^n_{\ell-1,1}) = \vartheta(e^n_{i,j}) < \vartheta(e^n_{i,j+1}) = \vartheta(\widehat{e}^n_{\ell,1}) \leq \vartheta(\widehat{e}^n_{\ell',1})$ *for all $\ell' \geq \ell$;*
2. $\widehat{\Gamma}^n_\ell = \dfrac{B^n_{i,j} - \phi(\tau_n)}{A^n_{i,j}}$ *and* $\vartheta(\widehat{e}^n_{\ell,v^n_\ell}) = \vartheta(e^n_{i,j+1}) > \vartheta(e^n_{i,j}) = \vartheta(\widehat{e}^n_{\ell-1,v^n_{\ell-1}}) \geq \vartheta(\widehat{e}^n_{\ell',v^n_{\ell'}})$ *for all $\ell' < \ell$.*

Proof. From the proof of Theorem 1, if $\Gamma^n_i < \widehat{\Gamma}^n_\ell < \Gamma^n_{i+1}$ then there exists an index j such that $\widehat{\Gamma}^n_\ell = \frac{B^n_{i,j}}{A^n_{i,j} - \phi(\tau_n)}$ or $\widehat{\Gamma}^n_\ell = \frac{B^n_{i,j} - \phi(\tau_n)}{A^n_{i,j}}$.

If $\widehat{\Gamma}^n_\ell = \frac{B^n_{i,j}}{A^n_{i,j} - \phi(\tau_n)}$ and still from the proof of Theorem 1, we have that $\vartheta(\widehat{e}^n_{\ell-1,1}) = \vartheta(e^n_{i,j}) < \vartheta(e^n_{i,j+1}) = \vartheta(\widehat{e}^n_{\ell,1})$. Lemma 1 then ensures that for all $i' \geq i$ and all j' such that $(\vartheta(e^n_{i',j'}), \vartheta(e^n_{i',j'+1})] \cap (\vartheta(e^n_{i,j}), \vartheta(e^n_{i,j+1})] \neq \emptyset$, we have that $-A^n_{i',j'}\gamma' + B^n_{i',j'} \leq -A^n_{i,j}\gamma + B^n_{i,j}$ for all $\gamma' \in (\Gamma^n_{i'}, \Gamma^n_{i'+1}]$ and all $\gamma \in (\Gamma^n_i, \Gamma^n_{i+1}]$ with $\gamma' \geq \gamma$. It follows that, if $\vartheta(e^n_{i',j'}) \leq \vartheta(e^n_{i,j})$ then $-(A^n_{i',j'}\gamma' - \phi(\tau_n)) + B^n_{i',j'} \leq 0$ for all $\gamma' \geq \widehat{\Gamma}^n_\ell$. This implies that $\vartheta(\widehat{e}^n_{\ell,1}) \leq \vartheta(\widehat{e}^n_{\ell',1})$ for all $\ell' \geq \ell$.

The situation is symmetrical if $\widehat{\Gamma}^n_\ell = \frac{B^n_{i,j} - \phi(\tau_n)}{A^n_{i,j}}$. □

In plain English, Lemma 2 says that each time that an asymmetry-bound $\widehat{\Gamma}_\ell^n$ is created, either $\vartheta(\widehat{e}_{\ell-1,1}^n)$ is no longer required as state-bound for all $\gamma \geq \widehat{\Gamma}_\ell^n$, (i.e. $\vartheta(\widehat{e}_{\ell-1,1}^n)$ vanishes after $\widehat{\Gamma}_\ell^n$), or $\vartheta(\widehat{e}_{\ell,v_\ell^n}^n)$ was not a necessary state-bound for all $\gamma < \widehat{\Gamma}_\ell^n$ (i.e. $\vartheta(\widehat{e}_{\ell,v_\ell^n}^n)$ appears after $\widehat{\Gamma}_\ell^n$). Since all the state-bounds of f_n belong to $\vartheta(\mathcal{K}_n)$ and, from Lemma 2, appear and vanish at most once during the computation of $\widehat{f}_n$ from f_n, we have that

$$\widehat{u}_n - u_n \leq 2|\mathcal{K}_n|.$$

Theorem 3. *For all non-root nodes n of T, we have that*

$$\widehat{u}_n \leq 2|T_n|.|\mathcal{K}_n|.$$

Proof. Each asymmetry-bound Λ of $\widehat{f}_n$ is such that there exists an unknown node $q \in T_n$, such that Λ is an asymmetry-bound of $\widehat{f}_q$ but not of f_q. From Lemma 2, this implies that there exists a known node k in $\vartheta(\mathcal{K}_q)$ such that $\vartheta(k)$ is required or not as state-bound of $\widehat{f}_q$ according to whether the asymmetry-parameter is smaller to Λ or not (or conversely). In short, all asymmetry-bounds of $\widehat{f}_n$ involve an unknown node $q \in T_n$ and a known descendant k of q. Moreover, from Lemma 2, all known descendants k of a known node q can be involved with at most twice asymmetry-bounds of $\widehat{f}_q$ which are not asymmetry-bounds of f_q. Since each node of $\mathcal{K}_n$ has less than $|T_n|$ unknown ancestors in T_n, it follows that the total number of asymmetry-bounds of $\widehat{f}_n$ cannot exceed $2|T_n|.|\mathcal{K}_n|$. □

Corollary 2. *Under the assumption that the number of children of node is bounded independently of the size of the tree, the algorithmic complexity of the computation of the time-dependent-asymmetric-linear parametric reconstruction of a tree T with a set of known nodes $\mathcal{K}$ is $O(|T|^2.|\mathcal{K}|^2)$ both in time and memory space.*

Proof. For all non-root unknown nodes n of T, computing the stem cost function $\widehat{f}_n$ from the cost function f_n is linear with the total number of "pieces" required to define $\widehat{f}_n$ (i.e. its size, see Algorithm 1). From Theorem 3, less than $2|T_n|.|\mathcal{K}_n|$ asymmetry-bounds are required for f_n. Over all intervals between two successive asymmetry-bounds, there are at most $|\mathcal{K}_n|$ state-bounds required. It follows that the size of $\widehat{f}_n$ is smaller than $2|T_n|.|\mathcal{K}_n|^2$ and its computation from f_n is $O(|T_n|.|\mathcal{K}_n|^2)$.

Obtaining the cost function f_n for all the stem cost functions $\widehat{f}_m$ of its children m is performed by using a procedure similar to that merging sorted lists. Under the assumption that the number of children m is bounded independently of the size of the tree, this operation is linear with total size of the stem cost functions $\widehat{f}_m$. In short, this stage is $O(|T_n|.|\mathcal{K}_n|^2)$.

It follows that the computation of all the cost and stem cost functions of the unknown nodes of T is $O(|T|^2.|\mathcal{K}|^2)$ both in time and memory space.

The stage computing the parametric reconstructions from the cost and stem cost functions is linear with their total size (Algorithm 2), which gives us the overall complexity of the algorithm. □

5 Example

The TDALP was applied to the dataset of [10], which contains the phylogenetic tree of extant cetaceans (including branch lengths) and their (average) body sizes (this dataset was used as it was – we just pruned extant taxa of which the body size was not provided). The results of the parametric reconstruction of the body size of cetaceans are displayed in Figures 2 and 3. Figure 2 shows the whole phylogenetic tree with the quarter pie representation of the parametric reconstruction of all the nodes.

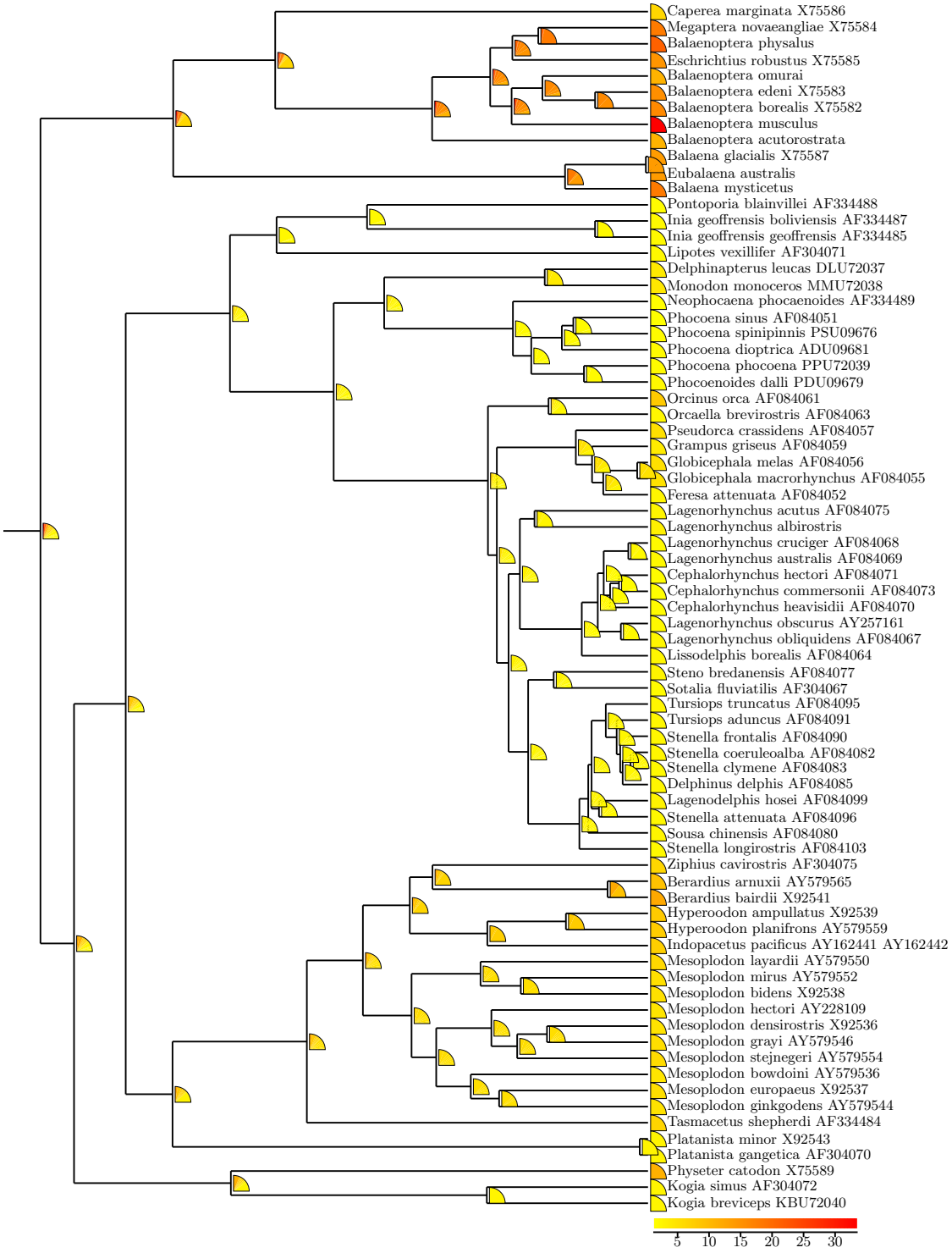


Figure 2: Parametric reconstruction of cetacean body size in meters, by taking into account branch lengths with $\phi(\tau) = \frac{1}{\tau}$. The reconstructed states are represented by colors in the quarter pies figuring the parametric reconstruction. The color scale is displayed at the bottom-right of the figure.

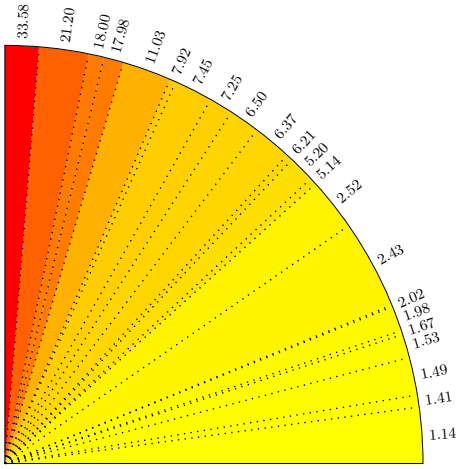


Figure 3: Detail of the parametric reconstruction of the root of the cetacean phylogenetic tree displayed in Figure 2.

Figure 3 details the parametric reconstruction of the most recent common ancestor of cetaceans (i.e. the root of the tree of Figure 2).

Figure 2 provides a synthetic representation of all the possible parsimonious reconstructions of the ancestral states. One can see at a glance that is little uncertainty with the reconstruction of some of the nodes while that of other ones is more mixed. This representation actually contains all the possible reconstructions from the TDALP with regard to the asymmetry parameter.

By focusing on a node of interest, for instance the root of the tree as displayed in Figure 3, one can take a closer look on which reconstructed states are possible, in what extent they are supported, and which are the corresponding evolutionary assumptions. In Figure 3, we do observe that a root state smaller than 2.52 m is supported by approximately half of the asymmetry parameters, which corresponds to trends ranging from highly positive to neutral. Conversely, an ancestral state greater than 8 m is reconstructed only for the uppermost quarter of the asymmetry parameters, which corresponds to the most negative evolutionary trends.

Remark that any assumption, or any evidence (e.g. some fossils), about one or several ancestral states directly translates into an assumption or an evidence about the asymmetry parameter, thus in a certain sense, about the nature and the intensity of the underlying evolutionary trend. This is basically done by checking which interval of the corresponding parametric reconstruction contains the assumed or known ancestral state.

References

[1] M. Csűrös. Ancestral Reconstruction by Asymmetric Wagner Parsimony over Continuous Characters and Squared Parsimony over Distributions. In C. Nelson and S. Vialette, editors, *Comparative Genomics*, volume 5267 of *Lecture Notes in Computer Science*, pages 72–86. Springer Berlin / Heidelberg, 2008.

[2] C. W. Cunningham. Some Limitations of Ancestral Character-State Reconstruction When Testing Evolutionary Hypotheses. *Systematic Biology*, 48(3):665–674, 1999.

[3] G. Didier. Parametric Maximum Parsimonious Reconstruction on Trees. *Bulletin of Mathematical Biology*, 73:1477–1502, 2011.

[4] J. S. Farris. Methods for Computing Wagner Trees. *Systematic Biology*, 19(1):83–92, 1970.

[5] J. B. Joy, R. H. Liang, R. M. McCloskey, T. Nguyen, and A. F. Y. Poon. Ancestral reconstruction. *PLoS Comput Biol*, 12(7):1–20, 07 2016.

[6] W. P. Maddison and D. Maddison. Mesquite: a modular system for evolutionary analysis. version 2.75., 2011. `http://mesquiteproject.org`.

[7] K. E. Omland. The Assumptions and Challenges of Ancestral State Reconstruction. *Systematic Biology*, 48(3):665–674, 1999.

[8] M. Royer-Carenzi and G. Didier. A comparison of ancestral state reconstruction methods for quantitative characters. *Journal of Theoretical Biology*, 404:126 – 142, 2016.

[9] M. Royer-Carenzi, P. Pontarotti, and G. Didier. Choosing the best ancestral character state reconstruction method. *Mathematical Biosciences*, 242:95–109, 2013.

[10] G. J. Slater, S. A. Price, F. Santini, and M. E. Alfaro. Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society of London B: Biological Sciences*, 2010.

[11] D. L. Swofford and W. P. Maddison. Reconstructing ancestral character states under wagner parsimony. *Mathematical Biosciences*, 87(2):199–229, 1987.

[12] D. L. Swofford and W. P. Maddison. Parsimony, character-state reconstructions, and evolutionary inferences. In R. L. Mayden, editor, *Systematics, Historical Ecology, and North American Freshwater Fishes*, pages 187–223. Stanford University Press, Stanford, California, 1992.

[13] A. Webster and A. Purvis. Testing the accuracy for reconstructing ancestral states of continuous characters. *Proc. R. Soc. Lond. B*, 269:143–149, 2002.