Syst. Biol. 74(4):672–684, 2025

https://doi.org/10.1093/sysbio/syaf025
Advance Access Publication April 17, 2025

PlaceMyFossils: An Integrative Approach to Analyze and Visualize the Phylogenetic Placement of Fossils Using Backbone Trees

SANTIAGO A. CATALANO[1,2,3,*], IGNACIO ESCAPA[4], KELSEY D. PUGH[2,5,6], ASHLEY S. HAMMOND[2,5], PABLO GOLOBOFF[1], AND SERGIO ALMÉCIJA[2,5,7]

[1]*Unidad Ejecutora Lillo, Consejo Nacional de Investigaciones Científicas y Técnicas – Fundación Miguel Lillo, Miguel Lillo 251, 4000 San Miguel de Tucumán, Argentina*
[2]*Division of Anthropology, American Museum of Natural History (AMNH), 200 Central Park West, New York, NY 10024, USA*
[3]*Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán, Miguel Lillo 205, 4000 San Miguel de Tucumán, Argentina*
[4]*CONICET - Museo Paleontológico Egidio Feruglio, Av. Fontana 140, 9100, Trelew, Argentina*
[5]*New York Consortium in Evolutionary Primatology, New York, NY 10024, USA*
[6]*Department of Anthropology, Brooklyn College, City University of New York, 2900 Bedford Avenue, Brooklyn, NY 11210, USA*
[7]*Institut Català de Paleontologia Miquel Crusafont (ICP-CERCA), Z Building c/ de les Columnes s/n, Campus de la Universitat Autònoma de Barcelona (UAB), 08193 Barcelona, Spain*
[*]*Correspondence to be sent to: Unidad Ejecutora Lillo, Consejo Nacional deInvestigaciones Científicas y Técnicas – Fundación Miguel Lillo, Miguel Lillo 251, 4000 San Miguel de Tucumán, Argentina; E-mail: sacatalano@gmail.com.*

Received 13 May 2024; reviews returned 27 February 2025; accepted 9 April 2025
Associate Editor: Seraina Klopfstein

Abstract.—In recent years, there has been a growing interest in using morphology to establish the placement of species on phylogenetic trees derived from molecular data. This is relevant in the case of recently extinct or fossil species, which are usually represented only by fragmentary morphology. In the latter case, constrained analyses using backbone trees have also proven helpful in evaluating the placement of fragmentary specimens on phylogenetic trees derived from morphological data. Consequently, several available phylogenetic programs now include functions to run constrained searches. However, a comprehensive evaluation of the obtained results is not readily available within existing analytical tools. Here, we present an integrated approach—*PlaceMyFossils*—specifically designed to (i) thoroughly evaluate the phylogenetic placement of given query species (especially fossils) on a reference tree, (ii) determine which characters and character partitions are most relevant in defining the phylogenetic placement, (iii) assess the confidence of the results, and (iv) define the optimal analytical conditions to place the query species. *PlaceMyFossils* combines several analyses implemented as an interactive script for *TNT* (Tree Analysis Using New Technologies software), a popular—and free—phylogenetic software that is widely used in paleontological studies. Finally, we demonstrate the utility and investigate the performance of *PlaceMyFossils* compared to other available tools using two disparate empirical datasets drawn from conifers and dinosaurs. While primarily designed for working with fossils, this tool also holds great potential for advancing morphological and molecular systematics. It offers a powerful resource for empirical systematists aiming to integrate molecular and morphological data. This is particularly relevant given the growing interest in morphological evolution as a complementary perspective on evolutionary processes and the drivers of diversification. [Keywords: backbone topology; error in placement; fossils; maximum likelihood; maximum parsimony; molecular scaffolds.]

Phylogenetic analyses using backbone trees involve placing one or more query species/specimen(s) on a given phylogenetic tree. Backbone-based analyses are commonly used in different scientific disciplines like virology (e.g., see Oude Munnink et al. 2021), metagenomics (e.g., Czech et al. 2022), and phylogenomics (e.g., Zuntini et al. 2024). Given the necessity to investigate the evolutionary role of species for which only morphology might be available (e.g., recently extinct or fossil species), a growing interest has recently emerged in using this approach to place query species according to morphological data in phylogenetic trees generated by molecular data (e.g., Pugh 2022; López-Martínez et al. 2023). This interest has been preceded by research in certain fields, such as paleobotany, where these types of analyses have been conducted for many years (e.g., Manos et al. 2007; Doyle and Endress 2010). Backbone-based analyses emerge as an important tool to complement unconstrained total evidence analyses by permitting several a posteriori evaluations that are hard or impossible to run using the total evidence approach and, simultaneously, with much lower computational requirements. In addition, phylogenetic analyses using backbone trees are also useful in studies exclusively based on morphological data, allowing quick analyses when scoring a new taxon and evaluating the differences in the phylogenetic signal among different character partitions.

There are several software tools available for conducting searches using a backbone tree, each operating within a different analytical framework: *TNT* (Tree Analysis Using New Technologies, Goloboff and Morales 2023) for Maximum Parsimony (*MP*), *RAxML-NG* (Kozlov et al. 2019) and *IQ-TREE 2* (Minh et al. 2020) for Maximum

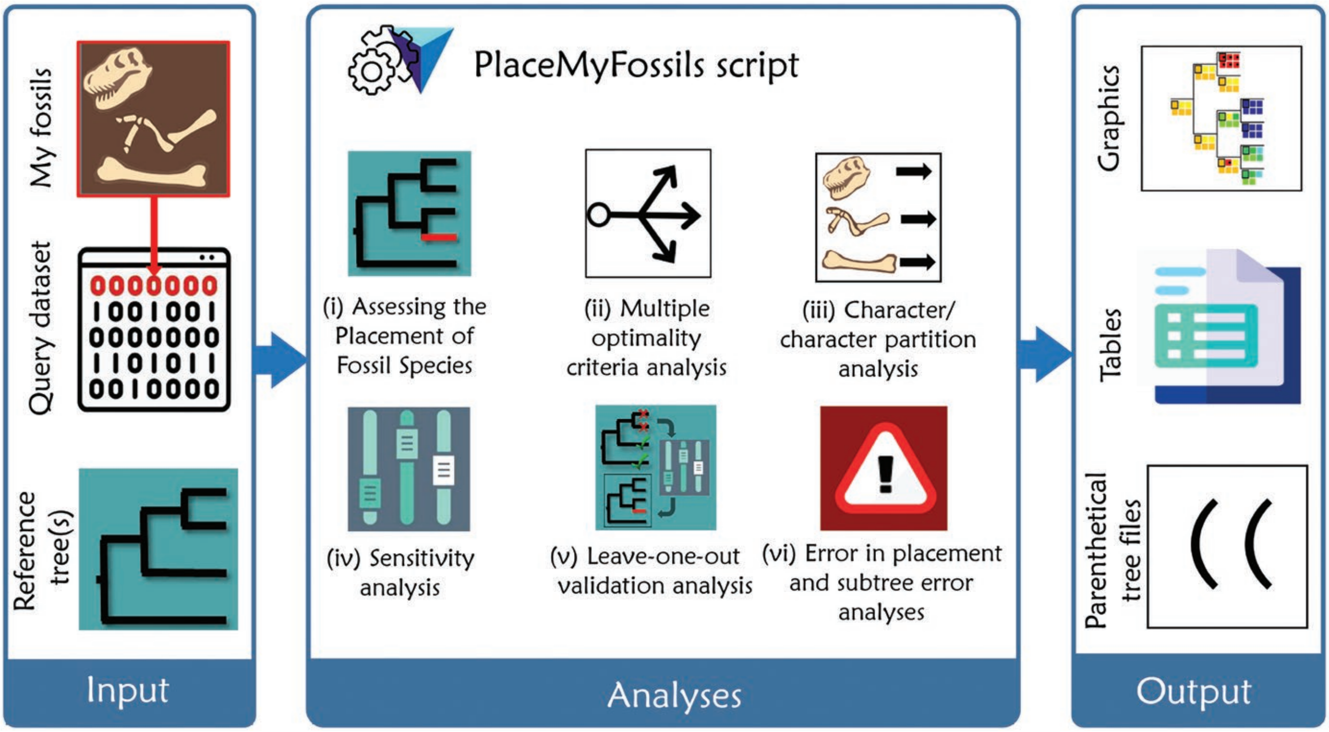

FIGURE 1. Workflow of *PlaceMyFossils TNT* script. The input data includes the query dataset (i.e., the dataset that will be used to place the query species), the reference tree, and the query species (included as part of the query dataset). Once the files are read in *TNT*, the script allows to (i) place query species (one or more) on a backbone tree and visualize the suboptimality values of every placement either for the complete dataset or (iii) individual characters/partitions; (ii) visual comparison of the results obtained under different optimality criteria; (iv) evaluation of the effect of different analytical parameters on the resulting query species placement; (v) determine the optimal analytical settings to place the query species; and (vi) calculate the error associated with the placement of the query species either for the complete dataset, and/or the different subtrees. The results are presented graphically in *svg* files, in tabulated *txt* files, and also in parenthetical tree files (*TNT* and Jplace format).

Likelihood (*ML*), *MrBayes* (Ronquist 2012), *Beast2* (Bouckaert et al. 2014) for Bayesian approaches, among others. Additionally, specific tools have been developed to place species/specimens on a reference tree. Among these, the Evolutionary Placement Algorithm (*EPA*; Berger et al. 2011; Barbera et al. 2019), an *ML* approach, is one of the most commonly used tools in paleontology (e.g., Bomfleur et al. 2015; Heckeberg 2020; Badano et al. 2021; Černý & Natale 2022). The *phytools* R package (Revell 2012) also implements functions to place fossil species on a reference tree under an *ML* (*locate.fossil*, Revell et al. 2015), and a similar approach has been recently developed based on a Bayesian criterion (*Cophymaru*, Parins-Fukuchi 2018). While not specifically designed to handle backbone trees, *RoguePlots*, a related tool developed by Klopfstein and Spasojevic (2019), provides the posterior probabilities (PP) of different placements of a fossil species on a majority consensus of trees generated on a Bayesian Analysis.

While all these tools are valuable for placing species on a reference tree, each has limitations for incorporating different types of morphological data and requires running multiple independent analyses or combining different analytical approaches to obtain comprehensive results. For example, *locate.fossil* and *Cophymaru* can only handle continuous characters; and the *EPA* approach does not generate straightforward visualizations of different optimality scores across the tree, nor does it handle continuous characters. In addition, none of the available approaches include functions to evaluate how different characters and character partitions support the optimal placement of the query species. To fill this analytical need, we present *PlaceMyFossils*, an integrative approach implemented in a new script for *TNT* (Goloboff and Morales 2023). This script combines various functions and analyses, enabling an in-depth evaluation of the phylogenetic placement of query species under different optimality criteria, assessing the evidence that supports the placements, and facilitating the visualization of the results. We illustrate the value of this approach with two empirical examples and compare the results with those obtained with other analytical tools.

PLACEMYFOSSILS DESCRIPTION

PlaceMyFossils is an interactive script for *TNT* that integrates several tools and functions to simplify and improve backbone-based analyses (Fig. 1). The script can be run in the GUI (Graphic Unit Interface) versions of *TNT* for Windows, Linux, and Mac operating systems.

TNT can be freely downloaded from https://www.lillo.org.ar/phylogeny/tnt. The script and a detailed user guide can be freely downloaded from https://github.com/sacatalano/PlaceMyFossils.

Although *PlaceMyFossils* is currently implemented as a *TNT* script, the analyses described in this work can be adapted to other analytical frameworks. For example, functions from R packages such as *ape* (Paradis and Schliep 2019) and *phytools* (Revell 2012) can be used to implement many of the analyses described in the manuscript.

Assessing and Visualizing the Placement of Fossil Species

Query species can be placed under three different optimality criteria: traditional *MP* (Farris 1970), Implied Weighting (*IW*; Goloboff 1993), and *ML* (Lewis 2001). It is possible to analyze datasets that include discrete, continuous (1-dimensional), and landmark (2- or 3-dimensional) data either alone or in any combination (except for analyses under *ML* that only allow discrete characters). The main output of the script is a figure (in scalable vector format, *svg*) showing the placement of one of the query species (called here *target species*) on the reference tree, color-coding the branches according to the degree of optimality of the alternative positions (Fig. 2a). Differing from other similar approaches (e.g., Revell et al. 2015; Parins-Fukuchi 2018), the present method is novel in allowing the inclusion of additional query species alongside the target species. This is particularly important when multiple species have uncertain placements (e.g., additional fossil species), as their inclusion may influence the optimal placement of the target species. As with a single query species, the scores represent the score of placing the target species on each branch of the reference tree. However, the calculations differ. When only the target species is analyzed, placement scores are directly obtained by positioning it on each branch of the reference tree and calculating the corresponding tree score. In contrast, when additional query species are included, the score for a given branch requires a phylogenetic search (TBR) with a backbone constraint that fixes the target species on the specified branch while allowing the remaining query species to be optimally placed. The placement of the query non-target species can be visualized in the parenthetical tree file generated by the script.

ML analyses are performed using a Markov model (Lewis 2001), having the option to use same/different branch length for each character partition (i.e., linked/unlinked models) and same/different number of maximum possible states for each character. The scores in *ML* analyses are, by default, expressed as $-\log L$ values. Alternatively, it is possible to express the values as Likelihood Weight Ratios, as implemented in the *EPA* approach (Berger et al. 2011). Due to the specific focus of *TNT*, the placement of multiple query species and the leave-one-out validation (*LOOV*) analyses are not available for *ML* in the current version of the script (see *PlaceMyFossils* User's Guide) but will be incorporated in future versions.

Multiple Optimality Criteria Analysis

PlaceMyFossils allows for comparison of the results obtained under different optimality criteria. The results are graphically displayed on the reference tree by plotting a colored checkerboard below each branch with colors representing the optimality of each placement of the target species (Fig. 2c).

Sensitivity Analysis

A sensitivity analysis can also be performed comparing the results under different analytical settings, which can either be defined from the menu or read from a text file (Fig. 2d). These settings may specify weights, state transformation costs, character activity, or any other settings defined by *TNT* commands.

Characters and Character Partitions Analysis

The script generates an *svg* file with a graphical representation of the placement scores for each partition on each branch using a colored checkerboard (Fig. 2b). This allows for evaluation of how the different characters/partitions are related to the optimal and suboptimal placements of the target species. The same information is also displayed in colored tables. For *ML* analyses, the interpretation of the results should consider that the presence of a character affects the rest of the scores by affecting the optimal branch length. Being aware of this, the character-by-character (and partition-by-partition) analysis is essential for understanding the results, as demonstrated in the examples below.

LOOV Analysis

When working with reference trees, it is possible to take advantage of the knowledge of accepted relationships to determine the optimal parameters to infer the placement of the query species. These settings may include model specifications, character settings (e.g., character weights, transformation costs, etc.), or other analytical conditions (as in the sensitivity approach). *PlaceMyFossils* implements a *LOOV* procedure to select the analytical conditions that minimize the error in placement (*EP*) (see "*EP and Subtree Error Analyses*") for the species included in the backbone tree (named here backbone species) and places the query species considering these optimal settings (Fig. 2d).

EP and Subtree Error Analyses

The evaluation of the placement scores along the reference tree gives important information to assess the confidence of the optimal placement. However, it is also essential to evaluate how accurate the query dataset is at placing the species in the expected position according

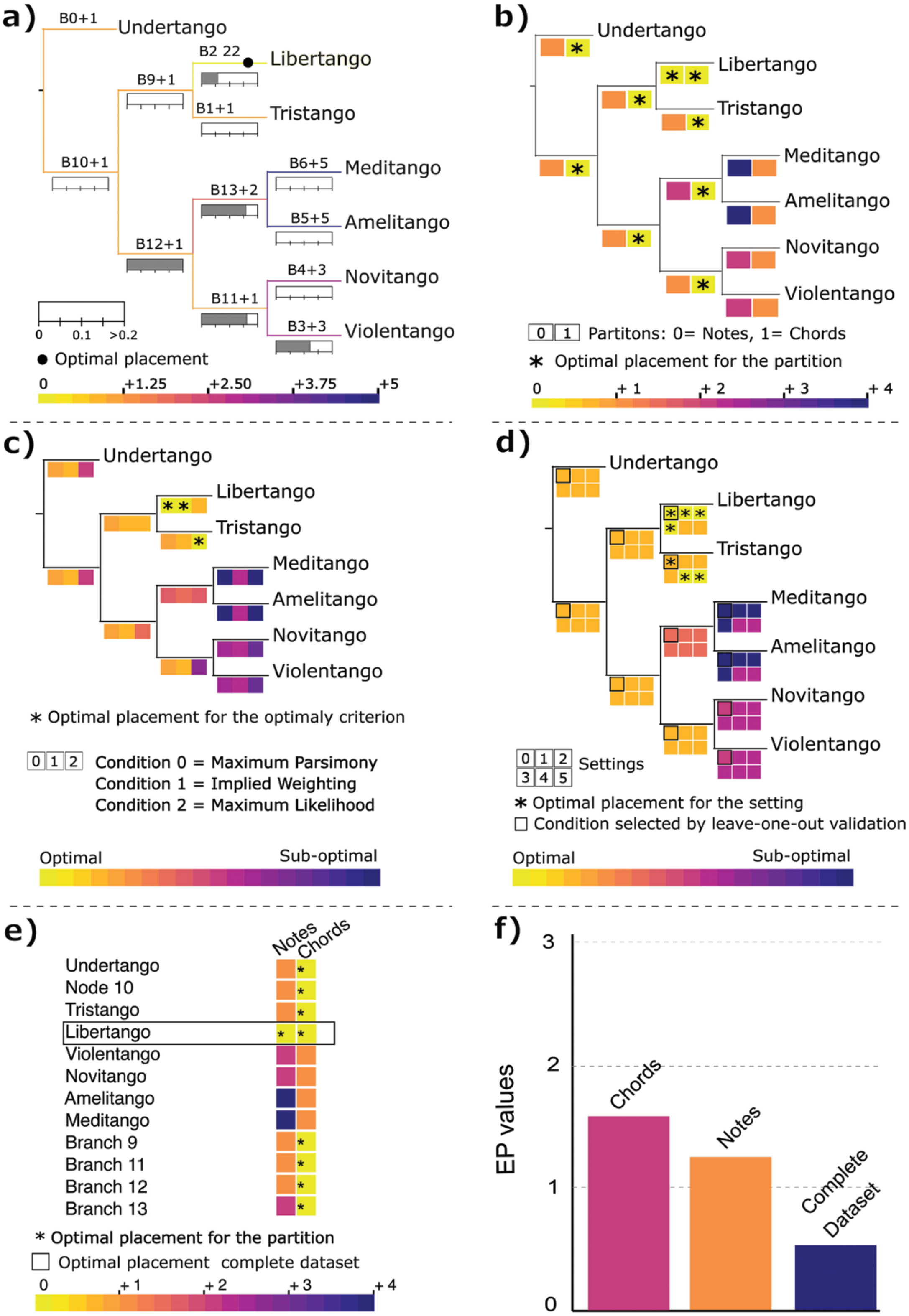

FIGURE 2. Graphical outputs of *PlaceMyFossils*. Color gradients in trees and tables represent the scores for the placement of the target species in a hypothetical example, using a musical metaphor where "chords" and "notes" represent different groups of characters. a) Main output showing the score for the hypothetical target species (*Nonino*) on each branch, the number of branch (B#), and (optionally) the in-subtree error values. b) Score for the placement of the target species on each tree branch for each character partition. c) Comparison of the results of three different optimality criteria. d) Results of *LOOV* and sensitivity analyses. e) Colored chart showing the differences in the score for each placement of the target species for each character partition. f) *EP* values for each character partition and the complete dataset.

to the evidence used to generate the reference tree. For example, consider a case where the phylogenetic signal of the morphological data is completely different from that of the genomic data used to infer the backbone tree. If there is a high congruence among morphological characters, the score of the best placement of the query species may be much better than the alternative placements. However, at the same time, this placement can be completely erroneous from the perspective of the molecular dataset considered to build the backbone tree. In real examples, the *EP* for the query species cannot be calculated directly because the required information is not available. *PlaceMyFossils* implements an approach to approximate the uncertainty associated with placing a query species in a backbone tree by calculating the *EP* for the backbone species. The underlying assumption is that the query species would behave similarly to the backbone species. Consequently, if the optimal placements based on the query dataset strongly differ from the original (i.e., the "correct") placements of the backbone species, the phylogenetic placement of the query species can be considered uncertain. *PlaceMyFossils* implements this calculation and evaluates the *EP* for the complete dataset and for each character partition (Fig. 2f). *EP* analysis is similar to that followed by Berger and Stamatakis (2010) and also by Ascarrunz et al. (2021), being the same metric followed in the LOOV analysis to select among analytical conditions. In addition, to determine that low levels of *EP* actually correspond to significant congruence in terms of placing species on the reference tree, our approach generates a null distribution of expected placement errors by pruning each backbone species, randomly placing those species on the tree, and calculating the distance between the original and the random position and summing up the distances for every backbone species.

Since *EP* is calculated for the whole tree, it is not possible to assess whether the dataset has different behaviors at different regions of the tree. To circumvent this limitation, we have also implemented an approach to evaluate the placement error for each subtree of the backbone tree. A subtree is defined here as including a node, its subtending branch, and all the descendant branches. The approach calculates the error for each subtree as the fraction of backbone species that are not part of that subtree in the backbone tree but that are optimally placed there when analyzing the query dataset. The rationale is that if many backbone species are wrongly placed by the query dataset for a certain subtree, the confidence in the placement of a query species (e.g., a fossil) on that particular subtree will be low. This error is called here "In-Subtree Error" (*ISE*) and is displayed along the branches of the reference tree (Fig. 2a). We express this error as a fraction of the maximum possible error for a given subtree. In addition, the script also calculates an Out-Subtree Error that establishes how common it is for the species from a given subtree to be placed outside this subtree. An extended explanation of the error in the placement approach, with worked examples, is provided in the printed Appendix.

When the target species has a character sampling that is deficient relative to that of the complete query dataset (as is usually the case in fossils), the *EP* may be underestimated. To better replicate the situation faced when placing the target species, it is possible to restrict the character sampling of each of the backbone species to be placed during the error calculation to the character sampling present in the target species. This is an option in *EP* and *LOOV* analyses.

Handling Multiple Reference Trees

PlaceMyFossils includes a function to work with multiple reference trees. This function is similar to that included in *RoguePlots* (Klopfstein and Spasojevic, 2019) for displaying the placement fossil species on a tree in a Bayesian framework. *RoguePlots* considers the placement of the fossil species on the sampled trees and displays the posterior probabilities on the majority rule consensus tree. In our function, the optimal placement for the target species on each reference tree is displayed (as black circles) on the strict consensus of all reference trees. The branches of the strict consensus tree are colored according to the minimum difference in score between the optimal placement for each reference tree and the score for placing the target species in the corresponding branch. In *RoguePlots* the posterior probabilities for branches that are not present in the majority rule consensus are displayed by coloring the vertical lines that connect branches. We follow the same approach here. The function to deal with multiple reference trees is implemented to evaluate the difference in scores of placing the target species along the branches of the trees, not for the rest of the analyses (e.g., character analysis). Other strategies may be followed to run the rest of the analyses under multiple reference trees, like using pruned consensuses or agreement subtrees (see the dinosaur example).

Effect of Query Species on Reference Topology

Backbone-based analyses assume that including the query species does not affect the rest of the phylogenetic relationships. This makes it possible to calculate the score of the query species at each branch of the reference tree. This assumption is reasonable when the backbone tree is derived from genomic data. However, when placing fragmented fossils in topologies derived from morphological data, this assumption need not always be true (e.g., see Gauthier et al. 1988). Since the reliability of the results of the analyses performed in *PlaceMyFossils* depends on this assumption, we have included a function that calculates the score of placing the target species in a reference tree while allowing the backbone species to change the position as well. Specifically, this function starts by considering as reference trees those trees derived from an unconstrained analysis that included the target species. Subsequently, the score of placing the target species on each branch of the tree is calculated by running a constrained search. In that search, the placement of the rest of the species is permitted to change but conserving the species composition for each tree partition defined at both ends of that branch. In other words, this approach summarizes the results of running multiple constrained searches, in each case placing the target species as sister of a clade present in the optimal trees. These results can be contrasted with those obtained using a fixed reference tree. If the scores calculated are congruent, the user can be confident in running the rest of the analyses (e.g., error analysis, character analysis) considering the reference topology as fixed. Details of this approach are presented in the Users' Guide of *PlaceMyFossils*.

Empirical Examples

To illustrate how *PlaceMyFossils* works, we provide two empirical studies representing different situations where the methodological tools proposed here are helpful. The first study analyses the placement of *Agathis immortalis,* an Araucariaceae (conifers) fossil from the Eocene of Patagonia, using a molecular phylogeny as a backbone tree and a morphological dataset (including discrete and continuous characters) as a query dataset (Escapa et al. 2018). The dataset analyzed included 63 morphological characters (ten continuous, 53 discrete), and 42 extant taxa. The characters classified in three different partitions: pollen cone (ten chars.), seed cone (29), and vegetative (24). In a second study, we showcase how *PlaceMyFossils* can be useful for analyzing the phylogenetic placement of a fragmentary specimen in a backbone tree derived from morphological data. In this example, we analyzed the placement of *Nyasasaurus parringtoni*, known only from a few vertebrae and a single humerus (Nesbitt et al. 2013). The dataset included 84 species and 457 discrete characters (Müller and Dias-da-Silva, 2019). The results obtained in both datasets were compared to those obtained by the *EPA* approach as implemented in *RAxML*. For this comparison, the continuous characters were excluded because *RAxML* cannot handle them. In addition, we compared the results obtained in the dinosaur dataset with those obtained using the *RoguePlots* function (Klopfstein and Spasojevic 2019) on the trees obtained in a Bayesian analysis run in *MrBayes* (Ronquist et al. 2012). Although more recent placement approaches have been published (Barbera et al. 2019; Balaban et al. 2020; Weddel et al. 2021), we preferred to compare our results with those obtained with tools previously used in paleontological studies. The same dinosaur dataset was considered to evaluate the *EP* for each of 9 character partitions representing different anatomical regions: skull (119 characters), mandible (27), teeth (39), vertebrae (51), pectoral (14), forelimb (41), pelvic (60), hindlimb (100), and dermal (6).

Araucariaceae Example

The *MP* optimal placement of *A. immortalis* on the molecular backbone tree was at the base of the *Agathis* clade, with a difference of 1.89 steps to the best suboptimal placements, all within *Agathis* (Fig. 3). The score of placing the target species clearly varies along the tree with a marked affinity to the agathioid clade, and a strong suboptimality in different lineages of *Araucaria* and in the outgroups. The same placement was obtained when five more fossil species were added as query species (Supplementary Results 1). The analysis of the *EP* showed that no backbone species outside *Agathis* was placed in that subtree (*ISE* = 0). The placement of the target species was also run under the *IW* criterion, choosing the concavity value using the *LOOV* approach. The result was the same as that obtained in the *MP* analysis (Supplementary Results 1). The placement obtained by *EPA* in *RAxML* is coincident with the *ML* placement obtained in *PlaceMyFossils* (Supplementary Results 1).

The character analysis showed that the seed cone partition supported the placement of the target species as stem lineage of *Agathis*. For the vegetative partition, multiple optimal placements were obtained, corresponding to every branch within *Agathis,* both crown and stem. In contrast, the pollen cone characters indicated a basal placement within Araucariaceae (Fig. 4a). The only character that had a different score between the branch subtending the agathioid clade and the branch subtending Araucariaceae node was the pollen cone width (Fig. 4b). The optimization of this character on the tree (not shown) reconstructed narrower pollen cones on basal branches and wider cones within *Agathis*, a clear synapomorphy for the genus. This indicates that the placement of *A. immortalis* at the base of Araucariaceae in the pollen cone partition was driven by this character. The highest *EP* was obtained for the pollen cone partition (Fig. 4c), though with small differences among partitions. The *ISE* for the vegetative and seed cone partitions for the subtree sustained by the branch where *A. immortalis* is placed was 0 (Supplementary Results 1), showing a high confidence in the placement of this species when both partitions were analyzed separately. The combined evaluation of the results obtained running *PlaceMyFossils* gives support for the placement of *Agathis immortalis* at the base of the genus.

Dinosaurs Example

The unconstrained search under *MP* produced multiple optimal trees. Hence, we followed different strategies to deal with multiple reference trees. First, we repeated the analyses considering 10 trees from more than 1000 optimal trees obtained. These 10 trees were selected by a *TNT* script (Supplementary Material) to be topologically as different as possible. Second, we used a maximum agreement subtree (i.e., subtree that has the largest number of species for which phylogenetic relationships are maintained in all source trees) as a backbone tree. From there, two different approaches were

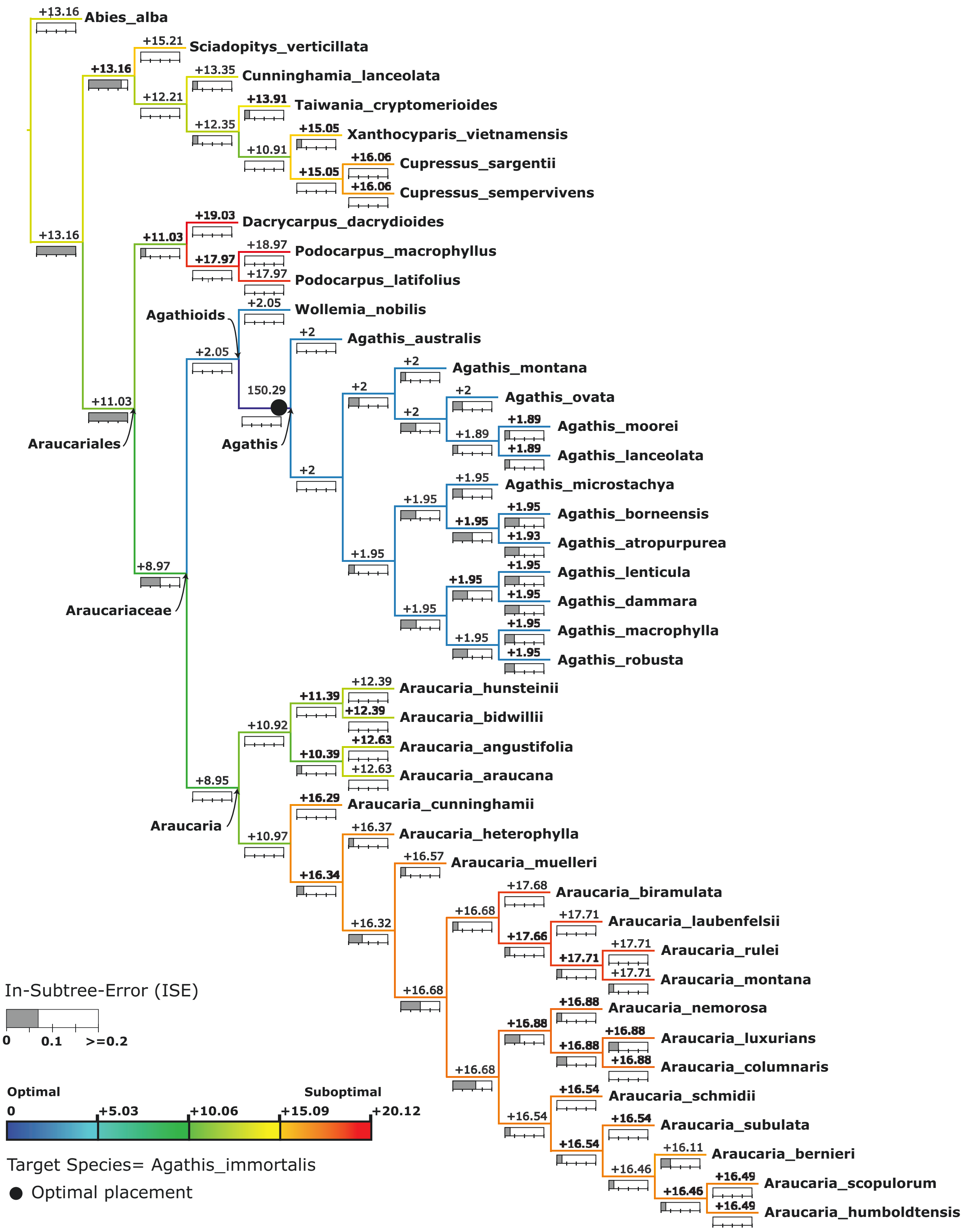

Figure 3. Results of running *PlaceMyFossils* for placing *A. immortalis* (Araucariaceae, conifers) on a molecular reference tree based on *MP* criterion. The tree is colored and labeled according to the optimality of the placement of the target species, showing the In-Subtree-Error value for each branch.

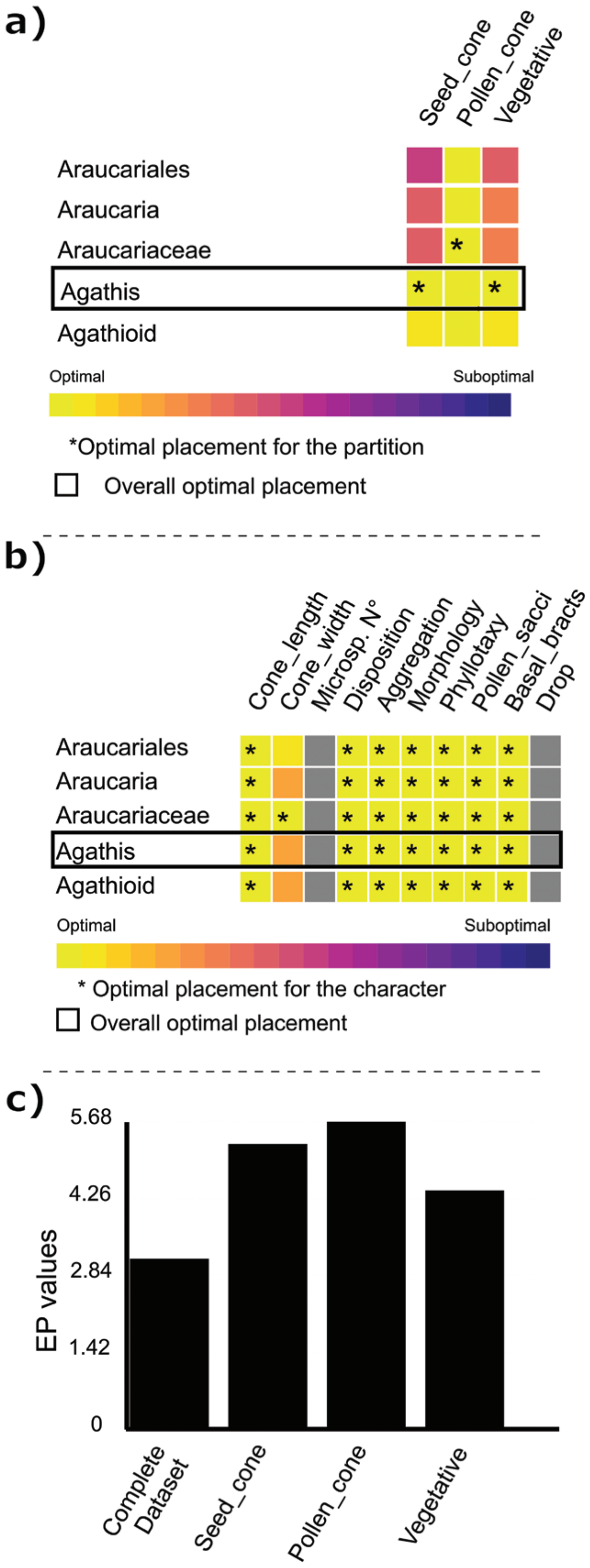

FIGURE 4. Results of running *PlaceMyFossils* for placing *Agathis immortalis* (Araucariaceae, conifers) on a molecular backbone tree. a) Colored table showing the *MP* scores for the placement of *A. immortalis* in branches subtending five relevant clades (see Fig. 3) for each character partition. b) Colored table showing the *MP* scores for the placement of *A. immortalis* in five relevant clades for each character of the pollen cone partition. Black rectangle indicates the optimal placement for the complete dataset. c) *EP* for each character partition when character sampling on *A. immortalis* is considered.

run. In the first, the pruned species were included in the query list (see *Multiple Optimal Reference Trees* section). In the second, the pruned species were directly excluded from the matrix, leaving *N. parringtoni* as the only species in the query list. We also calculated the placement scores considering 300 optimal trees randomly chosen and displaying the placement of the target species in the strict consensus. Finally, we ran an analysis letting the target species modify the relationships of the reference trees.

The analyses considering different optimality criteria (*ML*, *IW*, and *MP*) and different reference topologies all placed *N. parringtoni* within Sauropodomorpha (Fig. 5; Supplementary Results 2), except in some *IW* iterations (strong weighting + maximum agreement subtree as reference) where this species was placed outside Dinosaurs, as part of Silesauridae. Under the *MP* criterion, the score difference between the best placements within Sauropodomorpha and those within Silesauridae was 1–3 steps, depending on the backbone tree considered. For *ML*, *likelihood weight ratio* value calculated by the *EPA* approach summed up for the branches within Sauropodomorpha was 0.999 against 0.001 for branches within the Silesauridae clade (Supplementary Results 2). The optimal placement was coincident with that obtained in the *EPA* approach implemented in *RAxML* (Supplementary Results 2). When the placement of the backbone species was allowed to change, the placement of *N. parringtoni* was the same as the one obtained with the fixed backbones trees (Supplementary Results 2). The placements with the highest PP established by the *RoguePlots* function matched those obtained as optimal by *PlaceMyFossils* under the *MP* criterion. In addition, the placement with the highest PP was coincident with that obtained by the *EPA* approach (Supplementary Results 2).

The *ISE* for the subtree composed of Sauropodomorpha and its subtending branch was high (up to 0.10). However, this was not the case for the more shallow branches within Sauropodomorpha, where *N. parringtoni* was placed (*ISE* between 0 and 0.03). The analysis for each partition showed very high *ISE* values for the hindlimb partition for those subtrees, probably because that partition includes only four characters scored for the target species.

The optimal placements of *N. parringtoni*, considering the vertebrae characters, were within Sauropodomorpha. In contrast, the hindlimb, represented by only four characters, indicated two different regions of optimal placements: outside Dinosauria at the base of the tree and within Ceratosauria (Supplementary Results 2). This partitioned analysis indicates that the alternative (suboptimal) placement

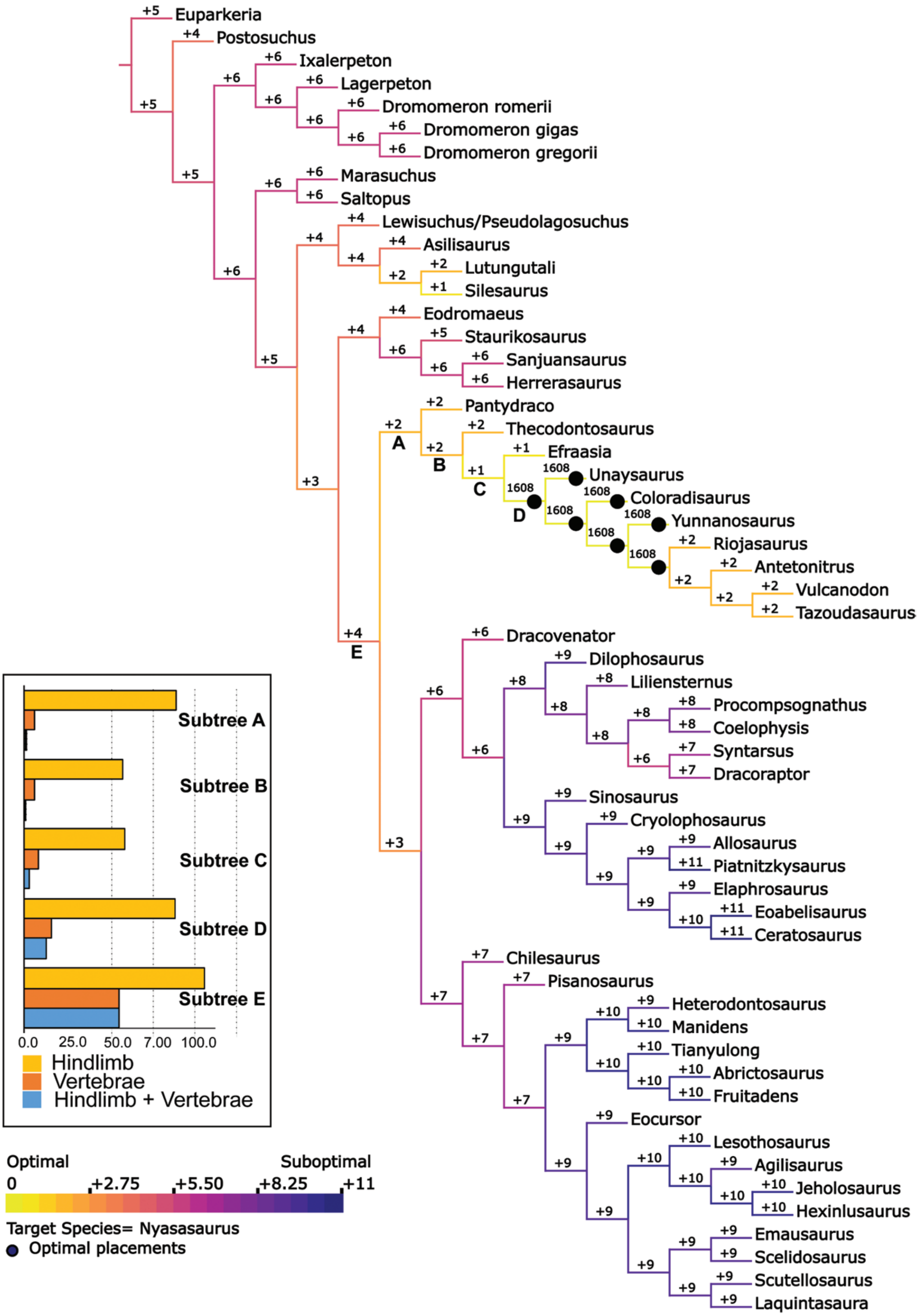

FIGURE 5. Placement of *N. parringtoni* in the dinosaur phylogeny established by *PlaceMyFossils* under Maximum Parsimony criterion (*MP*). The backbone tree corresponds to a maximum agreement subtree of all most parsimonious reference trees. The numbers on the branches represent the *MP* scores. The barplot shows the in-subtree error for relevant alternative placements for *N. parringtoni* calculated considering the character sampling of this species.

as part of Silesauridae obtained for the complete dataset was not due to alternative phylogenetic signals between the two partitions but due to discordance among the vertebrae characters.

The integrative evaluation of the results obtained using *PlaceMyFossils* casts some doubts about the placement of *N. parringtoni* as part of Dinosauria. First, under some analytical conditions, *N. parringtoni* is placed in a different clade (Silesauridae). Second, the score differences between the best placements within Sauropodomorpha and those within Silesauridae are small, at least in the case of *MP*, with a difference of just one step in some iterations. Third, the placement within Sauropodomorpha is supported by only one of the partitions (vertebrae) and not by the hindlimb. On the other hand, the *EPA* approach and the error analyses support the placement of *N. parringtoni* as a sauropodomorph. In any case, the approach presented here allows for the identification of the causes for uncertainty in the placement of the query species at different analytical dimensions.

We also analyzed the performance of the different partitions to place the backbone species on the reference trees. For this, we compare the *EP* values for each partition, considering the complete character sampling, not limiting the character sampling to those scored in *N. parringtoni*. The results showed large differences in *EP* values among partitions (Supplementary Results 2). The largest *EP* values were obtained for the pectoral and forelimb partitions. In contrast, the skull and hindlimb partitions presented the lowest *EP* values. The *EP* values were lower than expected by chance for all the partitions except the pectoral and the dermal. The leading cause for the different behavior of the different partitions is clearly related to the number of characters: the larger the number of characters in a partition, the lower the *EP* values (Supplementary Results 2).

Discussion and Final Remarks

We describe a new approach—*PlaceMyFossils*—to analyze the phylogenetic placement of query species on a reference tree. The utility of *PlaceMyFossils* was illustrated with two empirical examples (conifers and Dinosauria). The placements obtained were consistent with those generated by other analytical tools, such as *EPA* and *RoguePlots*. However, *PlaceMyFossils* provided additional relevant information for thoroughly evaluating the evidence supporting the results by defining the contribution of each character partition, establishing placement error, and determining the effect of different analytical conditions on the outcomes. Although *PlaceMyFossils* is not fully functional for *ML*, users interested in model-based analyses can find functionalities not available in other tools. For instance, it includes the visualization of *EPA* results, a feature not implemented in *RAxML*, that makes it necessary to use additional tools such as *gappa* (Czech et al. 2020), *ggtree* (Yu 2020), or *iTOL* (Letunic et al. 2024). Moreover, it supports discrete characters, which cannot be handled by other approaches like *locate.fossil* or *Cophymaru*. In addition, *PlaceMyFossils* allows the calculation and display of *ML* scores for groups of characters, a feature that distinguishes it from all other approaches. Hence, *PlaceMyFossils* complements the phylogenetic analysis toolbox available to morphologists, especially paleontologists. In this regard, *PlaceMyFossils* rather than aiming to replace traditional unconstrained analyses, it seeks to facilitate more informed backbone-based analyses for researchers who choose to use these methods.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.fbg79cp3s

Acknowledgments

We thank the Willi Hennig Society for subsidizing TNT and allowing it to be it freely available. We also thank Dr. Czech and Dr. Coiro for their valuable comments and suggestions that significantly improved the quality of the manuscript. Finally, we also thank D. Pol for suggesting the *Nyasasaurus* example, and Ambrosio Torres and Martin Morales for testing the script.

Funding

This work was supported by the Agencia Nacional de Promoción Científica y Técnica (PICT-BID-CONICET 2019-03675), the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 11220200102052), the Agencia Estatal de Investigación of the Spanish Ministerio de Ciencia e Innovación (AEI/10.13039/501100011033), and the Generalitat de Catalunya (CERCA Programme).

Data Availability

PlaceMyFossils script, along with its manual and working examples, can be found at https://github.com/sacatalano/PlaceMyFossils. Supplementary results, datasets for both empirical examples and *PlaceMyFossils* code are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.fbg79cp3s.

Error in placement Analyses

The error in placement (*EP*) is calculated as follows:

$$EP = \left(\sum_{i=0}^{i<B} \mathbf{Di} \right) / \mathrm{B} \qquad \text{(A1)}$$

Where B is the number of backbone species, and D_i is the number of intermediate nodes on the path connecting the original placement of species i in the backbone tree and the optimal placement of species i considering the query dataset. D may take different values in the presence of multiple optimal placements for the backbone species based on the query dataset. *PlaceMyFossils* displays the *EP* values calculated by summing up either the minimum, the maximum, or the average D values for each backbone species.

EP significance test

The *EP* value depends on the tree shape since the maximum possible D values also depend on the tree shape. This, in principle, would imply that the tree shape should be considered during *EP* calculation. However, since *EP* is used to compare the error in the placement of the different character partitions –or, alternatively, different conditions for the phylogenetic placement of the query species-, on the same topology, the possible effects of different tree shapes in the calculation are not relevant: the tree shape will affect the *EP* calculation for all partitions/conditions in the same way. While the tree shape does not affect the *EP* value when comparing the error among analytical conditions or partitions for a given topology, the absolute *EP* value of a given dataset is impacted. Hence, a null distribution of expected placement errors should be generated to determine that low *EP* values correspond to a significant congruence in terms of placing species on the reference topology. Our approach generates the latter by pruning each backbone species, randomly placing those species on the tree, calculating the distance between the original and the random positions, and summing up the distances for every backbone species. This process is repeated n times, and the p value is calculated as the fraction of the n iterations that present an *EP* value lower or equal than the *EP* observed for the query dataset. Since this procedure is done on the original topology, the shape of the tree is explicitly considered in the calculation.

Error in Placement along the tree: Subtree Error Analysis

EP analysis approximates the error that the query dataset has in placing query species on the reference topology. However, it is possible that the error varies in different tree regions. This is the most relevant information in the context of placing query species on a backbone tree: If many backbone species are wrongly placed by the query dataset at a certain subtree, the confidence in the placement of a query species on that particular

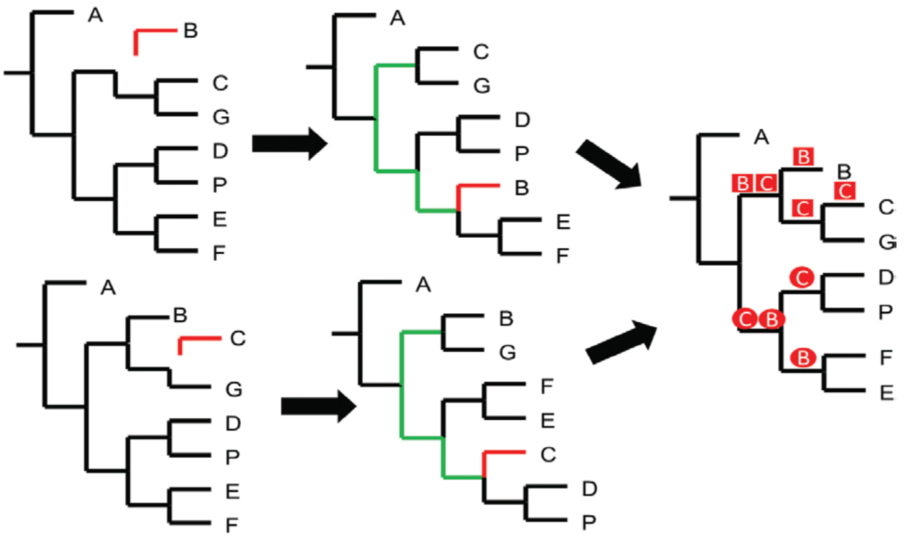

Figure A1. Rationale of the approach for calculating the error in placement for the complete tree (*EP*) and each “in-” and “out-” subtree (*ISE* and *OSE*, respectively). The backbone species are pruned from the tree (Left) and then placed back on the tree in the optimal placement(s) according to the query dataset. In this example, the optimal positions for species B and C differ from the correct original position (Middle). The green branches show the distance between the correct and the optimal placements. The total error for the complete tree considering the query dataset is calculated as the sum of these distances (measured as the number of intermediate nodes) divided by the number of backbone species. Assuming that the rest of the species are placed in the correct position, the error in this example would be (3 + 3) / 8 = 0.75. This error can be calculated for each character partition as well. The tree on the right shows how the *ISE* and *OSE* are calculated. Circles represent the basal branch of each subtree where the *ISE* is increased. Species C increases the *ISE* for the subtree formed by species D and P, and the subtree formed by D, P, F, and E. Species B increases the *ISE* of the subtrees formed by species F and E, as well as D, P, F, and E. Assuming that the rest of the species are placed in the correct position, the *ISE* for D + P and E + F would be 1 / 6 = 0.166, as one out of three species outside these subtrees are placed within the subtree. For the subtree (D P F E), the *ISE* would be 2 / 3 = 0.66. Squares represent subtrees where the *OSE* is increased. The *OSE* for (B C G) is 0.66 (two of the tree species corresponding to that subtree are incorrectly placed). The rest of the subtrees have an *OSE* of zero.

subtree will be low. Hence, in addition to the *EP* value, *PlaceMyFossils* calculates an error for every subtree of the backbone topology, where each subtree is represented by a node, its subtending branch, and all the descendant branches. The approach calculates the error for each subtree as the fraction of backbone species that are not part of that subtree in the backbone tree but that are optimally placed there when analysing the query dataset. Hence, the in-subtree error (*ISE*) for subtree *i* is calculated as:

$$ISE_{\mathrm{i}} = (\mathrm{O_i}/(\mathrm{B} - \mathrm{S_i} - 1)) \qquad \text{(A2)}$$

Where $\mathrm{O_i}$ is the number of backbone species that are wrongly placed in subtree *i*, B is the total number of species in the backbone, and $\mathrm{S_i}$ is the number of species in subtree *i*. The out-subtree error (*OSE*) is also calculated and describes how common it is for the species from a subtree in the backbone to end up outside this subtree. The *OSE* for the subtree *i* is calculated as:

$$OSE_{\mathrm{i}} = (\mathrm{C_i} \; / \; \mathrm{S_i}) \qquad \text{(A3)}$$

Where C is the number of species from the subtree *i* placed outside this subtree, and $\mathrm{S_i}$ is the number of species in subtree *i*. This error represents how commonly species from the backbone are wrongly placed as sisters of these terminal leaves. *OSE* is helpful regarding reaching conclusions about species not belonging to a specific clade. For instance, if we want to conclude that the query species is not a member of genus B, we can approximate the error associated with that statement by calculating how common species from genus B are placed in other nodes of the tree when the query dataset is considered to place the backbone species in the tree.

The *ISE* and *OSE* are calculated for each subtree, and as indicated each subtree includes at least one node, its subtending branch, and all the descendants of that node. Including the subtending branch as part of the subtree is justified based on the following rationale: Suppose that your analysis determines that the query species is placed on a branch "A" that subtends a node "A". We want to include every backbone species that is placed in the same branch in the error count because this will give us an idea of how confident we can be in the placement of the query species. Alternatively, if we consider that the branch subtending the subtree is not part of the subtree, we will erroneously consider that the error in placement where the query species was placed is zero.

The branch leading to a terminal node is a potential placement for the target species. Hence, it has its own *ISE*. This expresses how often species from the backbone are placed as the sister of species A when the query dataset is considered. If many species from the backbone are placed there, there is low confidence that the target species is the sister species of A. Similarly, the *OSE* is also calculated for the terminal nodes. In that case, the *OSE* indicates that this species is optimally placed elsewhere in the backbone when the query dataset is considered, not in its original position.

Interpreting error in placement results

The *EP*, *ISE*, and *OSE* values strictly describe and summarize the results of placing the backbone species on the reference tree. In that respect, the calculations are exact and objective measures. Whether this results can be extrapolated to the query species will depend on the particular dataset. To correctly interpret the *ISE* and *OSE* values it is important to consider the *ISE* values for the neighbour nodes. Consider a target species placed in a clade corresponding to a genus. An *ISE* equal to zero for this subtree indicates that we can be confident about the placement of the target species in that genus. However, the error analysis will also show confidence in placing the query species as part of the genus if any of the nodes between the optimal placement and the ancestor of the whole genus has a high *ISE* value. On the contrary, if the nodes along this path have a high error value, the placement of the species as part of the genus will be uncertain.

References

Ascarrunz E., Claude J., Joyce W.G. 2021. The phylogenetic relationships of geoemydid turtles from the Eocene Messel Pit Quarry: a first assessment using methods for continuous and discrete characters. PeerJ. 9:e11805.

Badano D., Fratini M., Maugeri L., Palermo F., Pieroni N., Cedola A., Haug J.T., Weiterschan T., Velten J., Mei M., Di Giulio M., Cerretti P. 2021. X-ray microtomography and phylogenomics provide insights into the morphology and evolution of an enigmatic Mesozoic insect larva. Syst. Entomol. 46:672–684.

Balaban M., Sarmashghi S., Mirarab S. 2020. APPLES: scalable distance-based phylogenetic placement with or without alignments. Syst. Biol. 69:566–578.

Barbera P., Kozlov A.M., Czech L., Morel B., Darriba D., Flouri T., Stamatakis A. 2019. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst. Biol. 68:365–369.

Berger S.A., Krompass D., Stamatakis A. 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol. 60:291–302.

Berger S.A., Stamatakis A. 2010. Accuracy of morphology-based phylogenetic fossil placement under maximum likelihood. In: ACS/IEEE International Conference on Computer Systems and Applications-AICCSA 2010: 1–9. IEEE.

Bomfleur B., Grimm G.W., McLoughlin S. 2015. *Osmunda pulchella sp. nov*. from the Jurassic of Sweden—reconciling molecular and fossil evidence in the phylogeny of modern ferns (Osmundaceae). BMC Evol. Biol. 15:1–25.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Drummond A.J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Černý D., Natale R. 2022. Comprehensive taxon sampling and vetted fossils help clarify the time tree of shorebirds (Aves, Charadriiformes). Mol. Phylogenet. Evol. 177:107620.

Czech L., Barbera P., Stamatakis A. 2020. Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. Bioinformatics. 36:3263–3265.

Czech L., Stamatakis A., Dunthorn M., Barbera P. 2022. Metagenomic analysis using phylogenetic placement—a review of the first decade. Front. Bioinform. 2:871393.

Doyle J.A., Endress P.K. 2010. Integrating Early Cretaceous fossils into the phylogeny of living angiosperms: magnoliidae and eudicots. J. Syst. Evol. 48:1–35.

Escapa I.H., Iglesias A., Wilf P., Catalano S.A., Caraballo-Ortiz M.A., Cúneo R. 2018. *Agathis* trees of Patagonia's Cretaceous-Paleogene death landscapes and their evolutionary significance. Am. J. Bot. 105:1345–1368.

Farris J.S. 1970. Methods for computing Wagner trees. Syst. Biol. 19:83–92.

Gauthier J., Kluge A.G., Rowe T. 1988. Amniote phylogeny and the importance of fossils. Cladistics. 4:105–209.

Goloboff P.A. 1993. Estimating character weights during tree search. Cladistics. 9:83–91.

Goloboff P.A., Morales M.E. 2023. TNT version 1.6, with a graphical interface for MacOS and Linux, including new routines in parallel. Cladistics. 39:144–153.

Heckeberg N.S. 2020. The systematics of the Cervidae: a total evidence approach. PeerJ 8:e8114.

Klopfstein S., Spasojevic T. 2019. Illustrating phylogenetic placement of fossils using RoguePlots: an example from ichneumonid parasitoid wasps (Hymenoptera, Ichneumonidae) and an extensive morphological matrix. PLoS One. 14:e0212942.

Kozlov A.M., Darriba D., Flouri T., Morel B., Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 35:4453–4455.

Letunic I., Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 52:W78–W82.

Lewis P.O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

López-Martínez A.M., Schönenberger J., von Balthazar M., González-Martínez C.A., Ramírez-Barahona S., Sauquet H., Magallón S. 2023. Integrating fossil flowers into the angiosperm phylogeny using molecular and morphological evidence. Syst. Biol. 72:837–855.

Manos P.S., Soltis P.S., Soltis D.E., Manchester S.R., Oh S.H., Bell C.D., Dilcher D.L., Stone D.E. 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. Syst. Biol. 56:412–430.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., Von Haeseler A., Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37:1530–1534.

Müller R.T., Dias-da-Silva S. 2019. Taxon sample and character coding deeply impact unstable branches in phylogenetic trees of dinosaurs. Hist. Biol. 31:1089–1092.

Nesbitt S.J., Barrett P.M., Werning S., Sidor C.A., Charig A.J. 2013. The oldest dinosaur? A middle triassic dinosauriform from Tanzania. Biol. Lett. 9:20120949.

Oude Munnink B.B., Worp N., Nieuwenhuijse D.F., Sikkema R.S., Haagmans B., Fouchier R.A., Koopmans M. 2021. The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology. Nat. Med. 27:1518–1524.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Parins-Fukuchi C. 2018. Bayesian placement of fossils on phylogenies using quantitative morphometric data. Evolution. 72:1801–1814.

Pugh K.D. 2022. Phylogenetic analysis of Middle-Late Miocene apes. J. Hum. Evol. 165:103140.

Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Revell L.J., Mahler D.L., Reynolds R.G., Slater G.J. 2015. Placing cryptic, recently extinct, or hypothesized taxa into an ultrametric phylogeny using continuous character data: a case study with the lizard *Anolis roosevelti*. Evolution 69:1027–1035.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539–542.

Wedell E, Cai Y, Warnow T. 2021. Scalable and accurate phylogenetic placement using pplacer-XR. In: Martín-Vide C., Vega-Rodríguez M.A., Wheeler T., editors. Algorithms for computational biology. AlCoB. Lecture notes in computer science, Vol. 12715. Cham: Springer. p. 94–105.

Yu G. 2020. Using ggtree to visualize data on tree-like structures. Curr. Protoc. Bioinform. 69:e96.

Zuntini A.R., Carruthers T., Maurin O., Bailey P.C., Leempoel K., Brewer G.E., Epitawalage N., Françoso E., Gallego-Paramo B., McGinnie C., Negrão R., Roy S.R., Simpson L., Toledo Romero E., Barber V.M.A., Botigué L., Clarkson J.J., Cowan R.S., Dodsworth S., Johnson M.G., Kim J.T., Pokorny L., Wickett N.J., Antar G.M., DeBolt L., Gutierrez K., Hendriks K.P., Hoewener A., Hu A.-Q., Joyce E.M., Kikuchi I.A.B.S., Larridon I., Larson D.A., de Lírio E.J., Liu J.-X., Malakasi P., Przelomska N.A.S., Shah T., Viruel J., Allnutt T.R., Ameka G.K., Andrew R.L., Appelhans M.S., Arista M., Ariza M.J., Arroyo J., Arthan W., Bachelier J.B., Bailey C.D., Barnes H.F., Barrett M.D., Barrett R.L., Bayer R.J., Bayly M.J., Biffin E., Biggs N., Birch J.L., Bogarín D., Borosova R., Bowles A.M.C., Boyce P.C., Bramley G.L.C., Briggs M., Broadhurst L., Brown G.K., Bruhl J.J., Bruneau A., Buerki S., Burns E., Byrne M., Cable S., Calladine A., Callmander M.W., Cano A., Cantrill D.J., Cardinal-McTeague W.M., Carlsen M.M., Carruthers A.J.A., de Castro Mateo A., Chase M.W., Chatrou L.W., Cheek M., Chen S., Christenhusz M.J.M., Christin P.-A., Clements M.A., Coffey S.C., Conran J.G., Cornejo X., Couvreur T.L.P., Cowie I.D., Csiba L., Darbyshire I., Davidse G., Davies N.M.J., Davis A.P., van Dijk K.-jent, Downie S.R., Duretto M.F., Duvall M.R., Edwards S.L., Eggli U., Erkens R.H.J., Escudero M., de la Estrella M., Fabriani F., Fay M.F., Ferreira Paola de L., Ficinski S.Z., Fowler R.M., Frisby S., Fu L., Fulcher T., Galbany-Casals M., Gardner E.M., German D.A., Giaretta A., Gibernau M., Gillespie L.J., González C.C., Goyder D.J., Graham S.W., Grall A., Green L., Gunn B.F., Gutiérrez D.G., Hackel J., Haevermans T., Haigh A., Hall J.C., Hall T., Harrison M.J., Hatt S.A., Hidalgo O., Hodkinson T.R., Holmes G.D., Hopkins H.C.F., Jackson C.J., James S.A., Jobson R.W., Kadereit G., Kahandawala I.M., Kainulainen K., Kato M., Kellogg E.A., King G.J., Klejevskaja B., Klitgaard B.B., Klopper R.R., Knapp S., Koch M.A., Leebens-Mack J.H., Lens F., Leon C.J., Léveillé-Bourret E., Lewis G.P., Li D.-Z., Li L., Liede-Schumann S., Livshultz T., Lorence D., Lu M., Lu-Irving P., Luber J., Lucas E.J., Luján M., Lum M., Macfarlane T.D., Magdalena C., Mansano V.F., Masters L.E., Mayo S.J., McColl K., McDonnell A.J., McDougall A.E., McLay T.G.B., McPherson H., Meneses R.I., Merckx V.S.F.T., Michelangeli F.A., Mitchell J.D., Monro A.K., Moore M.J., Mueller T.L., Mummenhoff K., Munzinger J., Muriel P., Murphy D.J., Nargar K., Nauheimer L., Nge F.J., Nyffeler R., Orejuela A., Ortiz E.M., Palazzesi L., Peixoto A.L., Pell S.K., Pellicer J., Penneys D.S., Perez-Escobar O.A., Persson C., Pignal M., Pillon Y., Pirani J.R., Plunkett G.M., Powell R.F., Prance G.T., Puglisi C., Qin M., Rabeler R.K., Rees P.E.J., Renner M., Roalson E.H., Rodda M., Rogers Z.S., Rokni S., Rutishauser R., de Salas M.F., Schaefer H., Schley R.J., Schmidt-Lebuhn A., Shapcott A., Al-Shehbaz I., Shepherd K.A., Simmons M.P., Simões A.O., Simões A.R.G., Siros M., Smidt E.C., Smith J.F., Snow N., Soltis D.E., Soltis P.S., Soreng R.J., Sothers C.A., Starr J.R., Stevens P.F., Straub S.C.K., Struwe L., Taylor J.M., Telford I.R.H., Thornhill A.H., Tooth I., Trias-Blasi A., Udovicic F., Utteridge T.M.A., Del Valle J.C., Verboom G.A., Vonow H.P., Vorontsova M.S., de Vos J.M., Al-Wattar N., Waycott M., Welker C.A.D., White A.J., Wieringa J.J., Williamson L.T., Wilson T.C., Wong S.Y., Woods L.A., Woods R., Worboys S., Xanthos M., Yang Y., Zhang Y.-X., Zhou M.-Y., Zmarzty S., Zuloaga F.O., Antonelli A., Bellot S., Crayn D.M., Grace O.M., Kersey P.J., Leitch I.J., Sauquet H., Smith S.A., Eiserhardt W.L., Forest F., Baker W.J. 2024. Phylogenomics and the rise of the angiosperms. Nature. 629:843–850.