



HAL
open science

Parsimony-based test for identifying changes in evolutionary trends for quantitative characters: implications for the origin of the amniotic egg

Gilles Didier, Olivier Chabrol, Michel Laurin

► **To cite this version:**

Gilles Didier, Olivier Chabrol, Michel Laurin. Parsimony-based test for identifying changes in evolutionary trends for quantitative characters: implications for the origin of the amniotic egg. *Cladistics*, 2019, 35 (5), pp.576-599. 10.1111/cla.12371 . hal-01609238

HAL Id: hal-01609238

<https://hal.science/hal-01609238>

Submitted on 19 Nov 2020

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.

Parsimony-based test for identifying changes in evolutionary
trends for quantitative characters: implications for the origin of
the amniotic egg

Gilles Didier¹, Olivier Chabrol² and Michel Laurin³

¹ IMAG, Univ Montpellier, CNRS, Montpellier, France

² Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

³ CR2P, Sorbonne Universités, CNRS/MNHN/UPMC, Muséum National d'Histoire Naturelle, 75005 Paris, France

November 19, 2020

Abstract

The origin of the amniotic egg was a major event in vertebrate evolution and is thought to have contributed to the spectacular evolutionary radiation of amniotes. We test one of the most popular scenarios proposed by Carroll in 1970 to explain the origin of the amniotic egg using a novel method based on an asymmetric version of linear parsimony (aka Wagner parsimony) for identifying the most parsimonious split of a tree into two parts between which the evolution of the character is allowed to differ. The new method evaluates the cost of splitting a phylogenetic tree at a given node as the integral, over all pairs of asymmetry parameters, of the most parsimonious costs that can be achieved by using the first parameter on the subtree pending from this node and the second parameter elsewhere. By testing all the nodes, we then get the most parsimonious split of a tree with regard to the character values at its tips. Among the nine trees and two characters tested, our method yields a total of 517 parsimonious trend changes in Permo-Carboniferous stegocephalians, a single one of which occurs in a part of the tree (among stem-amniotes) where Carroll's scenario predicts that there should have been distinct changes in body size evolutionary trends. This refutes the scenario because the amniote stem does not appear to have elevated rates of evolutionary trend shifts. Our nodal body size estimates offer less discriminating power, but they likewise fail to find strong support for Carroll's scenario.

Contents

1	Introduction	4
2	New method: parsimonious detection of trend changes	8
2.1	Asymmetric parsimony	8
2.1.1	Definitions and Notations	8
2.1.2	Asymmetric parsimonious cost	8
2.2	Outline of the approach	9
2.3	Partial cost functions	10
2.4	Integrating parsimonious costs over asymmetry parameter(s)	12
2.4.1	Single parameter case: no-split cost	12
2.4.2	Two parameters case: A- and B-split costs	14
3	Empirical tests of the new method	15
3.1	Two contrived examples	15
3.2	Two biological datasets previously analyzed with stochastic methods	15
4	New empirical study: stegocephalian body size evolution and origin of the amniotic egg	16
4.1	Dataset on stegocephalian body size	16
4.2	Analytical methods	17
4.3	Detected changes in evolutionary trends	18
4.4	Nodal estimates of body size in selected clades	19
5	Discussion	19
A	Proof of Theorem 2	25
B	Complexity of Algorithm 1 – Proof of Theorem 3	27
B.1	Proof of Theorem 3	30
C	Integrating parsimonious costs over two asymmetry parameters	31

1 Introduction

The origin of the amniotic egg has long been taken as an example of a major evolutionary innovation among vertebrates. This egg, which is adapted to be laid on emerged land, allowed amniotes to become fully terrestrial (Romer 1957, Tihen 1960, Carroll 1970, 1988, Laurin 2010), though some view its importance as potentially overstated (Skulan 2000). Several scenarios have been proposed to explain how the amniotic egg may have appeared, and how this is related to the move onto land by amniotes or their ancestors.

Romer (1957) suggested that the first amniotes were still mostly aquatic animals and that they came to land only or mostly to lay eggs, because the land was safer (inhabited by fewer predators) than the water and because he thought that stem-amniotes experienced seasonal drought, which would be disadvantageous for aquatic larvae. Romer's (1957) scenario no longer reflects the current consensus, partly because his interpretations about the lifestyle of the earliest amniotes have not been confirmed by recent research. For instance, ophiacodontids, which he viewed as some of the basalmost amniotes and the most representative of their lifestyle, appear to have been much much more terrestrial than he suggested, based on both bone microanatomy (Laurin and de Buffr enil 2016) and morphology (Felice and Angielczyk 2014). Similarly, the sediments in which early amniotes and their presumed relatives have been found are no longer believed to have been deposited in environments with seasonal drought (Laurin *et al.* 2007).

Tihen (1960) proposed a modification of Romer's (1957) scenario by suggesting that the terrestrial egg was first laid in a very humid, probably tropical and swampy environment, which would minimize dehydration problems for the eggs. Similarly, Goin and Goin (1962), suggested that stem-amniote terrestrial egg-laying is likely to have occurred in a humid montane environment. This was based on the observation that many extant lissamphibians that lay terrestrial eggs inhabit the vicinity of cool, rapidly flowing streams. These ideas were expanded by Szarski (1968), who suggested that the evolutionary pressure that drove development of the extra-embryonic membranes in stem-amniotes was the need to store urea into the allantois, and the osmoregulatory advantages of doing so (namely, increased ability to retain and even absorb water from the surroundings).

Carroll (1970, 1991) developed these ideas further by suggesting that the aquatic larval stage that characterized most early stegocephalians (Carroll 1988, Laurin 2010) was lost before the amniotic egg appeared. More importantly, contrary to his predecessors, he suggested that the earliest amniotes had already become terrestrial, which is compatible with the current ideas about the lifestyle of the earliest amniotes. Comparisons with eggs of some plethodontid salamanders (which lay eggs on land and have terrestrial hatchlings) suggest that the egg could not have measured more than 9 or 10 mm in diameter because of the initial lack of specialized surfaces to help gas exchange. In amniotic eggs, the chorion, allantois, and even (sometimes) the yolk sac facilitate gas exchange (Ferner and Mess 2011), whereas in anurans, the tail may be used for this, and in gymnophionans, external gills are used. Carroll (1970, 1991) also observed a correlation between egg size and adult body size, at least between species of urodeles

having a similar reproductive mode (one of which occurs only in plethodontid salamanders and consists of laying eggs on land) and among squamates. Because of this, Carroll (1970, 1991) suggested that the adults of the stem-amniotes that laid anamniotic eggs on land measured no more than 8-10 cm in snout-vent length. This implies a decrease in body size, compared to that of older (possibly stem-tetrapod, or simply older stem-amniote) ancestors, given what we know about body size of early tetrapods from the fossil record. After the appearance of the amniotic egg (selective pressures to improve gas exchange may have led to the appearance or further development of the extra-embryonic membranes), the constraint on egg size would have been released, and body size would consequently have increased.

Kohring (1995) suggested a scenario opposite to Romer's (1957). Namely, he hypothesized that the predecessor of the amniotic egg was laid in water and that even the first amniotes, which he inferred to have been terrestrial, continued laying their eggs in water for some time. In this context, he hypothesized that the amniotic extra-embryonic membranes appeared in an aquatic egg, and that only the outer egg membranes, which are more water-proof in amniotes than in amphibians, evolved in the terrestrial environment. Indeed, extant amniotes still show much variability in the degree of mineralization of the external membrane and in water permeability, both of which are intimately linked (Oftedal 2002). However, the universal development of the amniotic egg in the terrestrial environment despite large outer membrane permeability (Oftedal 2002) militates against Kohring's (1995) hypothesis. Similarly, the fact that mesosaurs, the first amniotes to have returned to a predominantly aquatic lifestyle, display extended embryo retention, perhaps in the form of viviparity (Piñeiro *et al.* 2012), suggests that very early in amniote evolution, the amniotic egg was adapted to be laid on emerged land.

While the fossil record yields few data directly relevant to origin of the amniotic egg because this event involves mostly the appearance of new structures that do not fossilize, the amnion, chorion, and allantois (Kohring 1995), Carroll's (1970, 1991) scenario (contrary to the others) makes predictions about how various stages of its development must have affected body size of the ancestral stem-amniotes, and these predictions can be tested using the fossil record. These predictions concern both changes in the direction of evolutionary trends and maximum body size. Namely, these predictions are (listed in chronological order in which they should be observed in the fossil record):

- H1, There was a change from the absence of trend, or possibly a slight trend towards increase in body size, to a relatively short-lived trend towards lower body size associated with some stem-amniotes, associated with egg-laying on land and the elimination of the aquatic larval stage.
- H2, Some stem-amniotes measured less than 10 cm in snout-vent length.
- H3, Body size started increasing shortly after appearance of the amniotic egg (or simultaneously with this even), prior to the divergence between synapsids and sauropsids.

These predictions were tested only once, to our knowledge (Laurin 2004), using a dataset comprised of 107 Permo-Carboniferous taxa, through parsimony squared-change optimization of body size to infer

the length of hypothetical stem-amniote ancestors on paleontological timetrees, and through phylogeny-informed regressions of size vs geological time, to assess the presence of trends. These methods were not ideal for this test. Squared-change parsimony optimization rests on a hypothesis that the studied characters evolved according to a Brownian motion model, which stipulates that there is no trend, whereas the tested scenario suggests that evolutionary trends should have been present during part of the evolutionary history of amniotes and of their ancestors. Indeed, the analyses of body size data suggest that there was a trend towards body size increase, at least in the amniote total clade (Laurin 2004, tables 9 and 10). Carroll's (1970, 1991) scenario also stipulates that at some point, there must have been a decrease in size of early stem-amniotes. While Laurin (2004) detected no evidence of this transient decrease in body size, this may well result from the paucity of the fossil record in this part of the tree, at least under most of the tested phylogenies; one of the topologies used in our tests, inspired by Panchen and Smithson (1988) suggests a far richer fossil record of stem amniotes than the other topology, which reflects Vallin and Laurin (2004). In the presence of trends, squared-change parsimony is expected to yield biased estimates, which hampers testing H2. The phylogeny-informed regressions performed in Laurin (2004) can assess the presence of a trend towards body size increase in early amniotes, but not the transient change in trends in body size evolution predicted above (H1).

New methodological developments described below allow a more rigorous test of this scenario, for two reasons. First, the new method that we describe below allows us to test where on the tree the most significant changes in trends occur. This is the most direct way of testing H1 and H3. Second, a recent linear parsimony optimization that incorporates branch length information allows character values at internal nodes to be inferred in the presence of trends (Didier 2017). This should allow a more rigorous test of H2. If our results falsified Carroll's (1970, 1991) scenario, they would indirectly favor (to an extent) alternative scenarios, such as those proposed by Kohring (1995), which we cannot test directly because they do not make clear predictions about body size evolution in amniotes and their predecessors.

We propose here a novel parsimony-based approach to assess where shifts in evolutionary trends occur. This approach, which we use to test Carroll's (1970, 1991) scenario on amniotic egg origin, does not require a priori hypotheses on where shifts may occur. Our approach is based on an asymmetric version of the linear parsimony introduced in Csűrös (2008) and further studied in Didier (2017), in both cases in an ancestral reconstruction context. This version of linear parsimony is called *asymmetric* because it allows different costs to be assigned to similarly-sized increases and decreases of a continuous character. To this end, the approach uses a parameter, referred to as the *asymmetry parameter*, which is defined as the ratio between the cost of an increase and that of a decrease of the same extent in Didier (2017) (we shall consider a slightly different parametrization here). Moreover, the version studied in Didier (2017) accounts for branch lengths in the evolutionary cost as a multiplicative factor defined as a function of the duration associated with a branch.

All the approaches developed so far for studying shifts in evolutionary trends are based on stochastic models of evolution for continuous characters, mainly Brownian motion or Ornstein-Uhlenbeck processes

(Bastide *et al.* 2017, Fuentes-G. *et al.* 2016, Cressler *et al.* 2015, Eastman *et al.* 2011, Khabbazian *et al.* 2016, Ingram and Mahler 2013, Landis *et al.* 2013, O’Meara *et al.* 2006, Revell 2008, Revell and Collar 2009, Slater 2013). The method presented in Butler and Losos (1997) used parsimony approaches but only for reconstructing ancestral states, which were next compared with simulations under Brownian models. Other approaches addressed the related question of testing the significance of splits given *a priori* (e.g., from environmental or dietary considerations) with regard to stochastic models (Thomas *et al.* 2006, 2009, Slater *et al.* 2010).

The problem that we are studying here can be formally stated as follows. Being given a phylogenetic tree and a continuous character which is known only for the tips of the branches, we aim to identify the node of the tree at which the evolution of the character starts to follow a different trend from the rest of tree, if such a node exists.

A parsimonious framework enables us to compute the minimal evolutionary cost (which reflects the amount of change implied) of the extant character values. This cost plays the role of the likelihood in probabilistic models. In the asymmetric linear parsimony case, the evolutionary cost depends on the asymmetry parameter α . In order to evaluate the relevance of putting an evolutionary shift at a node n , we consider evolutionary costs obtained by using an asymmetry parameter on the subtree \mathcal{T}_n and another parameter on the rest of the tree. Testing all the nodes of the tree enables us to identify the subtree on which the evolution of the character differs the most. We first provide a polynomial algorithm for computing such costs and for integrating them over all pairs of asymmetry parameters. This way we get a parsimonious method for finding the subtree on which evolution differs the most.

We developed *ParSplit*, a computer program that performs the parsimonious split of a phylogenetic tree at all possible locations with regard to tip values of characters written in C language. Source code of the software is available at <https://github.com/gilles-didier/ParSplit>.

Thus, Section 2 presents definitions, notations and asymmetric linear parsimony, then outlines the approach. In the same section, we define the *partial cost functions*, which form the basis of our computations and show how to integrate the split costs over all the possible values of the asymmetry parameter(s).

The approach is illustrated on two contrived examples and on two biological datasets in Section 3. The two contrived examples show that considering more than one asymmetric parameter does not always lower the evolutionary cost. The two biological datasets, from Slater *et al.* (2010) and Thomas and Freckleton (2012), were previously studied using stochastic methods to assess shifts in trends. They are used to demonstrate that our method yields sensible results often congruent with those of stochastic methods. Shifts detected with our approach are consistent with those obtained from methods based on stochastic models and separate well clades with specific evolutionary patterns like baleen whales (mysticetes) from other cetaceans. Our main results are presented in Section 4, in which a dataset about body size evolution of early stegocephalians is used to assess Carroll’s (1970, 1991) scenario about the appearance of the amniotic egg. For this purpose, we use the body size data and phylogenies previously used by Laurin (2004) to test this scenario using other methods. Results are discussed in Section 5.

2 New method: parsimonious detection of trend changes

2.1 Asymmetric parsimony

2.1.1 Definitions and Notations

We use the same notations as Didier (2017). The cardinal of any finite set \mathcal{S} is noted $|\mathcal{S}|$. In what follows, \mathcal{T} designates a rooted tree which may or may not be binary. As it should lead to no confusion, we still write \mathcal{T} for its set of nodes. For all nodes $n \in \mathcal{T}$, we put

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

Let us consider a subset \mathcal{K} of nodes of \mathcal{T} and 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*. For all nodes n of \mathcal{T} , we put \mathcal{K}_n for the subset of known nodes of the subtree \mathcal{T}_n , i.e. $\mathcal{K}_n = \mathcal{K} \cap \mathcal{T}_n$. Though the general case can be treated as in Didier (2017), we assume that all and only the tips of the tree have known values in order to lighten the statements. In plain English, \mathcal{K} (resp. \mathcal{K}_n) is the set of tips of \mathcal{T} (resp. of \mathcal{T}_n).

The values of $\{\vartheta(k) \mid k \in \mathcal{K}\}$ are the *known values* of \mathcal{T} . For all nodes n , we put $\vartheta(\mathcal{K}_n)$ for the set $\{\vartheta(k) \mid k \in \mathcal{K}_n\}$.

A ϑ -assignment of \mathcal{T} is a map ξ from \mathcal{T} to the set of real numbers which extends ϑ (i.e. such that $\xi(n) = \vartheta(n)$ for all nodes $n \in \mathcal{K}$). The set of all ϑ -assignments of \mathcal{T} is noted Ξ_ϑ .

2.1.2 Asymmetric parsimonious cost

A parsimony framework is based on a way of computing the evolutionary cost of an assignment. In the time-dependent-asymmetric-linear parsimony (*TDALP*, Didier 2017) 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 nonnegative 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 $n \in \mathcal{T}$). The case where ϕ is constant with $\phi(\tau) = 1$ for all τ corresponds to the standard parsimony scheme (Csűrös 2008, Farris 1970). In an evolutionary context, it makes sense to choose a decreasing function ϕ , i.e., the cost of a certain amount of change decreases with the time during which it occurs, but from a computational point de view, any positive function can be used. In the current implementation of the approach, users can choose between $\phi(\tau) = \frac{1}{\tau}$ and $\phi(\tau) = 1$.

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

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

Let us first remark that multiplying both λ and γ by a positive constant factor just leads to multiply the assignment costs by the same factor. Since computing an evolutionary cost with $\gamma = \lambda = 0$ makes no sense, we can assume $\gamma + \lambda > 0$ and divide both parameters by $\gamma + \lambda$ in order to normalize the parsimonious costs. From now on, we assume that $\gamma + \lambda = 1$. The cost of an assignment then only depends on a single parameter $\alpha = \frac{\gamma}{\gamma + \lambda}$ which belongs to $[0, 1]$, and is called α -cost:

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

Below, α will be referred to as the *asymmetry parameter*. Reconstructing with $\alpha < \frac{1}{2}$ (resp. with $\alpha = \frac{1}{2}$, with $\alpha > \frac{1}{2}$) makes the cost of an increase smaller than (resp. equal as, greater than) that of a decrease of the same amount. Intuitively, it corresponds to the assumption that the character evolves with a positive trend (resp. without trend, with a negative trend). Note that a slightly different parametrization is considered in Didier (2017).

The (*parsimonious*) α -cost of the pair (\mathcal{T}, ϑ) is defined as the minimal cost which can be achieved by a ϑ -assignment:

$$\check{\Delta}_{\alpha}(\mathcal{T}, \vartheta) = \min_{\xi \in \Xi_{\vartheta}} \Delta_{\alpha}(\xi). \quad (2)$$

2.2 Outline of the approach

Our aim is to find the most parsimonious way of splitting a phylogenetic tree \mathcal{T} into two parts according to an initial function ϑ (i.e., the character values of tips). In what follows, we will consider the two ways of splitting a phylogenetic tree at a node n displayed in Figure 1 and below referred to as *A-split* and *B-split*, which split the tree either just below a node or just above it, on one of the daughter-branches, respectively. Devising a parsimonious approach for identifying evolutionary shifts first requires to be able to associate an evolutionary cost to a given split.

In a parsimonious context, a natural idea should be to define the cost of a split of a tree into two parts as the minimum cost, over all pairs $(\alpha, \alpha') \in [0, 1]^2$, which can be achieved by using parameter α on a part of the tree and parameter α' on the other part. Unfortunately, even without splitting the tree, minimizing the parsimonious cost over all parameters α leads to trivial solutions: by setting $\alpha = 0$ (resp. $\alpha = 1$) the assignment which associates the greatest (resp. the smallest) of the known values to all unknown nodes has a null α -cost!

In order to overcome this issue, we first define the *no-split cost* of (\mathcal{T}, ϑ) as the sum over all parameters $\alpha \in [0, 1]$ of the α -parsimonious costs $\check{\Delta}_{\alpha}(\mathcal{T}, \vartheta)$. Note that this cost is parameter-free and depends only on \mathcal{T} and ϑ . It somehow reflects the disparity of the character with regard to the tree. We then define the cost of a split of \mathcal{T} into two parts as the sum over all pairs of parameters $(\alpha, \alpha') \in [0, 1]^2$ of the smallest cost which can be achieved by using parameter α on the first part and parameter α' on the

second one. The most parsimonious split of a phylogenetic tree with regard to a given initial function is either the whole tree if the no-split cost is lower than all the costs of the A - and B -splits, or the A - or B - split with the smallest cost otherwise.

The next two subsections show how to compute the no-, A - and B -splits costs.

2.3 Partial cost functions

For all nodes n of \mathcal{T} , we shall define and study the partial cost functions $f_n(\alpha, x)$, $\hat{f}_n(\alpha, x)$, $h_n(\alpha, x)$ and $\hat{h}_n(\alpha, x)$ that give the minimal costs which be obtained by associating x to n or its direct ancestor and by using parameter α on the corresponding parts of the tree as displayed in Figure 1.

Namely, following Csűrös (2008) and Didier (2017), for all nodes n of \mathcal{T} , we put $f_n(\alpha, x)$ for the minimal cost which can be achieved by an assignment ξ of the subtree \mathcal{T}_n (red part of Figure 1-left) such that $\xi(n) = x$. For all nodes n of \mathcal{T} but its root, we put $\hat{f}_n(\alpha, x)$ for the minimal cost which can be achieved by an assignment ξ of the tree consisting of m , the direct ancestor of n and of the subtree \mathcal{T}_n , which will be referred to as the *branch-based subtree of n* and noted $\hat{\mathcal{T}}_n$ (called *stem subtree of n* in Didier (2017), red part of Figure 1-right), verifying $\xi(m) = x$.

Cost functions $h_n(\alpha, x)$ and $\hat{h}_n(\alpha, x)$ give the complementary costs of $\hat{f}_n(\alpha, x)$ and $f_n(\alpha, x)$. Namely, for all nodes n of \mathcal{T} , we put $h_n(\alpha, x)$ for the minimum reconstruction cost of \mathcal{T} without n , the branch bearing n and the subtree rooted at n (i.e. the complementary of the branch-based subtree of n in the whole tree, blue part of Figure 1-right), which associates x to the direct ancestor of n . Last, we put $\hat{h}_n(\alpha, x)$ for the minimal cost which can be achieved by an assignment ξ of the whole tree with n and the branch bearing n but without the subtree rooted at n (blue part of Figure 1-left) which is such that $\xi(n) = x$.

The two following theorems provide the form of the partial cost functions (i.e., they are piecewise linear) and some of their properties, notably their convexity. Let us first recall and slightly adapt a Theorem from Didier (2017).

Theorem 1 (Didier 2017). *Let n be an unknown node of \mathcal{T} . The maps f_n and \hat{f}_n are piecewise-linear and continuous and, for all $\alpha > 0$, the maps $x \rightarrow f_n(\alpha, x)$ and $x \rightarrow \hat{f}_n(\alpha, x)$ are both convex.*

More precisely, if all the nodes of \mathcal{T}_n are unknown, then $f_n(\alpha, x) = 0$ for all $\alpha > 0$ 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 $(\kappa_{i,j}^n)_{1 \leq j \leq v_i^n}$ of known values of \mathcal{T}_n (i.e. in $\vartheta(\mathcal{K}_n)$) verifying $\kappa_{i,j}^n < \kappa_{i,j'}^n \Leftrightarrow j < j'$; by convention we set $\kappa_{i,0}^n = -\infty$ and $\kappa_{i,v_i^n+1}^n = +\infty$,
- two nonnegative 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 $\Gamma_{u_n+1}^n = 1$ and for all $0 \leq i \leq u_n$, all $0 \leq j \leq v_i^n$, all $\alpha \in (\Gamma_i^n, \Gamma_{i+1}^n]$ and all $x \in (\kappa_{i,j}^n, \kappa_{i,j+1}^n]$, we have

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

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

In the same way, if all the nodes of \mathcal{T}_n are unknown, then $\hat{f}_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise there exist \hat{u}_n , $(\hat{\Gamma}_i^n)_{1 \leq i \leq \hat{u}_n}$, $(\hat{v}_i^n)_{0 \leq i \leq u_n}$, $(\hat{\kappa}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 1 \leq j \leq \hat{v}_i^n}$, $(\hat{A}_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq \hat{v}_i^n}$, $(\hat{B}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$, $(\hat{C}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$ and $(\hat{D}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$ verifying the same properties as their f_n -counterparts except that we have $\hat{A}_{i,0}^n = \hat{A}_{i,\hat{v}_i^n}^n = \hat{B}_{i,\hat{v}_i^n}^n \leq \phi(\tau_n)$ and $\hat{B}_{i,0}^n = 0$, and such that, for all $0 \leq i \leq \hat{u}_n$, all $0 \leq j \leq \hat{v}_i^n$, all $\alpha \in (\hat{\Gamma}_i^n, \hat{\Gamma}_{i+1}^n]$ and all $x \in (\hat{\kappa}_{i,j}^n, \hat{\kappa}_{i,j+1}^n]$, we have

$$\hat{f}_n(\alpha, x) = -\hat{A}_{i,j}^n \alpha x + \hat{B}_{i,j}^n x + \hat{C}_{i,j}^n \alpha + \hat{D}_{i,j}^n.$$

Proof. In Didier (2017), this theorem was established by considering another parametrization of the parsimony cost, namely

$$\Delta_\gamma(x, y, \tau_m) = \begin{cases} \gamma \phi(\tau_m)(y - x) & \text{if } x < y, \\ \phi(\tau_m)(x - y) & \text{if } x \geq y. \end{cases} \quad (3)$$

The version above was obtained by dividing the cost just above and the corresponding partial cost functions by $\gamma + 1$ and by applying the substitution $\alpha = \frac{\gamma}{\gamma + 1}$ to transform the coefficients and the bounds of these cost functions. \square

A similar result can be stated for h_n and \hat{h}_n .

Theorem 2. Let n be a non-root unknown node of \mathcal{T} . The maps h_n and \hat{h}_n are piecewise-linear and continuous and, for all $\alpha > 0$, the maps $x \rightarrow h_n(\alpha, x)$ and $x \rightarrow \hat{h}_n(\alpha, x)$ are both convex.

More precisely, if all the nodes of \mathcal{T}_n are unknown then $h_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise, there exist:

- an integer w_n and a strictly increasing positive real sequence $(\Upsilon_i^n)_{1 \leq i \leq w_n}$,
- an integer sequence $(z_i^n)_{0 \leq i \leq w_n}$ and for, all $0 \leq i \leq w_n$, a sequence $(\rho_{i,j}^n)_{1 \leq j \leq z_i^n}$ of known values of $\mathcal{T}_{\setminus \mathcal{T}_n}$ (i.e. in $\vartheta(\mathcal{K}_{\setminus \mathcal{K}_n})$) verifying $\rho_{i,j}^n < \rho_{i,j'}^n \Leftrightarrow j < j'$; by convention we set $\rho_{i,0}^n = -\infty$ and $\rho_{i,z_i^n+1}^n = +\infty$,
- four real bi-sequences $(O_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$, $(P_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$, $(Q_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$ and $(R_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$

such that, by setting $\Upsilon_0^n = 0$ and $\Upsilon_{w_n+1}^n = 1$ and for all $0 \leq i \leq w_n$, all $0 \leq j \leq z_i^n$, all $\alpha \in (\Upsilon_i^n, \Upsilon_{i+1}^n]$ and all $x \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$, we have

$$h_n(\alpha, x) = -O_{i,j}^n \alpha x + P_{i,j}^n x + Q_{i,j}^n \alpha + R_{i,j}^n,$$

all the coefficients being such that h_n is continuous. Moreover, the sequence $(-O_{i,j}^n \alpha + P_{i,j}^n)_{0 \leq j \leq z_i^n}$ (i.e. the x -coefficients of h_n) is increasing with $-O_{i,z_i^n}^n \alpha + P_{i,z_i^n}^n \geq 0$ for all $\alpha \in (\Upsilon_i^n, \Upsilon_{i+1}^n]$.

In the same way, if all the nodes of \mathcal{T}_n are unknown, then $\hat{h}_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise there exist $\hat{w}_n, (\hat{Y}_i^n)_{1 \leq i \leq \hat{w}_n}, (\hat{z}_i^n)_{0 \leq i \leq w_n}, (\hat{\rho}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 1 \leq j \leq \hat{z}_i^n}, (\hat{O}_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq \hat{z}_i^n}, (\hat{P}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}, (\hat{Q}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}$ and $(\hat{R}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}$ verifying the same properties as their h_n -counterparts and such that, for all $0 \leq i \leq \hat{w}_n$, all $0 \leq j \leq \hat{z}_i^n$, all $\alpha \in (\hat{Y}_i^n, \hat{Y}_{i+1}^n]$ and all $x \in (\hat{\rho}_{i,j}^n, \hat{\rho}_{i,j+1}^n]$, we have

$$\hat{h}_n(\alpha, x) = -\hat{O}_{i,j}^n \alpha x + \hat{P}_{i,j}^n x + \hat{Q}_{i,j}^n \alpha + \hat{R}_{i,j}^n.$$

Proof. See Appendix A. □

Theorem 2 suggests the procedure sketched in Algorithm 1 for computing the functions h_n and \hat{h}_n for all non-root nodes n . The complexity of Algorithm 1 is studied in Appendix B.

Theorem 3. *Under the assumption that the number of children of any node is bounded independently of the size of the tree, the algorithmic complexity of the computation of the cost functions h_n and \hat{h}_n for all nodes n of a tree \mathcal{T} with a set of known nodes \mathcal{K} is $O(|\mathcal{T}|^2 \cdot |\mathcal{K}|^2)$ both in time and memory space.*

Proof. See Appendix B.1. □

2.4 Integrating parsimonious costs over asymmetry parameter(s)

2.4.1 Single parameter case: no-split cost

We show here how to sum the parsimonious α -cost $\check{\Delta}_\alpha(\mathcal{T}, \vartheta)$ over all asymmetry parameters α . Let us start by recalling Remark 3 from Didier (2017).

Remark 1 (Didier (2017)). *Let r be the root of \mathcal{T} and $(\dot{\Gamma}_k^r)_{0 \leq k \leq w_r+1}$ be the elements of*

$$\left(\bigcup_{1 \leq i \leq u_r} \left(\left\{ \frac{B_{i,j}^r}{A_{i,j}^r} \mid 0 \leq j \leq v_i^r \right\} \cap (\Gamma_i^r, \Gamma_{i+1}^r) \right) \right) \cup \{\Gamma_i^r \mid 0 \leq i \leq u_r + 1\},$$

indexed in increasing order. There exists a sequence $(\rho_k^r)_{0 \leq k \leq w_r}$ of increasing values in $\vartheta(\mathcal{K})$ such that for all $0 \leq k \leq w_r$ and all $\dot{\Gamma}_k^r \leq \alpha < \dot{\Gamma}_{k+1}^r$, an α -parsimonious reconstruction associates ρ_k^r to r . If $\dot{\Gamma}_k^r < \alpha < \dot{\Gamma}_{k+1}^r$, then all α -parsimonious reconstructions associate ρ_k^r to r .

An algorithm computing $(\dot{\Gamma}_k^r)_{0 \leq k \leq w_r+1}$ and $(\rho_k^r)_{0 \leq k \leq w_r}$ from f_r is provided in Didier (2017). Its complexity is linear with the total number of bounds required for defining f_r .

From Theorem 1, we have that the most parsimonious α -cost is, for all $\dot{\Gamma}_k^r < \alpha < \dot{\Gamma}_{k+1}^r$,

$$\begin{aligned} \check{\Delta}_\alpha(\mathcal{T}, \vartheta) &= \min_x f_n(\alpha, x) \\ &= \left(-A_{\mathbf{i}(k), \mathbf{j}(k)}^r \alpha + B_{\mathbf{i}(k), \mathbf{j}(k)}^r \right) \rho_k^r + C_{\mathbf{i}(k), \mathbf{j}(k)}^r \alpha + D_{\mathbf{i}(k), \mathbf{j}(k)}^r \end{aligned}$$

where $\mathbf{i}(k)$ is the index i such that $(\dot{\Gamma}_k^r, \dot{\Gamma}_{k+1}^r] \subseteq (\Gamma_i^r, \Gamma_{i+1}^r]$ and $\mathbf{j}(k)$ is the index j such that $\rho_k^r \in (\kappa_{\mathbf{i}(k), \mathbf{j}(k)}^r, \kappa_{\mathbf{i}(k), \mathbf{j}(k)+1}^r]$.

Integrating the parsimonious cost of (\mathcal{T}, ϑ) over all possible values of the asymmetry parameter can then be done explicitly:

```

Function compute_h_and_hat_h
  input : a tree  $\mathcal{T}$  of root  $r$ 
  /* Compute functions  $h_k$  and  $\hat{h}_k$  for all children  $k$  of the root */
  forall  $k \in \mathcal{C}_r$  do
     $h_k \leftarrow \sum_{c \in (\mathcal{C}_r) \setminus \{k\}} \hat{f}_c$ ;
    recursion_h_and_hat_h( $k$ );

```

```

Function recursion_h_and_hat_h
  input : a node  $n$  of  $\mathcal{T}$ 
  /* Call function to_stem on  $h_n$  for computing  $\hat{h}_n$ , use it for computing  $h_k$  for all children  $k$  of
   $n$  then recursively treat all children  $k$  */
   $\hat{h}_n \leftarrow \text{to\_stem}(h_n)$ ;
  forall  $k \in \mathcal{C}_n$  do
     $h_k \leftarrow \hat{h}_n + \sum_{c \in (\mathcal{C}_n) \setminus \{k\}} \hat{f}_c$ ;
    recursion_h_and_hat_h( $k$ );

```

```

Function to_stem
  input : a piecewise linear function  $h_n$  given under notations of Theorem 2
  output: the piecewise linear function  $\hat{h}_n$  given under notations of Theorem 2
  /* Compute  $\hat{h}_n$  from  $h_n$  */
   $\hat{w}_n \leftarrow -1$ ;  $\hat{\Upsilon}_0^n \leftarrow \Upsilon_0^n$ ;
  for  $i \leftarrow 0$  to  $w_n$  do
     $j^+ \leftarrow \min \left\{ j \mid \frac{P_{i,j-1}^n + \phi(\tau_n)}{O_{i,j-1}^n + \phi(\tau_n)} > \Upsilon_i^n \right\}$ ;  $M_+ \leftarrow \max \left\{ j \mid \frac{P_{i,j-1}^n + \phi(\tau_n)}{O_{i,j-1}^n + \phi(\tau_n)} < \Upsilon_{i+1}^n \right\}$ ;
     $j^- \leftarrow \min \left\{ j \mid \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)} > \Upsilon_i^n \right\}$ ;  $M_- \leftarrow \max \left\{ j \mid \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)} < \Upsilon_{i+1}^n \right\}$ ;
    while  $j^- \leq M_-$  or  $j^+ \leq M_+$  do
       $\hat{w}_n \leftarrow \hat{w}_n + 1$ ;
      for  $\hat{\Upsilon}_{\hat{w}_n}^n < \gamma < \hat{\Upsilon}_{\hat{w}_n+1}^n$ , set
        
$$\hat{h}_n(\alpha, x) = \begin{cases} (1 - \alpha)\phi(\tau_n)(\rho_{i,j^+}^n - x) + h_n(\alpha, \rho_{i,j^+}^n) & \text{if } x < \rho_{i,j^+}^n, \\ h_n(\alpha, x) & \text{if } \rho_{i,j^+}^n \leq x < \rho_{i,j^-}^n, \\ \alpha\phi(\tau_n)(x - \rho_{i,j^-}^n) + h_n(\alpha, \rho_{i,j^-}^n) & \text{if } x \geq \rho_{i,j^-}^n. \end{cases}$$

      if  $j^- \leq M_-$  or  $j^+ \leq M_+$  then
         $\hat{\Upsilon}_{\hat{w}_n+1}^n \leftarrow \min \left\{ \frac{P_{i,j^+-1}^n + \phi(\tau_n)}{O_{i,j^+-1}^n + \phi(\tau_n)}, \frac{P_{i,j^-}^n}{O_{i,j^-}^n + \phi(\tau_n)} \right\}$ ;
        if  $\frac{P_{i,j^+-1}^n + \phi(\tau_n)}{O_{i,j^+-1}^n + \phi(\tau_n)} \leq \hat{\Upsilon}_{\hat{w}_n}^n$  then  $j^+ \leftarrow j^+ + 1$ ;
        if  $\frac{P_{i,j^-}^n}{O_{i,j^-}^n + \phi(\tau_n)} \leq \hat{\Upsilon}_{\hat{w}_n}^n$  then  $j^- \leftarrow j^- + 1$ ;
      else  $\hat{\Upsilon}_{\hat{w}_n+1}^n \leftarrow \Upsilon_{i+1}^n$ ;
  return  $\hat{h}_n$ 

```

Algorithm 1: Computation of h_n and \hat{h}_n for all non-root nodes n of \mathcal{T} .

$$\begin{aligned}
\int_0^1 \check{\Delta}_\alpha(\mathcal{T}, \vartheta) d\alpha &= \sum_{0 \leq k \leq w_r} \int_{\dot{\Gamma}_k^r}^{\dot{\Gamma}_{k+1}^r} \check{\Delta}_\alpha(\mathcal{T}, \vartheta) d\alpha \\
&= \sum_{0 \leq k \leq w_r} \left[\left(-A_{\mathbf{i}(k), \mathbf{j}(k)}^r \rho_k^r + C_{\mathbf{i}(k), \mathbf{j}(k)}^r \right) \left(\frac{(\dot{\Gamma}_{k+1}^r)^2 - (\dot{\Gamma}_k^r)^2}{2} \right) \right. \\
&\quad \left. + \left(B_{\mathbf{i}(k), \mathbf{j}(k)}^r \rho_k^r + D_{\mathbf{i}(k), \mathbf{j}(k)}^r \right) \left(\dot{\Gamma}_{k+1}^r - \dot{\Gamma}_k^r \right) \right]
\end{aligned}$$

The quantity thus obtained will be referred to as the *no-split cost* of (\mathcal{T}, ϑ) .

2.4.2 Two parameters case: A- and B-split costs

In order to test a trend change at a node n of \mathcal{T} , we shall consider the two ways of splitting \mathcal{T} at n displayed in Figure 1, namely, the most parsimonious costs obtained with

- (A) an asymmetry parameter α' on the subtree \mathcal{T}_n (i.e., the node-based subtree of n) and an asymmetry parameter α on the rest of \mathcal{T} (i.e., with α' on the red part and with α on the blue part of Figure 1-a).
- (B) an asymmetry parameter α' on the subtree consisting of the direct ancestral lineage of n , the branch bearing n and \mathcal{T}_n (i.e., the branch-based subtree of n) and an asymmetry parameter α on the rest of \mathcal{T} (i.e., with α' on the red part and with α on the blue part of Figure 1-b);

In Case A, given two asymmetry parameters α and α' , we are able to compute the smallest cost that can be achieved by associating x to n with asymmetry parameters α' on \mathcal{T}_n and α on the rest of \mathcal{T} . Namely, this cost is

$$\Omega_n^A(\alpha, \alpha', x) = \hat{h}_n(\alpha, x) + f_n(\alpha', x).$$

In the same way, for Case B, the smallest cost that can be achieved by associating x to the direct ancestral lineage of n and by using the asymmetry parameter α' on the branch-based subtree of n and α on the rest of \mathcal{T} is given by

$$\Omega_n^B(\alpha, \alpha', x) = h_n(\alpha, x) + \hat{f}_n(\alpha', x).$$

Let us define $\check{\Omega}_n^A(\alpha, \alpha')$ (resp. $\check{\Omega}_n^B(\alpha, \alpha')$) as the smallest cost obtained with asymmetry parameter α' on \mathcal{T}_n (resp. on $\hat{\mathcal{T}}_n$) and with asymmetry parameter α elsewhere, namely,

$$\check{\Omega}_n^A(\alpha, \alpha') = \min_x \Omega_n^A(\alpha, \alpha', x) \quad \left(\text{resp. } \check{\Omega}_n^B(\alpha, \alpha') = \min_x \Omega_n^B(\alpha, \alpha', x) \right).$$

The *A-split cost* (resp. the *B-split cost*) of (\mathcal{T}, ϑ) at n is the integral of $\check{\Omega}_n^A$ (resp. of $\check{\Omega}_n^B$) over all pairs of asymmetry parameters $(\alpha, \alpha') \in [0, 1]^2$, namely,

$$\int_0^1 \int_0^1 \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha \quad \left(\text{resp. } \int_0^1 \int_0^1 \check{\Omega}_n^B(\alpha, \alpha') d\alpha' d\alpha \right).$$

In Appendix C, we show that the space of parameters to which belong (α, α') can be split into trapezoids on which the map Ω_n^A (resp. Ω_n^B) is linear with respect to α and α' , yielding an explicit evaluation of the integral over all the pairs of parameters (α, α') of the most parsimonious cost of A-splits (resp. of B-splits), i.e., the A- and B-split costs.

3 Empirical tests of the new method

For all examples and datasets below, the parsimonious split identification was performed by setting $\phi(\tau) = \frac{1}{\tau}$ for computing the evolutionary costs.

3.1 Two contrived examples

Figure 2 and Table 1 illustrate the fact that considering more than one asymmetry parameter on a tree may or may not lead to lower the parsimonious cost. The two examples displayed in Figure 2 are based on the same tree but with different character values at tips (in brackets just after idents of the tips). Table 1 displays the A- and B-split costs for all nodes of the tree with regard to the character values of Figure 2.

In the example of Figure 2-a, the lowest split cost observed is the A-split cost at node ‘n’. It is smaller than the no-split cost (i.e., the A-split at the root ‘r’, first line of Table 1-a). Conversely, in the example of Figure 2-b, there is no A- or B-split cost lower than the no-split cost.

From a general point of view, it may be worth paying interest not only to the most parsimonious split but also to the other splits with costs smaller than the no-split cost.

3.2 Two biological datasets previously analyzed with stochastic methods

Figures 3 and 4 display all the splits that lead to a cost lower than the no-split one. Splits are represented with ‘x’ at the corresponding nodes (only A-splits were found for these datasets). Figures also display the ranks of the splits with regard to their costs (ordered from lowest to highest). The color of the subtree pending from a split depends on the corresponding split cost, more precisely on the cost of the “most recent” split including it. The color scale starts from blue, which corresponds to the no-split cost, to red for the lowest cost observed (most parsimonious split). These two figures were output by our software *ParSplit*.

For the two datasets studied here, all the most parsimonious splits are A-splits. Moreover, no B-split costs were smaller than the no-split cost. This is not a general behavior since, in some datasets (e.g., Section 4 and Figure 5), we did encounter cases in which some of the B-split costs were lower than the no-split cost and even smaller than the A-split costs at the same nodes.

We observe in Figure 3, which shows average adult female body length in meters, that the split with the lowest cost occurs at the most recent common ancestor of the mysticetes (i.e., *Balaenidae*, *Neobalaenidae* and *Balaenopteridae*). First, this split is consistent with a probable shift in diet type, since *mysticetes* are filter feeders. Second, Slater *et al.* (2010) also distinguished this clade and found that “The largest phylogenetic mean body size was recovered for *Mysticeti*...” by using approaches based on heterogeneous stochastic models from Thomas *et al.* (2006, 2009).

Figure 4 displays the parsimonious splits of the tree of *Anolis* squamates according to the body size of males (maximal snout-vent length, in mm). This dataset comes from Thomas and Freckleton (2012),

in which two methods, called *trait medusa 1* and *2* and both based on stochastic models, were used for identifying rate shifts. These methods were inspired by a method for detecting shifts in diversification rates (Rabosky 2006, Alfaro *et al.* 2009; “MEDUSA” stands for Modelling Evolutionary Diversification Using Stepwise AIC). Method *trait medusa 1* detected only two shifts at the same nodes as those of our two splits with the smallest costs, ranked in the same order as with our approach. Method *trait medusa 2* identified a third shift at the same node as that of the split ranked 5 with our method, for which the costs of split of ranks 3, 4 and 5 are very close.

4 New empirical study: stegocephalian body size evolution and origin of the amniotic egg

4.1 Dataset on stegocephalian body size

We used the dataset presented in Laurin (2004) to study body size in early stegocephalians. These data comprise 107 terminal taxa ranging in age from Frasnian (Late Devonian) to the Late Permian, but most taxa are not more recent than Early Permian. The data thus comprises taxa spanning about 125 Ma, from about 380 Ma to 255 Ma. This includes some finned tetrapodomorphs (*Eusthenopteron*, *Panderichthys*), but most (105) taxa are limbed vertebrates (stegocephalians) representing the tetrapod stem, stem-amphibians, and early stem-amniotes, as well as 37 terminal taxa of crown-amniotes. Exact counts vary between hypotheses because there is a controversy about the origin of lissamphibians and amniotes (e.g., Ruta and Coates 2007, Marjanović and Laurin 2018), but under all hypotheses, stem-amphibians and amniotes are well-represented.

Body size is represented by cranial and presacral length (defined as the length from anterior surface of atlantal centrum to anterior surface of first sacral centrum), both in mm, and presented here (Supplementary Online Material [SOM] 1) log-transformed (\ln) because body size distribution of most taxa is closer to a log-normal distribution than to a normal distribution, so evolutionary analyses of body size are better performed on log-transformed values than on raw values (e.g., Bokma *et al.* 2016). This presumably reflects the fact that the amplitude of body size variations are proportional to body size itself ; thus, for instance, it is far easier for a 1 m long animal to increase its length by 1 cm than for another animal measuring only 5 cm.

Recent work has resulted in a slight shift in the established soft consensus on stegocephalian phylogeny, but to facilitate comparisons with the results of Laurin (2004), we reanalyzed the data using our new methods on the same set of phylogenies, to better highlight the differences in results reflecting the methodological innovations introduced by Didier (2017) and here (see below). In addition, we produced two new trees that reflect more recent analyses that place *Solenodonsaurus* either at the base of lepospondyls (Danto *et al.* 2012) or stemward of seymouriamorphs (Marjanović and Laurin 2018). Other than that, only minor changes would need to be made to reflect the current consensus; recent

research has confirmed that the trees used here still adequately represent the current views on early stegocephalian phylogeny (Marjanović and Laurin 2013). The main controversy concerns the origin of lissamphibians (Marjanović and Laurin 2013, 2018). In most of our trees (all except for trees 13 and 14; see Table 2 and SOM 2), we assume that it was among lepospondyls and label the relevant clades (Tetrapoda, Amphibia, and Reptiliomorpha or pan-amniotes) as such, but note that assuming an origin of lissamphibians among temnospondyls (as is done under trees 13 and 14; see Table 2 and SOM 2) does not alter our results or conclusions (see below). As in Laurin (2004), we tested six variants of the main reference tree, by varying minimal internal branch lengths (3 or 5 Ma), and by placing the controversial taxon *Westlothiana* in three positions: as sister-group of the clade that we interpret as the tetrapod crown (lepospondyls, *Solenodonsaurus*, and cotylosaurs), as the basalmost amphibian (lepospondyl), or as the basalmost reptiliomorph (pan-amniote). The topology of Panchen and Smithson (1988), which we use as the seventh tree (trees 13 and 14 in Table 2 and SOM 2), has not been recovered in recent analyses, but we retained it because it reflects the long-held conception that lepospondyls are amphibians, which is still correct according to some studies, including those that we favor (Marjanović and Laurin 2013, 2018). It also reflects the hypothesis that embolomeres and seymouriamorphs are stem-amniotes, which is also favored by most recent studies (e.g., Ruta and Coates 2007), though this is incompatible with our preferred topology (Marjanović and Laurin 2013, 2018). These seven trees have been used as such to assess evolution of skull length, and they were pruned to assess evolution of the presacral vertebral length, because the latter was available in a subset of taxa. This yields 14 trees representing those previously used by Laurin (2004), on which our tests were carried out. In addition, we modified the variant of the main reference tree with 3 ma minimal internal branch lengths and with *Westlothiana* as a basal amphibian (lepospondyl) by placing *Solenodonsaurus* closer to lepospondyls than to *Westlothiana* (Danto *et al.* 2012) or stemward of seymouriamorphs (Marjanović and Laurin 2013, 2018), which gave two more trees used only for analysis of cranial characters. This results in a total of 16 analyses tied to as many trees, which are presented in SOM 2 (and also present in SOM 1).

4.2 Analytical methods

We use the new method described above to determine if there is evidence of a higher rate of changes in trends in body size evolution in and near the amniote stem than in other parts of the tree, or if changes there are more probable than elsewhere. This is predicted by H1 and H3 (see introduction). We looked for trend changes on the amniote stem and in its vicinity, expanding the search both stemward (along much of the tetrapod stem) and along the basalmost branches of Amphibia, in case that some changes actually occurring on the tetrapod stem might be slightly misplaced by our analysis. The changes are numbered according to their significance rank, with rank 1 being the most significant.

Given the ability of asymmetric linear parsimony to account for trends in character evolution (Didier 2017), some nodal reconstructions are more plausible than others. Thus, for the nodal reconstructions where our previous analyses (Laurin 2004) found trends for body size increase (in reptiliomorphs), we

distinguish between all equiparsimonious reconstructions and the most plausible ones. In practice, this means that for Amniota and nodes just below (Cotylosauria) or above (Synapsida and Sauropsida), the lower half of the nodal values are the most plausible (because these reflect estimates that assume a trend towards increasing values over time). Under the topology derived from Panchen and Smithson (1988), the amniote stem is very long, but we apply this preference (for lower half of the estimates) only from the node basal to *Solenodonsaurus* and cotylosaurs and upwards. This is because the small size of *Westlothiana* (located immediately stemward of this) suggests that the amniotic egg had not appeared at the node just below *Westlothiana*, and because the presence of aquatic larvae in seymouriamorphs (Laurin 2000), and the presence of lateral-line canals in embolomeres (e.g., Holmes and Carroll 2010) also indicate that these taxa did not lay amniotic eggs and exhibited the biphasic reproductive lifestyle retained in many lissamphibians (Duellman and Trueb 1986). However, if Carroll’s (1970, 1991) scenario is correct, the estimates near the base of stem-amniotes should on the contrary be closer to those yielded by the asymmetry parameter favoring decrease in body size (upper range of size estimates).

4.3 Detected changes in evolutionary trends

Among the nine trees and two characters tested, our method yields a single change (but it is only the seventh most significant possible change in skull length on that tree) that may add limited support to Carroll’s (1970, 1991) scenario (Table 2). In the tree with the main topology in which *Westlothiana* is placed among stem-amniotes (tree 5, T1swr in Table 2 and SOM 2), we detect a shift in trends in skull length associated with Tetrapoda, the node that includes cotylosaurs and lepospondyls, according to some recent studies (Marjanović and Laurin 2013, 2018). This presumably reflects a shift towards decreasing size, given the small size of most lepospondyls and that of *Westlothiana*. This is a surprisingly low number of shifts congruent with Carroll’s (1970, 1991) scenario, as shown by a thorough examination of the set of possible shifts detected over one of the trees, selected as a representative example. Under our main reference tree, which we illustrate (Fig. 5; this matches tree 1, T1s in Table 2 and SOM 2) to facilitate comparisons with the results of Laurin (2004), our new method finds 42 possible locations of trend changes in the skull, but none of them are located on the amniote stem or in the basalmost branches of Amniota. The five most significant shifts are within sphenacodontid synapsids, apparently linked with a changes induced by the comparatively small *Tetraceratops*, the putative first therapsid (Amson and Laurin 2011). The next several changes are likewise in branches within Amniota fairly far from its root. The only change located close to the amniote stem (though not directly on it) is the 21st most significant, which is located at the base of lepospondyls, and which presumably reflects their fairly small size. This pattern is typical of the nine trees over which we performed our analyses for cranial length; we found an average of 42 changes per tree. For presacral length, we found fewer changes (an average of 20 per trees), which is unsurprising given that we had fewer data for this character. Thus, out of 517 possible trend changes identified by our analyses, a single one may support Carroll’s (1970, 1991) scenario. Remarkably, this remains true even if we interpret temnospondyls, rather than lepospondyls,

as the amphibian stem.

4.4 Nodal estimates of body size in selected clades

Our results on nodal estimates of body size are more ambiguous than those on trend changes (Tables 3-5), even if we consider only the lower half of the estimates on parts of the tree where increase in body size is expected (see Section 4.2). All trees yield nodal reconstructions that are compatible with the hypothesis that at least some stem-amniotes may have measured less than 10 cm in snout-vent length (Fig. 6; Table 5), and this is congruent with Carroll’s (1970, 1991) scenario. However, in all cases, the smallest hypothetical ancestor is the last common ancestor of lepospondyls and amniotes (Tables 3-5), which we interpret as the origin of the tetrapod crown. After that, body size increases, and only the node Cotylosauria (Fig. 6) may still have a snout-vent length inferior to 10 cm. However, in all cases, only the minimal size estimate (which implies a strong trend towards body size increase) is congruent with Carroll’s (1970, 1991) scenario. There is substantial ambiguity in the nodal reconstructions, which reflects the wide range in asymmetry coefficient considered, and the maximal snout-vent length of the ancestors mentioned above systematically reaches at least 1.64 m, so uncertainty about the size of these hypothetical ancestors spans more than one order of magnitude, which prevents us from drawing firm conclusions on this basis. Again, interpreting temnospondyls rather than lepospondyls as the amphibian stem does not change our results because inferred nodal values in the basal parts of the tree tend to be higher than in what we interpret as the tetrapod crown.

5 Discussion

Changes in the rate and/or the trend of character evolution are often associated with major biological phenomena such as adaptation or evolutionary convergence (Ingram and Mahler 2013). Detecting shifts in the evolution of a trait is thus an important question, which has been widely addressed, notably in the case of quantitative characters. The performance of the new method that we presented above seems to be good. It is worth noting that, as observed on the two biological datasets presented in Section 3.2 (on cetacean and *Anolis* squamate body size), our parsimony-based approach identifies the evolutionary shifts detected using methods based on stochastic models of evolution. In all the situations where parsimony is preferred, our approach provides a parsimonious solution for identifying evolutionary shifts.

We emphasize the fact that the various splits displayed in Figures 3, 4 and 5 have to be considered as different alternatives for splitting the tree into two parts and absolutely not as multiple, simultaneous splits of the tree into more than two parts. Finding the most parsimonious split of the tree into $k > 2$ parts in the same sense as above could be performed with ideas similar as those developed in this work. The algorithmic complexity increases exponentially with k , but computations would be still feasible for small k (e.g. smaller than 5 for usual phylogenetic trees). In particular, one could seek this for the most

parsimonious cost obtained with a same asymmetry parameter on k subtrees and another parameter for the rest of the tree, following the ideas of Ingram and Mahler (2013) for studying convergence with Ornstein-Uhlenbeck processes.

Our results refute Carroll's (1970, 1991) scenario about the origin of the amniotic egg. Of the 16 tested combinations of trees and characters, only one yielded a change that may lend additional support to Carroll's (1970, 1991) scenario. This change (in skull length) is located at the base of a clade that includes lepospondyls and cotylosaurs. We interpret this as the base of Tetrapoda, and the change is probably towards a reduction in size near the base of this clade. This is not entirely congruent with Carroll's (1970, 1991) scenario because that scenario does not postulate a decrease in size in early amphibians. Lepospondyls are sometimes interpreted as reptiliomorphs (Ruta and Coates 2007), though we consider it more likely that they are stem-amphibians (Marjanović and Laurin 2013, 2018). However, even if lepospondyls are considered as reptiliomorphs, this change is still not really congruent with Carroll's (1970, 1991) scenario because some lepospondyls, such as *Microbrachis*, retained aquatic larvae with external gills (Olori 2015), whereas others were neotenic with a well-developed hyobranchial apparatus, and as such, may have retained external gills through adulthood, although this latter point is uncertain (Wellstead 1991). This suggests that this node (the smallest clade that includes lepospondyls and amniotes) is not associated with the origin of the amniotic egg. Thus, under this alternative interpretation of the position of lepospondyls (as reptiliomorphs rather than amphibians), the only way to reconcile the decrease in body size with Carroll's (1970, 1991) scenario is to suppose that this node is associated with the shift towards terrestrial egg-laying that preceded the appearance of the amniotic egg, but that in lepospondyls (at least; reproductive mode is unknown in diadectomorphs), aquatic larvae were retained. This is possible if eggs were laid close to the water, which would have allowed the larvae to reach it within minutes after hatching. Finally, the fact that this possible change in trends in skull length evolution is only the seventh most significant one weakens further the support that this may lend to Carroll's (1970, 1991) scenario. Also, it is striking that out of 517 possible shifts in trends in body size evolution identified by our method, a single one may lend some support to Carroll's (1970, 1991) scenario. Obviously, our results suggest that the amniote stem was not a part of the tree with a higher rate of change in body size trends than the background rate found in Permo-Carboniferous stegocephalians, whether that stem originates at the divergence between amniotes and lepospondyls (our preferred hypothesis) or between amniotes and temnospondyls.

Our results about nodal body size estimates show less discriminating power. The minimal estimates for some nodes are compatible with Carroll's (1970, 1991) scenario, but the maximal parsimonious estimates are far above the size predictions consistent with that scenario. However, we can state that our nodal estimates do not provide substantial support for that scenario, without refuting it strongly either. In this respect, they are congruent with our previous analysis based on squared-change parsimony (Laurin 2004).

The new analyses presented here reinforce the confidence that we can have in the mismatch between

Carroll’s (1970, 1991) predictions and our inferences about body size evolution around the origin of the amniotic egg because our new analyses do not rest on a hypothesis that body size (log-transformed) evolved according to a Brownian motion model, though Laurin (2004) found that this was a reasonable assumption, and because the method that we used here (Didier 2017) allows us to estimate nodal values in the presence of strong trends in body size evolution (which is not possible using squared-change parsimony). In addition, our new method allowed us to check if significant shifts in trends in body size evolution occurred in the parts of the tree in which such shifts should occur, under Carroll’s (1970, 1991) scenario, something that was not attempted by Laurin (2004). Nevertheless, our results overlap broadly, which is not surprising given that both methods yield a fairly broad range of values. Another reason for the lack of obvious discrepancy in the results of both studies is the fact that we did not find evidence of drastic trend changes in the critical part of the tree, coupled with the fact that Laurin (2004: tables 9, 10) had found evidence of a trend in only a small part of the tree (amniotes and their stem), and only in some trees. The only remaining caveat is that this part of the tree is poorly represented in the fossil record under our preferred topology (but not if temnospondyls are assumed to be stem-amphibians), and in this respect, the situation has not changed significantly in the last two decades, even though the fossil record of amniote embryos has recently been pushed back from the Triassic (Cheng *et al.* 2004) to the Early Permian (Piñeiro *et al.* 2012). However, fossils are discovered every day, and it is possible that in a few years, a more robust test of Carroll’s (1970, 1991) scenario will be possible.

Our study cannot discriminate between the other scenarios about the origin of the amniotic egg because they make no predictions about body size evolution in early reptiliomorphs, but research on the reproductive biology of extant tetrapods (e.g., Shine *et al.* 2018) and on the physiological constraints applying to tetrapod eggs (e.g., Skulan 2000, Oftedal 2002) should continue shedding new light on this major evolutionary innovation.

References

- Alfaro, M. E., Santini, F., Brock, C., Alamillo, H., Dornburg, A., Rabosky, D. L., Carnevale, G., and Harmon, L. J. (2009). Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences*, **106**(32), 13410–13414.
- Amson, E. and Laurin, M. (2011). On the affinities of *Tetraceratops insignis*, an Early Permian synapsid. *Acta Palaeontologica Polonica*, **56**(2), 301–312.
- Bastide, P., Mariadassou, M., and Robin, S. (2017). Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **79**(4), 1067–1093.
- Bokma, F., Godinot, M., Maridet, O., Ladevèze, S., Costeur, L., Solé, F., Gheerbrant, E., Peigné, S.,

- Jacques, F., and Laurin, M. (2016). Testing for Depéret’s rule (body size increase) in mammals using combined extinct and extant data. *Systematic Biology*, **65**(1), 98–108.
- Butler, M. A. and Losos, J. B. (1997). Testing for unequal amounts of evolution in a continuous character on different branches of a phylogenetic tree using linear and squared-change parsimony: an example using lesser antillean *Anolis* lizards. *Evolution*, **51**(5), 1623–1635.
- Carroll, R. (1970). Quantitative aspects of the amphibian-reptilian transition. *Form. Funct.*, **3**, 165–178.
- Carroll, R. L. (1988). *Vertebrate paleontology and evolution*. W. H. Freeman, New York.
- Carroll, R. L. (1991). The origin of reptiles. In H.-P. Schultze and L. Trueb, editors, *Origins of the higher groups of tetrapods: controversy and consensus*, pages 331–353. Cornell University Press, Ithaca.
- Cheng, Y.-n., Wu, X.-c., and Ji, Q. (2004). Triassic marine reptiles gave birth to live young. *Nature*, **432**(7015), 383.
- Cressler, C. E., Butler, M. A., and King, A. A. (2015). Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. *Systematic Biology*, **64**(6), 953–968.
- Csűrös, M. (2008). 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.
- Danto, M., Witzmann, F., and Müller, J. (2012). Redescription and phylogenetic relationships of *Solenodonsaurus janenschii* Broili, 1924, from the Late Carboniferous of Nýřany, Czech Republic. *Fossil Record*, **15**(2), 45–59.
- Didier, G. (2017). Time-dependent-asymmetric-linear-parsimonious ancestral state reconstruction. *Bulletin of Mathematical Biology*, **79**(10), 2334–2355.
- Duellman, W. E. and Trueb, L. (1986). *Biology of amphibians*. McGraw-Hill, New York.
- Eastman, J. M., Alfaro, M. E., Joyce, P., Hipp, A. L., and Harmon, L. J. (2011). A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution*, **65**(12), 3578–3589.
- Farris, J. S. (1970). Methods for computing Wagner trees. *Systematic Biology*, **19**(1), 83–92.
- Felice, R. N. and Angielczyk, K. D. (2014). Was *Ophiacodon* (Synapsida, Eupelycosauria) a swimmer? A test using vertebral dimensions. In C. F. Kammerer, K. D. Angielczyk, and J. Fröbisch, editors, *Early evolutionary history of the Synapsida*, pages 25–51. Springer, Dordrecht.
- Ferner, K. and Mess, A. (2011). Evolution and development of fetal membranes and placentation in amniote vertebrates. *Respiratory Physiology & Neurobiology*, **178**(1), 39–50.

- Fuentes-G., J. A., Housworth, E. A., Weber, A., and Martins, E. P. (2016). Phylogenetic ANCOVA: Estimating changes in evolutionary rates as well as relationships between traits. *The American Naturalist*, **188**(6), 615–627.
- Goin, O. B. and Goin, C. J. (1962). Amphibian eggs and the montane environment. *Evolution*, **16**(3), 364–371.
- Holmes, R. B. and Carroll, R. L. (2010). An articulated embolomere skeleton (Amphibia: Anthracosauria) from the Lower Pennsylvanian (Bashkirian) of Nova Scotia. *Canadian Journal of Earth Sciences*, **47**(3), 209–219.
- Ingram, T. and Mahler, D. (2013). SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, **4**(5), 416–425.
- Khabbazian, M., Kriebel, R., Rohe, K., and Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, **7**(7), 811–824.
- Kohring, R. R. (1995). Reflections on the origin of the amniote egg in the light of reproductive strategies and shell structure. *Historical Biology*, **10**(3), 259–275.
- Landis, M. J., Schraiber, J. G., and Liang, M. (2013). Phylogenetic analysis using Lévy processes: Finding jumps in the evolution of continuous traits. *Systematic Biology*, **62**(2), 193–204.
- Laurin, M. (2000). Seymouriamorphs. In H. Heatwole and R. L. Carroll, editors, *Amphibian biology*, pages 1064–1080. Surrey Beatty & Sons Chipping Norton.
- Laurin, M. (2004). The evolution of body size, Cope’s rule and the origin of amniotes. *Systematic Biology*, **53**(4), 594–622.
- Laurin, M. (2010). *How vertebrates left the water*. University of California Press, Berkeley.
- Laurin, M. and de Buffrénil, V. (2016). Microstructural features of the femur in early ophiacodontids: A reappraisal of ancestral habitat use and lifestyle of amniotes. *C. R. Palevol*, **15**(1-2), 115–127.
- Laurin, M., Meunier, F. J., Germain, D., and Lemoine, M. (2007). A microanatomical and histological study of the paired fin skeleton of the Devonian sarcopterygian *Eusthenopteron foordi*. *Journal of Paleontology*, **81**(1), 143–153.
- Marjanović, D. and Laurin, M. (2013). The origin(s) of extant amphibians: a review with emphasis on the “lepospondyl hypothesis”. *Geodiversitas*, **35**(1), 207–272.
- Marjanović, D. and Laurin, M. (2018). Reproducibility in phylogenetics: reevaluation of the largest published morphological data matrix for phylogenetic analysis of Paleozoic limbed vertebrates. *PeerJ Preprints*, **6**, e1596v3.

- Oftedal, O. T. (2002). The origin of lactation as a water source for parchment-shelled eggs. *Journal of Mammary Gland Biology and Neoplasia*, **7**(3), 253–266.
- Olori, J. C. (2015). Skeletal morphogenesis of *Microbrachis* and *Hyloplecion* (Tetrapoda: Lepospondyli), and implications for the developmental patterns of extinct, early tetrapods. *PloS one*, **10**(6), e0128333.
- O’Meara, B. C., Ané, C., Sanderson, M. J., and Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. *Evolution*, **60**(5), 922–933.
- Panchen, A. L. and Smithson, T. R. (1988). The relationships of the earliest tetrapods. In M. J. Benton, editor, *The phylogeny and classification of the tetrapods*, volume 1, pages 1–32. Oxford University Press (Clarendon) Oxford.
- Piñeiro, G., Ferigolo, J., Meneghel, M., and Laurin, M. (2012). The oldest known amniotic embryos suggest viviparity in mesosaurs. *Historical Biology*, **24**(6), 620–630.
- Rabosky, D. L. (2006). Likelihood methods for detecting temporal shifts in diversification rates. *Evolution*, **60**(6), 1152–1164.
- Revell, L. (2008). On the analysis of evolutionary change along single branches in a phylogeny. *The American Naturalist*, **172**(1), 140–147. PMID: 18554139.
- Revell, L. J. and Collar, D. C. (2009). Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, **63**(4), 1090–1100.
- Romer, A. S. (1957). Origin of the amniote egg. *The Scientific Monthly*, **85**(2), 57–63.
- Ruta, M. and Coates, M. I. (2007). Dates, nodes and character conflict: addressing the lissamphibian origin problem. *Journal of Systematic Palaeontology*, **5**(1), 69–122.
- Shine, R., Wapstra, E., and Olsson, M. (2018). Seasonal shifts along the oviparity-viviparity continuum in a cold-climate lizard population. *Journal of Evolutionary Biology*, **31**(1), 4–13.
- Skulan, J. (2000). Has the importance of the amniote egg been overstated? *Zoological Journal of the Linnean Society*, **130**(2), 235–261.
- Slater, G. J. (2013). Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the Cretaceous-Palaeogene boundary. *Methods in Ecology and Evolution*, **4**(8), 734–744.
- Slater, G. J., Price, S. A., Santini, F., and Alfaro, M. E. (2010). Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society of London B: Biological Sciences*.
- Szarski, H. (1968). The origin of vertebrate foetal membranes. *Evolution*, **22**(1), 211–214.
- Thomas, G. H. and Freckleton, R. P. (2012). MOTMOT: models of trait macroevolution on trees. *Methods in Ecology and Evolution*, **3**(1), 145–151.

- Thomas, G. H., Freckleton, R. P., and Székely, T. (2006). Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proceedings of the Royal Society of London B: Biological Sciences*, **273**(1594), 1619–1624.
- Thomas, G. H., Meiri, S., and Phillimore, A. B. (2009). Body size diversification in *Anolis*: novel environment and island effects. *Evolution*, **63**(8), 2017–2030.
- Tihen, J. A. (1960). Comments on the origin of the amniote egg. *Evolution*, **14**(4), 528–531.
- Vallin, G. and Laurin, M. (2004). Cranial morphology and affinities of *Microbrachis*, and a reappraisal of the phylogeny and lifestyle of the first amphibians. *Journal of Vertebrate Paleontology*, **24**(1), 56–72.
- Wellstead, C. F. (1991). Taxonomic revision of the Lysorophia, Permo-Carboniferous lepospondyl amphibians. In *Bulletin of the AMNH 209*, pages 1–90. New York: American Museum of Natural History.

A Proof of Theorem 2

Proof. The proof follows the same ideas as that of Theorem 1 from Didier (2017), a noticeable difference being that here the induction goes from the root to the tips. Namely, we shall prove the theorem by proving the three following assertions:

1. if n is a child of the root, then h_n satisfies the properties of the theorem;
2. for all nodes n , if h_n satisfies the properties of the theorem, then so it is for \hat{h}_n ;
3. for all nodes n , if \hat{h}_n satisfies the properties of the theorem, then so it is for all h_k with k child of n .

By construction, if n is a direct descendant of the root r , we have that

$$h_n(\alpha, x) = \sum_{k \in (\mathcal{C}_r) \setminus \{n\}} \hat{f}_k(\alpha, x). \quad (4)$$

Assertion 1 then follows from Theorem 1.

In order to prove Assertion 2, let us remark that for all nodes n , by setting

$$g_x(y) = \Delta_\alpha(y, x, \tau_n) + h_n(\alpha, y),$$

we have

$$\hat{h}_n(\alpha, x) = \inf_{y \in \mathbb{R}} g_x(y).$$

Let us assume that h_n satisfies the properties of the theorem. In particular, h_n is piecewise linear with, for all $0 \leq i < w_n$, all $\alpha \in [\Upsilon_i^n, \Upsilon_{i+1}^n)$, all $0 \leq j \leq z_i^n$ and all $y \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$:

$$h_n(\alpha, y) = -O_{i,j}^n \alpha y + P_{i,j}^n y + Q_{i,j}^n \alpha + R_{i,j}^n.$$

Let us first consider the case where $y \geq x$ for which we have $\Delta_\alpha(x, y, \tau_n) = (1 - \alpha)\phi(\tau_n)(y - x)$.

For all $0 \leq i < w_n$, all $\alpha \in [\Upsilon_i^n, \Upsilon_{i+1}^n)$, all $0 \leq j \leq z_i^n$ and all $y \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$, we get that

$$\begin{aligned} g_x(y) &= (1 - \alpha)\phi(\tau_n)(y - x) - O_{i,j}^n \alpha y + P_{i,j}^n y + Q_{i,j}^n \alpha + R_{i,j}^n \\ &= [-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n)] y - (1 - \alpha)\phi(\tau_n)x + Q_{i,j}^n \alpha + R_{i,j}^n \end{aligned}$$

Since h_n satisfies the properties of the theorem, it is convex with respect to x . The sequence $(-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n))_{0 \leq j \leq z_i^n}$ thus increases with j . Let j_α^+ be defined as

- $j_\alpha^+ = 0$ if $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n) \geq 0$ for all $0 \leq j \leq \widehat{z}_i^n$;
- the greatest integer smaller or equal to z_i^n and such that $-(O_{i,j_\alpha^+}^n + \phi(\tau_n))\alpha + P_{i,j_\alpha^+}^n + \phi(\tau_n) < 0$ otherwise. In other words, $j_\alpha^+ - 1$ is the index of the last interval on which g_x strictly decreases with $y \geq x$ and is the greatest index such that $\alpha \leq \frac{P_{i,j_\alpha^+}^n + \phi(\tau_n)}{O_{i,j_\alpha^+}^n + \phi(\tau_n)}$.

Since g_x decreases before $\rho_{i,j_\alpha^+}^n$ and increases after this value, we get that

$$\inf_{y \geq x} g_x(y) = \begin{cases} g_x(\rho_{i,j_\alpha^+}^n) & \text{if } x < \rho_{i,j_\alpha^+}^n, \\ g_x(x) & \text{otherwise,} \end{cases} \quad (5)$$

and that

$$g_x(\rho_{i,j_\alpha^+}^n) \leq g_x(x) \quad \text{for all } x < \rho_{i,j_\alpha^+}^n. \quad (6)$$

Let us now consider the case where $y < x$ for which we have $\Delta_\alpha(x, y, \tau_n) = \alpha\phi(\tau_n)(x - y)$. For all $0 \leq i < w_n$, all $\alpha \in [\Upsilon_i^n, \Upsilon_{i+1}^n)$, all $0 \leq j \leq z_i^n$ and all $y \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$, we get that

$$\begin{aligned} g_x(y) &= \alpha\phi(\tau_n)(x - y) - O_{i,j}^n \alpha y + P_{i,j}^n y + Q_{i,j}^n \alpha + R_{i,j}^n \\ &= [-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n] y + \alpha\phi(\tau_n)x + Q_{i,j}^n \alpha + R_{i,j}^n \end{aligned}$$

Let us define j_α^- as

- $j_\alpha^- = z_i^n + 1$ if $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n < 0$ for all $0 \leq j \leq \widehat{z}_i^n$;
- the smallest index such that $-(O_{i,j_\alpha^-}^n + \phi(\tau_n))\alpha + P_{i,j_\alpha^-}^n \geq 0$ otherwise. In other words, j_α^- corresponds to the first interval on which g_x does not decrease with $y < x$ and is the smallest index such that $\alpha \geq \frac{P_{i,j_\alpha^-}^n}{O_{i,j_\alpha^-}^n + \phi(\tau_n)}$.

We have that

$$\inf_{y < x} g_x(y) = \begin{cases} g_x(x) & \text{if } x < \rho_{i,j_\alpha^-}^n, \\ g_x(\rho_{i,j_\alpha^-}^n) & \text{otherwise.} \end{cases} \quad (7)$$

and that

$$g_x(\rho_{i,j_\alpha^-}^n) \leq g_x(x) \quad \text{for all } x \geq \rho_{i,j_\alpha^-}^n. \quad (8)$$

Since $\phi(\tau_n)$ is positive, $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n \geq 0$ implies that $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n) \geq 0$ for all i and j . It follows that that

$$j_\alpha^+ \leq j_\alpha^- \quad \text{and} \quad \rho_{i,j_\alpha^+}^n \leq \rho_{i,j_\alpha^-}^n.$$

The fact that

$$\begin{aligned}\widehat{h}_n(\alpha, x) &= \inf_{y \in \mathbb{R}} g_x(y) \\ &= \min\{\inf_{y \geq x} g_x(y), \inf_{y < x} g_x(y)\},\end{aligned}$$

altogether with Equations 5 and 7 and Inequalities 6 and 8 imply that

$$\widehat{h}_n(\alpha, x) = \begin{cases} (1 - \alpha)\phi(\tau_n)(\rho_{i,j_\alpha^+}^n - x) + h_n(\alpha, \rho_{i,j_\alpha^+}^n) & \text{if } x < \rho_{i,j_\alpha^+}^n, \\ h_n(\alpha, x) & \text{if } \rho_{i,j_\alpha^+}^n \leq x < \rho_{i,j_\alpha^-}^n, \\ \alpha\phi(\tau_n)(x - \rho_{i,j_\alpha^-}^n) + h_n(\alpha, \rho_{i,j_\alpha^-}^n) & \text{if } x \geq \rho_{i,j_\alpha^-}^n. \end{cases} \quad (9)$$

We assume here that h_n satisfies the properties of the theorem. This first implies that for all α , the map $x \rightarrow \widehat{h}_n(\alpha, x)$ is piecewise linear and continuous. Next, the map $x \rightarrow h_n(\alpha, x)$ is convex, in particular between $\rho_{i,j_\alpha^+}^n$ and $\rho_{i,j_\alpha^-}^n$. Moreover, definitions of j_α^+ and j_α^- ensure that if $j_\alpha^+ > 0$ then $-O_{i,j_\alpha^+}^n \alpha + P_{i,j_\alpha^+}^n \geq (1 - \alpha)\phi(\tau_n)$ and if $j_\alpha^- \leq z_i^n$ then $-O_{i,j_\alpha^-}^n \alpha + P_{i,j_\alpha^-}^n \leq \alpha\phi(\tau_n)$. It follows that the map $x \rightarrow \widehat{f}_n(\alpha, x)$ is still convex. Moreover, if $j_\alpha^- \leq z_i^n$, then the x -coefficient of \widehat{h}_n for $x \geq \rho_{i,j_\alpha^-}^n$ is $\alpha\phi(\tau_n)$, which is nonnegative. Otherwise, we have in particular that $-\widehat{O}_{i,z_i^n}^n \alpha + \widehat{P}_{i,z_i^n}^n = -O_{i,z_i^n}^n \alpha + P_{i,z_i^n}^n > \alpha\phi(\tau_n) \geq 0$, which implies that the x -coefficient of \widehat{h}_n for $x \in (\rho_{i,z_i^n}^n, \infty]$ is positive.

Let us put Φ_1, \dots, Φ_p for the elements of

$$\left(\left\{ \frac{P_{i,j}^n + \phi(\tau_n)}{O_{i,j}^n + \phi(\tau_n)} \middle| 0 \leq j \leq z_i^n \right\} \cup \left\{ \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)} \middle| 0 \leq j \leq z_i^n \right\} \right) \cap (\Upsilon_i^n, \Upsilon_{i+1}^n)$$

indexed in increasing order. By construction, the indices j_α^+ and j_α^- are both constant over all the sub-intervals $(\Upsilon_i^n, \Phi_1]$, $(\Phi_1, \Phi_2]$, \dots , $(\Phi_{p-1}, \Phi_p]$, $(\Phi_p, \Upsilon_{i+1}^n]$. The fact that h_n satisfies the properties of the theorem and Equation 9 imply that for all values x , the map $\alpha \rightarrow \widehat{h}_n(\alpha, x)$ is linear over all the sub-intervals above. We get that the map \widehat{h}_n is piecewise linear and its continuity with regard to α is straightforward to verify at all bounds Φ_k for $1 \leq k \leq p$.

Last, Assertion 3 follows straightforwardly from Theorem 1 and from the fact that, for all children k of n , we have

$$h_k(\alpha, x) = \widehat{h}_n(\alpha, x) + \sum_{c \in (\mathcal{C}_n) \setminus \{k\}} \widehat{f}_c(\alpha, x), \quad (10)$$

which ends the proof. \square

B Complexity of Algorithm 1 – Proof of Theorem 3

Like the algorithm computing $(f_n)_{n \in \mathcal{T}}$ and $(\widehat{f}_n)_{n \in \mathcal{T}}$ provided in Didier (2017), the complexity of Algorithm 1 depends on the total number of bounds (over asymmetry parameters and over character values) required for defining $(h_n)_{n \in \mathcal{T}}$ and $(\widehat{h}_n)_{n \in \mathcal{T}}$. The number of bounds over character values, below referred to as x -bounds, is, by construction, smaller than $|\mathcal{K}|$ (we don't count the two extrema $-\infty$ and $+\infty$). We shall bound the number of intervals over asymmetry parameters, below referred to as α -bounds, following the same ideas as in Didier (2017).

From now on, we make the implicit assumption that all bounds of the partial cost functions are actually required to define it. Namely, under the notations of Theorem 2 and for all nodes n of \mathcal{T} , the cost function h_n is such that:

- for all $0 \leq i \leq w_n$ and all $0 \leq j < z_i^n$, we have that

$$(O_{i,j}^n, P_{i,j}^n, Q_{i,j}^n, R_{i,j}^n) \neq (O_{i,j+1}^n, P_{i,j+1}^n, Q_{i,j+1}^n, R_{i,j+1}^n);$$

- for all $0 \leq i < w_n$, there exists $0 \leq j \leq z_i^n$ and $0 \leq j' \leq z_{i+1}^n$ such that

$$(\rho_{i,j}^n, \rho_{i,j+1}^n) \cap (\rho_{i+1,j'}^n, \rho_{i+1,j'+1}^n) \neq \emptyset \text{ and } (O_{i,j}^n, P_{i,j}^n, Q_{i,j}^n, R_{i,j}^n) \neq (O_{i+1,j'}^n, P_{i+1,j'}^n, Q_{i+1,j'}^n, R_{i+1,j'}^n);$$

and that the same holds for the partial cost functions \hat{h}_n , f_n and \hat{f}_n .

Remark that since h_n is continuous, the first assertion implies that for all $0 \leq i \leq w_n$ and all $0 \leq j < z_i^n$, we have that

$$(O_{i,j}^n, P_{i,j}^n) \neq (O_{i,j+1}^n, P_{i,j+1}^n);$$

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

Proof. We shall follow the same three steps as for the proof of Theorem 2 and the same ideas as for that of Lemma 1 from Didier (2017). Let us first recall that Lemma 1 from Didier (2017) ensures that the x -coefficient of \hat{h}_k decreases with α for all children k of the root different from n . Actually, a slightly different parametrization is used in Didier (2017) but the same change of variable as for Theorem 1 shows that this lemma remains true with the parametrization used here.

First, for all children n of the root of \mathcal{T} , Equation 4 and Lemma 1 from Didier (2017) ensure that the lemma holds for h_n .

Second, let us prove that if the lemma is true for h_n then it is true for \hat{h}_n . Let $\alpha \in [0, 1]$ be an asymmetry parameter and i be the index such that $\Upsilon_{i-1}^n < \alpha \leq \Upsilon_i^n$. From the proof of Theorem 2, there exists an index $0 \leq j \leq z_i^n$ such that, by setting $\pi_\alpha^+ = \rho_{i,j}^n$, the x -coefficient of $\hat{h}_n(\alpha, x)$ is equal to $-(1-\alpha)\phi(\tau_n)$ for $x < \pi_\alpha^+$ and greater than $-(1-\alpha)\phi(\tau_n)$ otherwise (π_α^+ is possibly equal to $-\infty$).

Symmetrically, there exists an index $0 \leq j \leq z_i^n$ such that, by setting $\pi_\alpha^- = \rho_{i,j}^n$, the x -coefficient of $\hat{h}_n(\alpha, x)$ is equal to $\alpha\phi(\tau_n)$ for $x > \pi_\alpha^-$ and smaller than $\alpha\phi(\tau_n)$ otherwise.

By construction and from the induction hypothesis, we get that both π_α^+ and π_α^- increase with α .

Let us now consider two asymmetry parameters $0 \leq \alpha' \leq \alpha'' \leq 1$. We then have $\pi_{\alpha'}^+ \leq \pi_{\alpha''}^+$, $\pi_{\alpha'}^- \leq \pi_{\alpha''}^-$, $\pi_{\alpha'}^+ \leq \pi_{\alpha''}^+$ and $\pi_{\alpha'}^- \leq \pi_{\alpha''}^-$. The first two inequalities come from construction (see the proof of Theorem 2) and the two last ones from the remark just above. This leaves only two cases to investigate:

1. $\pi_{\alpha'}^+ \leq \pi_{\alpha'}^- \leq \pi_{\alpha''}^+ \leq \pi_{\alpha''}^-$ and
2. $\pi_{\alpha'}^+ \leq \pi_{\alpha''}^+ \leq \pi_{\alpha'}^- \leq \pi_{\alpha''}^-$.

By checking all the possible positions of x with regard to the bounds $\pi_{\alpha'}^+$, $\pi_{\alpha'}^-$, $\pi_{\alpha''}^+$ and $\pi_{\alpha''}^-$, as in the proof of Lemma 1 of Didier (2017), we verify that the x -coefficient of $\widehat{h}_n(\alpha', x)$ is always greater than that of $\widehat{h}_n(\alpha'', x)$, either directly or from the induction hypothesis both in Cases 1 and 2. The x -coefficient of $\widehat{h}_n(\alpha, x)$ does decrease with α .

Third, we remark that if the x -coefficient of $\widehat{h}_n(\alpha, x)$ decreases with α then, since the same holds for all children k of n from Lemma 1 of Didier (2017), Equation 10 implies that the x -coefficient of $\widehat{h}_c(\alpha, x)$ decreases with α for all children c of n . \square

Proposition 1. *Under the notations of Theorem 2 and for all nodes n of \mathcal{T} , we have that*

$$\widehat{w}_n - w_n \leq 2|\mathcal{K} \setminus \mathcal{K}_n|.$$

Proof. Let n be a node of \mathcal{T} with $\widehat{w}_n > w_n$. There exists an index ℓ in $\{1, \dots, \widehat{w}_n\}$ such that $\widehat{\Upsilon}_\ell^n \notin \{\Upsilon_k^n \mid 1 \leq k \leq w_n\}$. Let i be the index such that $\Upsilon_i^n < \widehat{\Upsilon}_\ell^n < \Upsilon_{i+1}^n$. From the proof of Theorem 2, there exists an index $0 \leq j \leq z_i^n$ such that at least one of the following assertions holds:

1. $\widehat{\Upsilon}_\ell^n = \frac{P_{i,j}^n + \phi(\tau_n)}{O_{i,j}^n + \phi(\tau_n)}$,
2. $\widehat{\Upsilon}_\ell^n = \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)}$.

In other words, there are two types of α -bounds of \widehat{h}_n which are not α -bounds of h_n .

If $\widehat{\Upsilon}_\ell^n = \frac{P_{i,j}^n + \phi(\tau_n)}{O_{i,j}^n + \phi(\tau_n)}$ (Type 1) and still from the proof of Theorem 2, we have that $\rho_{\ell-1,1}^n = \rho_{i,j}^n < \rho_{i,j+1}^n = \rho_{\ell,1}^n$. Lemma 1 then ensures that for all $i' \geq i$ and all j' such that $(\rho_{i',j'}^n, \rho_{i',j'+1}^n] \cap (\rho_{i,j}^n, \rho_{i,j+1}^n] \neq \emptyset$, we have that $-O_{i',j'}^n \alpha' + P_{i',j'}^n \leq -O_{i,j}^n \alpha + P_{i,j}^n$ for all $\alpha' \in (\Upsilon_{i'}^n, \Upsilon_{i'+1}^n]$ and all $\alpha \in (\Upsilon_i^n, \Upsilon_{i+1}^n]$ with $\alpha' \geq \alpha$. It follows that, if $\rho_{i',j'}^n \leq \rho_{i,j}^n$, then $-(O_{i',j'}^n \alpha' - \phi(\tau_n)) + P_{i',j'}^n \leq 0$ for all $\alpha' \geq \widehat{\Upsilon}_\ell^n$. This implies that $\rho_{\ell,1}^n \leq \rho_{\ell',1}^n$ for all $\ell' \geq \ell$. In plain English, each time that an α -bound $\widehat{\Upsilon}_\ell^n$ of Type 1 appears, there is a x -bound which is no longer required for defining $\widehat{\Upsilon}_\ell^n(\alpha, x)$ for all $\alpha > \widehat{\Upsilon}_\ell^n$.

The situation is symmetrical if $\widehat{\Upsilon}_\ell^n = \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)}$. Each time that an α -bound $\widehat{\Upsilon}_\ell^n$ of Type 2 appears, there is a x -bound of h_n which is no longer required for defining $\widehat{\Upsilon}_\ell^n(\alpha, x)$ for all $\alpha < \widehat{\Upsilon}_\ell^n$.

Since all x -bounds used in the definition of h_n belong to $\vartheta(\mathcal{K} \setminus \mathcal{K}_n)$ and can be involved at most once in the appearance of α -bounds of Type 1 and at most once in the appearance of α -bounds of Type 2, we get that the number of “new” α -bounds of \widehat{h}_n is smaller than $2|\mathcal{K} \setminus \mathcal{K}_n|$. \square

Corollary 1. *Under the notations of Theorem 2 and for all nodes n of \mathcal{T} , both w_n and \widehat{w}_n are smaller than $4|\mathcal{T}| \cdot |\mathcal{K}|$*

Proof. From Equation 4 and Theorem 3 from Didier (2017), if n is a child of the root then $w_n \leq 2|\mathcal{T}_n| \cdot |\mathcal{K} \setminus \mathcal{K}_n| \leq 2|\mathcal{T}| \cdot |\mathcal{K}|$. Let m is a non-root node of \mathcal{T} , there exists a path k_1, k_2, \dots, k_ℓ of \mathcal{T} with $\ell < |\mathcal{T}|$ going from a child of the root k_0 to $k_\ell = m$. The number w_m of α -bounds of h_m is obtained by adding all the increases along the path k_1, k_2, \dots, k_ℓ to w_{k_1} :

$$w_m = w_{k_1} + \sum_{i=1}^{\ell-1} (\widehat{w}_{k_i} - w_{k_i}) + \sum_{i=1}^{\ell-1} (w_{k_{i+1}} - \widehat{w}_{k_i}).$$

From Equation 4 and Theorem 3 from Didier (2017) and since k_1 is a child of the root, we have

$$w_{k_1} \leq 2|\mathcal{T}_{k_1}| \cdot |\mathcal{K}_{\setminus \mathcal{K}_{k_1}}|.$$

Proposition 1 ensures that

$$\widehat{w}_{k_i} - w_{k_i} \leq 2|\mathcal{K}_{\setminus \mathcal{K}_{k_i}}|, \quad \text{for all } 1 \leq i \leq \ell.$$

Equation 10 and Theorem 3 of Didier (2017) imply that

$$w_{k_{i+1}} - \widehat{w}_{k_i} \leq \sum_{c \in (\mathcal{C}_{k_i}) \setminus \{k_{i+1}\}} 2|\mathcal{T}_c| \cdot |\mathcal{K}_c|, \quad \text{for all } 1 \leq i < \ell.$$

We get that

$$\begin{aligned} w_m &\leq 2|\mathcal{T}_{k_1}| \cdot |\mathcal{K}_{\setminus \mathcal{K}_{k_1}}| + \sum_{i=1}^{\ell-1} 2|\mathcal{K}_{\setminus \mathcal{K}_{k_i}}| + \sum_{i=1}^{\ell-1} \left(\sum_{c \in \mathcal{C}_{k_i} \setminus \{k_{i+1}\}} 2|\mathcal{T}_c| \cdot |\mathcal{K}_c| \right) \\ &\leq 2|\mathcal{T}| \left(|\mathcal{K}_{\setminus \mathcal{K}_{k_1}}| + \sum_{i=1}^{\ell-1} \left(\sum_{c \in (\mathcal{C}_{k_i}) \setminus \{k_{i+1}\}} 2|\mathcal{T}_c| \cdot |\mathcal{K}_c| \right) \right) + \sum_{i=1}^{\ell-1} 2|\mathcal{K}_{\setminus \mathcal{K}_{k_i}}| \\ &\leq 2|\mathcal{T}| \cdot |\mathcal{K}_{\setminus \mathcal{K}_m}| + 2(\ell-1)|\mathcal{K}| \\ &\leq 4|\mathcal{T}| \cdot |\mathcal{K}|. \end{aligned}$$

Since $\widehat{w}_m - w_m \leq 2|\mathcal{K}_{\setminus \mathcal{K}_m}|$ (Proposition 1), we have that

$$\begin{aligned} \widehat{w}_m &\leq 2|\mathcal{T}| \cdot |\mathcal{K}_{\setminus \mathcal{K}_m}| + 2\ell|\mathcal{K}| \\ &\leq 4|\mathcal{T}| \cdot |\mathcal{K}|. \end{aligned}$$

□

B.1 Proof of Theorem 3

Proof. From Theorem 1, less than $4|\mathcal{T}| \cdot |\mathcal{K}|$ α -bounds are required for defining \widehat{h}_n . Over intervals between two successive α -bounds, \widehat{h}_n requires at most $|\mathcal{K}_n|$ x -bounds. It follows that the total number of bounds required for defining \widehat{h}_n is $O(|\mathcal{T}| \cdot |\mathcal{K}|^2)$. Since the part of Algorithm 1 computing \widehat{h}_n from h_n is linear with the total number of bounds required to define \widehat{h}_n , its complexity is $O(|\mathcal{T}| \cdot |\mathcal{K}|^2)$ both in time and memory space.

The cost function h_n is obtained by summing the partial cost functions \widehat{h}_m of its direct ancestor m and the partial cost functions \widehat{f}_c of its siblings c , which is performed by using a procedure similar to that of 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 the sum of all the bounds of these cost functions cost functions which is $O(|\mathcal{T}| \cdot |\mathcal{K}|^2)$.

In conclusion, the computation of all the partial cost functions of \mathcal{T} is $O(|\mathcal{T}|^2 \cdot |\mathcal{K}|^2)$ both in time and memory space. □

C Integrating parsimonious costs over two asymmetry parameters

We shall detail only Case A, Case B being very similar.

From Theorems 1 and 2, the map $\Omega_n^A : (\alpha, \alpha', x) \rightarrow \hat{h}_n(\alpha, x) + f_n(\alpha', x)$ is piecewise linear and continuous. Namely, under the notations of these theorems, for all $0 \leq i \leq \hat{w}_n$, all $0 \leq j \leq u_n$ and by putting $(\pi_{i,j,k}^n)_{0 \leq k \leq c_{i,j}^n}$ for the elements of $\{\hat{\rho}_{i,k}^n \mid 0 \leq k \leq \hat{z}_i^n\} \cup \{\kappa_{i,k}^n \mid 0 \leq k \leq v_j^n\}$ sorted in increasing order, we have that, for all $0 \leq i \leq \hat{w}_n$, all $\alpha \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n]$, all $0 \leq j \leq u_n$, $\alpha' \in (\Gamma_j^n, \Gamma_{j+1}^n]$, all $0 \leq k \leq c_{i,j}^n$ and all $x \in (\pi_{i,j,k}^n, \pi_{i,j,k+1}^n]$,

$$\begin{aligned} \Omega_n^A(\alpha, \alpha', x) &= (-\hat{O}_{i,k_{\hat{h}}}^n \alpha + \hat{P}_{i,k_{\hat{h}}}^n)x + \hat{Q}_{i,k_{\hat{h}}}^n \alpha + \hat{R}_{i,k_{\hat{h}}}^n + (-A_{i,k_f}^n \alpha + B_{i,k_f}^n)x + C_{i,k_f}^n \alpha + D_{i,k_f}^n \\ &= (-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n)x + W_{i,j,k}^n \alpha + X_{i,j,k}^n \alpha' + Y_{i,j,k}^n, \end{aligned}$$

where $k_{\hat{h}}$ is the index such that $\pi_{i,j,k}^n = \hat{\rho}_{i,k_{\hat{h}}}^n$, k_f is the index such that $\pi_{i,j,k}^n = \kappa_{i,k_f}^n$, $T_{i,j,k}^n = \hat{O}_{i,k_{\hat{h}}}^n$, $U_{i,j,k}^n = A_{i,k_f}^n$, $V_{i,j,k}^n = \hat{P}_{i,k_{\hat{h}}}^n + B_{i,k_f}^n$, $W_{i,j,k}^n = \hat{Q}_{i,k_{\hat{h}}}^n$, $X_{i,j,k}^n = C_{i,k_f}^n$, $Y_{i,j,k}^n = \hat{R}_{i,k_{\hat{h}}}^n + D_{i,k_f}^n$.

Let us remark that, from Theorems 1 and 2, for all asymmetry parameters α and α' , the map $x \rightarrow \Omega_n^A(\alpha, \alpha', x)$ is convex.

The smallest cost obtained with asymmetry parameter α' on \mathcal{T}_n and with asymmetry parameter α elsewhere is

$$\check{\Omega}_n^A(\alpha, \alpha') = \min_x \Omega_n^A(\alpha, \alpha', x).$$

Since Ω_n^A is convex with respect to x , for all $0 \leq i \leq \hat{w}_n$, all $\alpha \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n]$, all $0 \leq j \leq u_n$ and all $\alpha' \in (\Gamma_j^n, \Gamma_{j+1}^n]$, we have that

$$\check{\Omega}_n^A(\alpha, \alpha') = (-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n) \pi_{i,j,\ell}^n + W_{i,j,\ell}^n \alpha + X_{i,j,\ell}^n \alpha' + Y_{i,j,\ell}^n,$$

where ℓ is the smallest index such that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$. Such an index ℓ always exists since Theorems 1 and 2 imply that $-T_{i,j,c_{i,j}^n}^n \alpha - U_{i,j,c_{i,j}^n}^n \alpha' + V_{i,j,c_{i,j}^n}^n \geq 0$.

For all indices $0 \leq i \leq \hat{w}_n$, $0 \leq j \leq u_n$, the lines of $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $0 \leq k \leq c_{i,j}^n$ partition $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into subsets on which the smallest index ℓ such that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ remains the same, i.e. in which $\check{\Omega}_n^A$ is linear with respect to α and α' (see Figure 7 without taking into account the dotted line).

Proposition 2. *For all indices $0 \leq i \leq \hat{w}_n$, $0 \leq j \leq u_n$, if there exists $0 \leq \ell \leq c_{i,j}^n$ such that $U_{i,j,\ell}^n \neq 0$ and such that the line of $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n$ intersects $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$, then the points (α, α') for which $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ are those which are below the line of $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n = 0$.*

Proof. Since from Theorem 1, $U_{i,j,k}^n$ is nonnegative for all $0 \leq k \leq c_{i,j}^n$, we have that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n \geq 0$ if and only if $\alpha' \leq \frac{-T_{i,j,k}^n \alpha + V_{i,j,k}^n}{U_{i,j,k}^n}$, i.e., the point (α, α') is below the line of $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n = 0$ (Fig. 7). \square

Proposition 3. For all indices $0 \leq i \leq \hat{w}_n$, $0 \leq j \leq u_n$ and $0 \leq k < k' \leq c_{i,j}^n$, the lines of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ and $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n = 0$ do not intersect each other at any point in $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n)$.

Proof. Since $x \rightarrow \Omega_n^\Delta(\alpha, \alpha', x)$ is convex, we have that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n \leq -T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n$. It follows that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n > 0$ implies $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n > 0$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n)$.

From the implicit assumption that all the bounds of partial functions are required, if $k \neq k'$ then $(T_{i,j,k}^n, U_{i,j,k}^n, V_{i,j,k}^n) \neq (T_{i,j,k'}^n, U_{i,j,k'}^n, V_{i,j,k'}^n)$, i.e., two lines intersect each other at a single point. If the lines of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ and $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n = 0$ intersect each other at a point of $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n)$, they split $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into four non-empty parts among which at least one contains points (α, α') such that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n > 0$ and $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n < 0$ which is in contradiction with the convexity of Ω_n^Δ with respect to x since we assume $k < k'$. \square

Proposition 4. Let i and j be two indices with $0 \leq i \leq \hat{w}_n$ and $0 \leq j \leq u_n$.

- a. If there exists $0 \leq \ell \leq c_{i,j}^n$ such that both $T_{i,j,\ell}^n = U_{i,j,\ell}^n = 0$ and $V_{i,j,\ell}^n \geq 0$, then there is no line of $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ intersecting $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- b. If there exists $0 \leq \ell \leq c_{i,j}^n$ such that $T_{i,j,\ell}^n \neq 0$ and $U_{i,j,\ell}^n = 0$, then
 - if $T_{i,j,\ell}^n > 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} > \hat{\Upsilon}_i^n$, then there is no line of $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ intersecting $(\hat{\Upsilon}_i^n, \min\{\hat{\Upsilon}_{i+1}^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}) \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
 - if $T_{i,j,\ell}^n < 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} < \hat{\Upsilon}_{i+1}^n$, then there is no line of $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ intersecting $(\max\{\hat{\Upsilon}_i^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- c. For all $0 \leq k < k' \leq c_{i,j}^n$, the part of the line corresponding to k' intersecting $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$ is above that of k on $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$.

Proof. These three properties are again consequences of the convexity of the map $x \rightarrow \Omega_n^\Delta(\alpha, \alpha', x)$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$, i.e., of the fact that the sequence $(-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n)_{0 \leq k \leq c_{i,j}^n}$ increases with k . In particular, this implies that if there exist a part $I \subseteq (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$ and an index ℓ such that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in I$, then there is no line of $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ splitting I into two non-empty parts, since this would imply the existence of points $(\alpha, \alpha') \in I$ with $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n < 0$.

The three assertions are proved as follows.

- a. If $T_{i,j,\ell}^n = U_{i,j,\ell}^n = 0$ and $V_{i,j,\ell}^n \geq 0$, then we have that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- b. Let us assume that $T_{i,j,\ell}^n \neq 0$ and $U_{i,j,\ell}^n = 0$.

- If $T_{i,j,\ell}^n > 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} > \widehat{\Upsilon}_i^n$, then we have that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in (\widehat{\Upsilon}_i^n, \min\{\widehat{\Upsilon}_{i+1}^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- If $T_{i,j,\ell}^n < 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} < \widehat{\Upsilon}_{i+1}^n$, then we have that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in (\max\{\widehat{\Upsilon}_i^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}, \widehat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.

c. The property directly follows from Proposition 2 and from the opening remark of this proof. \square

The A -split cost of (\mathcal{T}, ϑ) at n is the integral of $\check{\Omega}_n^A$ over all pairs $(\alpha, \alpha') \in [0, 1]^2$:

$$\int_0^1 \int_0^1 \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha = \sum_{i=0}^{\widehat{w}_n} \sum_{j=0}^{u_n} \int_{\widehat{\Upsilon}_i^n}^{\widehat{\Upsilon}_{i+1}^n} \int_{\Gamma_j^n}^{\Gamma_{j+1}^n} \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha.$$

For all $0 \leq i \leq \widehat{w}_n$ and all $0 \leq j \leq u_n$, computing

$$\int_{\widehat{\Upsilon}_i^n}^{\widehat{\Upsilon}_{i+1}^n} \int_{\Gamma_j^n}^{\Gamma_{j+1}^n} \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha$$

can be performed by splitting $(\widehat{\Upsilon}_i^n, \widehat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into trapezoids of the form $\{(\alpha, \alpha') \mid \beta < \alpha \leq \omega \text{ and } a\alpha + b < \alpha' \leq c\alpha + d\}$ (i.e., with two vertical sides) in which the coefficients of $\check{\Omega}_n^A$ are constant and of the form:

$$\check{\Omega}_n^A(\alpha, \alpha') = (-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n) \kappa_{i,j,k}^n + W_{i,j,k}^n \alpha + X_{i,j,k}^n \alpha' + Y_{i,j,k}^n,$$

for a certain index k . The different lines of Figure 7, here including the dotted one, partition $(\widehat{\Upsilon}_i^n, \widehat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into such trapezoids (the triangle below the line of $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ is seen as a degenerated trapezoid with a left side of length 0).

The algorithm partitioning $(\widehat{\Upsilon}_i^n, \widehat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into trapezoids is not presented here since it is quite basic thanks to Propositions 2, 3 and 4, which limit the number of situations that we have to deal with.

The integral of $\check{\Omega}_n^A$ over a trapezoid $\{(\alpha, \alpha') \mid \beta < \alpha \leq \omega \text{ and } a\alpha + b < \alpha' \leq c\alpha + d\}$ can be explicitly computed:

$$\begin{aligned} \int_{\beta}^{\omega} \int_{a\alpha+b}^{c\alpha+d} \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha &= \int_{\beta}^{\omega} \int_{a\alpha+b}^{c\alpha+d} ((-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n) \kappa_{i,j,k}^n + W_{i,j,k}^n \alpha + X_{i,j,k}^n \alpha' \\ &\quad + Y_{i,j,k}^n) d\alpha' d\alpha \\ &= \frac{1}{2} \left[\frac{\omega^3 - \beta^3}{3} (2(-T_{i,j,k}^n \kappa_{i,j,k}^n + W_{i,j,k}^n)(c-a) + (-U_{i,j,k}^n \kappa_{i,j,k}^n + X_{i,j,k}^n)(c^2 - a^2)) \right. \\ &\quad + (\omega^2 - \beta^2) ((-T_{i,j,k}^n \kappa_{i,j,k}^n + W_{i,j,k}^n)(d-b) + (V_{i,j,k}^n \kappa_{i,j,k}^n + Y_{i,j,k}^n)(c-a) \\ &\quad \quad \quad \left. + (-U_{i,j,k}^n \kappa_{i,j,k}^n + X_{i,j,k}^n)(cd - ab)) \right. \\ &\quad \left. + (\omega - \beta) (2(V_{i,j,k}^n \kappa_{i,j,k}^n + Y_{i,j,k}^n)(d-b) + (-U_{i,j,k}^n \kappa_{i,j,k}^n + X_{i,j,k}^n)(d^2 - b^2)) \right] \end{aligned}$$

Symmetrically, by setting

$$\check{\Omega}_n^B(\alpha, \alpha') = \min_x \Omega_n^B(\alpha, \alpha', x),$$

we define the *B-split cost* of (\mathcal{T}, ϑ) at n as

$$\int_0^1 \int_0^1 \check{\Omega}_n^{\text{B}}(\alpha, \alpha') d\alpha' d\alpha = \sum_{i=0}^{w_n} \sum_{j=0}^{u_n} \int_{\Upsilon_i^n}^{\Upsilon_{i+1}^n} \int_{\hat{\Gamma}_j^n}^{\hat{\Gamma}_{j+1}^n} \check{\Omega}_n^{\text{B}}(\alpha, \alpha') d\alpha' d\alpha.$$

All the remarks stated above about A-split costs still hold for B-split costs which are computed in the very same way.

List of Figures

1	Two different ways of splitting a phylogenetic tree at node n	36
2	A tree with two sets of character values (in brackets) for tips. a: A A-split at node ‘n’ improves the parsimonious cost of the whole tree. b: No split improves the parsimonious cost of the whole tree (see Table 1).	37
3	Parsimonious splits of the tree of cetaceans according to their body sizes with their ranks. All the parsimonious splits are A-splits. Dataset from Slater <i>et al.</i> (2010). Color shading reflects the strength of the signal for change in trends in body size, with increasing values from blue (weak) to red (strong).	38
4	Parsimonious splits of the tree of <i>Anolis</i> squamates according to their body sizes with their ranks. All the parsimonious splits are A-splits. Dataset from Thomas and Freckleton (2012). Color shading reflects the strength of the signal for change in trends in body size, with increasing values from blue (weak) to red (strong).	39
5	Changes in skull length (ln of skull length, measured in mm, is shown here) evolutionary trends detected by our new method. There are both parsimonious A- and B-splits (A-splits are presented with ‘×’ at the corresponding nodes and B-splits with ‘+’ inside the corresponding branches and their ranks in italic). The test is illustrated here using the first (main) tree, for cranial length. Note that none of the parsimonious changes are located on the amniote stem, which extends between the nodes Tetrapoda and Amniota, under our preferred hypothesis. However, even placing the origin of that stem basal to Temnospondyly does not alter our conclusions: no parsimonious shifts are detected on the amniote stem. Color shading reflects the strength of the signal for change in trends in the natural logarithm of skull length, in mm, with increasing values from blue (weak) to red (strong).	40
6	Left: Schematic of the representation of all the possible parsimonious reconstructions for a given node with asymmetry parameter α going from 0 to 1 (Didier 2017). Each part of the quarter pie is colored according to the most parsimonious reconstruction (also displayed in front of it) obtained from the asymmetry parameters corresponding to the slopes delineating it. Right: Nodal estimates of the natural logarithm of cranial length (in mm) for the node Cotylosauria under tree 1 (see Figure 5) for all asymmetry parameter $0 \leq \alpha \leq 1$. The curve with a slope 1 (green line) represents the assumption of no trend in the evolution of this character. Values below that axis assume a positive trend and are indicated by a “+” and an arc with arrows at both ends and are shaded yellow to orange; the similar arc above the green axis, associated with a “-” delimits the values coherent with a trend towards size decrease and these are colored orange to red. Values below that curve are consistent with the assumption that there is a positive trend, whereas those above the curve are congruent with the assumption that the character value decreases through time. Thus, under Carroll’s (1970, 1991) scenario, size should increase around that node, so the most plausible estimates range from 3.09 to 3.98 (values below the curve); transforming these back to lengths through exponential transformation yields values of 22.0 to 53.5 mm. Carroll’s (1970, 1991) scenario implies that at least some stem-amniotes measured less than 100 mm in snout-vent length, and this translates into a maximal ln-transformed value of 3.22, if we assume that the skull represented about a quarter of the snout-vent length. Hence, only the smallest of the inferred nodal values (3.09) is compatible with Carroll’s (1970, 1991) scenario.	41
7	Partition of $(\hat{Y}_i^n, \hat{Y}_{i+1}^n) \times (\Gamma_j^n, \Gamma_{j+1}^n)$ resulting from the lines of equation $-T_{i,j,k+1}^n \alpha - U_{i,j,k+1}^n \alpha' + V_{i,j,k+1}^n = 0$ and $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ (the top and the bottom solid lines respectively).	42

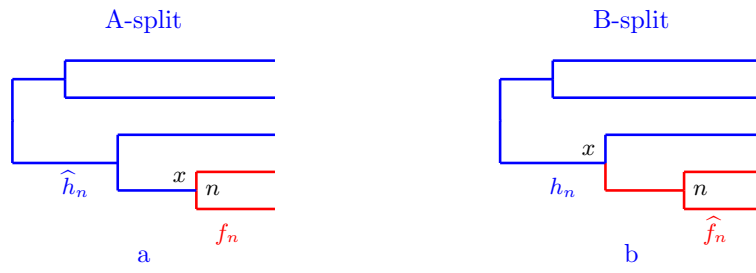


Figure 1: Two different ways of splitting a phylogenetic tree at node n .



Figure 2: A tree with two sets of character values (in brackets) for tips. a: A A-split at node 'n' improves the parsimonious cost of the whole tree. b: No split improves the parsimonious cost of the whole tree (see Table 1).

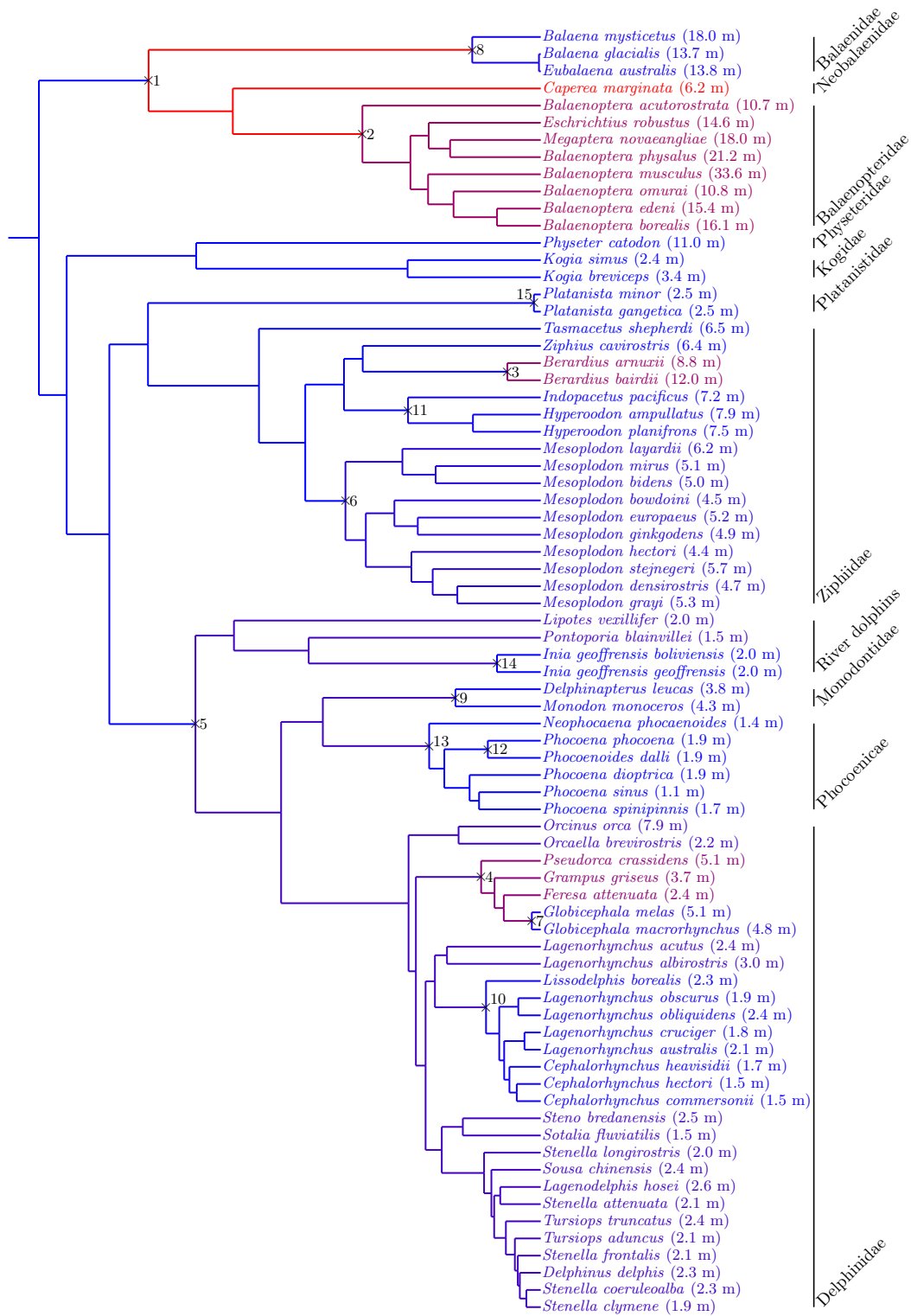


Figure 3: Parsimonious splits of the tree of cetaceans according to their body sizes with their ranks. All the parsimonious splits are A-splits. Dataset from Slater *et al.* (2010). Color shading reflects the strength of the signal for change in trends in body size, with increasing values from blue (weak) to red (strong).

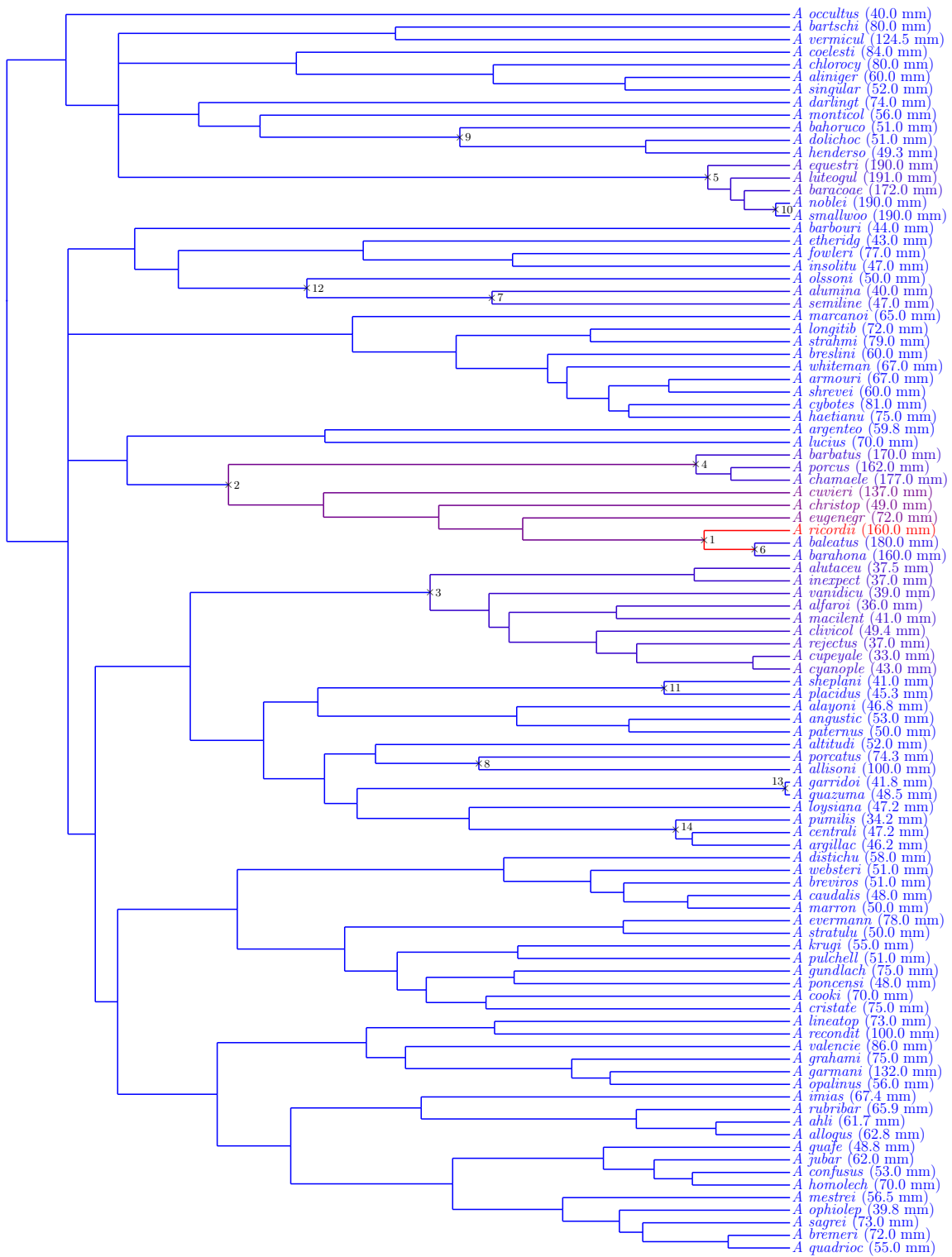


Figure 4: Parsimonious splits of the tree of *Anolis* squamates according to their body sizes with their ranks. All the parsimonious splits are A-splits. Dataset from Thomas and Freckleton (2012). Color shading reflects the strength of the signal for change in trends in body size, with increasing values from blue (weak) to red (strong).

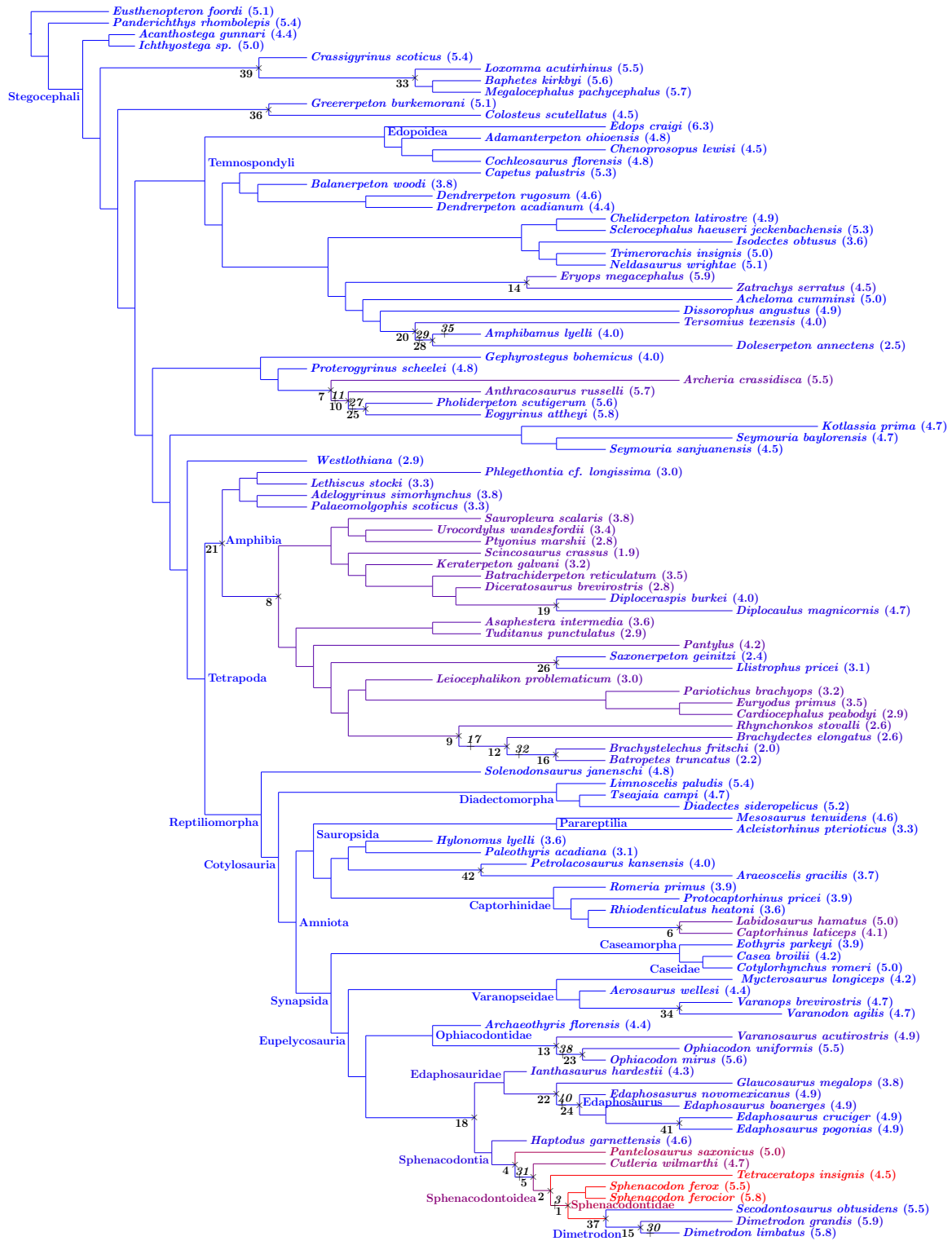


Figure 5: Changes in skull length (ln of skull length, measured in mm, is shown here) evolutionary trends detected by our new method. There are both parsimonious A- and B-splits (A-splits are presented with ‘x’ at the corresponding nodes and B-splits with ‘+’ inside the corresponding branches and their ranks in italic). The test is illustrated here using the first (main) tree, for cranial length. Note that none of the parsimonious changes are located on the amniote stem, which extends between the nodes Tetrapoda and Amniota, under our preferred hypothesis. However, even placing the origin of that stem basal to Temnospondyli does not alter our conclusions: no parsimonious shifts are detected on the amniote stem. Color shading reflects the strength of the signal for change in trends in the natural logarithm of skull length, in mm, with increasing values from blue (weak) to red (strong).

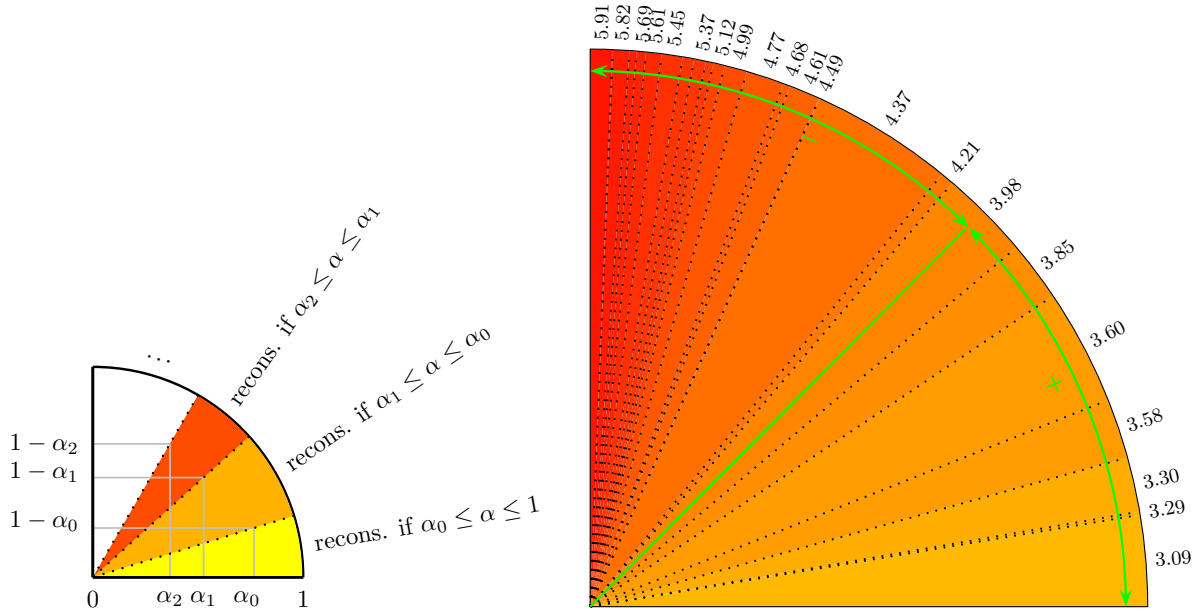


Figure 6: Left: Schematic of the representation of all the possible parsimonious reconstructions for a given node with asymmetry parameter α going from 0 to 1 (Didier 2017). Each part of the quarter pie is colored according to the most parsimonious reconstruction (also displayed in front of it) obtained from the asymmetry parameters corresponding to the slopes delineating it. Right: Nodal estimates of the natural logarithm of cranial length (in mm) for the node Cotylosauria under tree 1 (see Figure 5) for all asymmetry parameter $0 \leq \alpha \leq 1$. The curve with a slope 1 (green line) represents the assumption of no trend in the evolution of this character. Values below that axis assume a positive trend and are indicated by a “+” and an arc with arrows at both ends and are shaded yellow to orange; the similar arc above the green axis, associated with a “-” delimits the values coherent with a trend towards size decrease and these are colored orange to red. Values below that curve are consistent with the assumption that there is a positive trend, whereas those above the curve are congruent with the assumption that the character value decreases through time. Thus, under Carroll’s (1970, 1991) scenario, size should increase around that node, so the most plausible estimates range from 3.09 to 3.98 (values below the curve); transforming these back to lengths through exponential transformation yields values of 22.0 to 53.5 mm. Carroll’s (1970, 1991) scenario implies that at least some stem-amniotes measured less than 100 mm in snout-vent length, and this translates into a maximal \ln -transformed value of 3.22, if we assume that the skull represented about a quarter of the snout-vent length. Hence, only the smallest of the inferred nodal values (3.09) is compatible with Carroll’s (1970, 1991) scenario.

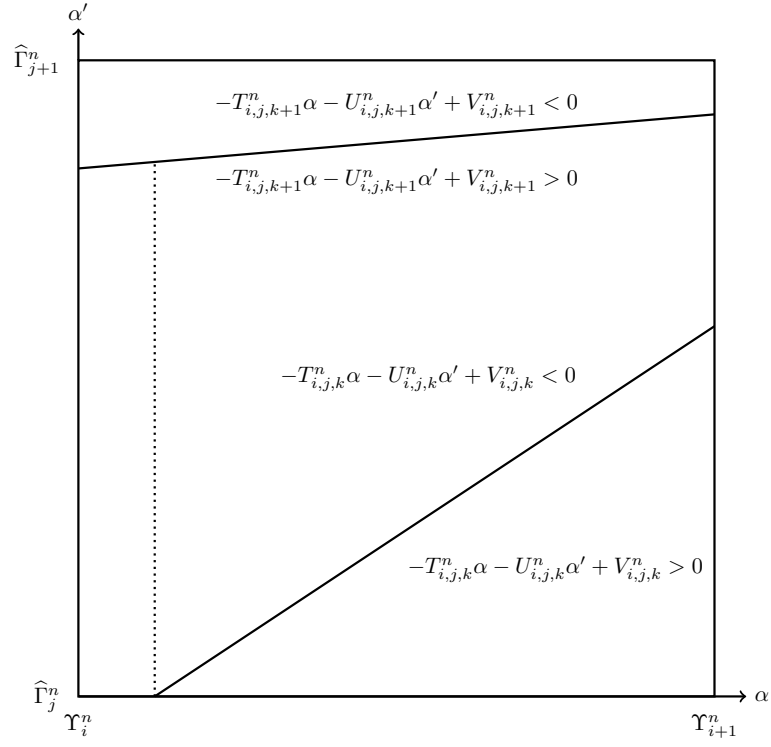


Figure 7: Partition of $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ resulting from the lines of equation $-T_{i,j,k+1}^n \alpha - U_{i,j,k+1}^n \alpha' + V_{i,j,k+1}^n = 0$ and $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ (the top and the bottom solid lines respectively).

List of Tables

1	Left: (resp. Right:) A- and B-split costs for all nodes of the tree with character values of Figure 2-a (resp. of Figure 2-b). The first line of the tables displays the A-split cost of the root 'r', i.e., the no-split cost computed as presented in Section 2.4.1.	44
2	Shifts in evolutionary trends detected by our new asymmetric parsimony-based method. The columns give, from left to right, the tree number and name (from the Mesquite Nexus file available as on-line supplement), whether or not one of the detected shifts supports Carroll's (1970, 1991) scenario, the number of changes found, and the location of the shift closest to where Carroll's (1970, 1991) scenario predicts changes should be detected. For the latter, the name of the corresponding taxon is given; when no such name is available, the names of two taxa representing both daughter-branches are given. Note that out of 517 changes, not a single one unambiguously supports Carroll's (1970, 1991) scenario, though one could possibly be interpreted as such. Legend of tree names: a, amphibians; p, presacral length; r, reptiliomorphs; s, skull; w, <i>Westlothiana</i> ; Tree 1 is for skull; Tree 2 is for presacral length; Tree 5 is for skull but has <i>Westlothiana</i> among reptiliomorphs. Some trees differ only by minimal internal branch length, a simple tree scaling method that was used in Laurin (2004) to account for the fact that many internal branches had a minimal length of 0 but were presumably not really of 0 length.	45
3	Estimates of cranial length (mm) of selected hypothetical ancestors according to the full possible range of asymmetry coefficients. For a legend of the meaning of tree names, see Table 2. Maximal plausible values reflect assumptions on trends. Thus, for Tetrapoda, given that Laurin found no evidence for trends, all values can be considered. For Amniota and nodes just below or above, given the detected increase in size, only the lower half of the values need to be considered.	46
4	Estimates of presacral length (mm) of selected hypothetical ancestors according to the full possible range of asymmetry coefficients. For a legend of the meaning of tree names, see Table 2. Given that the presacral length of <i>Solenodonsaurus</i> is unknown, Reptiliomorpha and Cotylosauria correspond to the same node for this analysis.	47
5	Estimates of snout-vent length (mm) of selected hypothetical ancestors according to the full possible range of asymmetry coefficients. This length is obtained by adding the length of the skull from Table 3 and the presacral length from Table 4; this has been done only for nodes for which both lengths could be inferred. Only values in bold type are congruent with Carroll's (1970, 1991) ideas. For a legend of the meaning of tree names, see Table 2.	48

node	A-split cost	B-split cost	node	A-split cost	B-split cost
r	0.32	-	r	0.50	-
m	0.32	0.40	m	0.52	0.59
n	0.31	0.35	n	0.54	0.59
a	-	0.34	a	-	0.62
b	-	0.34	b	-	0.62
c	-	0.35	c	-	0.57
p	0.34	0.40	p	0.54	0.59
d	-	0.35	d	-	0.53
e	-	0.35	e	-	0.53

Table 1: Left: (resp. Right:) A- and B-split costs for all nodes of the tree with character values of Figure 2-a (resp. of Figure 2-b). The first line of the tables displays the A-split cost of the root ‘r’, i.e., the no-split cost computed as presented in Section 2.4.1.

Index in Mesquite file Trees used for the analyses		Presence of significant shift on stem- amniote branches	Number of possible changes found	Rank and location of the most significant shift in a relevant part of the tree
1	T1s, 107 cranial internal 3 (main analysis)	No	42	21, Amphibia
2	T1p, Postcranial ln length 63 internal 3	No	17	1, Nectridea, microsaurus
3	T2s, 107 cranial internal 5 (variant 1, longer internal branches)	No	40	15, Synapsida
4	T2p, Postcranial ln internal 5 (variant 1, longer internal branches)	No	21	7, Synapsida; 4, Nectridea, microsaurus
5	T1swr, 107 cranial internal 3 Westlo reptiliomorph, all taxa	Possibly	43	7, Tetrapoda
6	T1pwr, Postcranial ln length 63 internal 3 Westlo reptiliomorph, all taxa	No	20	1, Nectridea, microsaurus
7	T2swr, 107 cranial internal 5 Westlo reptiliomorph, all taxa	No	40	26, Amphibia
8	T2pwr, Postcranial ln internal 5 Westlo reptiliomorph, all taxa	No	24	3, Synapsida
9	T1swa, 107 cranial internal 3 Westlo amphibian, all taxa	No	44	5, Amphibia
10	T1pwa, Postcranial ln length 63, internal 3 Westlo amphibian, all taxa	No	17	1, Nectridea, microsaurus
11	T2swa, 107 cranial internal 5 Westlo amphibian, all taxa	No	41	13, Amphibia; 30, Synapsida
12	T2pwa, Postcranial ln internal 5 Westlo amphibian, all taxa	No	22	15, Synapsida
13	T3s, 107 internal 5 (Smithson)	No	40	2, Archeria, Eogyrinus
14	T3p, Internal 5 (Smithson) presacral	No	22	8, Protorothyrididae
15	T1swa, <i>Solenodonsaurus</i> stem-tetrapod, 107 cranial internal 3	No	43	6, Amphibia
16	T1swa, <i>Solenodonsaurus</i> amphibian 107 cranial internal 3	No	41	12, Amphibia
Average number of shifts			32.3	
Average number of shifts for skull			41.5	
Average number of shifts for presacral length			20.4	
Total			517	

Table 2: Shifts in evolutionary trends detected by our new asymmetric parsimony-based method. The columns give, from left to right, the tree number and name (from the Mesquite Nexus file available as on-line supplement), whether or not one of the detected shifts supports Carroll's (1970, 1991) scenario, the number of changes found, and the location of the shift closest to where Carroll's (1970, 1991) scenario predicts changes should be detected. For the latter, the name of the corresponding taxon is given; when no such name is available, the names of two taxa representing both daughter-branches are given. Note that out of 517 changes, not a single one unambiguously supports Carroll's (1970, 1991) scenario, though one could possibly be interpreted as such. Legend of tree names: a, amphibians; p, presacral length; r, reptiliomorphs; s, skull; w, *Westlothiana*; Tree 1 is for skull; Tree 2 is for presacral length; Tree 5 is for skull but has *Westlothiana* among reptiliomorphs. Some trees differ only by minimal internal branch length, a simple tree scaling method that was used in Laurin (2004) to account for the fact that many internal branches had a minimal length of 0 but were presumably not really of 0 length.

		Tetrapoda	Reptiliomorpha	Cotylosauria	Amniota	Sauropsida	Synapsida
1, T1s	Minimum	6.7	22.0	22.0	22.0	22.0	44.7
	Maximum	368.7	368.7	368.7	368.7	151.4	368.7
	Maximal plausible	368.7	368.7	53.5	53.5	53.5	67.4
3, T2s	Minimum	6.7	22.0	22.0	22.0	83.1	223.6
	Maximum	368.7	368.7	368.7	368.7	290.0	1274.1
	Maximal plausible	368.7	368.7	53.5	53.5	223.6	223.6
5, T1swr	Minimum	6.7	18.9	22.0	22.0	22.0	44.7
	Maximum	368.7	368.7	368.7	368.7	162.4	368.7
	Maximal plausible	368.7	368.7	36.6	36.6	36.6	67.4
6, T2swr	Minimum	6.7	18.9	22.0	22.0	22.0	44.7
	Maximum	368.7	368.7	368.7	368.7	151.4	368.7
	Maximal plausible	368.7	368.7	53.5	53.5	53.5	79.0
9, T1swa	Minimum	6.7	22.0	22.0	22.0	22.0	44.7
	Maximum	368.7	368.7	368.7	368.7	151.4	368.7
	Maximal plausible	368.7	368.7	67.4	67.4	67.4	67.4
11, T2swa	Minimum	6.7	22.0	22.0	22.0	22.0	44.7
	Maximum	368.7	368.7	368.7	368.7	151.4	368.7
	Maximal plausible	368.7	368.7	53.5	53.5	53.5	79.0
13, T3s	Minimum	6.7	18.9	22.0	22.0	22.0	44.7
	Maximum	539.2	368.7	368.7	368.7	151.4	368.7
	Maximal plausible	539.2	368.7	83.1	83.1	53.5	83.1
T1swa <i>Solenodonsaurus</i> stem-tetrapod 107 cranial internal 3	Minimum	6.7	6.7	6.7	22.0	22.0	44.7
	Maximum	368.7	539.2	539.2	368.7	151.4	368.7
	Maximal plausible	368.7	83.1	83.1	53.5	53.5	67.4
T1swa <i>Solenodonsaurus</i> amphibian 107 cranial internal 3	Minimum	6.7	6.7	22.0	22.0	22.0	44.7
	Maximum	368.7	539.2	368.7	368.7	151.4	368.7
	Maximal plausible	368.7	539.2	53.5	53.5	53.5	67.4

Table 3: Estimates of cranial length (mm) of selected hypothetical ancestors according to the full possible range of asymmetry coefficients. For a legend of the meaning of tree names, see Table 2. Maximal plausible values reflect assumptions on trends. Thus, for Tetrapoda, given that Laurin found no evidence for trends, all values can be considered. For Amniota and nodes just below or above, given the detected increase in size, only the lower half of the values need to be considered.

		Tetrapoda	Reptiliomorpha Cotylosauria	Amniota	Sauropsida	Synapsida
2, T1p	Minimum	27.7	83.1	83.1	83.1	223.6
	Maximum	1274.1	1274.1	1274.1	244.7	1274.1
	Maximal plausible	1274.1	1274.1	223.6	223.6	223.6
4, T2p	Minimum	27.7	27.7	83.1	83.1	223.6
	Maximum	1274.1	1571.8	1274.1	290.0	1274.1
	Maximal plausible	1274.1	1571.8	219.2	244.7	314.2
7, T1pwr	Minimum	27.7	27.7	83.1	83.1	223.6
	Maximum	1274.1	1571.8	1274.1	290.0	1274.1
	Maximal plausible	1274.1	1571.8	223.6	223.6	223.6
8, T2pwr	Minimum	27.7	27.7	83.1	83.1	223.6
	Maximum	1274.1	1571.8	1274.1	290.0	1274.1
	Maximal plausible	1274.1	1571.8	184.9	184.9	223.6
10, T1pwa	Minimum	27.7	27.7	83.1	83.1	223.6
	Maximum	1274.1	1571.8	1274.1	290.0	1274.1
	Maximal plausible	1274.1	1571.8	223.6	223.6	223.6
12, T2pwa	Minimum	27.7	27.7	83.1	83.1	223.6
	Maximum	1274.1	1571.8	1274.1	290.0	1274.1
	Maximal plausible	1274.1	1571.8	219.2	219.2	223.6
14, T3p	Minimum	27.7	27.7	83.1	83.1	223.6
	Maximum	1571.8	1571.8	1274.1	290.0	1274.1
	Maximal plausible	1571.8	1571.8	254.7	244.7	314.2

Table 4: Estimates of presacral length (mm) of selected hypothetical ancestors according to the full possible range of asymmetry coefficients. For a legend of the meaning of tree names, see Table 2. Given that the presacral length of *Solenodonsaurus* is unknown, Reptiliomorpha and Cotylosauria correspond to the same node for this analysis.

		Tetrapoda	Reptiliomorpha Cotylosauria	Amniota	Sauropsida	Synapsida
1, T1s	Minimum	34.3	105.1	105.1	105.1	268.3
	Maximum	1642.8	1642.8	1642.8	396.1	1642.8
	Maximal plausible	1642.8	1327.6	277.1	277.1	291.0
3, T2s	Minimum	34.3	49.6	105.1	166.2	447.3
	Maximum	1642.8	1940.5	1642.8	580.1	2548.2
	Maximal plausible	1642.8	1625.4	272.7	468.3	537.8
5, T1swr	Minimum	34.3	49.6	105.1	105.1	268.3
	Maximum	1642.8	1940.5	1642.8	452.4	1642.8
	Maximal plausible	1642.8	1608.4	260.2	260.2	291
6, T2swr	Minimum	34.3	49.6	105.1	105.1	268.3
	Maximum	1642.8	1940.5	1642.8	441.4	1642.8
	Maximal plausible	1642.8	1625.4	238.5	238.5	302.7
9, T1swa	Minimum	34.3	49.6	105.1	105.1	268.3
	Maximum	1642.8	1940.5	1642.8	441.4	1642.8
	Maximal plausible	1642.8	1639.2	291.0	291.0	291.0
11, T2swa	Minimum	34.3	49.6	105.1	105.1	268.3
	Maximum	1642.8	1940.5	1642.8	441.4	1642.8
	Maximal plausible	1642.8	1625.4	272.7	272.7	302.7
13, T3s	Minimum	34.3	49.6	105.1	105.1	268.3
	Maximum	2111.0	1940.5	1642.8	441.4	1642.8
	Maximal plausible	2111.0	1654.9	337.8	298.2	397.3

Table 5: Estimates of snout-vent length (mm) of selected hypothetical ancestors according to the full possible range of asymmetry coefficients. This length is obtained by adding the length of the skull from Table 3 and the presacral length from Table 4; this has been done only for nodes for which both lengths could be inferred. Only values in bold type are congruent with Carroll's (1970, 1991) ideas. For a legend of the meaning of tree names, see Table 2.