LETTER

https://doi.org/10.1038/s41586-018-0093-3

The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps

Tiago R. Simões[1]*, Michael W. Caldwell[1,2], Mateusz Tałanda[3], Massimo Bernardi[4,5], Alessandro Palci[6], Oksana Vernygora[1], Federico Bernardini[7,8], Lucia Mancini[9] & Randall L. Nydam[10]

Modern squamates (lizards, snakes and amphisbaenians) are the world's most diverse group of tetrapods along with birds[1] and have a long evolutionary history, with the oldest known fossils dating from the Middle Jurassic period—168 million years ago[2–4]. The evolutionary origin of squamates is contentious because of several issues: (1) a fossil gap of approximately 70 million years exists between the oldest known fossils and their estimated origin[5–7]; (2) limited sampling of squamates in reptile phylogenies; and (3) conflicts between morphological and molecular hypotheses regarding the origin of crown squamates[6,8,9]. Here we shed light on these problems by using high-resolution microfocus X-ray computed tomography data from the articulated fossil reptile *Megachirella wachtleri* (Middle Triassic period, Italian Alps[10]). We also present a phylogenetic dataset, combining fossils and extant taxa, and morphological and molecular data. We analysed this dataset under different optimality criteria to assess diapsid reptile relationships and the origins of squamates. Our results re-shape the diapsid phylogeny and present evidence that *M. wachtleri* is the oldest known stem squamate. *Megachirella* is 75 million years older than the previously known oldest squamate fossils, partially filling the fossil gap in the origin of lizards, and indicates a more gradual acquisition of squamatan features in diapsid evolution than previously thought. For the first time, to our knowledge, morphological and molecular data are in agreement regarding early squamate evolution, with geckoes—and not iguanians—as the earliest crown clade squamates. Divergence time estimates using relaxed combined morphological and molecular clocks show that lepidosaurs and most other diapsids originated before the Permian/Triassic extinction event, indicating that the Triassic was a period of radiation, not origin, for several diapsid lineages.

Megachirella preserves traits that indicate that it is a lepidosaurian reptile, such as the presence of a well-developed quadrate conch, an ectepicondylar foramen in the humerus and pleurodont dentition. Some of these features led previous authors to recognize the lepidosauromorph affinities of *Megachirella*, which was previously considered as a non-squamate lepidosauromorph, although no definitive conclusions on its phylogenetic placement had ever been reached[5,10,11]. Yet, the unique condition of *Megachirella* as one of the very few articulated and well-preserved Triassic lepidosauromorphs hints at its potential to help resolve important aspects of lepidosaur evolution. Here we provide substantial new information on *Megachirella*, based on personal observations and high-resolution microfocus X-ray computed tomography (micro-CT) scans, which reveal several previously unnoticed features in *Megachirella* (Fig. 1, Extended Data Figs. 1, 2 and Supplementary Discussion). Results from the micro-CT scans include a combination of features that are found uniquely in squamates: a triradiate squamosal (not tetraradiate as in most other diapsids, including rhynchocephalians); the squamosal lacks an anteriorly concave articulatory facet for the postorbital; a well-developed alar process of the prootic; a well-developed radial condyle on the humerus; an ulnar patella; a secondary curvature of the clavicles; and an expanded epiphysis of the first metacarpal along with the absence of the first distal carpal (suggesting its fusion with the first metacarpal, as observed in modern squamates[12]). Finally, the micro-CT scans indicate that *Megachirella* has features that are absent in all rhynchocephalians (the sister lineage to squamates), including the earliest forms such as *Gephyrosaurus*: the presence of a splenial; the ectopterygoids are directed anteriorly (not laterally as in rhynchocephalians); the presacral pleurocentra lack a notochordal canal; and dorsal (coronoid) expansion of the surangular and dentary bones is absent. The new information presented here, along with our extensive revision of diapsid and early squamate phylogeny, unambiguously resolves the placement of *Megachirella* as the oldest known squamate. As expected for a squamate that is 85 million years (Myr) older than the oldest previously known articulated squamates for which the osteology is well known—*Eichstaettisaurus* and *Ardeosaurus* from the Late Jurassic of Germany[8,13]—*Megachirella* retains numerous plesiomorphic features. These features are observed in other diapsid reptiles, and some are retained in rhynchocephalians, but they are almost entirely lost in crown squamates. These include amphicoelic vertebrae (although present in geckoes and *Huehuecuetzpalli*), a small quadratojugal, gastralia and an entepicondylar foramen in the humerus.

Assessing the phylogenetic position of *Megachirella* and other lepidosauromorph reptiles is challenging because there has never been a phylogenetic dataset comprising a rich sampling of both non-lepidosaurian diapsid reptiles and squamates. Almost invariably, broad-scale reptile phylogenies have represented the nearly 10,000 extant species and the hundreds of fossil species of squamates as a single operational taxonomic unit[14–16] (for more examples, see Supplementary Methods). This approach oversimplifies the enormous diversity of phenotypes and genotypes in squamates. Conversely, studies focused on squamate phylogeny never include more than a few taxa outside the Squamata to serve as outgroups[9,17]. Here we create the first morphological phylogenetic dataset comprising all the main branches of the diapsid tree of life, including extant taxa and fossils from all major lineages of rhynchocephalians (for example, tuataras) and squamates at the species level (Supplementary Data 1–5). We also focused on primary data collection, personally observing numerous specimens covering 100% of the taxa included in this dataset. We performed a meticulous revision of reptile and squamate phylogenetic characters (and created new characters) to avoid issues caused by logical or biological biases in morphological characters[18]. Owing to the rich sampling of extant squamate species, we also included molecular data from 16 loci (13 nuclear and 3 mitochondrial). The analyses performed include morphological and combined evidence (morphological and molecular data) analyses of diapsid and lepidosaurian relationships, carried out under multiple phylogenetic inference methods (see Methods).

[1]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada. [2]Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, Alberta, Canada. [3]Department of Palaeobiology and Evolution, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland. [4]MUSE, Museo delle Scienze di Trento, Trento, Italy. [5]School of Earth Sciences, University of Bristol, Bristol, UK. [6]College of Science and Engineering, Flinders University, Adelaide, South Australia, Australia. [7]Museo Storico della Fisica e Centro di Studi e Ricerche Enrico Fermi, Roma, Italy. [8]The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy. [9]Elettra Sincrotrone Trieste S.C.p.A., Trieste, Italy. [10]Department of Anatomy, Arizona College of Osteopathic Medicine, Midwestern University, Glendale, AZ, USA. *e-mail: tsimoes@ualberta.ca

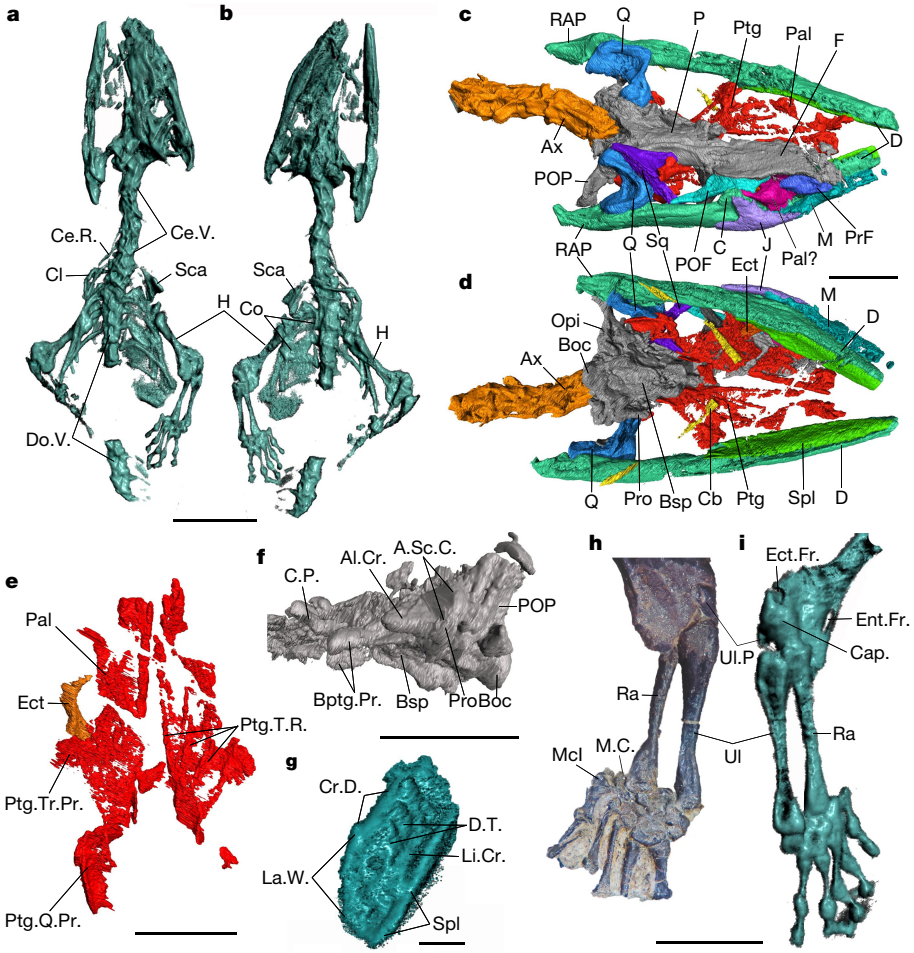


Fig. 1 | Holotype of *M. wachtleri* (PZO 628). a, **b**, Whole skeleton dorsal and ventral views. **c**, **d**, Skull in dorsal (**c**) and ventral (**d**) views. **e**, Palatal region in ventral view. **f**, Braincase in left lateral view. **g**, Dentary in cross-section. **h**, **i**, Right forelimb in dorsal (**h**) and ventral (**i**) views. Abbreviations: Al.Cr., prootic alar crest; A.Sc.C., anterior semicircular canal; Ax, axis; Boc, basioccipital; Bptg.Pr., basipterygoid process; Bsp, basisphenoid; C, coronoid; Cap, capitulum; Cb, ceratobranchial; Ce.R., cervical rib; Ce.V., cervical vertebrae; Cl, clavicle; Co, coracoid; C.P., cultriform process; Cr.D., crista dorsalis; D, dentary; Do.V., dorsal vertebrae; D.T., dentary teeth; Ect, ectopterygoid; Ect.Fr., ectepicondylar foramen; Ent.Fr., entepicondylar foramen; F, frontal; H, humerus; J, jugal; La.W., labial wall; Li.Cr., lingual crest; M, maxilla; M.C., medial centrale; McI, metacarpal I; P, parietal; Opi, opisthotics; Pal, palatine; POF, postorbitofrontal; POP, paraoccipital process; PrF; prefrontal; Pro, prootic; Ptg, pterygoid; Ptg.Q.Pr, pterygoid quadrate process; Ptg.T.R., pterygoid tooth rows; Ptg.Tr.Pr., pterygoid transverse process; Q, quadrate; Ra, radius; RAP, retroarticular process; Sca, scapula; Spl, splenial; Sq, squamosal; Ul, ulna; Ul.P., ulnar patella. Scale bars, 10 mm (**a**, **b**), 5 mm (**c**–**f**, **h**, **i**) and 1 mm (**g**).

Despite the difference in the datasets used (that is, morphology versus combined evidence) and phylogenetic optimality criteria, all results converge on *Megachirella* representing a stem squamate along with *Marmoretta oxoniensis*, from the Middle Jurassic of Britain, and *Huehuecuetzpalli mixtecus*, from the Early Cretaceous period of Mexico (Fig. 2 and Extended Data Figs. 3–8). This resolution is particularly well supported in the combined evidence analysis, in which *Megachirella* has a leaf stability above the overall mean (Extended Data Fig. 9). In analyses with maximum parsimony, *Sophineta cracoviensis* also falls within the Squamata stem, but this is not recovered in the remaining analyses. This indicates that some taxa previously proposed to be early-evolving lepidosauromorphs (for example, *Megachirella* and *Marmoretta*)[5,10,11] actually represent the oldest known squamates, partially filling the supposed 70-Myr fossil gap in the early history of the clade. Other taxa also considered to be early lepidosauromorphs by previous studies (for example, kuehneosaurids and *Saurosternon*[5]) are consistently found in our results to be nested in other parts of the diapsid tree outside the Lepidosauromorpha. Additionally, all previous morphology- and molecular-based squamate phylogenies available in the literature disagree with each other concerning the earliest-evolving crown group squamates: iguanians for morphology-based analyses[17,19], but dibamids and gekkotans for molecular analyses[7,20,21] (see also Supplementary Methods). The results of the combined evidence analyses typically match those of the molecular data alone[6,9]; however, our results show unprecedented agreement between morphological and molecular data, in placing geckoes instead of iguanians among the earliest-evolving squamates. Iguanians are consistently found further crownward in the tree, nested either with anguimorphs and snakes (clade Toxicofera, Extended Data Figs. 3, 5–8), or with teiioids (Extended Data Fig. 4). This unprecedented agreement between molecular and morphological data with regards to the early evolution of squamates might be a consequence of our broad sampling of taxa outside squamates (thus affecting character polarity and branch length parameters) and strict criteria for morphological dataset construction.

Megachirella provides unique insights into the early acquisition of squamatan features, as it is the first unequivocal squamate from the Triassic. *Megachirella*, and also *Huehuecuetzpalli*[22], show that features that are commonly attributed to squamates characterize crown squamates, but were not yet present in stem squamates. For instance, *Megachirella* and *Huehuecuetzpalli* still retain amphicoelic vertebrae, an entepicondylar foramen, and lack a ball-like distal epiphysis of the ulna. *Megachirella* further indicates that the loss of the quadratojugal and gastralia occurred within squamates, and not at the point of divergence from rhynchocephalians. The same pattern occurs in

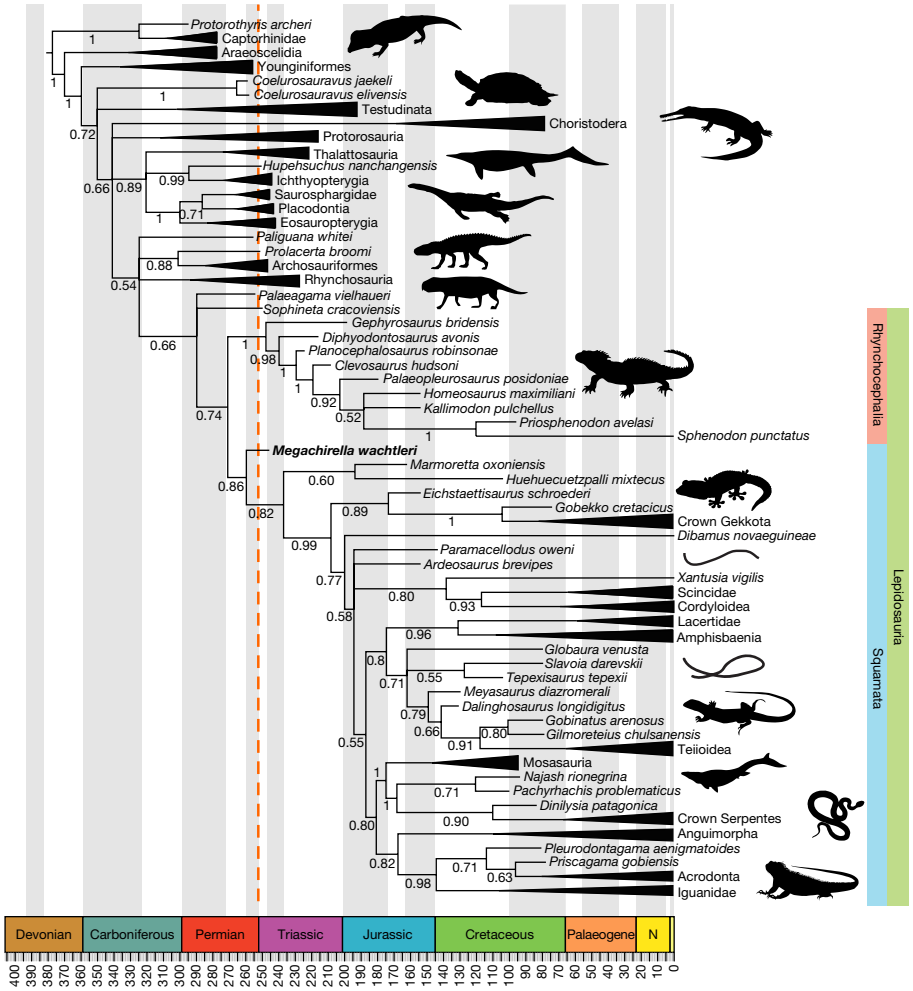


Fig. 2 | Combined evidence relaxed-clock Bayesian inference analysis with total evidence tip and node dating using the fossilized birth–death tree model. Summary of the majority rule consensus tree depicting the median divergence time estimates for the major diapsid and squamate lineages against a geological time scale. Numbers at nodes indicate posterior probabilities and the orange dashed line represents the Permian/Triassic mass extinction event. For the full tree and 95% highest posterior density on divergence times see Extended Data Fig. 8. N, Neogene period.

rhynchocephalians, for which Triassic and Early Jurassic fossils were previously known[23], and which retain plesiomorphic features (such as the pleurodont dentition) that are absent in most of the later members of that group.

Previous molecular-clock estimates have placed the squamate crown divergence time between the Late Triassic and Early Jurassic[6,7,24], and lepidosaurs originating at some point in the Triassic[5,6] or the Middle Permian period[7,25]. Our time-calibrated Bayesian inference analyses combine information from both the molecular and morphological relaxed-clocks on lepidosaurs and other diapsid lineages (Fig. 2 and Extended Data Fig. 8), providing a more holistic approach to the divergence time of squamates, lepidosaurs and other diapsids. Our estimates indicate lepidosaurs originated 269 Myr ago (median estimate) in the Middle Permian, and crown squamates 206 Myr ago in the Late Triassic (thus agreeing with recent phylogenomic analyses[7]). Furthermore, our morphological sampling allows a more precise estimate of the origin of the squamate root by the inclusion of fossils now recognized as stem squamates, and thus the age of origin of all squamates can be set at 257 Myr ago, close to the Permian/Triassic mass extinction (PTME).

Some of the oldest known fossils for certain diapsid lineages are known from the earliest Triassic, including ichthyosaurs[16], sauropterygians[26] and archosaurs[27], with more recent fossil evidence already suggesting the presence of archosauriforms in the Late Permian[28], strongly suggesting their divergence preceded the PTME. In accordance, our divergence time estimates for almost all major diapsid lineages (such as lepidosaurs, archosauriforms and marine reptiles) are in the Permian (Fig. 2 and Extended Data Fig. 8) and not the Triassic (the period from which their oldest known fossils are known). This corresponds to the general expectation that the oldest known fossil of a lineage is likely to be much younger than the actual divergence time for that same lineage[29].

The origin of lepidosaurs and other major diapsid lineages before the PTME contradicts previous ideas suggesting that those groups originated in the aftermath of the greatest mass extinction in Earth's history[30]. Instead, our results indicate those lineages already existed, but radiated in the Triassic. It is likely that the PTME opened new niches and opportunities to lineages previously restricted in diversity, thus enabling their radiation in the Triassic into numerous forms and sizes, occupying all major biomes on the planet.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0093-3

Received: 5 December 2017; Accepted: 28 February 2018;
Published online 30 May 2018.

1. Uetz, P. & Hošek, J. *The Reptile Database* http://www.reptile-database.org (2017).
2. Nessov, L. Late Mesozoic amphibians and lizards of Soviet Middle Asia. *Acta Zool. Cracov.* **31**, 475–486 (1988).

3. Fedorov, P. & Nessov, L. A lizard from the boundary of the Middle and Late Jurassic of north-east Fergana. *Bull. St Petersburg Univ. Geol. Geog.* **3**, 9–14 (1992).
4. Evans, S. Crown group lizards (Reptilia, Squamata) from the Middle Jurassic of Britain. *Palaeontographica. A* **250**, 123–154 (1998).
5. Jones, M. E. H. et al. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol. Biol.* **13**, 208 (2013).
6. Pyron, R. A. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Syst. Biol.* **66**, 38–56 (2017).
7. Irisarri, I. et al. Phylotranscriptomic consolidation of the jawed vertebrate timetree. *Nat. Ecol. Evol.* **1**, 1370–1378 (2017).
8. Simões, T. R., Caldwell, M. W., Nydam, R. L. & Jiménez-Huidobro, P. Osteology, phylogeny, and functional morphology of two Jurassic lizard species and the early evolution of scansoriality in geckoes. *Zool. J. Linn. Soc.* **180**, 216–241 (2017).
9. Reeder, T. W. et al. Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PLoS ONE* **10**, e0118199 (2015).
10. Renesto, S. & Posenato, R. A new lepidosauromorph reptile from the Middle Triassic of the Dolomites (Northern Italy). *Riv. Ital. Paleontol. Stratigr.* **109**, 463–474 (2003).
11. Renesto, S. & Bernardi, M. Redescription and phylogenetic relationships of *Megachirella wachtleri* Renesto et Posenato, 2003 (Reptilia, Diapsida). *Palaontol. Z.* **88**, 197–210 (2014).
12. Carroll, R. L. in *Problems in Vertebrate Evolution* Vol. 4 (eds Andrews, S. M. et al.) 1–28 (Academic Press, London, 1977).
13. Mateer, N. Osteology of the Jurassic lizard *Ardeosaurus brevipes* (Meyer). *Palaeontology* **25**, 461–469 (1982).
14. Chen, X. H., Motani, R., Cheng, L., Jiang, D.-Y. & Rieppel, O. The enigmatic marine reptile *Nanchangosaurus* from the lower triassic of Hubei, China and the phylogenetic affinities of Hupehsuchia. *PLoS ONE* **9**, e102361 (2014).
15. Müller, J. in *Recent Advances in the Origin and Early Radiation of Vertebrates* (eds Arratia, G. et al.) 379–408 (F. Pfeil, Münich, 2004).
16. Motani, R., Minoura, N. & Ando, T. Ichthyosaurian relationships illuminated by new primitive skeletons from Japan. *Nature* **393**, 255–257 (1998).
17. Conrad, J. L. Phylogeny and systematics of Squamata (Reptilia) based on morphology. *Bull. Am. Mus. Nat. Hist.* **310**, 1–182 (2008).
18. Simões, T. R., Caldwell, M. W., Palci, A. & Nydam, R. L. Giant taxon-character matrices: quality of character constructions remains critical regardless of size. *Cladistics* **33**, 198–219 (2017).
19. Gauthier, J. A., Kearney, M., Maisano, J. A., Rieppel, O. & Behlke, A. D. B. Assembling the squamate tree of life: perspectives from the phenotype and the fossil record. *Bull. Peabody Mus. Nat. Hist.* **53**, 3–308 (2012).
20. Pyron, R. A., Burbrink, F. T. & Wiens, J. J. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* **13**, 93 (2013).
21. Vidal, N. & Hedges, S. B. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *C. R. Biol.* **328**, 1000–1008 (2005).
22. Reynoso, V.-H. *Huehuecuetzpalli mixtecus* gen. et sp. nov: a basal squamate (Reptilia) from the Early Cretaceous of Tepexi de Rodríguez, Central México. *Phil. Trans. R. Soc. Lond. B* **353**, 477–500 (1998).
23. Evans, S. E. The skull of a new eosuchian reptile from the Lower Jurassic of South Wales. *Zool. J. Linn. Soc.* **70**, 203–264 (1980).
24. Zheng, Y. & Wiens, J. J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol. Phylogenet. Evol.* **94**, 537–547 (2016).
25. Hugall, A. F., Foster, R. & Lee, M. S. Y. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. *Syst. Biol.* **56**, 543–563 (2007).
26. Jiang, D.-Y. et al. The Early Triassic Eosauropterygian *Majiashanosaurus discocoracoidis*, gen. et sp. nov. (Reptilia, Sauropterygia), from Chaohu, Anhui Province, People's Republic of China. *J. Vertebr. Paleontol.* **34**, 1044–1052 (2014).
27. Butler, R. J. et al. The sail-backed reptile *Ctenosauriscus* from the latest Early Triassic of Germany and the timing and biogeography of the early archosaur radiation. *PLoS ONE* **6**, e25693 (2011).
28. Bernardi, M., Klein, H., Petti, F. M. & Ezcurra, M. D. The origin and early radiation of archosauriforms: integrating the skeletal and footprint record. *PLoS ONE* **10**, e0128449 (2015).
29. Ho, S. Y. & Phillips, M. J. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst. Biol.* **58**, 367–380 (2009).
30. Chen, Z.-Q. & Benton, M. J. The timing and pattern of biotic recovery following the end-Permian mass extinction. *Nat. Geosci.* **5**, 375–383 (2012).

Acknowledgements We are grateful for funding from the Vanier Canada and the Izaak Walton Killam Memorial PhD scholarships to T.R.S.; Euregio Science Fund (call 2014, IPN16) to M.B.; Midwestern University Intramural Funds to R.L.N.; Natural Science and Engineering Research Council of Canada Discovery Grant to M.W.C. (23458); Alberta Ukrainian Centennial Scholarship to O.V.; and National Science Centre grant 2014/13/N/NZ8/02467 to M.T.; and E. Kustatscher for access to the holotype of *M. wachtleri*.

Reviewer information *Nature* thanks M. Baron, J.-C. Rage and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.R.S. conducted phylogenetic data collection and analyses; M.W.C. conceived the project; F.B., L.M., M.B., T.R.S. and A.P. conducted micro-CT scans and computed tomography segmentations; T.R.S. and O.V. performed molecular sequence alignment; T.R.S., M.B., M.T., A.P. and R.L.N. performed morphological description; all authors contributed to writing and discussions.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0093-3.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0093-3.

Correspondence and requests for materials should be addressed to T.R.S.

METHODS

Micro-CT. The holotype of *Megachirella wachtleri* was analysed by micro-CT at the Multidisciplinary Laboratory of the Abdus Salam International Centre of Theoretical Physics (Trieste, Italy), using a system specifically designed in collaboration with Elettra-Sincrotrone Trieste (Basovizza, Italy) for the study of palaeontological and archaeological materials[31]. The micro-CT acquisition of the complete specimen was carried out by using a sealed X-ray source (Hamamatsu L8121-03) at a voltage of 150 kV, a current of 100 μA and with a focal spot size of 20 μm. The X-ray beam was filtered by a 1.5-mm-thick aluminium absorber. A set of 2,400 projections of the sample were recorded over a total scan angle of 360° by a flat panel detector (Hamamatsu C7942SK-25) with an exposure time of 2.0 s. The resulting micro-CT slices were reconstructed in 16-bit format using the commercial software DigiXCT (DIGISENS) and an isotropic voxel size of 42.51 μm. Additionally, the proximal part of the sample was re-analysed (voltage 150 kV, current 100 μA, 1-mm copper filter, exposure time/projection 3.0 s and 1,800 projections over 360°) setting an effective pixel size of 18 μm and reconstructed using the same software to achieve a higher spatial resolution.

Morphological dataset construction. All taxa used in this study were personally observed by at least one of us, and more than half by two or more of the co-authors. The new dataset presented herein includes a large sample of species of squamates, as well as a broad variety of non-squamatan lepidosaurs and non-lepidosaurian diapsid species, representing all of the major clades of diapsid reptiles. Characters were assessed based on primary homology assessment and according to strict criteria for character construction, to avoid biases owing to logical or biological dependencies across characters, overweighting of any anatomical attributes and many other issues that may affect the morphological component of phylogenetic datasets[18]. We selected *Protorothyris archeri* as the outgroup to our analyses and all morphological characters were treated as unordered (see Supplementary Methods for additional details).

Molecular dataset alignment, model selection and partitions. The molecular dataset consists of 16 genetic markers (13 nuclear and 3 mitochondrial loci) for 38 extant taxa. A complete list of sampled loci and sequence lengths is provided in Supplementary Table 1. Sequence data for the selected coding regions were obtained from GenBank (Supplementary Data 2). For three ingroup taxa, *Liolaemus signifer*, *Pristidactylus scapulatus* and *Stenocercus scapularis*, for which molecular data were not available, we used sequences of the congeneric species, *L. ornatus*, *P. torquatus* and *S. guentheri*, respectively. Sequences were aligned in the MAFFT 7.245[32] online server using the global alignment strategy with iterative refinement and consistency scores. For the protein-coding genes, alignments were verified by translating nucleotide sequences to amino acids. The final multiple sequence alignment was concatenated and visually examined in Mesquite 3.04[33]. Molecular sequences from all extant taxa were analysed for the best partitioning scheme and model of evolution using PartitionFinder2[34] under Akaike information criterion.

Equal weights maximum parsimony analysis. Analyses were conducted in TNT v.1.1[35] using the new technology search algorithms. This strategy enables the sampling of trees from a broader spectrum of local optima than is allowed by the heuristic search with ratchet runs in PAUP* v.4.0 beta 10, especially for large datasets[35,36]. Tree searches were conducted using 1,000 initial trees by random addition sequences with 100 iterations or rounds for each of the four NTS algorithms: sectorial search, ratchet, drift and tree fusing. The output trees were used as the starting trees for subsequent runs, using 1,000 iterations/rounds of each of the new technology search algorithms. The latter step was repeated once, and the final output trees were filtered for all the most parsimonious trees (MPTs). A total of 621 MPTs were obtained with 2,268 steps each.

Implied weights maximum parsimony analysis. Analyses were also conducted in TNT, using the implied weighting algorithm[37], with a $K = 12$ and collapsing all branches with support = 0. Tree searches were conducted as performed for the equal weights parsimony analysis. Larger K values than the default (3.0) are indicated to perform better for large datasets[38]. A total of five best fit trees were obtained (fit = 91.768892) and used to calculate the strict consensus tree.

Bayesian inference analyses. Analyses were conducted using Mr. Bayes v.3.2.6[39] using the Cedar computer cluster made available through Compute Canada and the CIPRES Science Gateway v.3.3[40]. Molecular partitions were analysed using the models of evolution obtained from PartitionFinder2 (see dataset), and the morphological partition was analysed with the MkV model[41].

The distribution for rate heterogeneity was tested for best fit to the data under both γ and log-normal distributions, as it was recently demonstrated that a log-normal distribution may better fit morphological data for a large variety of datasets[42,43]. Fit to the data was assessed using Bayes factors $[B_{10}]$[44,45] calculated with the marginal model likelihoods obtained from the stepping-stone sampling method[46]. The interpretation of the results of the model fit to the data was performed as previously described[45]: when $2\log_e(B) > 2$ (positive evidence against model M_0); when $2\log_e(B) > 6$ (strong evidence against model M_0); when $2\log_e(B) > 10$ (very strong evidence against model M_0). However, $2\log_e(B)$ was less than one between the γ and log-normal runs, indicating that there was no significant difference in fit to the morphological data between both distributions. The morphological partition was thus analysed under the γ model for all subsequent analyses.

Time-calibrated relaxed-clock Bayesian inference analyses. We implemented 'total-evidence-dating' using the fossilized birth–death tree model with sampled ancestors, under a relaxed-clock model in Mr. Bayes v.3.2.6[47–49]. The chosen relaxed-clock model is the independent γ rate relaxed-clock model[50]. This is a continuous uncorrelated relaxed-clock model using a gamma distribution to assess clock rate variation across lineages. The latter is compatible with the fossilized birth–death tree model, unlike the compound Poisson process relaxed-clock model[48]. The base clock rate was given an informative prior, which was derived from the non-clock Bayesian inference analysis: the median value for tree height in substitutions from the entire posterior trees sample divided by the age of the tree, which is based on the median of the distribution for the root prior: 25.1658/325.45 = 0.0773, in natural log scale = −2.560061. We chose to use the exponent of the mean to provide a broad standard deviation ($e^{0.0773} = 1.080366$) as previously recommended[6]. The sampling strategy was set to diversity, which is more appropriate when extant taxa are sampled in a manner that maximizes diversity (as performed herein) and fossils are sampled randomly[47,48]. Diversity sampling is very common in higher-level phylogenies, and not accounting for it has a deep effect on tree inference, pushing divergence times further back and creating unreasonably older and more variable divergence times[48,51]. This is a considerable advantage of using Mr. Bayes for divergence time estimates over current implementations available in the software package BEAST[52].

The wealth of fossil taxa in our dataset, including some of the oldest known taxa for many clades, provided numerous calibration points. Therefore, the vast majority of our calibrations were based on tip dating, which accounts for the uncertainty in the placement of fossil taxa and avoids the issue of bound estimates for node-based age calibrations[47]. The fossil ages used for tip dating correspond to the uniform prior distributions on the age range of the stratigraphic occurrence of the fossils (available in Supplementary Table 2). However, it has recently been demonstrated that using tip dates only can contribute to unrealistically older divergence time estimates for some clades[53,54]. Therefore, when we lacked the oldest known fossils for any of the clades in our analysis (namely, captorhinids, choristoderes, snakes and rhynchocephalians), we used node-age calibrations with a soft lower bound as long as the age of the oldest known fossil was well-established and there was overwhelming support in the literature (and in all our other analyses) for their monophyletism. Combined with the diversity sampling strategy, the latter dating protocol can ensure reliable divergence time estimates.

The age of the root was set with a soft lower bound, which gives a low (but non-zero) likelihood of the age being older than the lower bound value. Minimum and maximum root bounds were placed as follows. The minimum age was set at the oldest possible age for the oldest known reptile, *Hylonomus* (from the Joggins Formation in Nova Scotia, Canada), which comes from the late Bashkirian Stage (early Pennsylvanian, Late Carboniferous) and is between 318 and 315 Myr old[55]. Considering *Petrolacosaurus* may be as much as 307 Myr old, placing the minimum age at 318 Myr seems consistent, as the most recent common ancestor of diapsids and captorhinids must have been at least a few million years older than *Petrolacosaurus*. The maximum age was based on the maximum soft age for the reptile–synapsid split[56], 332.9 Ma.

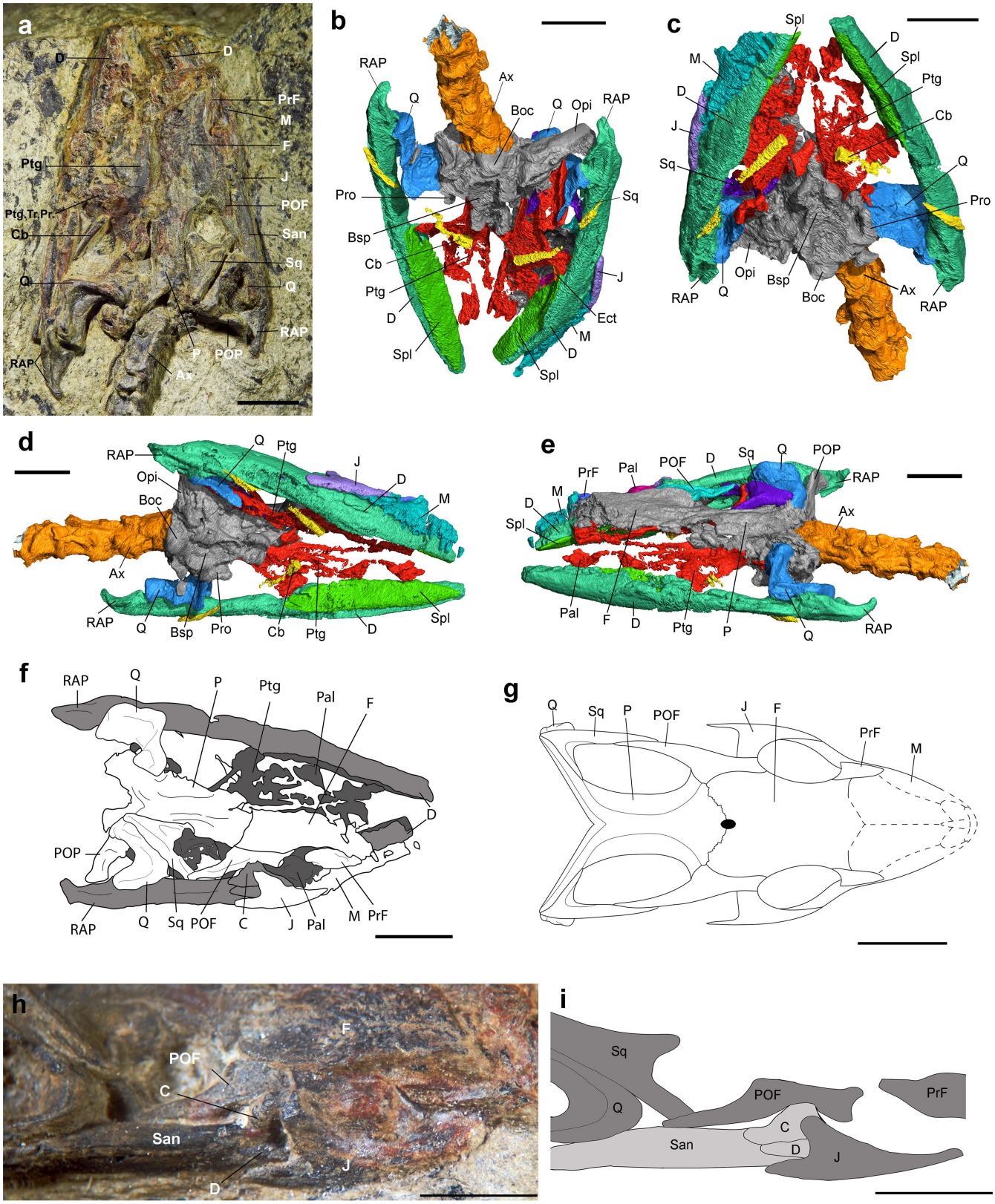
Convergence of independent runs was assessed using an average standard deviation of split frequencies of approximately 0.01, potential scale reduction factors of approximately 1 for all parameters[57] and an effective sample size greater than 200 for each parameter.

Leaf stability. Leaf stability was assessed using RogueNaRok[58], which allows assessing the difference between the highest and the second highest support values for alternative resolutions of each taxon quartet or triplet in the dataset (LSdif)[59]. We applied this method to the posterior trees from the Bayesian inference analysis including both the morphological and molecular data. Because of the large number of taxa and large number of trees, it was necessary to downsample the total number of posterior trees from each analysis (100,000 trees after discarding burn-in). The final sample consisted of 10,000 trees (selecting one at every 10 trees) using the Burntrees script for Perl (https://github.com/nylander/Burntrees). Taxon names and raw data relating to each number depicted in Extended Data Fig. 9 can be found in Supplementary Table 3.

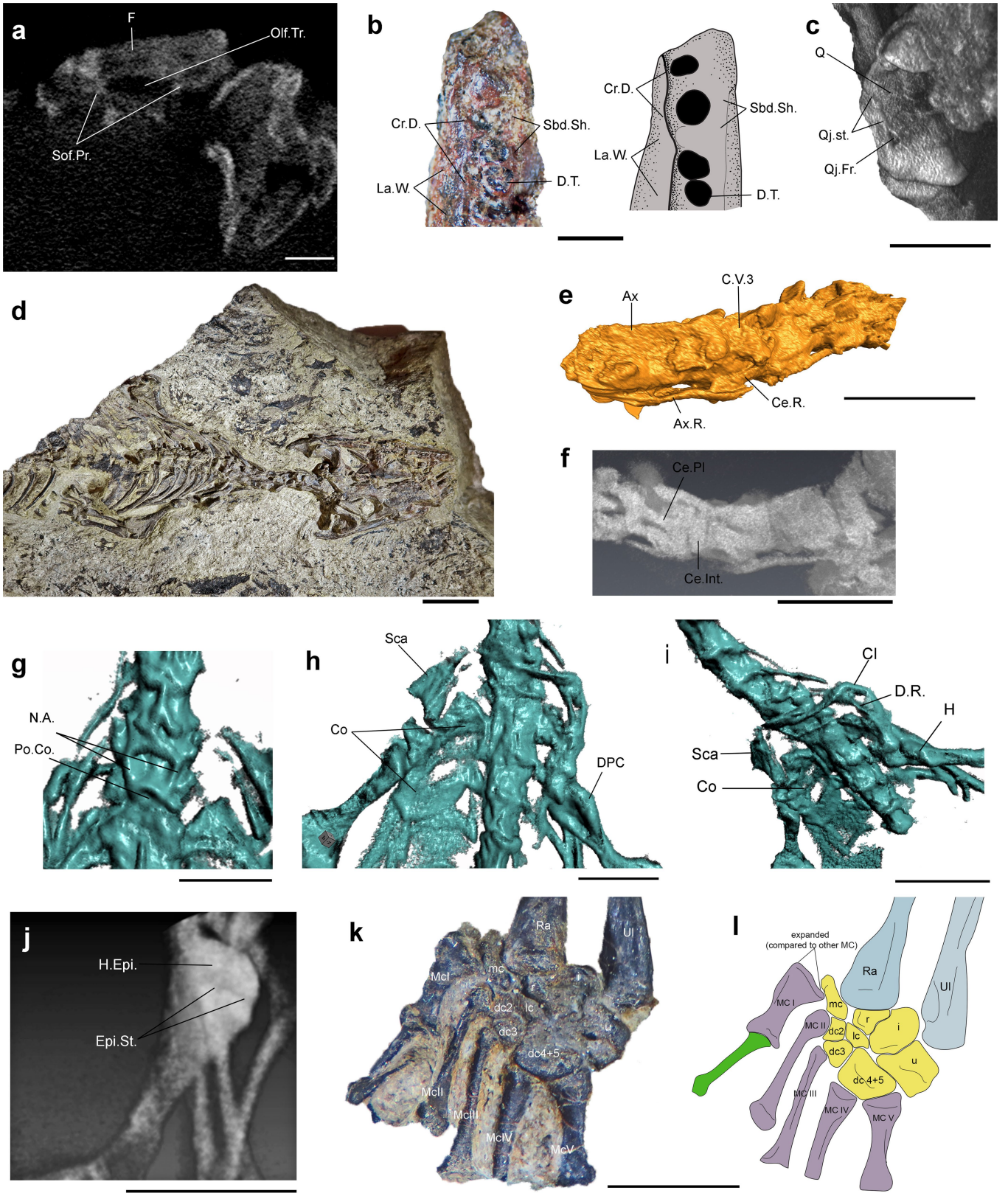
Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. The micro-CT scan data are available from the authors upon reasonable request. The morphological and molecular datasets for the phylogenetic analyses, including the Mr. Bayes parameters block, are available as Supplementary Information.

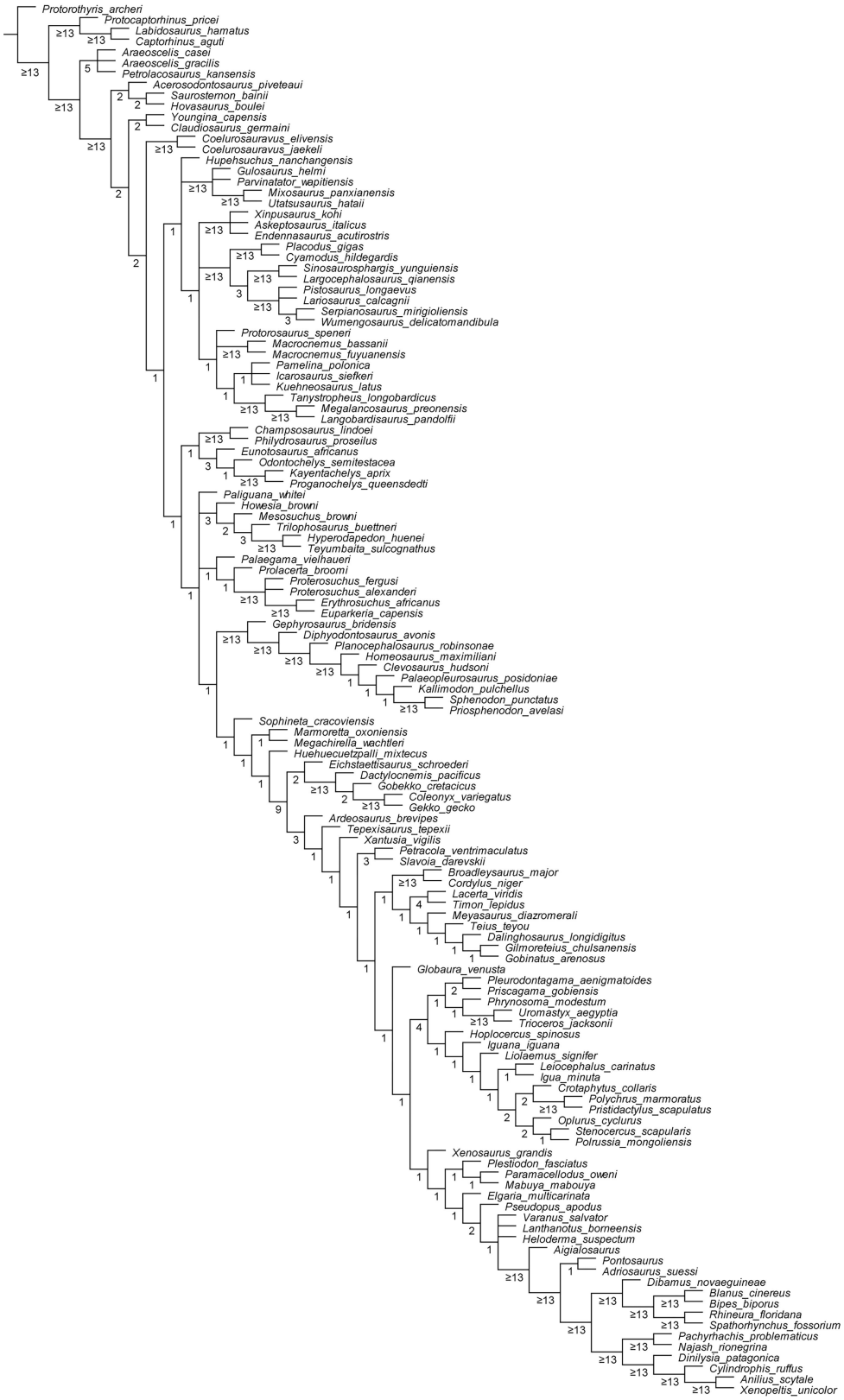
31. Tuniz, C. et al. The ICTP-Elettra X-ray laboratory for cultural heritage and archaeology. *Nucl. Instrum. Methods Phys. Res. A* **711**, 106–110 (2013).
32. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
33. Maddison, W. P. & Maddison, D. R. *Mesquite: a Modular System for Evolutionary Analysis v.3.04* http://mesquiteproject.org (2015).
34. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
35. Goloboff, P. A., Farris, J. S. & Nixon, K. C. TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786 (2008).
36. Goloboff, P. A. Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* **15**, 415–428 (1999).
37. Goloboff, P. A., Carpenter, J. M., Arias, J. S. & Esquivel, D. R. M. Weighting against homoplasy improves phylogenetic analysis of morphological data sets. *Cladistics* **24**, 758–773 (2008).
38. Goloboff, P. A., Torres, A. & Arias, J. S. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics* https://doi.org/10.1111/cla.12205 (2017).
39. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
40. Miller, M. A., Pfeiffer, W. & Schwartz, T. in *Proceedings of the 2010 Gateway Computing Environments Workshop (GCE)* (IEEE, 2010).
41. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 (2001).
42. Harrison, L. B. & Larsson, H. C. Among-character rate variation distributions in phylogenetic analysis of discrete morphological characters. *Syst. Biol.* **64**, 307–324 (2015).
43. Wagner, P. J. Modelling rate distributions using character compatibility: implications for morphological evolution among fossil invertebrates. *Biol. Lett.* **8**, 143–146 (2012).
44. Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P. & Nieves-Aldrey, J. L. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**, 47–67 (2004).
45. Kass, R. E. & Raftery, A. E. Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995).
46. Xie, W., Lewis, P. O., Fan, Y., Kuo, L. & Chen, M.-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst. Biol.* **60**, 150–160 (2011).
47. Ronquist, F. et al. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.* **61**, 973–999 (2012).
48. Zhang, C., Stadler, T., Klopfstein, S., Heath, T. A. & Ronquist, F. Total-evidence dating under the fossilized birth–death process. *Syst. Biol.* **65**, 228–249 (2016).
49. Stadler, T. Sampling-through-time in birth–death trees. *J. Theor. Biol.* **267**, 396–404 (2010).
50. Lepage, T., Bryant, D., Philippe, H. & Lartillot, N. A general comparison of relaxed molecular clock models. *Mol. Biol. Evol.* **24**, 2669–2680 (2007).
51. Höhna, S., Stadler, T., Ronquist, F. & Britton, T. Inferring speciation and extinction rates under different sampling schemes. *Mol. Biol. Evol.* **28**, 2577–2589 (2011).
52. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **10**, e1003537 (2014).
53. O'Reilly, J. E., Dos Reis, M. & Donoghue, P. C. J. Dating tips for divergence-time estimation. *Trends Genet.* **31**, 637–650 (2015).
54. O'Reilly, J. E. & Donoghue, P. C. Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. *Biol. Lett.* **12**, 20150975 (2016).
55. Ogg, J. G., Ogg, G. & Gradstein, F. M. *A Concise Geologic Time Scale* (Elsevier, Amsterdam, 2016).
56. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electronica* **18**, 1–106 (2015).
57. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
58. Aberer, A. J., Krompass, D. & Stamatakis, A. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst. Biol.* **62**, 162–166 (2013).
59. Wilkinson, M. Identifying stable reference taxa for phylogenetic nomenclature. *Zool. Scr.* **35**, 109–112 (2006).



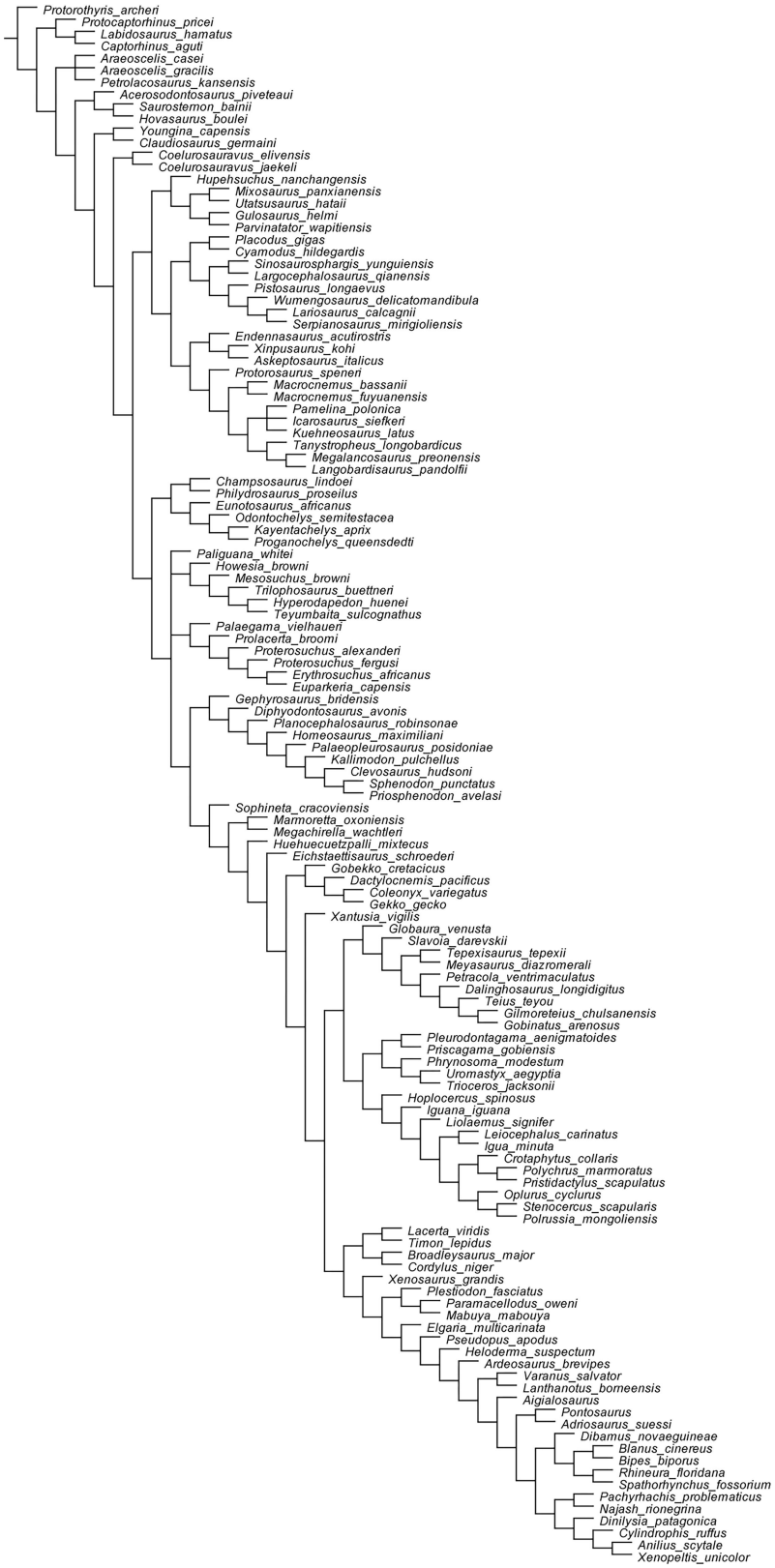
Extended Data Fig. 1 | Cranial anatomy of *M. wachtleri* (PZO 628) based on personal examination and micro-CT scan data. a, Skull in dorsal view. **b**, Skull in posteroventral view. **c**, Skull in anteroventral view. **d**, Skull in right ventrolateral view. **e**, Skull in left dorsal lateral view. **f**, Line drawing of the skull in dorsal view. **g**, Reconstruction of the skull in dorsal view. **h**, Detailed view of right lateral side of the skull. **i**, Drawing of the view in **h**. San, surangular. Scale bars, 5 mm (**a**–**g**).



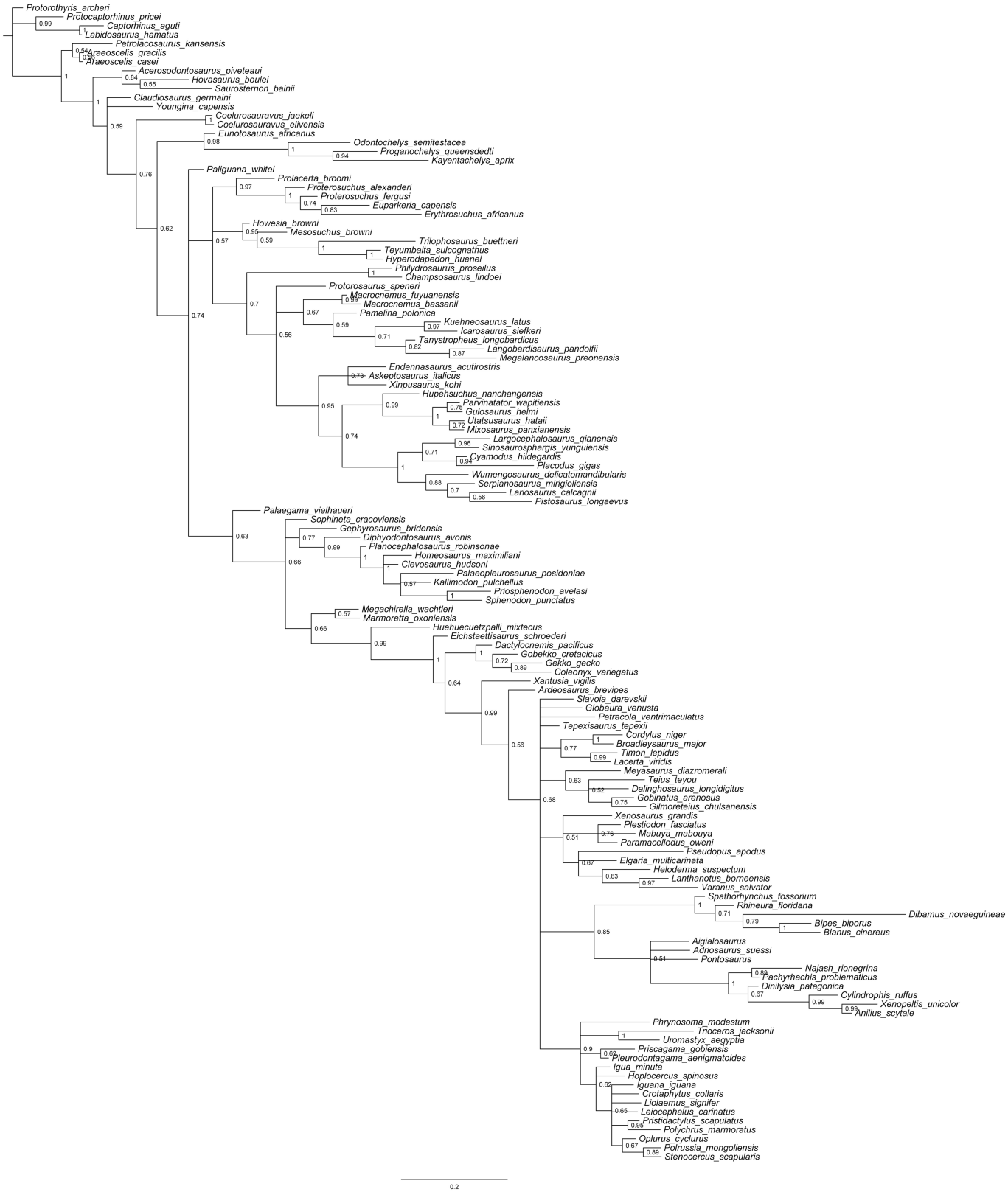
Extended Data Fig. 2 | Cranial and postcranial anatomy of *M. wachtleri* (PZO 628) based on personal examination and micro-CT scan data. **a**, Cross-section of the skull at the level of the frontals in anterior view. **b**, Details of the anterior end of the left dentary in occlusal view. **c**, Left quadrate. **d**, Whole body of the holotype as preserved in the slab (dorsal view). **e**, Anterior cervical vertebrae in left lateral view. **f**, Longitudinal section of the anterior cervicals in ventral view. **g**, Last cervicals and anterior dorsals in dorsal view. **h**, Pectoral girdle in ventral view. **i**, Pectoral girdle in left ventrolateral view. **j**, Right humerus in ventral view. **k**, Right manus in dorsal view. **l**, Line drawing of right manus in dorsal view. Ax.R., axis rib; Ce.Pl., cervical, pleurocentrum; Co, cotyle; C.V.3, third cervical vertebra; dc2–5, distal carpals 2–5; DPC, deltopectoral crest; D.R., dorsal rib; D.T., dentary teeth; Epi.St., epiphysial suture; H.Epi., humeral epiphysis; i, intermedium; lc, lateral centrale; McI–V, metacarpals I–V; N.A., neural arch; Olf.Tr., olfactory tract; Po.Co., posterior cotyle; Qj.Fr., quadratojugal foramen; Qj.St., quadratojugal suture; r, radiale; Sbd.Sh., subdentary shelf; Sof.Pr., subolfactory processes; u, ulnare. Scale bars, 1 mm (**a**, **b**), 5 mm (**c**, **e**–**h**, **j**–**l**), 10 mm (**d**, **i**).



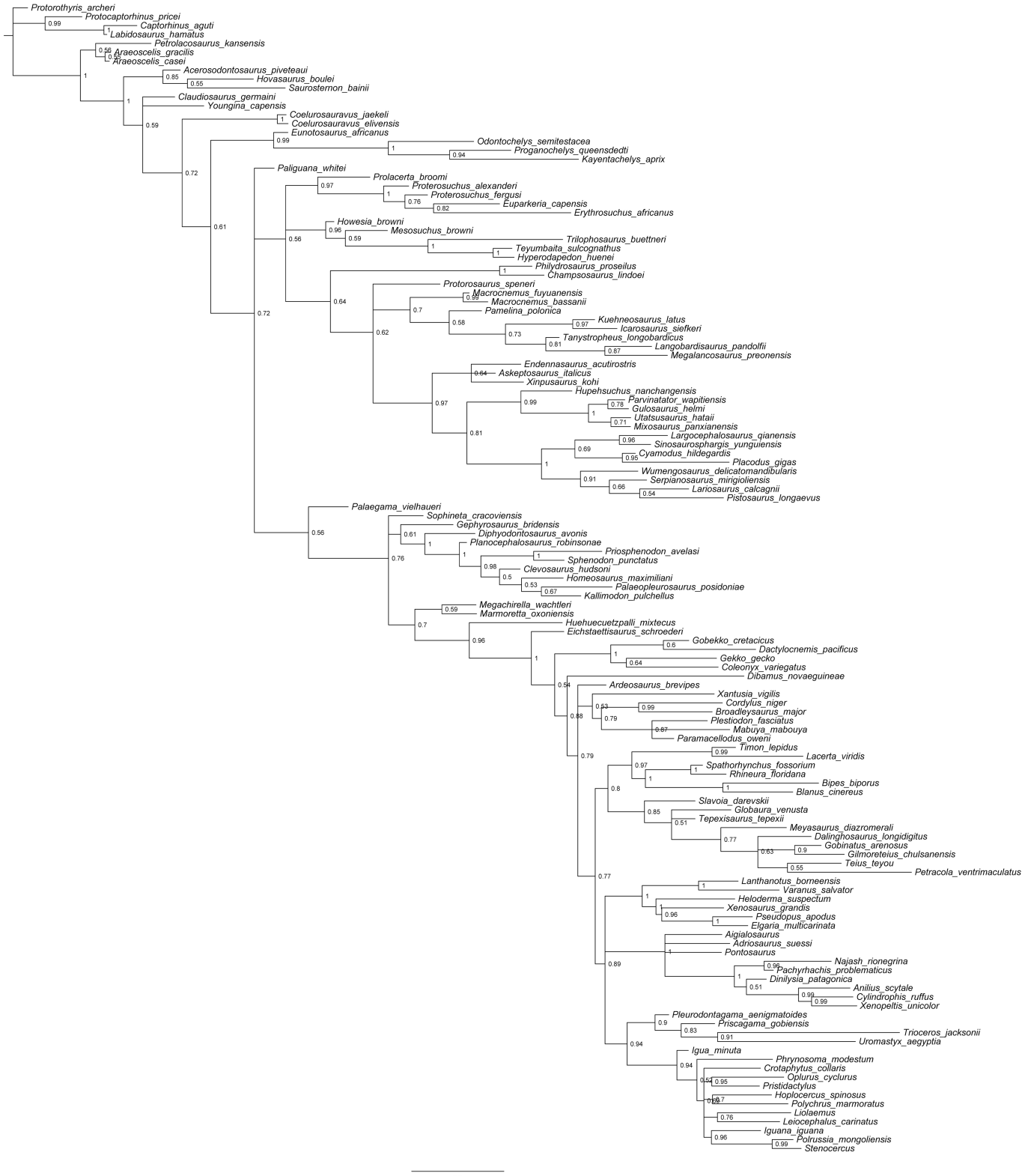
Extended Data Fig. 3 | Equal weights maximum parsimony analysis, morphological data only. Strict consensus of 621 most parsimonious trees (2,268 steps each). Numbers at nodes indicate Bremer indices.



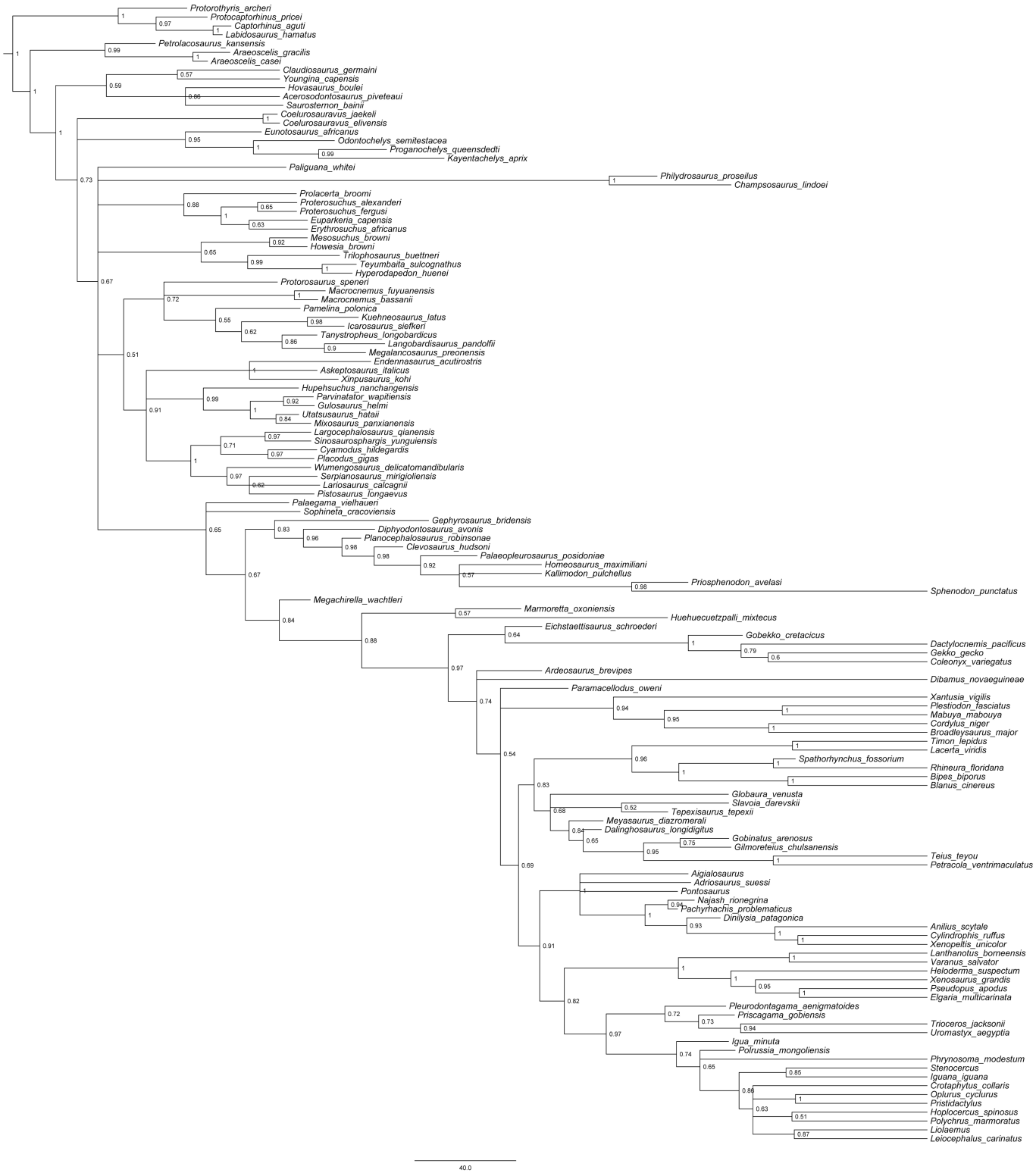
Extended Data Fig. 4 | Implied weighting maximum parsimony analysis, morphological data only. Strict consensus of the five best feet trees (fit = 91.768892).



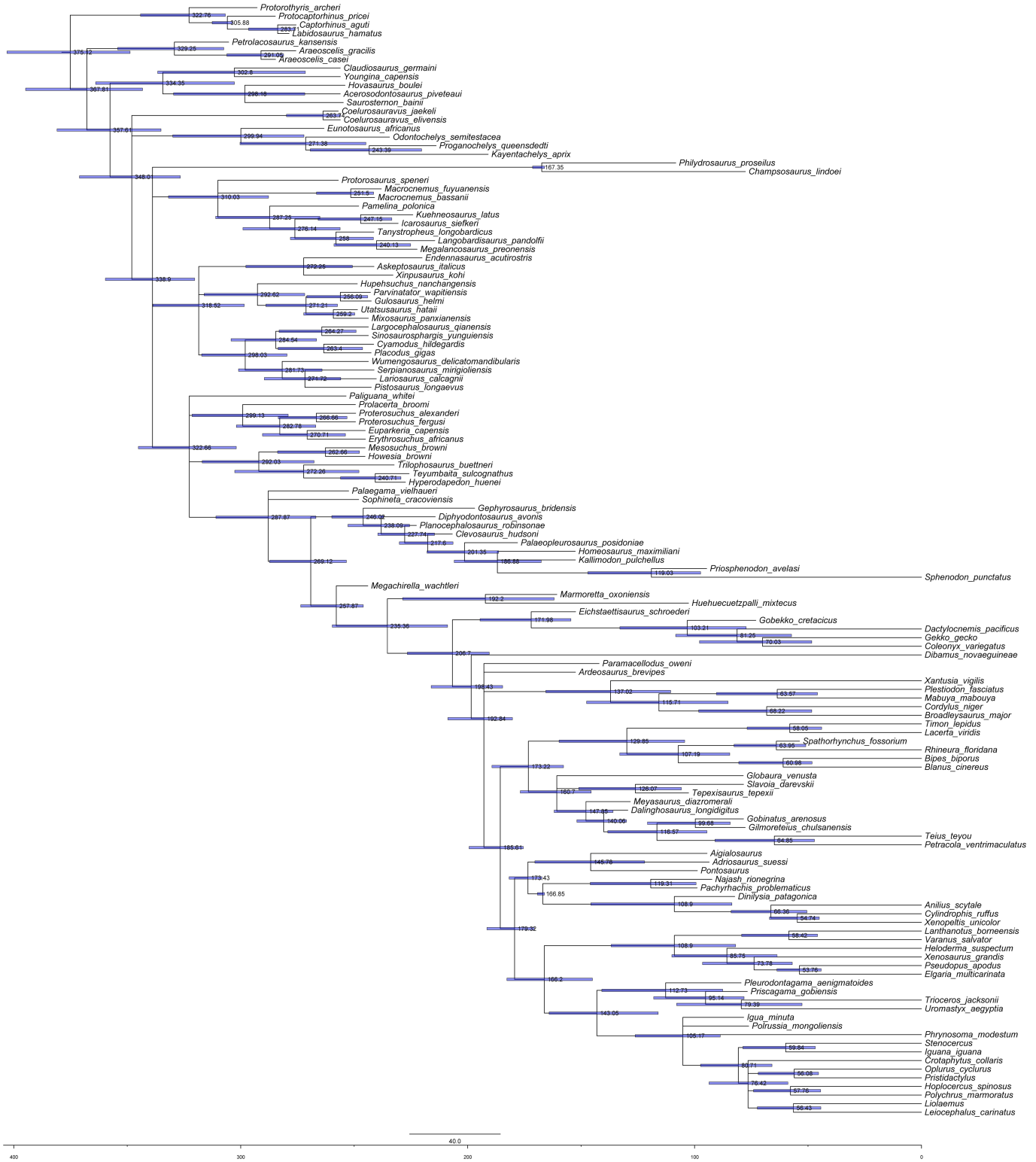
Extended Data Fig. 5 | Bayesian inference analysis, morphological data only. Bayesian majority-rule consensus tree. Numbers at nodes indicate posterior probabilities.



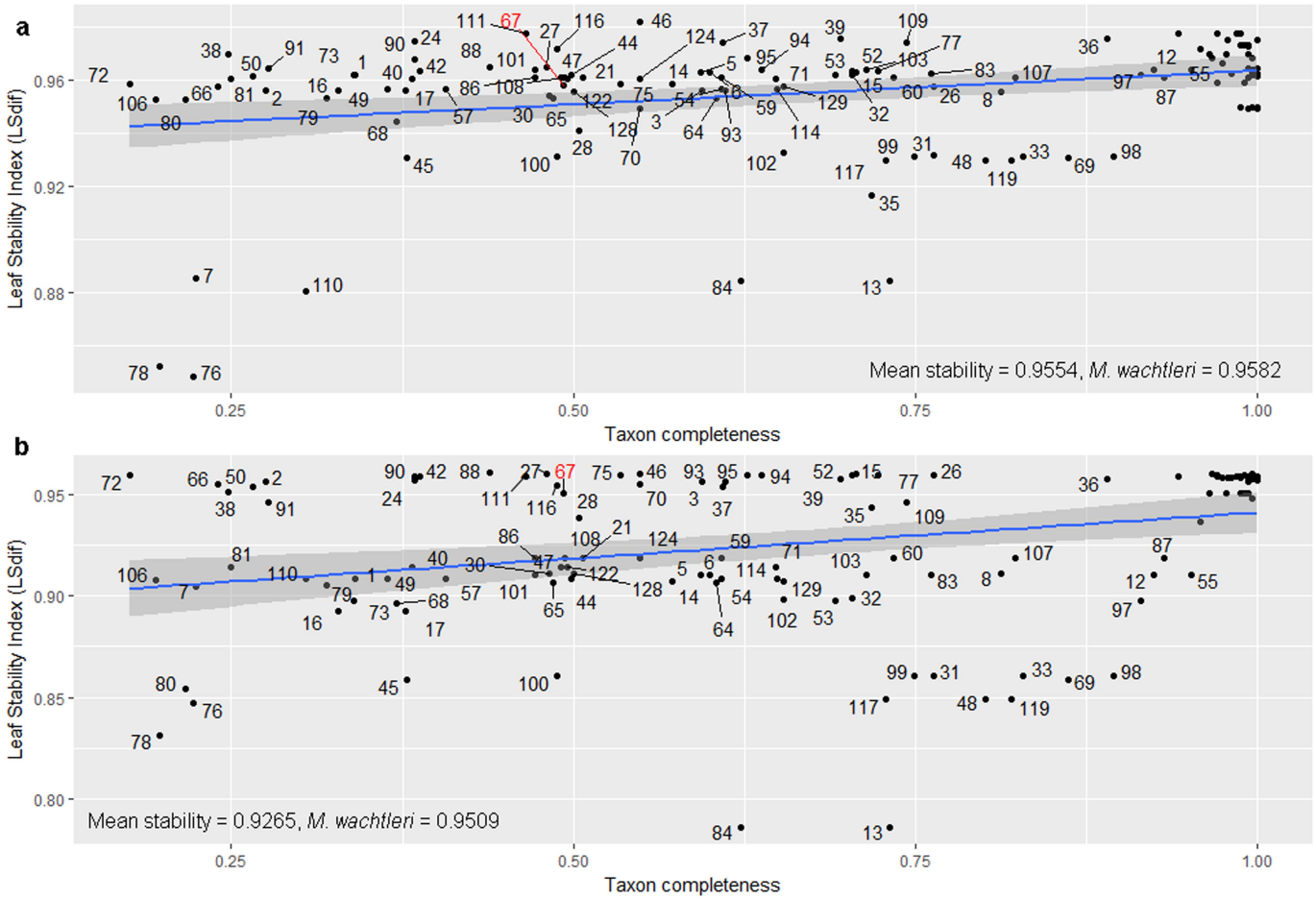
Extended Data Fig. 6 | Bayesian inference analysis, combined morphological and molecular data. Bayesian majority-rule consensus tree. Numbers at nodes indicate posterior probabilities.



Extended Data Fig. 7 | Relaxed-clock Bayesian inference analysis with total-evidence tip dating using the fossilized birth–death tree model, combined morphological and molecular data. Bayesian majority-rule consensus tree. Numbers at nodes indicate posterior probabilities.



Extended Data Fig. 8 | Relaxed-clock Bayesian inference analysis with total-evidence tip and node dating using the fossilized birth–death tree model, combined morphological and molecular data. Bayesian majority-rule consensus tree. Numbers at nodes indicate median estimates for the divergence times, and node bars indicate the 95% highest posterior density for divergence times.



Extended Data Fig. 9 | Taxon stability plotted against taxon completeness in the analysis combining both morphological and molecular data. a, Taxon stability in uncalibrated Bayesian inference analysis. **b**, Taxon stability in relaxed-clock Bayesian inference analysis with tip dating. Taxon stability increases directly proportional to taxon completeness. *M. wachtleri* (taxon 67, in red) has a stability slightly above average for uncalibrated Bayesian inference, and well above average for Bayesian inference with tip dating. All taxa are identified in Supplementary Table 3 ($n = 129$ taxa). Regression line in blue and 95% confidence interval in grey. Labels for extant taxa (~100% completeness) are omitted for simplicity.

nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample sizes for phylogenetic analysis are not statistically determined. However, we designed this study to cover different taxa from every major taxonomic group of relevance to our purposes.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded from our phylogenetic analyses. The posterior trees from the Bayesian inference analyses were downsampled to 10,000 trees in order to calculate the leaf stability index, because of the data size constraints of RogueNaRok.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   Reproducibility is assured by an extremely detailed account of every taxon included in our analyses in the supplementary information (e.g. list of the observed specimens, age, locality, main anatomical and taxonomic bibliography, etc.). The characters utilized in our analysis are also detailed in the supplementary information. Data matrices and Bayesian analytical data blocks are all available as supplementary data.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   All organisms fall into a single experimental group for the phylogenetic analyses.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   This is not relevant to our study because all taxa were included in our analyses.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Our data was compiled in Mesquite (v. 3.04); molecular alignments were performed with MAFFT (Multiple Sequence Alignment Software Version 7); our analysis were performed in the phylogenetic software TNT (v. 1.1), Mr, Bayes (v. 3.2.6). Leaf stability was calculated with RogueNaRok (online server). Tree filtering was performed with Burntrees script for Perl, and linear regression performed in R.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

There are no restrictions on the availability of unique materials. All specimens considered here are housed in publicly available museum or university collections, and our raw CT scan data is fully available upon request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No commonly misidentified cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Only fossil data, or skeletonized specimens housed in museum collections were used. No specific data about their age or sex was available for those materials at the time of data collection.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Our study did not involve human research participants.