








## TECHNICAL REPORT

Fast production of large, time-calibrated, informal  
supertrees with `tree.merger`

by SILVIA CASTIGLIONE<sup>1\*</sup> , CARMELA SERIO<sup>2</sup> ,  
ALESSANDRO MONDANARO<sup>1,3</sup> , MARINA MELCHIONNA<sup>1</sup>  and  
PASQUALE RAIA<sup>1</sup> 

<sup>1</sup>Department of Earth Sciences, Environment & Resources, University of Naples Federico II, 80138, Napoli, Italy; [silvia.castiglione@unina.it](mailto:silvia.castiglione@unina.it)

<sup>2</sup>Research Centre in Evolutionary Anthropology & Palaeoecology, School of Biological & Environmental Sciences, Liverpool John Moores University, Liverpool, UK

<sup>3</sup>Department of Earth Sciences, University of Florence, 50121, Firenze, Italy

\*Corresponding author

Typescript received 24 May 2021; accepted in revised form 28 October 2021

**Abstract:** Assembling informal supertrees inclusive of extinct species is a useful but particularly long and complex procedure. We introduce a new, interactive, piece of software that allows the fast production of large, time-calibrated informal supertrees, single-handedly mixing multiple phylogenetic information from different sources. The software, embodied in the `tree.merger` function available as part of the RRphylo R package, allows the merging of different trees into one or adding individual species to a target phylogeny. Time calibration is implemented automatically within the function according to user-specified, optional age values that can be

provided for nodes and/or tips. We applied `tree.merger` to two different case-studies. The first, hypothetical, case study pertains to the tree of odontocete cetaceans. The second case study regards the expansion of the new, higher-level phylogeny of dinosaurs proposing the sister clade relationship between Ornithoscelida and Sauropodomorpha from *c.* 50 to a 357 species tree. The two case studies took less than five seconds each to complete, on a regular personal computer.

**Key words:** informal supertree, phylogenetics, dinosaur, `tree.merger`, RRphylo.

THERE is recent, widespread appreciation that including fossil information in phylogenetic analyses increases the power of inference on trait evolution and provides better estimates of ancestral states (Slater & Harmon 2013; Chira & Thomas 2016; Castiglione *et al.* 2020; Koch & Parry 2020). However, reaping such benefits requires producing phylogenetic trees inclusive of extinct species, that maintain both the correct tree topology and the exact ages for tips and nodes (Bininda-Emonds 2004). This task is often time-consuming, requires extensive manipulation, and is easily frustrated by the chance of adding new phylogenetic information from the publication of new and updated phylogenies, while the process is still ongoing (Aze *et al.* 2011; Cascini *et al.* 2019; Koch & Parry 2020). Although the tree topology can easily be changed with interactive software interfaces such as Mesquite (Maddison & Maddison 2019) and TreeGraph (Stöver & Müller 2010), until now, no software has allowed the straightforward merging of different phylogenies while holding correct species and nodes ages. Here, we present such a tool, the `tree.merger` application, embedded in the R package RRphylo (Castiglione

*et al.* 2018). We present `tree.merger` using a hypothetical tree manipulation procedure regarding the phylogeny of odontocete cetaceans (to illustrate the full functionalities of the tool), and apply the algorithm to a real case study: the phylogeny of dinosaurs. Dinosaurs were traditionally viewed as including two major sister clades: those with a bird-like pelvis, Ornithischia (duck-billed, armoured and frilled dinosaurs), and those with crocodylian-like pelvis, Saurischia (including theropods and birds, and the sauropods). A third, rather enigmatic Triassic clade of typically South American dinosaurs, known as Herrerasauridae, is considered to be allied to Saurischia, although with uncertain phylogenetic positioning (Bittencourt *et al.* 2014). This quite stable phylogenetic arrangement was challenged by Baron *et al.* (2017). By analysing 457 phenotypic characters from 74 different taxa Baron *et al.* proposed a brand-new high-level phylogeny of dinosaurs. The authors found Saurischia to be paraphyletic, whereas theropods and Ornithischia were proposed to be sister taxa, and collapsed under the clade Ornithoscelida (a name originally coined by Huxley in 1870 but later superseded by Seeley's

classic partitioning of dinosaurs into a saurischian and an ornithischian clade in 1888). In the Baron *et al.* (2017) tree, basal dinosauroomorpha such as herrerasaurids were placed with Sauropodomorpha in the revisited Saurischia. However, herrerasaurids were later moved outside the Ornithoscelida + Saurischia clade by Baron & Williams (2018). The trees of Baron *et al.* (2017) and Baron & Williams (2018) include *c.* 50 species. While perfectly feasible to present new, high-level phylogenetic hypotheses, the Baron *et al.* (2017) and Baron & Williams (2018) trees do not even approach the known taxonomic and morphological diversity of dinosaurs, so that the impact of the new topologies on trait evolution is unknown. A 441 species-wide informal supertree of dinosaurs was published by Benson *et al.* (2014) and converted into a 326 species tree correlated with body size data in Castiglione *et al.* (2018) after adding information on avian dinosaurs from Lee *et al.* (2014). Here, we use `tree.merger` to translate the Castiglione *et al.* (2018) tree into Baron *et al.*'s (2017) higher-level topology, effectively converting the Baron *et al.* phylogeny into a much larger, informal supertree for which body size data are available.

## MATERIAL AND METHOD

`tree.merger` is meant to combine phylogenetic information derived from different sources into a single informal supertree. It works by adding new lineages to a pre-existing 'backbone' phylogeny and calibrating internal (i.e. nodes) and terminal (i.e. species) ages according to a mixture of user-specified values and backbone tree ages.

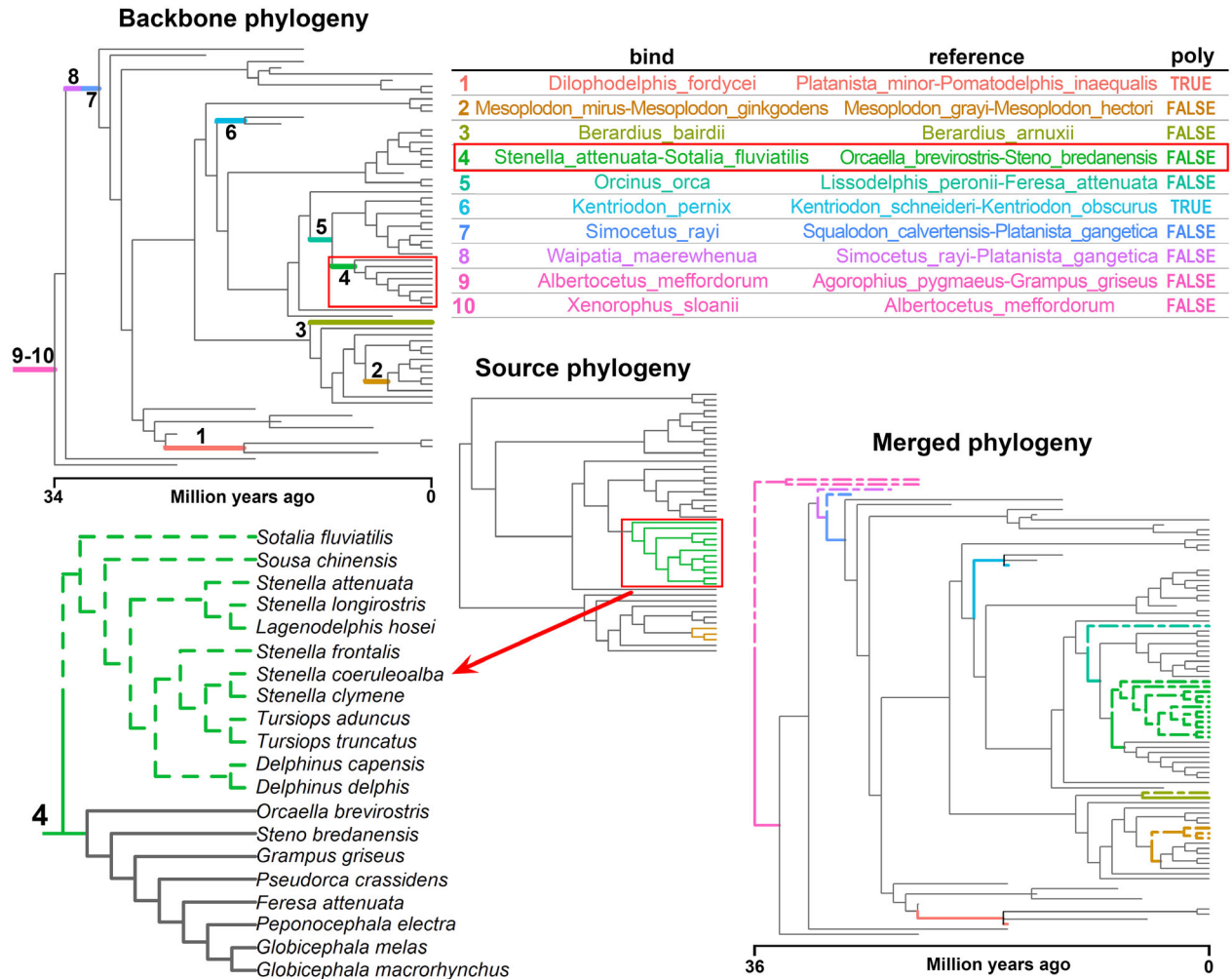
Under `tree.merger`, entire clades can be extracted from a source tree and added to the backbone phylogeny. Alternatively, individual species can be added singularly (whether or not they belong to a source tree), just indicating their name and desired position on the backbone. The information about *what* is attached, meaning whether it is individual species or entire clades, *where* on the backbone phylogeny the tree branches are transferred and *how* they are attached to the backbone (i.e. whether they will form a dichotomous or polytomous clade) are provided in the form of a specification dataset (Fig. 1; Table 1). This dataset and, optionally, vectors of species and node ages are the only objects the user must prepare to perform the tree merging. In the specification dataset, a 'bind' column refers to new species or clades which are taken from the source tree and transferred to the backbone as sister to species/clades indicated in the 'reference' column (Fig. 1; Table 1). The reference clade on the backbone or the binding clade derived from the source phylogeny are identified by indicating a pair of species, such that the most recent common ancestor of the pair includes all the other species within the clade. For instance, given a clade **A** defined by the most recent common ancestor of species **sp1** and **sp2** and all of

its descendants on the source tree, and a clade **B** defined by the most recent common ancestor of species **sp3** and **sp4** and all of its descendants on the backbone tree which will be sister to **A** in the resulting, combined supertree, the user has just to specify 'sp1-sp2' in the bind column, and 'sp3-sp4' in the reference column (Fig. 1). In the case of the addition of individual species, the entry information in the bind column reduces to the species name (e.g. 'sp1'). In the reference column, the user will indicate either a species name on the backbone (which will be sister to the new species), or a pair of species referring to a common ancestor in the reference column if the new species is meant to be sister to a clade. Regardless of whether the new taxa are added as individual species or as a part of a clade, once bound they may serve as reference for further additions, so that the user does not need to add any further statements to make `tree.merger` work iteratively. For instance, suppose the user wish to attach a single species **sp5** to the clade **A+B** produced as above (i.e. by adding 'sp1-sp2' in the bind, and 'sp3-sp4' in the reference column). The user may add a line to the specification dataset reading sp5 in the bind column and 'sp1-sp3' (that is the most recent common ancestor of the **A+B** clade) in the reference column. The user does not even have to specify the sp5/'sp1-sp3' statement before the 'sp1-sp2'/'sp3-sp4' statement in the specification dataset, since `tree.merger` reorders the sequence of species addition automatically.

Since binding information in the specification dataset is conceived as a mean to update the backbone tree, if a species indicated in the specification dataset is also present on the backbone it will be dropped from the backbone and then re-merged according to the specification.

Once all species and clades present in the specification dataset rows are attached, nodes and species ages are calibrated by means of the `RRphylo` function `scaleTree`. Given vectors of species and nodes ages (meant as time distance from the recent) indicated by the user, `scaleTree` rescales branches and leaves of the tree to accommodate the specified ages. If the specified extinction (or origination) age for a given species (clade) predates the age of its direct ancestor, the function automatically moves the ancestors and all the nodes above it back in time, while keeping the tree height fixed. A 'min.branch' argument value is used to set the minimum length of the branches intervening between the shifted nodes. This way the tree is well-conformed, the ancestor-descendants relationships remain unchanged, and possible changes to the original tree topology only pertain to the branch lengths on the path from the root to along the 'calibrated' species.

By default, `tree.merger` assumes the backbone to be correctly calibrated, hence it will retain its nodes and tips ages (including the age of the tree root) if the user does not indicate otherwise. For each clade added to the backbone, the time distances between the most recent common ancestor of the clade and its descendant nodes are



**FIG. 1.** tree.merger machinery. Numbered and coloured branches on the backbone phylogeny (upper left corner) indicate where matching species/clades in the specification dataset (upper right corner) are attached. Binding clades at lines 2 and 4 of the specification dataset are extracted from the source phylogeny (at the centre of the figure). Species and clades in the ‘bind’ column of the specification dataset are attached on the backbone tree as sisters to the species/clades in the reference column. If ‘poly’ is set to TRUE the species/clades deriving from the source tree are bound to the backbone in the form of a polytomy. Enlargement of the binding procedure for clade 4 is plotted at the lower left corner: solid grey branches represent the reference clade on the backbone; dashed green branches represent the ‘bind’ clade dropped from the source phylogeny and attached as sister to the reference. The final merged tree is shown on the lower right corner. Coloured solid branches correspond to the branches on the backbone. Coloured dashed branches represent newly attached species/clades. As two species are added to the backbone tree root, the total height for the merged tree increases. Appendix S1 and Appendix S3 include instructions and data to reproduce the example in the figure.

kept fixed. All the new species, irrespective of whether they are attached as part of a clade or individually, are placed at the maximum distance from the tree root. As for the specification dataset, each node is identified by a pair of species whose most recent common ancestor is the node itself. Also, if the most recent age (i.e. the maximum distance from the tree root) differs between the source and the backbone trees, the difference between them in this exact order (source minus backbone) must be supplied as the ‘age.offset’ argument. This should be positive when the backbone tree attains

younger age than the source tree, and vice-versa. If the source tree is uncalibrated or lacks branch lengths, tree.merger may still be applied to add the source phylogenetic information to the backbone. In this case, the user should just give unit branch lengths to the source, attach the source to the backbone, and then rescale node ages of the resulting tree with scaleTree. In other words, although it is not feasible to use a cladogram tree (i.e. with no branch lengths) as the source tree as it is, an easy and fast workaround allows the use of cladograms anyway.

**TABLE 1.** Specification dataset to perform `tree.merger` on dinosaurs.

Bind	Reference	Poly
<i>Chungkingosaurus_jiangbeiensis-Sauropelta edwardsi</i>	<i>Lesothosaurus_diagnosticus</i>	FALSE
<i>Homalocephale_calathocercos-Hypacrosaurus_altispinus</i>	<i>Hexinlusaurus_multidens</i>	FALSE
<i>Stormbergia_dangershoeki</i>	<i>Hexinlusaurus_multidens-Hypacrosaurus_altispinus</i>	FALSE
<i>Chuandongocoelurus_primitivus-Zuolong_salleei</i>	<i>Cryolophosaurus_elliotti</i>	FALSE
<i>Elaphrosaurus_bambergi-Skorpiovenator_bustingorryi</i>	<i>Cryolophosaurus_elliotti-Zuolong_salleei</i>	FALSE
<i>Coelophysis_rhodesiensis</i>	<i>Coelophysis_bauri</i>	FALSE
<i>Kotasaurus_yamanpalliensis-Saltasaurus_loricatus</i>	<i>Vulcanodon_karibaensis</i>	FALSE
<i>Lessemsaurus_sauropoides</i>	<i>Antetonitrus_ingenipes</i>	FALSE
<i>Anchisaurus_polyzelus</i>	<i>Antetonitrus_ingenipes-Vulcanodon_karibaensis</i>	TRUE
<i>Jingshanosaurus_xinwaensis</i>	<i>Yunnanosaurus_huangi-Antetonitrus_ingenipes</i>	TRUE
<i>Massospondylus_carinatus-Adeopapposaurus_mognai</i>	<i>Lufengosaurus_magnus</i>	FALSE
<i>Sarhsaurus_aurifontanalisis</i>	<i>Lufengosaurus_magnus-Vulcanodon_karibaensis</i>	FALSE
<i>Eucnemesaurus_fortis</i>	<i>Riojasaurus_incertus</i>	FALSE
<i>Ruehleia_bedheimensis</i>	<i>Plateosaurus_engelhardti-Sarhsaurus_aurifontanalisis</i>	FALSE
<i>Plateosaurus_cullingworthi</i>	<i>Ruehleia_bedheimensis-Plateosaurus_engelhardti</i>	FALSE
<i>Chromogisaurus_novasi</i>	<i>Saturnalia_tupiniquim</i>	FALSE

The ‘bind’ column indicates clades or species to bind as sister to corresponding clades or species in the ‘reference’ column. ‘poly’ is used to indicate whether the ‘bind’ and the ‘reference’ taxa will form a polytomous clade on the combined tree. In both ‘bind’ and ‘reference’ columns clades must be identified by indicating a pair of species, such that the most recent common ancestor of the pair includes all the other species within the clade.

Once the merging procedure is complete, the user may choose to scan an interactive plot which allows the visualization one by one of the species/clades specified in the ‘bind’ column in their final position in the combined tree, distinguishing original and new branches of the combined tree according to different colour codes. This option, which only works on RStudio (RStudio Team 2020), allows the user to zoom in and visualize the product of `tree.merger`. Upon specification, `tree.merger` also stores a pdf file showing the final combined tree.

#### Case study 1: *Odontocetes*

We derived the *Odontoceti* tree from the cetacean phylogeny published in Castiglione *et al.* (2020) which is embedded in RRphylo. The tree includes 91 odontocete species, both extinct (23) and extant (68). The clade origin (i.e. the tree root) is set at 36 Ma. Here, we show how such a tree could be assembled by means of `tree.merger`, starting from hypothetical backbone and source trees. The tree supports most of the sister-group relationships within odontoceti and mysticeti, with paraphyletic ‘archaeocetes’ placed basally (Lloyd & Slater 2021).

We reduced the size of the original tree to 68 species by removing the species within the Stenoninae and Delphininae subfamilies, the clade including *Mesoplodon europaeus* + *M. ginkgodens* + *M. mirus*, *Berardius bairdii*, and *Orcinus orca* from the living species. We further dropped *Dilophodelphis fordycei*, *Kentriodon pernix*,

*Simocetus rayi*, *Waipatia maerewhenua*, *Albertocetus meffordorum*, and *Xenorophus sloanii* from the extinct species. By dropping the most basal species within our odontocetes tree (*Albertocetus meffordorum* and *Xenorophus sloanii*), we also reduced the total height of the tree to 34.03 Ma. The resulting phylogeny was used as the backbone tree (Fig. 1).

The source tree was produced by assembling a hypothetical living phylogeny including Delphinoidea, Inoidea, Platanistoidea, Lipotoidea and Ziphioidea. Specifically, Stenoninae, Delphininae and Lissodelphininae were included among Delphinidae, while Orcininae and Globicephalinae were omitted. *Berardius bairdii* and the clade including *Mesoplodon hectori*, *M. stejnegeri*, *M. densirostris*, *M. peruvianus*, *M. perrini* and *M. grayi* were removed from Ziphioidea. The final source tree included 47 living species (Fig. 1).

We assembled the specification dataset so that all species/clades were added to the backbone in the exact same position they were on the original tree (Fig. 1). The Stenoninae + Delphininae clade was attached as sister to Globicephalinae + Orcininae (*partim.*) clade. We added the clade including *Mesoplodon europaeus* + *M. ginkgodens* + *M. mirus* as sister to the clade including *M. hectori* + *M. stejnegeri* + *M. densirostris* + *M. peruvianus* + *M. perrini* + *M. grayi*. *Orcinus orca* was placed as sister to the clade defined by the most recent common ancestor of *Lissodelphis peronii* and *Feresa attenuata*. *Berardius bairdii* was attached as sister to the co-generic *B. arnuxii*. We bound *Kentriodon pernix* to form a polytomous clade with

*K. schneideri* and *K. obscurus*. *Dilophodelphis fordycei* was placed in a polytomous clade with *Pomatodelphis inaequalis* and species within the *Platanista* genus. *Simocetus rayi* and then *Waipatia maerewhenua* were added outside the clade including all superfamilies of living odontocetes + *Squaloziphius emlongi* and *Squalodon calvertensis*. Finally, *Albertocetus mefforum* and *Xenorophus sloanii* were collapsed into a single clade placed as sister to the entire tree.

We supplied the vector of extinction ages for all newly added extinct species. Since the addition of species at the tree root (i.e. *Albertocetus mefforum* + *Xenorophus sloanii*) shifts the origin of the tree back in time, we further provided the vector of node ages setting the tree root at 36 Ma. Both the backbone and the source phylogenies end at the Recent, therefore the ‘age.offset’ was not specified. The case study can be implemented by using the data and code provided as Appendix S3 and Appendix S1, respectively.

#### Case study 2: Dinosaurs

We integrated the dinosaur phylogenetic information from Baron *et al.* (2017), Baron & Williams (2018) and Castiglione *et al.* (2018) into a single informal supertree. Our goal was to widen the phylogeny from Baron *et al.* (2017) with species included in Castiglione *et al.*'s tree, that is in turn derived from trees by Benson *et al.* (2014) and Lee *et al.* (2014). We placed theropods and ornithischians as descending from a single ornithoscelid clade sister to sauropodomorphs (Baron *et al.* 2017), but with herrerasaurids outside the resulting Ornithoscelida + Saurischia clade (Baron & Williams 2018) (Fig. 2).

The phylogeny from Castiglione *et al.* (2018) is embedded in RRphylo. After pruning Pterosauria, the tree retains 326 species: 94 Ornithischia, 162 Theropoda inclusive of early birds, 2 Herrerasauridae, and 68 Sauropodomorpha (Fig. 2). Species extinction ages span from 234.83 Ma to 70 Ma. The tree root is set at 246 Ma, so that the total tree height is 176 myr (246–70 Ma).

We transcribed the phylogenetic tree topology from Baron *et al.* (2017) into Newick format. The tree includes 56 dinosaur species: 14 Ornithischia, 16 Theropoda, 4 Herrerasauridae, and 22 Sauropodomorpha (Fig. 2). The origin of the tree is set at 247 Ma, species extinction ages span from 234.83 Ma to 123.6 Ma. The total tree height is 123.4 myr. Extinction ages were not given in Baron *et al.* (2017) and are therefore derived from Castiglione *et al.* (2018) for the 25 species shared by both trees and downloaded from the Paleobiology Database (<http://paleodb.org>) otherwise.

The phylogeny from Baron *et al.* (2017) was used as the backbone and the tree from Castiglione *et al.* (2018) as the source. The specification dataset was assembled by cross-referencing species between the trees and checking

whether and where taxa appearing solely on the source phylogeny could be added to the backbone (see Appendix S2). We anticipate, though do not claim, that all our phylogenetic choices are correct; our goal here was the production of the informal supertree, not to revise the phylogeny of dinosaurs.

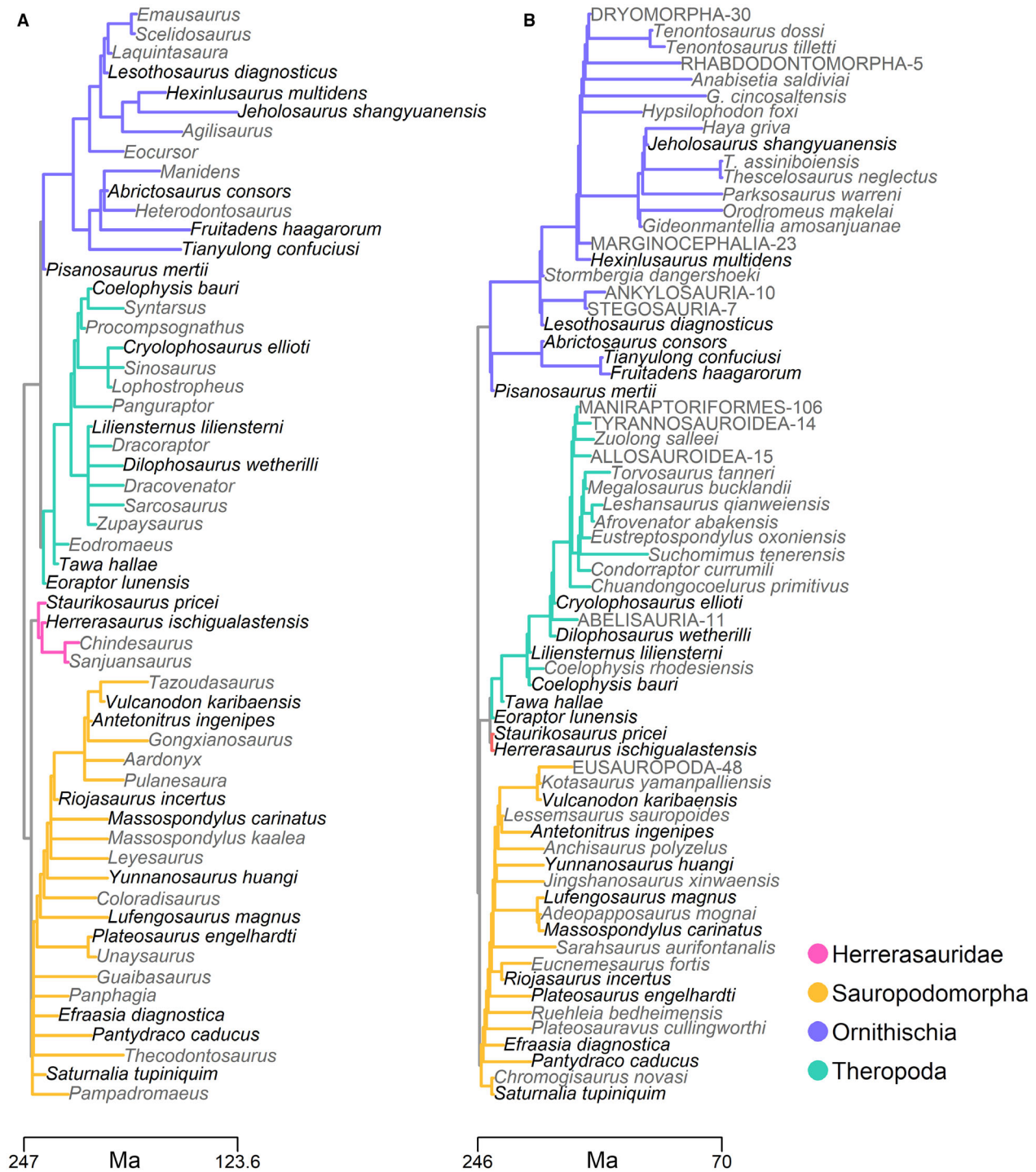
In Ornithischia, we added the clade including Thyreophora as sister to *Lesothosaurus diagnosticus*. Cerapoda were placed sister to *Hexinlusaurus multidentis*, and *Stombergia dangershoeki* as sister to the clade including Cerapoda and *Hexinlusaurus multidentis*. We notice that *Stombergia* and *Lesothosaurus* have been proposed to represent different ontogenetic stages of a single species, meaning that *Lesothosaurus* should be possibly removed from the supertree.

Cerapoda on the source tree include *Jeholosaurus shangyuanensis*, which is sister to *Hexinlusaurus multidentis* on the backbone phylogeny. In this case, the function removes *Jeholosaurus shangyuanensis* from its current position on the backbone tree and attaches it (together with all Cerapoda) in its updated position within Cerapoda (see Appendix S2 for details).

Within Theropoda, the Tetanurae were attached as sister to *Cryolophosaurus ellioti*, which is the only tetanuran shared by the two trees. We bound Abelisauria sister to Tetanurae, and *Coelophysus rhodesiensis* to the co-generic *C. bauri* (see Appendix S2 for details).

Within Sauropodomorpha, we added the clade *Kotasaurus yamanpalliensis* + Eusauropoda as sister to *Vulcanodon karibaensis*. *Lessemsaurus sauropoides* was placed sister to *Antetonitrus ingenipes*, *Anchisaurus polyzelus* in a polytomy with Sauropoda. *Jingshanosaurus xinwaensis* falls in a polytomous clade with other Sauropodiformes. We designed Massopoda other than Sauropoda as follows. The clade *Massospondylus carinatus* + *Adeopapposaurus mognai* was attached as sister to *Lufengosaurus magnus*; *Sarhsaurus aurifontanalis* as sister to the clade including Massospondylidae + Sauropodiformes; *Eucnemesaurus fortis* as sister to *Riojasaurus incertus*. We placed *Ruehleia bedheimensis* sister to Massopoda and then *Plateosaurus cullingworthi* sister to both. *Chromogisaurus novasi* was placed as sister to *Saturnalia tupiniquim* among basal Sauropodomorpha (see Appendix S2 for details).

All of this binding information was collected and properly tabulated into the specification dataset shown in Table 1. We further provided the function with the vector of extinction ages for all the species extracted from the source tree. A second vector includes the ages for the tree root (placed at 247.2 Ma), the node at the origin of Ornithischia (236 Ma), the node at the origin of Theropoda (236 Ma), and the node at the origin of Sauropodomorpha (243 Ma). The zero-reference point for all ages is the age of the youngest species from both trees:



**FIG. 2.** Phylogenetic trees used in this study, from: A, Baron *et al.* (2017); B, Castiglione *et al.* (2018). Some of the main clades on the tree in B are condensed for graphical purposes (the clade numerosity is indicated). Black labels indicate species appearing on both trees. Axis labels indicate age in millions of years.

70 Ma. Additionally, since the trees do not end at the same age (123.6 Ma for the backbone and 70 Ma for the source), we indicated the difference between the youngest age from the source tree and the youngest age from the

backbone (−53.6) as the ‘age.offset’ argument (see Appendix S2 for details).

To move Herrerasauridae from their position as sister to Sauropodomorpha to the outside of the entire



Dinosauria clade (according to Baron & Williams 2018), we performed `tree.merger` a second time. We used the newly generated phylogeny derived by merging trees from Baron *et al.* and Castiglione *et al.* as backbone, and the tree from Baron *et al.* (2017) as source. We assembled a simple one-line specification dataset indicating the clade identified by the most recent common ancestor of *Sanjuansaurus* and *Staurikosaurus\_prici* as ‘bind’, the clade including all other dinosaurs (set to be the most recent common ancestor of *Pampadromaeus* and *Emausaurus*) as ‘reference’, and ‘poly’ to be false. Extinction ages for species within Herrerasauridae were derived from the merged tree itself. This way, herrerasaurids were removed from their position as sister to Sauropodomorpha and correctly placed as sister to Dinosauria, while keeping extinction ages unaltered. Since no age estimate is available for the separation between Dinosauria and Herrerasauria (as named in Baron & Williams 2018), we arbitrarily set the tree root at 248.2 Ma, that is, one million years older than the origin of Dinosauria. As for the first `tree.merger` run, since the combined tree ends at 70 Ma while Baron *et al.*’s (2017) tree ends at 123.6 Ma, we indicated the difference between these ages as ‘age.offset’ argument (source tree minus backbone tree = 53.6).

## RESULTS

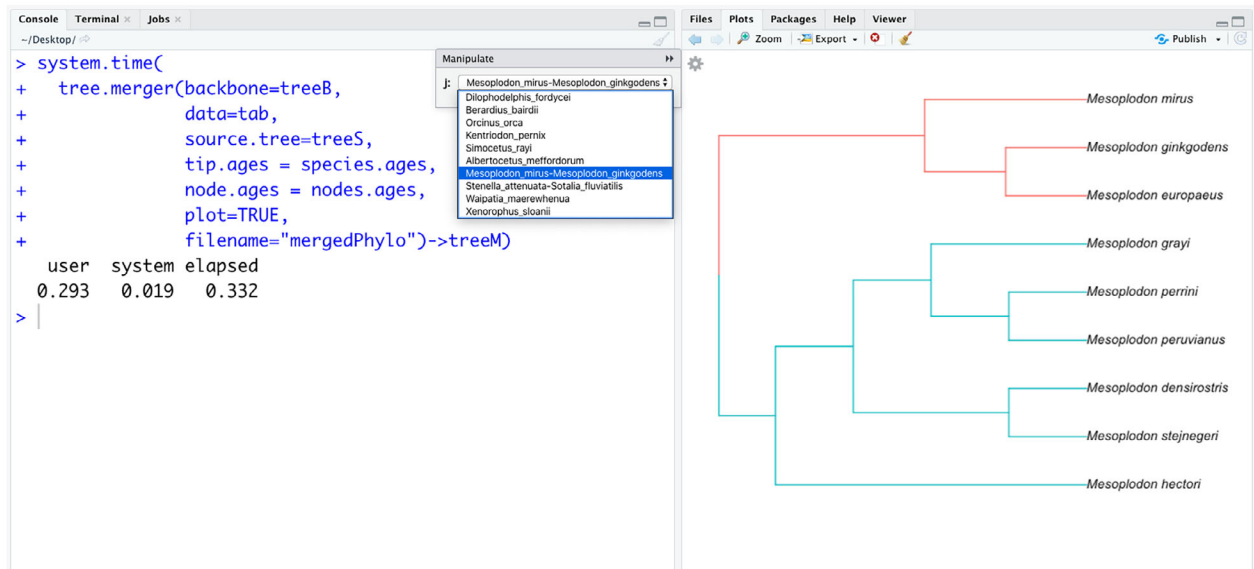
The topology of the odontocete phylogeny assembled by `tree.merger` matches the topology of the starting tree.

There is no difference between imposed and realized species ages. The function required 0.25 s to complete (Fig. 3).

As applied to dinosaurs, `tree.merger` returned a time-calibrated phylogeny including 357 species (Fig. 4). Species and node ages were correctly calibrated according to the provided age vectors. The difference between imposed and realized ages at nodes is 4500 years (0.0045 myr). The same figure for species ages is  $-2.35 \times 10^{-16}$ , on average. The final tree includes: 101 Ornithischia, 172 Theropoda, 4 Herrerasauridae, and 80 Sauropodomorpha species. Performing `tree.merger` twice (i.e. first to merge the phylogenies of Baron *et al.* and Castiglione *et al.*, and then to move herrerasaurids) required 4.16 and 5.15 s, respectively, to complete.

## DISCUSSION

Evolutionary investigations on phenotypic trait change involving the use of phylogenetic comparative methods require using a phylogenetic tree inclusive of the species under scrutiny. In the realm of palaeontology, producing such trees often implies synthesizing phylogenetic information coming from different sources into a single informal supertree (Bininda-Emonds 2004). Such practice is becoming commonplace (Brace *et al.* 2015; Chávez-Hoffmeister 2020; Melchionna *et al.* 2020; Prieto-Márquez *et al.* 2020; Fabbri *et al.* 2021; MacLaren 2021; Medina *et al.* 2021; Varnham *et al.* 2021; Yu *et al.* 2021) which is



**FIG. 3.** Screenshot of RStudio console showing the visual output of `tree.merger` as performed on odontocetes. The drop-down menu allows the user to choose which of the ‘bind’ species/clades to visualize on the plot device. Blue branches refer to species/clades from the backbone phylogeny. Red branches pertain to newly added species/clades.

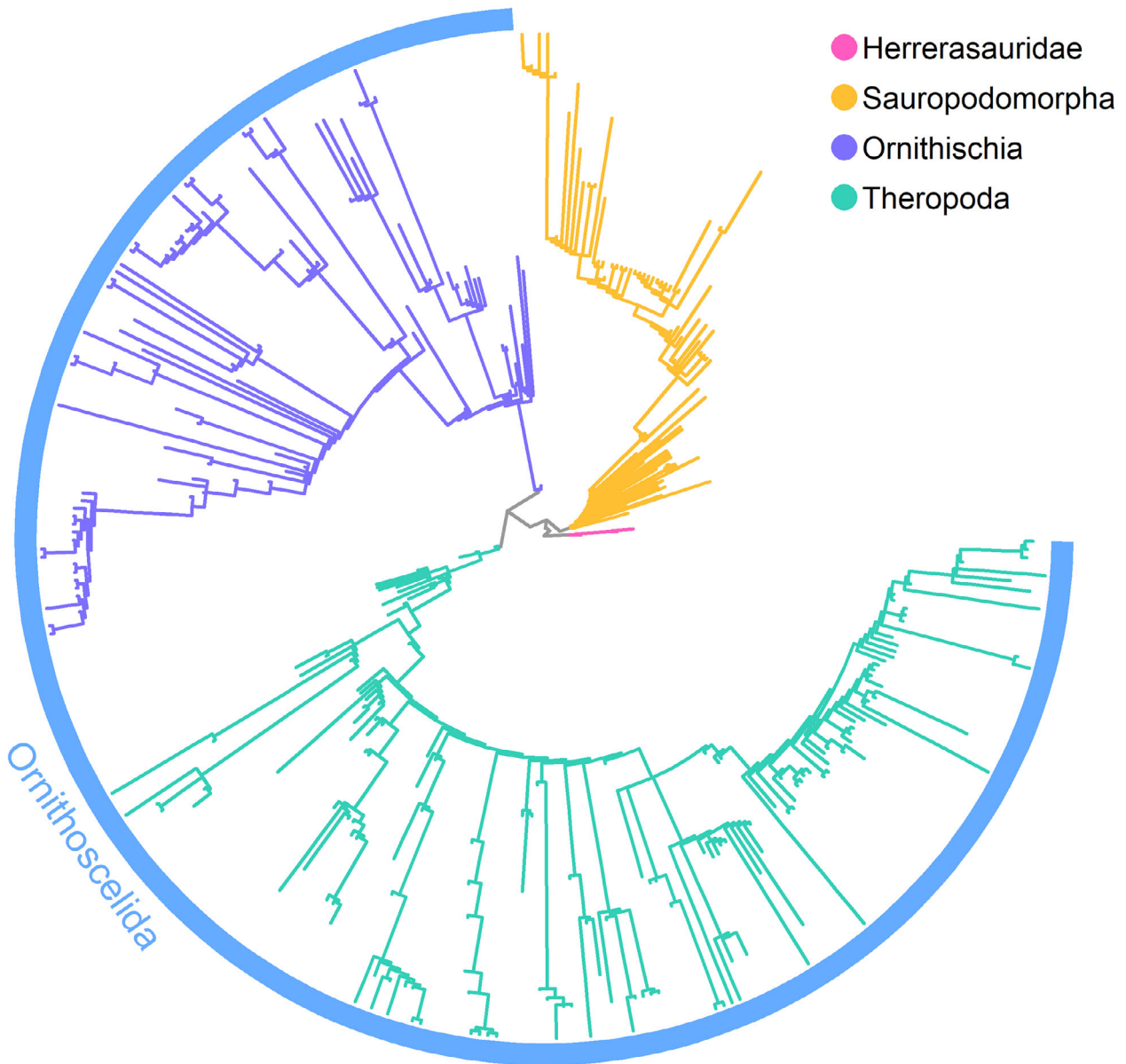


FIG. 4. Phylogenetic tree of dinosaurs resulting from `tree.merger`.

welcome since the fossil record provides better-informed reconstruction of trait evolution as compared to using trees of living species only (Slater *et al.* 2010; Schnitzler *et al.* 2017; Castiglione *et al.* 2020; Koch & Parry 2020; Lloyd & Slater 2021). Unfortunately, the assembly of informal supertrees (Bininda-Emonds 2004) can be a particularly long and time-consuming procedure, especially when it comes to specifying age estimates for the tree tips and internal nodes. Further, the age estimates come with considerable uncertainty (Lloyd *et al.* 2016), and conflicting phylogenetic evidence must be accounted for as in the formal supertree approach (Bininda-Emonds 2004; Lloyd &

Slater 2021), which requires several alternative trees and their attached age/topological uncertainty must be considered. We developed a new R function, named `tree.merger`, that allows us to produce such informal supertrees rapidly, and to calibrate branch lengths according to age estimates indicated by the user. The function also produces an interactive view of its product which allows for real-time, visual checking of the specific branches of the tree which have been modified.

The outcome of `tree.merger` is an informal supertree. As such, it represents a mere phylogenetic hypothesis. The likelihood that such a hypothesis is correct and



the influence of making such an assumption on further hypothesis testing (e.g. on the inference of the tempo and mode of trait evolution; Lloyd *et al.* 2016) should be carefully assessed. In RRphylo, we provide specific tools to fulfil this goal. For instance, in RRphylo we provide the function `swapONE` to produce an alternative phylogeny starting from a given tree by altering its topology and branch lengths randomly swapping the position of a user-specified percentage of the tree species and moving in age a user-specified percentage of nodes between the ages of their parental and daughter nodes. A second function, `fix.poly`, randomly resolves (or creates upon indication) polytomies to non-zero branch lengths, either for the entire tree or at specific locations indicated by the user.

We tested `tree.merger` on two different case studies. The first, purely hypothetical case regards the evolution of odontocete cetaceans. However hypothetical, we observe that the cetacean phylogeny has changed several times during the last two decades, thanks to the recognition of their close relationship to Artiodactyla (Spaulding *et al.* 2009), and especially to the discovery of several new archaeoceti and extinct crown cetaceans, shedding new light on the origin and evolution of the clade (Lambert *et al.* 2010; Fitzgerald 2012; Lambert *et al.* 2017; Fordyce & Marx 2018; Lloyd & Slater 2021) that suggests `tree.merger` could be particularly indicated in this case.

The higher-level phylogeny of dinosaurs has remained stable for some 150 years. Yet, in 2017 the publication of a new phylogenetic arrangement, resurrecting the Ornithoscelida (Theropoda + Ornithischia) clade changed things (Baron *et al.* 2017). Whether or not this hypothesis should be preferred over the classic Saurischia/Ornithischia dichotomy is not at issue here. We translated Baron *et al.*'s 56 species tree into a 357 species tree correlated with body size data (Benson *et al.* 2014; Castiglione *et al.* 2018) amenable to drawing new inferences on body size (or other trait) evolution in the group. Under both circumstances, the computational time ranges from less than 5 s to half a minute, and the difference between age estimates at nodes and tips is negligible. This means that `tree.merger` is at once extremely fast and accurate. We propose that the function will help to expand the use of fossil information in studies of trait evolution and phylogenetic inference, helping to bypass the limitation to this task posed by the often troublesome and time-consuming manipulation of large trees with conventional tools.

## CONCLUSION

Including fossil species in phylogenetic trees and merging different trees into a single supertree increases the power and reliability of phylogenetic analyses of trait evolution.

However, assembling such supertrees could be a particularly long and complex procedure, further frustrated by the need to account for age uncertainty and conflicting phylogenetic views. We produced a new algorithm, embedded in the RRphylo package function `tree.merger` which allows the fast production of time-calibrated supertrees. We tested the function on two different case studies, proving `tree.merger` to be both fast and accurate, thus promising as an aid for expanding the scope of phylogenetic analyses of trait evolution to the fuller inclusion of fossil species.

*Acknowledgements.* We are grateful to Katie E. Davis and an anonymous reviewer for their insightful reviews.

*Author contributions.* SC and PR conceived the method. SC, PR and CS wrote the R function. MM and AM tested the function on real cases. All authors contributed to writing the paper and preparing supporting information files.

## DATA ARCHIVING STATEMENT

This is Paleobiology Database official publication number 419.

*Editor.* Robert Sansom

## SUPPORTING INFORMATION

Additional Supporting Information can be found online (<https://doi.org/10.1111/pala.12588>):

**Appendix S1.** html vignette to reproduce Case Study 1: Odontocetes.

**Appendix S2.** html vignette to reproduce Case Study 2: Dinosaurs.

**Appendix S3.** R workspace including data to reproduce the cetacean case study.

**Appendix S4.** Phylogenetic tree from Baron *et al.* (2017) in Newick format.

## REFERENCES

- AZE, T., EZARD, T. H. G., PURVIS, A., COXALL, H. K., STEWART, D. R. M., WADE, B. S. and PEARSON, P. N. 2011. A phylogeny of Cenozoic macroperforate planktonic foraminifera from fossil data. *Biological Reviews*, **86**, 900–927.
- BARON, M. G. and WILLIAMS, M. E. 2018. A re-evaluation of the enigmatic dinosauriform *Caseosaurus Crosbyensis* from the Late Triassic of Texas, USA and its implications for early dinosaur evolution. *Acta Palaeontologica Polonica*, **63**, 129–145.
- BARON, M. G., NORMAN, D. B. and BARRETT, P. M. 2017. A new hypothesis of dinosaur relationships and early dinosaur evolution. *Nature*, **543**, 501–506.

- BENSON, R. B. J., CAMPIONE, N. E., CARRANO, M. T., MANNION, P. D., SULLIVAN, C., UPCHURCH, P. and EVANS, D. C. 2014. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biology*, **12**, e1001853.
- BININDA-EMONDS, O. R. P. 2004. The evolution of supertrees. *Trends in Ecology & Evolution*, **19**, 315–322.
- BITTENCOURT, J. S., ARCUCCI, A. B., MARSICANO, C. A. and LANGER, M. C. 2014. Osteology of the Middle Triassic archosaur *Lewisuchus admixtus* Romer (Chañares Formation, Argentina), its inclusivity, and relationships amongst early dinosauromorphs. *Journal of Systematic Palaeontology*, **13**, 189–219.
- BRACE, S., TURVEY, S. T., WEKSLER, M., HOOGLAND, M. L. P. and BARNES, I. 2015. Unexpected evolutionary diversity in a recently extinct Caribbean mammal radiation. *Proceedings of the Royal Society B*, **282**, 20142371.
- CASCINI, M., MITCHELL, K. J., COOPER, A. and PHILLIPS, M. J. 2019. Reconstructing the evolution of giant extinct kangaroos: comparing the utility of DNA, morphology, and total evidence. *Systematic Biology*, **68**, 520–537.
- CASTIGLIONE, S., TESONE, G., PICCOLO, M., MELCHIONNA, M., MONDANARO, A., SERIO, C., DI FEBBRARO, M. and RAI, P. 2018. A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology & Evolution*, **9**, 974–983.
- CASTIGLIONE, S., SERIO, C., MONDANARO, A., MELCHIONNA, M., CAROTENUTO, F., DI FEBBRARO, M., PROFICO, A., TAMAGNINI, D. and RAI, P. 2020. Ancestral state estimation with phylogenetic ridge regression. *Evolutionary Biology*, **47**, 220–232.
- CHÁVEZ-HOFFMEISTER, M. 2020. Bill disparity and feeding strategies among fossil and modern penguins. *Paleobiology*, **46**, 176–192.
- CHIRA, A. M. and THOMAS, G. H. 2016. The impact of rate heterogeneity on inference of phylogenetic models of trait evolution. *Journal of Evolutionary Biology*, **29**, 2502–2518.
- FABBRI, M., NAVALÓN, G., KOCH, N. M., HANSON, M., PETERMANN, H. and BHULLAR, B.-A. 2021. A shift in ontogenetic timing produced the unique sauropod skull. *Evolution*, **75**, 819–831.
- FITZGERALD, E. M. G. 2012. Archaeocete-like jaws in a baleen whale. *Biology Letters*, **8**, 94–96.
- FORDYCE, R. E. and MARX, F. G. 2018. Gigantism precedes filter feeding in baleen whale evolution. *Current Biology*, **28**, 1670–1676.e2.
- HUXLEY, T. H. 1870. On the classification of the Dinosauria, with observations on the Dinosauria of the Trias. *Quarterly Journal of the Geological Society*, **26**, 32–51.
- KOCH, N. M. and PARRY, L. A. 2020. Death is on our side: paleontological data drastically modify phylogenetic hypotheses. *Systematic Biology*, **69**, 1052–1067.
- LAMBERT, O., BIANUCCI, G., POST, K., DE MUIZON, C., SALAS-GISMONDI, R., URBINA, M. and REUMER, J. 2010. The giant bite of a new raptorial sperm whale from the Miocene epoch of Peru. *Nature*, **466**, 105–108.
- LAMBERT, O., MARTÍNEZ-CÁCERES, M., BIANUCCI, G., DI CELMA, C., SALAS-GISMONDI, R., STEURBAUT, E., URBINA, M. and DE MUIZON, C. 2017. Earliest Mysticete from the late Eocene of Peru sheds new light on the origin of baleen whales. *Current Biology*, **27**, 1535–1541.e2.
- LEE, M. S. Y., CAU, A., NAISH, D. and DYKE, G. J. 2014. Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science*, **345**, 562–566.
- LLOYD, G. T. and SLATER, G. J. 2021. A total-group phylogenetic metatree for Cetacea and the importance of fossil data in diversification analyses. *Systematic Biology*, **70**, 922–939.
- LLOYD, G. T., BAPST, D. W., FRIEDMAN, M. and DAVIS, K. E. 2016. Probabilistic divergence time estimation without branch lengths: dating the origins of dinosaurs, avian flight and crown birds. *Biology Letters*, **12**, 20160609.
- MACLAREN, J. A. 2021. Biogeography a key influence on distal forelimb variation in horses through the Cenozoic. *Proceedings of the Royal Society B*, **288**, 20202465.
- MADDISON, W. P. and MADDISON, D. R. 2019. Mesquite: a modular system for evolutionary analysis. Version 3.61. <http://www.mesquiteproject.org/>
- MEDINA, M. C., SOUSA-BAENA, M. S., PRADO, E., ACEVEDO-RODRÍGUEZ, P., DIAS, P. and DEMARCO, D. 2021. Laticifers in Sapindaceae: structure, evolution and phylogenetic importance. *Frontiers in Plant Science*, **11**, 2247.
- MELCHIONNA, M., MONDANARO, A., SERIO, C., CASTIGLIONE, S., DI FEBBRARO, M., ROOK, L., DINIZ-FILHO, J. A. F., MANZI, G., PROFICO, A., SANSALONE, G. and RAI, P. 2020. Macroevolutionary trends of brain mass in Primates. *Biological Journal of the Linnean Society*, **129**, 14–25.
- PRIETO-MÁRQUEZ, A., GARCIA-PORTA, J., JOSHI, S. H., NORELL, M. A. and MAKOVICKY, P. J. 2020. Modularity and heterochrony in the evolution of the ceratopsian dinosaur frill. *Ecology & Evolution*, **10**, 6288–6309.
- RSTUDIO TEAM 2020. *RStudio: integrated development for R*. RStudio Inc, Boston, MA. <http://www.rstudio.com/>
- SCHNITZLER, J., THEIS, C., POLLY, P. D. and ERO-NEN, J. T. 2017. Fossils matter - understanding modes and rates of trait evolution in Musteloidea (Carnivora). *Evolutionary Ecology Research*, **18**, 187–200.
- SEELEY, H. G. 1888. I. On the classification of the fossil animals commonly named Dinosauria. *Proceedings of the Royal Society of London*, **43**, 258–265.
- SLATER, G. J. and HARMON, L. J. 2013. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods in Ecology & Evolution*, **4**, 699–702.
- 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 B*, **277**, 3097–3104.
- SPAULDING, M., O'LEARY, M. A. and GATESY, J. 2009. Relationships of Cetacea (Artiodactyla) among mammals: increased taxon sampling alters interpretations of key fossils and character evolution. *PLoS One*, **4**, e7062.

- STÖVER, B. C. and MÜLLER, K. F. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, **11**, 7.
- VARNHAM, G. L., MANNION, P. D. and KAMMERER, C. F. 2021. Spatiotemporal variation in completeness of the early cynodont fossil record and its implications for mammalian evolutionary history. *Palaeontology*, **64**, 307–333.
- YU, Y., ZHANG, C. and XU, X. 2021. Deep time diversity and the early radiations of birds. *Proceedings of the National Academy of Sciences*, **118**, e2019865118.