Running head: PLACING FOSSILS WITH MORPHOMETRIC DATA

Title: Bayesian and likelihood placement of fossils on phylogenies from quantitative morphometrics

Caroline Parins-Fukuchi

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109, USA.

E-mail: cfukuchi@umich.edu

Telephone: (734) 474-7241

Acknowledgements I thank James Saulsbury, Gregory Stull, Joseph Walker, and Stephen Smith for helpful comments that improved the manuscript. I also greatly thank Iju Chen and Steven Manchester for providing the Vitaceae measurement data, and Gregory Stull for coordinating and helping to prepare the data for analysis.

Supplementary Data All supplemental material, including scripts, newick files, morphological and molecular alignments used in this study are available as a GitHub repository (https://github.com/carolinetomo/fossil_placement_tests).

Abstract

Jointly developing a comprehensive tree of life from living and fossil taxa has long been a fundamental issue in evolutionary biology.One major complication has stemmed from difficulties in merging evidence from extant and extinct organisms. While many fields have reached varying stages of synthesis, efforts have been hindered by their dependence on qualitative descriptions of morphology. Though rarely applied to phylogenetic inference, morphometrics can improve these issues by generating more rigorous ways to quantify variation in morphological structures, and have the added benefit of facilitating the rapid and objective aggregation of large morphological datasets. I describe a new Bayesian method that identifies the positions of fossil taxa on reference trees using morphometric data. This method reconstructs branch lengths based on morphological disparity that do not require scaling to absolute time scale, suggesting a new framework though which to construct true Bayesian node calibration priors for molecular dating that explicitly incorporates uncertainty and a new option through which to explore patterns in morphological disparity across taxa that simplify existing phylogenetic comparative methods. I am hopeful that the approach described here will help to facilitate a deeper integration of neo- and paleontological data to move morphological phylogenetics further into the genomic era.

Introduction: The role of fossil data in reconstructing phylogeny among living organisms has long been a central, yet contentious, topic in evolutionary biology. This has manifested over the past decade in the rapid proliferation of 'total-evidence' methods that seek to simultaneously reconstruct the relationships and divergence times between living and fossil taxa using cladistic morphological matrices. These approaches, based upon probabilistic models of molecular and morphological character, have increased understanding of evolutionary tempo across large clades, and provide compelling evidence in favor of incorporating fossils in phylogenetic analyses (Pyron

2011; Ronquist *et al.* 2012).

Researchers' enthusiasm for reconstructing a comprehensive tree of life has encouraged the integration of fossils with living taxa in phylogenetic analyses. Improving integration between fossil and living taxa has the capability to benefit both paleo- and neontological studies. In addition to the results of Berger and Stamatakis discussed above, the inclusion of fossils improves the reconstruction of ancestral states using phylogenetic comparative methods (Slater *et al.* 2012). As another example, increasing the rigor with which fossils are placed on phylogenies is expected to improve the accuracy and treatment of uncertainty in divergence time estimation (Guindon 2018).

A constant source of difficulty when jointly estimating phylogeny between living and extinct organisms is the unavailability of molecular data in nearly all fossil taxa. As a result, there has been a need to explore the compatibility of molecular with morphological data to better understand the capability of fossil and extant species to reciprocally inform reconstruction of phylogeny and divergence times. Previous work has sought to determine whether the inclusion of molecular data representing extant species can improve the reconstruction of relationships among fossils represented by morphology alone (Wiens 2009; Wiens *et al.* 2010). The result of these studies suggest that the inclusion of morphological characters comprising living and fossil species does not have a tendency to decrease the accuracy of phylogenetic reconstructions, and can improve estimation of fossil placements in well-behaved datasets. Expanding upon these observations, Berger and Stamatakis (2010) have shown that methods placing fossils on fixed molecular phylogenies can yield accurate results by filtering through conflicting signal. This shows that placing incomplete fossil data in a molecular context scaffolded by extant relationships can improve reconstruction among fossils.

Despite the abundance of both clear and subtle benefits to improving the integration of fossil and living taxa in phylogenetics, challenges have arisen from conflicting and noisy information presented by morphological data. Another issue that is noted by Berger and Stamatakis (cited above) stems from the reality that morphological alignments commonly contain very few sites, often 50-500, compared to molecular datasets, which can contain hundreds of thousands of sites. This can cause the likelihood of molecular partitions to dwarf those of morphological partitions, limiting the influence of morphology in reconstructions of topology and branch lengths. For these reasons, Berger and Stamatakis advocated fixing the relationships of extant taxa *a priori* using molecular reconstructions, and using the resulting scaffold to identify conflicting signal in morphological data.

The fragmentary nature of fossil data further exacerbates these challenges. These occur at multiple levels important to phylogenetic analysis. At one level, morphological data are frequently susceptible to displaying biased or misleading signal. This may often stem in part from the general practice of assigning discrete character states to taxa through qualitative assessment. The subjective nature of this process can cause major irreconcilable disagreement between results achieved from different researchers (Hauser and Presch 1991; Pleijel 1995; Wilkinson 1995; Hawkins *et al.* 1997; Scotland and Pennington 2000; Scotland *et al.* 2003; Brazeau 2011; Simões *et al.* 2017). As an added layer of potential bias, these matrices are also frequently filtered to include only characters that researchers consider before the fact to be accurately informative in phylogenetic reconstruction. At another level, the discrete character matrices most commonly employed in phylogenentics can often be difficult to adequately model. At present, researchers employing probabilistic methods generally use the so-called ‘Mk’ model (Lewis 2001). This is a generalization of the Jukes-Cantor model of nucleotide substitution that accommodates k possible

character states. Although previous work based upon simulated data has suggested that Mk-based approaches outperform parsimony (Wright and Hillis 2014), the extent and conditions under which this is the case in empirical datasets is unclear (Goloboff et al. 2017). Empirical datasets are also likely to depart significantly from the assumptions of the Mk model. This poor match between model assumptions and data can lead to erratic results and high uncertainty in posterior estimates of divergence times (Ronquist *et al.* 2016). The sensitivity of topological reconstruction to this frequent mismatch is fairly unclear at present.

For all of these reasons, continuous traits have been suggested as a feasible alternative (Felsenstein 1973, 1988; MacLeod 2002; Parins-Fukuchi 2017). Tools that quantify morphological size and shape have the capacity to alleviate many of the concerns relating to bias and subjectivity that occur with discrete characters. Approaches such as geometric morphometrics offer the potential to holistically incorporate all dimensions of shape to inform phylogeny. The continuous state space of morphometric data might also increase the amount of information that can be extracted from morphological datasets, which may be beneficial when analyzing poorly-sampled fossil data.

Traditional linear morphometric measurements have long been employed in morphological phylogenetics, but are typically discretized to more easily analyze them alongside present-absence data. However, these transformations may decrease the amount of information in continuous datasets by binning fine-scaled variation into shared discrete categories, and are susceptible to the difficulties in modeling under the Mk model described above. Geometric morphometric data have shown utility in several previous phylogenetic studies using parsimony-based methods (González-José *et al.* 2008; Catalano *et al.* 2010; Smith and Hendricks 2013), but have not gained substantial traction. This may be in part due to the lack of available

tools to analyze continuous trait data in a probabilistic framework.

The earliest studies investigating probabilistic methods of phylogenetic inference were developed using continuous characters modelled under Brownian Motion (BM) (Cavalli-Sforza and Edwards 1967; Felsenstein 1973). Due in part to the abundant discrete character data that became available with the emergence of DNA sequencing, these approaches were quickly overshadowed in popularity by discrete trait approaches based upon Markov nucleotide substitution models. Continuous trait models have since gained significant popularity in phylogenetic comparative methods, but still are rarely used for phylogenetic inference. As a result, few implementations exist, with only ContML in the PHYLIP package and RevBayes (through scripting) providing such functionality (Höhna *et al.* 2016). The approaches used in these packages are also fairly minimalistic, with no real tailoring to the challenges that might be presented by empirical datasets.

In this paper, I describe a new approach that places fossils on molecular trees using quantitative characters modeled under BM. This seeks to tackle some of the most pressing obstacles associated with the use of traditional and geometric morphometric data in phylogenetic inference. Using simulated data, I validate and explore the behavior of the implementation. I also analyze empirical datasets representing the Vitaceae family of flowering plants (citation?) and carnivoran mammals (Jones *et al.* 2015) comprised of traditional and geometric morphometric measurements, respectively. The method uses Markov chain Monte Carlo (MCMC) to infer the evolutionary placements of fossils and branch lengths. Although MCMC is generally associated with Bayesian inference, the implementation described here can perform inference both with and without the use of priors to enable exploration of a range of inferential paradigms to best accommodate diverse morphometric datasets.

Methods and Materials:

Brownian motion model

The approaches that I describe in this paper all rely upon the familiar BM model of evolution (Butler and King 2004; O'Meara *et al.* 2006) . Under BM, traits are assumed to be multivariate distributed, with variances between taxa defined by the product of their evolutionary distance measured in absolute time and the instantaneous rate parameter (σ):

$$dX(t) = \sigma dB(t) \tag{1}$$

where *dX(t)* is the time derivative of the change in trait *X* and *dB(t)* corresponding to normally distributed random variables with mean 0 and variance *dt*. This leads to the expectation that over time *t*,

$$E(X_t) = X_0 \tag{2}$$

with

$$Var(X_t) = \sigma t \tag{3}$$

where X_0 gives the trait value at t_0.

The methods that I describe use a slightly different parameterization and likelihood calculation than most conventional implementations used in modern phylogenetic comparative methods (PCMs). These generally construct a variance-covariance (VCV) matrix from a dated, ultrametric phylogeny to calculate the likelihood of the data, assuming a multivariate normal distribution (Butler and King 2004; O'Meara *et al.* 2006). Since these methods treat the topology

and branching times as known, the goal is typically to obtain the maximum likelihood estimate (MLE) of the rate parameter (σ) to examine evolutionary rate across clades.

One drawback to the use of this version of the phylogenetic BM model in the reconstruction of topology is its requirement that phylogenies be scaled to absolute time. Although it is possible to simultaneously estimate divergence times and topology while analyzing continuous traits, this can cause additional error and requires the specification of a tree prior that can accommodate non-ultrametric trees that include fossils. This requirement would also cause circularity in cases where researchers are interested in obtaining estimates and error in fossil placements in order to more rigorously inform molecular clock calibrations. To overcome the need for simultaneously estimating divergence times and fossil placements, I estimate the product σt together. As a result, rate and absolute time are confounded in the trait and tree models. Branch lengths, which reflect the morphological disparity between taxa, are thus measured in units of morphological standard deviations per site. This interpretation could be thought roughly of as a continuous analogue to the branch lengths obtained from discrete substitution models. Similarly to the discrete case, long branch lengths could reflect either a rapid rate of evolution or a long period of divergence (in absolute time) along that lineage.

Computation of the likelihood:

Rather than use the computationally expensive VCV likelihood calculation, I use the reduced maximum likelihood (REML) calculation described by Felsenstein (1973). In this procedure, the tree likelihood is computed from the phylogenetic independent contrasts (PICs) using a 'pruning' algorithm. In this procedure, each internal node is visited in a postorder traversal, and the log-likelihood, L_{node} is calculated as multivariate normal, with a mean equal to the contrast between the character states, x_1 and x_2 at each subtending edge and variance calculated as the sum of each

child edge, v_1 and v_2:

$$L_{node} = \frac{1}{2} * \frac{\log(2\pi) + \log(v_1 + v_2) + (x_1 - x_2)^2}{v_1 + v_2} \tag{4}$$

The PIC, $x_internal$, is calculated at each internal node and used as the character state representing the internal node during the likelihood computation at the parent node. The edge length of the internal node, $v_internal$ is also extended by averaging the lengths of the child nodes to allow the variance from the tips to propagate from the tips to the root:

$$x_{internal} = \frac{(x_1 * v_2) + (x_2 * v_1)}{v_1 + v_2} \tag{5}$$

$$v_{internal} = v_{internal} + \frac{(v_1 * v_2)}{(v_1 + v_2)} \tag{6}$$

This procedure is performed at each internal node. The total log-likelihood of the tree, L_{tree} is calculated by summing the log-likelihoods calculated at each of the n internal nodes.

$$L_{tree} = \sum_{node=1}^{n} L_{node} \tag{7}$$

Priors:

Since the estimation of branch lengths from continuous traits is relatively uncharted territory in phylogenetics, I implemented and tested three different branch length priors derived from the molecular canon: 1) flat (uniform), 2) exponential, and 3) a compound dirichlet prior after (Rannala *et al.* 2011). The compound dirichlet prior also offers an empirical Bayes option that uses an initial ML estimate of the branch lengths to specify the parameter corresponding to mean tree length.

Markov-chain Monte Carlo

This method uses a Metropolis-Hastings (MH) algorithm (Hastings 1970) to simulate the posterior or confidence distribution of fossil insertion points and branch lengths. Rearrangements of the topological positions of fossil taxa are performed by randomly pruning and reinserting a fossil taxon to generate a proposal. This is a specific case of the standard subtree pruning and regrafting (SPR) move for unrooted tees. Branch lengths are updated both individually and by randomly applying a multiplier to subclades of the tree. MH proposal ratios were derived from the unrooted SPR and multiplier moves described by Yang (2014).

Generating a rough ML starting tree:

To generate an initial estimate of fossil placements and branch lengths, I estimate an approximate ML starting tree. Initial placements are achieved using stepwise addition. Each fossil is individually inserted along all existing branches of the tree, with the insertion point that yields the highest likelihood retained. At each step, MLEs of the branch lengths are computed using the iterative procedure introduced by (Felsenstein 1981). In this procedure, the tree is rerooted along each node. PICs are calculated to each of the three edges subtending the new root, and then the MLE of each edge (v_i) is computed analytically by averaging the distances between the PICs calculated from each *j* of *n* sites in the character alignment (x_{ij}):

$$\hat{v_{1j}} = \frac{\sum_{j=1}^{n}(x_{1j} - x_{2j})(x_{1j} - x_{3j})}{n} \tag{8}$$

$$\hat{v_{2j}} = \frac{\sum_{j=1}^{n}(x_{2j} - x_{1j})(x_{2j} - x_{3j})}{n} \tag{9}$$

$$\hat{v_{3j}} = \frac{\sum_{j=1}^{n}(x_{3j} - x_{1j})(x_{3j} - x_{2j})}{n} \tag{10}$$

This process is iterated by successively rerooting on each node of the tree and calculating the branch lengths until their values and the likelihoods converge. Felsenstein (1981) gives a more detailed explanation of the algorithm, along with a complete derivation of the three-taxon MLEs.

Once the optimal placement of all of the fossils has been identified, the branch lengths are recalculated and can be used to inform branch length priors used during MCMC simulation. One problem with interpreting the results of the ML approach on their own is that it has a strong propensity to becoming trapped in local optima. As a result, it should be used and interpreted cautiously in messy datasets. In the applications here, the topologies achieved from this procedure are restricted to the construction of starting trees, while the branch lengths inform the specification of branch length priors. This procedure allows straightforward construction of non-random starting trees for the MCMC and priors that reflect the the dataset under analysis.

Filtering for concordant sites:

One major hurdle involved in the use of morphological data is their frequent tendency to display noisy and discordant signal. This problem might be expected to manifest even more intrusively in morphometric datasets than in discrete datasets, since traits are much less likely to be excluded *a priori* on the basis of perceived unreliability. As a result, there is a need to filter through noisy signal to favor more reliable sites. I developed a procedure adapted from Berger and Stamatakis (2010) for this purpose. This computes a set of weights based upon the concordance of each site with the reference tree. In this procedure, the likelihood (L_{ref}) of each site is calculated on the reference tree (excluding fossil taxa). Next, the likelihood (Ln) of each

site is calculated along each *n* of 100 randomly generated phylogenies. If the likelihood of the site is higher along the reference tree than the current random tree, the weight of the site is incremented by one. Thus, site *j* receives the integer weight:

$$\overrightarrow{W}_j^{int} = \sum_{n=1}^{100} \delta_n \tag{11}$$

where δnj = 1 if:

$$L_{ref} > L_n \tag{12}$$

and δnj = 0 if:

$$L_{ref} < L_n \tag{13}$$

This yields a weight vector that is the same length as the character matrix, with each site possessing a weight between 0 and 100. The sites are then weighted using one of three schemes: 1) whole integer values, where the weight equals the value obtained from equation 7, 2) a floating point value between 0 and 1, where the value generated from the random comparison is divided by 100, and 3) a binary value where the weight is equal to 1 if the site displayed a higher likelihood in the reference tree than 95 or more of the random trees, and 0 if less than 95:

$$\overrightarrow{W}_j^{binary} = 1 \tag{14}$$

if

$$\overrightarrow{W}_j^{int} > 95 \tag{15}$$

and

$$\overrightarrow{W}_j^{binary} = 0 \tag{16}$$

if

$$\overrightarrow{W}_j^{int} < 95 \tag{17}$$

In application, I found that integer weighting caused poor MCMC mixing, and so the floating and binary schemes are probably most practical in most cases. Since it filters out discordant sites completely, the binary scheme is enforces a harsher penalty than the floating and integer schemes, and so might be of greatest use in particularly noisy datasets. As an additional note, although these procedures share similar terminology to the site weights calculated during parsimony analysis of multi-state characters, they differ in their purpose. Parsimony site weights are intended to normalize the contribution of characters with differing state spaces to the overall tree length. In contrast, the site weighting approach deployed here is designed to decrease the contribution of sites that disagree with the reference topology to the overall tree likelihood, instead highlighting signal that is taken to be more reliable. As a result, the guide tree is used to identify sites that are most likely to reliably inform fossil placements.

Simulations:

To explore the behavior of these approaches under different settings and to validate the implementation, I performed a set of simulations. From a single simulated tree, I pruned five "fossil" taxa and estimated their positions along the tree using 100 datasets of 50 characters simulated under BM. The tree was simulated under a birth-death model, with a birth parameter of

1.0 and a death parameter of 0.5. Extinct lineages were retained. The final tree conained 41 taxa, leaving a 36-taxon reference tree when the five fossils were pruned. To explore the effect of conflicting and noisy signal, I also generated alignments consisting of 50 “clean” traits simulated along the true tree, and combined with partitions of “dirty” traits in intervals of 10, 25, and 50 generated along random trees. All simulations were performed using the phytools package (Revell 2012).

I restricted the simulations to a fairly small number of traits because this reflected a similar alignment size as the two empirical datasets. This level of sampling is fairly common among existing morphometric datasets, which are typically compiled from only one or two organs. In the future, I am hopeful that quantitative morphometric datasets will increase in size to encompass much broader sampling across organs, but at present, it seemed most sensible to examine the level of sampling expected from existing datasets. Unlike in a previous paper (Parins-Fukuchi 2017), I also did not simulate traits that display covariance among sites. This is because 1) I show in the previous study that covariance does not significantly handicap reconstructions from continuous traits, and 2) because in this study I was primarily interested in examining the effect of inducing random noise without the potentially confounding effect of covariance. Although covariance has been expressed as a major concern in morphometric phylogenetics (Felsenstein 1988, 2002), there is no reason to expect greater covariance between continuous traits than discrete traits, which, ideally, should describe similar aspects of morphology.

These simulated datasets were then used to reconstruct the placements of the five fossils. To explore the relative performance of weighting schemes, I performed reconstructions using both the binary and floating approaches. These were supplemented by analyses of the noisy datasets without applying site weights. MCMC simulations were run for 1,000,000 generations and

checked to ensure that the effective sample sizes (ESS) exceeded 200. The exponential branch length prior was employed for the simulated data with a mean of 1.0. To evaluate the accuracy of the placement method, I then calculated the distances between the true and reconstructed fossil placements. This was calculated by counting the number of nodes separating the true insertion branch from the reconstructed insertion branch. Placement accuracy was evaluated using the *maximum a posteriori* (MAP) summaries of tree distributions. MAP trees represent the single most sampled tree during the MCMC run. They are thus somewhat analogous to ML trees in presenting the point estimate that is assumed to have the highest asymptotic density over the likelihood or posterior (depending on paradigm) surface. Clade support values were also calculated. Tree summary and placement distances were calculated using custom Python scripts.

Empirical analyses:

I estimated the phylogenetic positions of fossils using a morphological matrix comprised of 51 continuous measurements gathered from pollen and seed specimens sampled across 147 extant and 8 fossil Vitaceae taxa. These data were acquired from Chen (2009). I constructed a guide tree for the extant taxa from 8 nuclear and chloroplast genes gathered from Genbank using the PHLAWD system (Soltis *et al.* 2011). Using this scaffolding, I analyzed the morphological data to estimate the positions of the fossil taxa. Individual runs were performed under all three branch length priors to assess stability across models. All analyses were run for 30,000,000 generations and visually checked for convergence. Analyses were performed with binary weights applied to the sites and compared to an unweighted analysis. To ensure that MCMC runs were not trapped in local optima, several runs were performed under each combination of settings. For each, the analysis with the highest mean likelihood was retained.

To explicitly test the informativeness of geometric morphometric data in fossil placement, I

also performed analyses on a dataset of 33 3D landmark coordinates representing 46 extant and 5 extinct fossil carnivoran crania (Jones *et al.* 2015). A reference tree composed of the 46 extant taxa was obtained from the data supplement of the original study. These coordinates were subjected to Procrustes transposition using MorphoJ (Klingenberg 2011). This yielded a matrix where each character represented the X, Y, or Z position of one landmark. The resulting traits displayed phylogenetic signal, but the transformed coordinates showed very low dispersion (variance) on an absolute scale. This appeared to flatten the likelihood surface, causing difficulites in achieving MCMC convergence. To remedy this, I scaled all of the traits to increase the absolute variance evenly across taxa evenly at each site while maintaining the original pattern of relative variances across taxa using the scale() function in R (R Core Team 2016). This procedure preserved the signal present in the original dataset, since the relative distances between taxa remained the same. Final analyses were performed on this transformed set of measurements. As with the Vitaceae dataset, I analyzed the canid data under all three branch length priors. MCMC simulations were run for 20,000,000 generations, and visually examined using Tracer v1.6 to assess convergence. Both empirical datasets achieved large ESS values.

For both datasets, I used starting trees and branch lengths generated from the rough ML method described above. Sites were weighted using the binary for the final analyses. Intermediate analyses using unweighted and float-weighted sites were also performed, and are presented in the data supplement. Dirichlet priors were assigned alpha parameters of 1.0 and beta parameters specified as the total tree length of the ML starting tree. Exponential branch length priors were assigned mean values of 1.0.

Since the empirical datasets were more complex than the simulated data, I summarized the tree distibutions as maximum clade credibility (MCC) summaries. These summaries maximize

the support of each clade. These were compared to the MAP estimates, however, and yielded generally concordant placements (supplementary material). MCC summaries were obtained using the SumTrees script that is bundled with the DendroPy package (Sukumaran and Holder 2010). Branch lengths were summarized as the mean across all sampled trees.

Software:

All analyses of fossil placements and branch lengths were performed using the new software package *cophymaru* written in the Go language. The source code is publicly available as free software at https://github.com/carolinetomo/cophymaru.

Results and Discussion

Simulations

Reconstructions of fossil placements from the simulated datasets showed that the method is generally accurate in placing fossil taxa (Table 1). In the absence of noisy traits, reconstruction is nearly always correct, with the reconstructed position of each fossil placed less than 0.1 nodes away from the true position on average. In the presence of random noise, the reconstructions are fairly accurate, except when noise becomes severe. Nevertheless, even in extreme cases, estimated positions tend to fall within the correct region of the tree, falling 1.85 and ~3 nodes away from the correct placement on average when alignments contain an equal number of clean and dirty sites.

Overall, binary weighting shows improved accuracy over float and unweighted analyses. However, despite the apparent advantage of binary weighting, it is possible that the float weighting scheme could remain beneficial in cases where the distribution of noise varies betwen different regions of trees. This is because the float weighting scheme limits the contribution of noisy sites to the likelihood rather than entirely excluding them. This possibility was not examined in this set of simulations, since the dirty traits were generated to reflect completely

random noise. However, in reality, noise may be structured to display . In these cases, continuous traits may display misleading signal among some subset of taxa, but correctly informative signal among other subsets. Further work will be needed to determine whether absolute float weight values scale predictably with the noisiness of particular sites across clades.

Overall, the simulations demonstrate the efficacy of the method for the phylogenetic placement of fossils and provide a validation of the computational implementation. The analysis of clean datasets shows that the method performs well, estimating fossil placements with very low error when signal is clear. The adaptation of Berger and Stamatakis' (2010) site weight calibration approach also appears to effectively filter through noisy datasets to improve estimation. The binary weight calibrations appear particularly effective at dealing with rampant misleading random noise, with improving accuracy by 2 to 20 times depending on the relative proportion of signal and noise compared to unweighted analyses. These results show promise toward the prospect of applying the method developed in this work to the analysis of large-scale morphometric datasets, where significant noise might be expected. Although introducing noise to unweighted analyses decreases reconstruction accuracy, the method performs predictably, and still manages to place fossils on average within the correct neighborhood. However, when weighting schemes are applied, the performance improves drastically, highlighting the promise of this method for the analysis of empirical datasets.

dataset	binary_weights	float_weights	unweighted
50 clean	0.065	0.067	0.059
50 clean + 10 dirty	0.156	1.234	2.516
50 clean + 25 dirty	0.965	2.623	3.105

dataset	binary_weights	float_weights	unweighted
50 clean + 50 dirty	1.85	3.069	3.348

Table 1. Mean distances of true and reconstructed fossil placements. Distances are measured as the average number of nodes separating reconstructed placements from their true positions across all 100 replicates of each dataset.

Vitaceae dataset:

Application of the fossil placement method to the Vitaceae dataset showed generally positive results (Fig. 1, Fig. S1). The weight calibration procedure revealed substantial noise in the dataset, with 10-12 of 51 sites failing to favor the molecular reference tree over the random trees at least 95% of the time across all runs. Despite this noise, the binary weighting scheme appeared to adequately filter through this noise to generate biologically reasonable results. *Vitis tiffneyi*, *Parthenocissus_clarnensis* , and *Ampelopsis rooseae* all share clades with the extant members of their respective genera. *Palaeovitis_paradoxa*, and *Cissocarpus jackesiae*, which represent genera with no extant species, both group confidently with separate, non-monophyletic groups of crown *Cissus*. *Ampelocissus wildei* is placed convidently along with crown *Cissus*, separated by only a node from *Palaeovitis paradoxa*. All six of these taxa are stable in their placements, grouping within the same clades across runs, and when both the exponential and empirical compound dirichlet priors are applied.

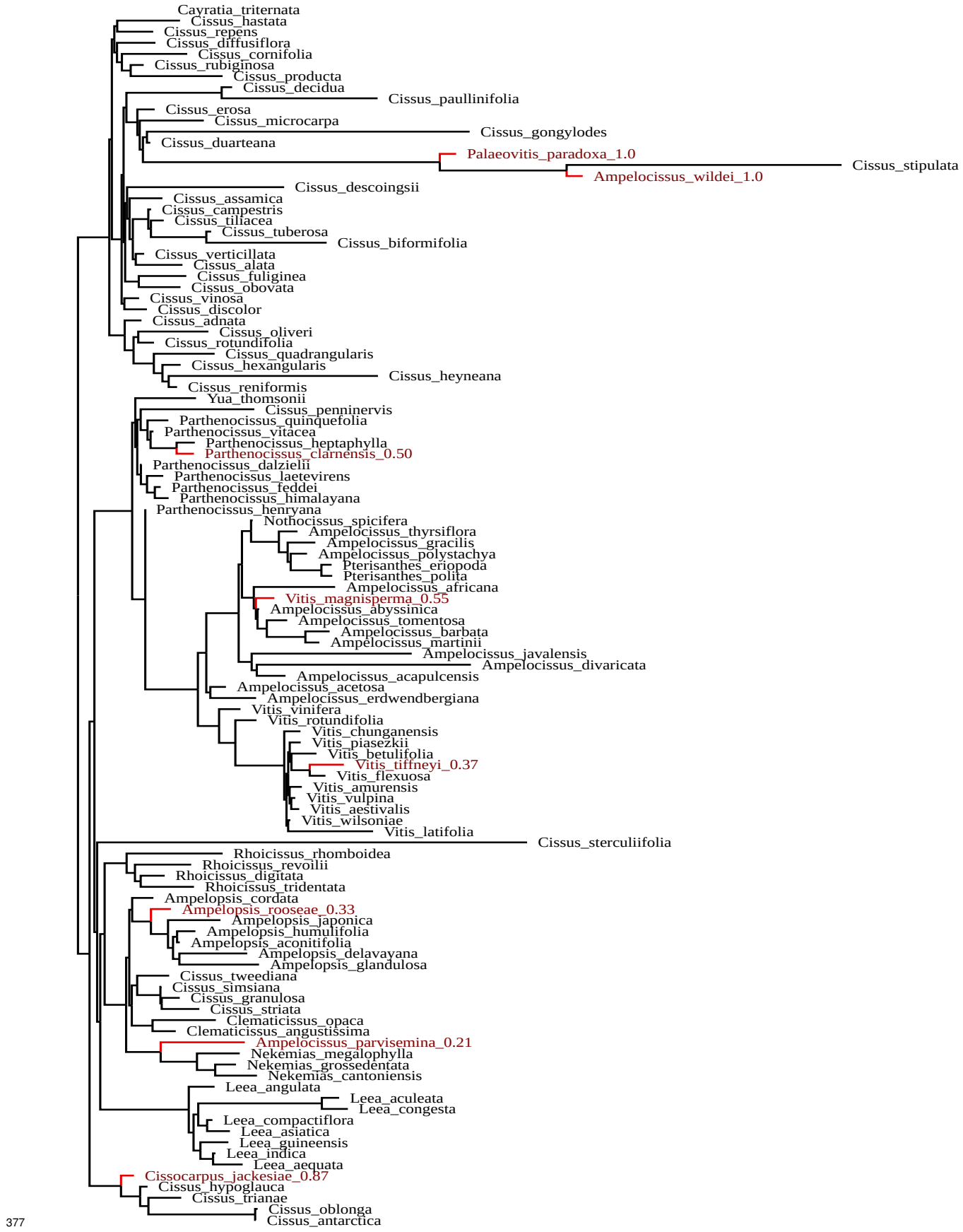

Cayratia_triternata
Cissus_hastata
Cissus_repens
Cissus_diffusiflora
Cissus_cornifolia
Cissus_rubiginosa
Cissus_producta
Cissus_decidua
Cissus_paullinifolia
Cissus_erosa
Cissus_microcarpa
Cissus_gongylodes
Cissus_duarteana
Palaeovitis_paradoxa_1.0
Cissus_stipulata
Ampelocissus_wildei_1.0
Cissus_descoingsii
Cissus_assamica
Cissus_campestris
Cissus_tiliacea
Cissus_tuberosa
Cissus_biformifolia
Cissus_verticillata
Cissus_alata
Cissus_fuliginea
Cissus_obovata
Cissus_vinosa
Cissus_discolor
Cissus_adnata
Cissus_oliveri
Cissus_rotundifolia
Cissus_quadrangularis
Cissus_hexangularis
Cissus_heyneana
Cissus_reniformis
Yua_thomsonii
Cissus_penninervis
Parthenocissus_quinquefolia
Parthenocissus_vitacea
Parthenocissus_heptaphylla
Parthenocissus_clarnensis_0.50
Parthenocissus_dalzielii
Parthenocissus_laetevirens
Parthenocissus_feddei
Parthenocissus_himalayana
Parthenocissus_henryana
Nothocissus_spicifera
Ampelocissus_thyrsiflora
Ampelocissus_gracilis
Ampelocissus_polystachya
Pterisanthes_eriopoda
Pterisanthes_polita
Ampelocissus_africana
Vitis_magnisperma_0.55
Ampelocissus_abyssinica
Ampelocissus_tomentosa
Ampelocissus_barbata
Ampelocissus_martinii
Ampelocissus_javalensis
Ampelocissus_divaricata
Ampelocissus_acapulcensis
Ampelocissus_acetosa
Ampelocissus_erdwendbergiana
Vitis_vinifera
Vitis_rotundifolia
Vitis_chunganensis
Vitis_piasezkii
Vitis_betulifolia
Vitis_tiffneyi_0.37
Vitis_flexuosa
Vitis_amurensis
Vitis_vulpina
Vitis_aestivalis
Vitis_wilsoniae
Vitis_latifolia
Cissus_sterculiifolia
Rhoicissus_rhomboidea
Rhoicissus_revoilii
Rhoicissus_digitata
Rhoicissus_tridentata
Ampelopsis_cordata
Ampelopsis_rooseae_0.33
Ampelopsis_japonica
Ampelopsis_humulifolia
Ampelopsis_aconitifolia
Ampelopsis_delavayana
Ampelopsis_glandulosa
Cissus_tweediana
Cissus_simsiana
Cissus_granulosa
Cissus_striata
Clematicissus_opaca
Clematicissus_angustissima
Ampelocissus_parvisemina_0.21
Nekemias_megalophylla
Nekemias_grossedentata
Nekemias_cantoniensis
Leea_angulata
Leea_aculeata
Leea_congesta
Leea_compactiflora
Leea_asiatica
Leea_guineensis
Leea_indica
Leea_aequata
Cissocarpus_jackesiae_0.87
Cissus_hypoglauca
Cissus_trianae
Cissus_oblonga
Cissus_antarctica
377

Figure 1. Vitaceae fossil placements inferred under the exponential branch length prior. Fossil taxa and branches are highlighted in red. Values following fossil tip labels indicate posterior support for placement. Topology is summarized from the posterior using the set of maximally credible clades (MCC). The placements are depicted only in the subclade containing all 6 fossils.

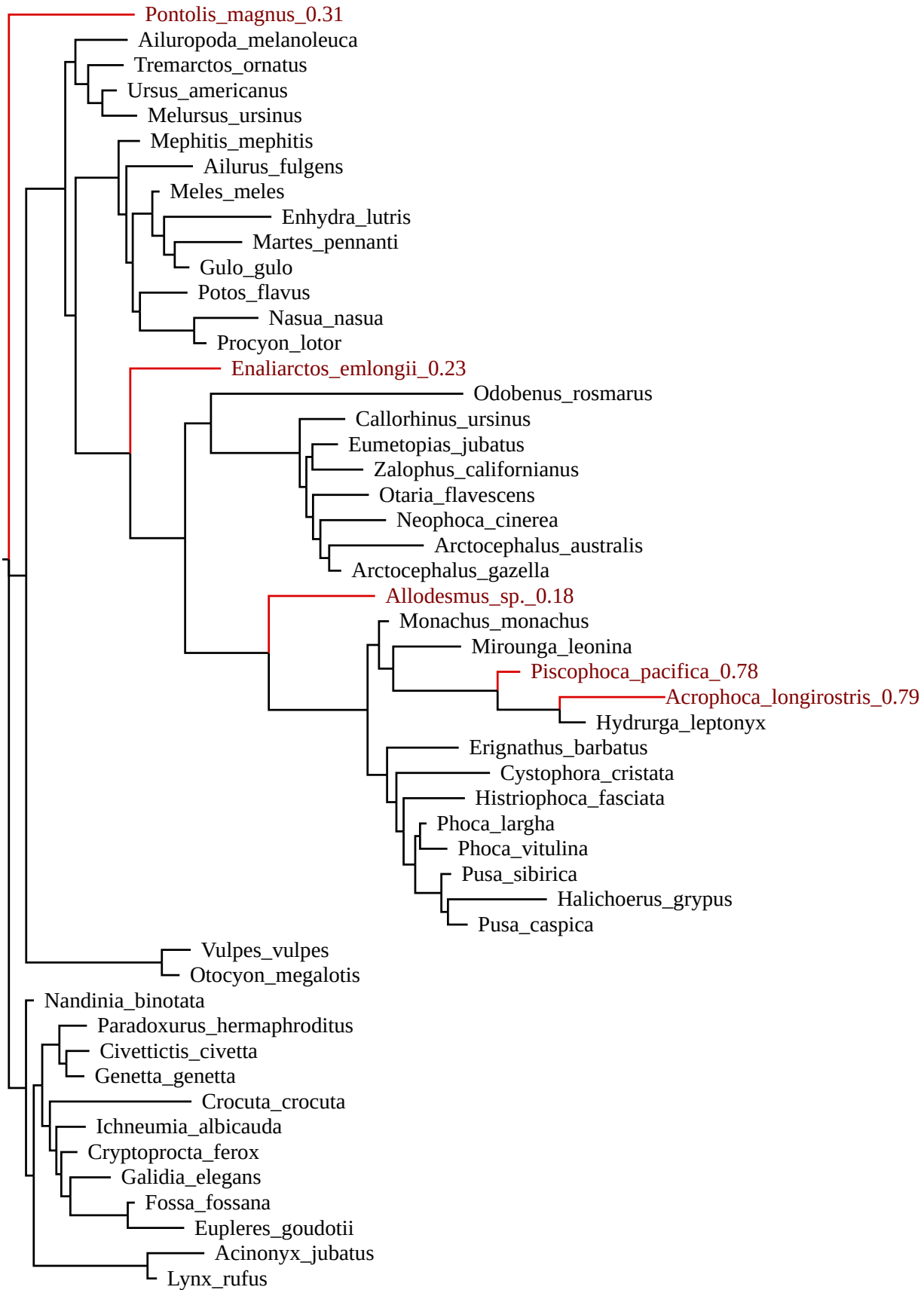
The remaining two fossils are substantially less stable in their placements. *Ampelocissus parvisemina* shows erratic placement, alternately occupying clades shared by crown *Vitis* or *Nekemias* in the best exponential and dirichlet prior runs, respectively. Its placement also changes across different runs using the same prior, and shows poor support in all cases. Under the exponential prior, the *Ampelocissus parvisemina* placement shows a 0.2 posterior probability, which decreases to 0.058 under the dirichlet prior (Fig. S1). Similarly, *Vitis magnisperma* alternately resolves into clades shared by crown *Cissus* and *Ampelocissus* under the exponential and dirichlet priors, with posterior support values of 0.23 and 0.54, respectively.

Despite the erratic behavior of the last two taxa, the low posterior support exhibited by their placements is reassuring. In many cases, fossils may simply present weak information due to shortcomings in geologic and taxonomic sampling. When this is occurs, it is unlikely that any greater certainty in their placement can be achieved except with increased data. Therefore, one major benefit to the approach described here is the clarity with which it describes uncertainty in the placement of fossils. Importantly, such uncertainty revealed by analysis under this method may improve the utility of erratic fossils. For instance, while such fossils are often omitted from analysis due to perceived unreliability, this approaches described here demonstrate more clearly the confidence and uncertainty surrounding a reconstructed set of possible placements. This uncertainty may allow the fossils to participate in node calibration for molecular dating using new methods that accommodate uncertainty in calibration placement (Heath *et al.* 2014; Guindon

2018). Since the present analyses quantify the uncertainty in fossil placements, this information may be used to construct non-arbitrary, true Bayesian priors on node ages.

Carnivoran dataset:

Analysis of the carnivoran dataset also yielded generally reasonable results (Fig. 2). The placements of *Piscophoca pacifica*, *Acrophoca longirostris*, *Enaliarctos emlongii*, and *Allodesmus* agree with previous results (Amson and de Muizon 2014; Jones *et al.* 2015). The placement of *Piscophoca pacifica* and *Acrophoca longirostris* differs slightly from the topology used by Jones et al., placing the two taxa in a more nested position. However, this placement is consistent with the results of Amson and Muison. *Enaliarctos emlongii* and *Allodesmus* resolve in positions identical to the topology used by Jones and colleagues. *Pontolis magnus* is more erratic in its placement, alternating between placement at the center of the unrooted topology, or grouping erroneously with *Vulpes* and *Otocyon*. Nevertheless, like the problem taxa in the Vitaceae example above, the placement of *Pontolis* displays reassuringly weak support, both in terms of its posterior density and in its tendency to group at the center of the tree. Interestingly, although the placements of *Enaliarctos emlongii* and *Allodesmus* remain stable across runs, both display weak support.


Pontolis_magnus_0.31
Ailuropoda_melanoleuca
Tremarctos_ornatus
Ursus_americanus
Melursus_ursinus
Mephitis_mephitis
Ailurus_fulgens
Meles_meles
Enhydra_lutris
Martes_pennanti
Gulo_gulo
Potos_flavus
Nasua_nasua
Procyon_lotor
Enaliarctos_emlongii_0.23
Odobenus_rosmarus
Callorhinus_ursinus
Eumetopias_jubatus
Zalophus_californianus
Otaria_flavescens
Neophoca_cinerea
Arctocephalus_australis
Arctocephalus_gazella
Allodesmus_sp._0.18
Monachus_monachus
Mirounga_leonina
Piscophoca_pacifica_0.78
Acrophoca_longirostris_0.79
Hydrurga_leptonyx
Erignathus_barbatus
Cystophora_cristata
Histriophoca_fasciata
Phoca_largha
Phoca_vitulina
Pusa_sibirica
Halichoerus_grypus
Pusa_caspica
Vulpes_vulpes
Otocyon_megalotis
Nandinia_binotata
Paradoxurus_hermaphroditus
Civettictis_civetta
Genetta_genetta
Crocuta_crocuta
Ichneumia_albicauda
Cryptoprocta_ferox
Galidia_elegans
Fossa_fossana
Eupleres_goudotii
Acinonyx_jubatus
Lynx_rufus

417

Figure 2. Fossil placements inferred from the carnivoran dataset using the compound Dirichlet prior. Placements are displayed as the maximum clade credibility summary of the posterior distribution of trees. Branch lengths represent morphological disparity. Values trailing fossil tip names display posterior support.

In both datasets, placement under the exponential branch length prior yields conservative estimates of uncertainty in the fossil placements, displaying generally low posterior support, except when placements are exceptionally stable such as with *Ampelocissus wildei*. This is especially important in 'rogue' taxa such as *Vitis magnisperma*. Branch support under the compound dirichlet prior is higher across several fossils in the Vitaceae dataset. The positions of the six taxa with stable behavior (listed above) do not change significantly under the compound dirichlet compared to the exponential prior. Closer examination is needed to better determine the significance of this apparent sensitivity of posterior support measures to prior choice observed in Vitaceae. The carnivoran dataset does not exhibit the same behavior, with both branch support and fossil placements similar across priors.

Moving forward, it will be important to explore the behavior of this method when applied to morphometric data collected under a variety of approaches and sampling schemes. The success of the weight calibrations on the simulated and empirical datasets suggests the possibility of applying the method to very large morphometric datasets by providing a means to filter through the noise that may occur when sampling densely across taxa and organs. Such a framework would facilitate the development of a more data-centric approach to morphological phylogenetics that reduces common sources of bias in morphological datasets. This would encourage an exploration of conflict and concordance in signal through quantitative data analysis rather than by attempting to filter during the stage of data collection. They provide a means to move forward, and should

encourage the development and exploration of more data sources to determine the most effective means to inform phylogeny using morphometry.

It is also worth noting that certain features can be uniquely described qualitatively, such as the loss and gain of structures. It would be possible, and in certain cases, useful to combine such discrete information into the morphometric framework described here. This would resemble the analysis of genomic rearrangements frequently used in large molecular phylogenetic studies. As progress in this area develops, it will be important to better understand the behavior of different sources of morphological data at different timescales, and the most appropriate ways to combine, model, and gather such datasets.

Conclusions:

The methods described here provides a new means for biologists to reliably and confidently place fossils in the tree of life. Although the simulated and empirical analyses show several imperfections and a need for further refinement of these methods, the overall accuracy and conservative assessment of uncertainty displayed in the examples appear encouraging. As molecular phylogenetics advances in its use of genomic data to answer fundamental questions across the tree of life, it will be important for morphological phylogenetics and paleontology to keep pace. Analysis of morphometric data using the approach shown here will help to improve issues surrounding subjectivity in character collection, and will help morphological datasets to scale better in the genomic era. New advances in the collection of morphometric data, combined with refinements to the approach developed here will better equip morphology to speak to major outstanding questions across the tree of life.

Supplementary figures:

Cissus_descoingsii
Cissus_assamica
Cissus_verticillata
Cissus_alata
Cissus_campestris
Cissus_tiliacea
Cissus_biformifolia
Cissus_tuberosa
Cissus_fuliginea
Cissus_obovata
Cissus_decidua
Cissus_paullinifolia
Cissus_erosa
Cissus_microcarpa
Cissus_gongylodes
Cissus_duarteana
Palaeovitis_paradoxa_1.0
Cissus_stipulata
Ampelocissus_wildei_1.0
Cissus_vinosa
Cissus_discolor
Cissus_hastata
Cissus_repens
Cissus_diffusiflora
Cissus_cornifolia
Cissus_rubiginosa
Cissus_producta
Cissus_adnata
Cissus_oliveri
Cissus_rotundifolia
Cissus_quadrangularis
Cissus_hexangularis
Cissus_heyneana
Cissus_reniformis
Cissus_trianae
Cissus_oblonga
Cissus_antarctica
Cissocarpus_jackesiae_0.55
Cissus_hypoglauca
Leea_angulata
Leea_guineensis
Leea_indica
Leea_aequata
Leea_aculeata
Leea_congesta
Leea_compactiflora
Leea_asiatica
Rhoicissus_rhomboidea
Rhoicissus_revoilii
Rhoicissus_digitata
Rhoicissus_tridentata
Nekemias_megalophylla
Nekemias_grossedentata
Nekemias_cantoniensis
Ampelopsis_cordata
Ampelopsis_japonica
Ampelopsis_delavayana
Ampelopsis_rooseae_0.42
Ampelopsis_glandulosa
Ampelopsis_humulifolia
Ampelopsis_aconitifolia
Clematicissus_angustissima
Clematicissus_opaca
Cissus_tweediana
Cissus_simsiana
Cissus_granulosa
Cissus_striata
Cissus_sterculiifolia
Vitis_magnisperma_1.0
Parthenocissus_henryana
Nothocissus_spicifera
Ampelocissus_thyrsiflora
Ampelocissus_gracilis
Ampelocissus_polystachya
Pterisanthes_eriopoda
Pterisanthes_polita
Ampelocissus_africana
Ampelocissus_abyssinica
Ampelocissus_tomentosa
Ampelocissus_barbata
Ampelocissus_martinii
Ampelocissus_javalensis
Ampelocissus_acapulcensis
Ampelocissus_divaricata
Ampelocissus_acetosa
Ampelocissus_erdwendbergiana
Vitis_vinifera
Vitis_rotundifolia
Vitis_chunganensis
Vitis_piasezkii
Vitis_betulifolia
Vitis_flexuosa
Vitis_tiffney_0.62
Ampelocissus_parvisemina_0.62
Vitis_amurensis
Vitis_vulpina
Vitis_aestivalis
Vitis_wilsoniae
Vitis_latifolia
Yua_thomsonii
Parthenocissus_dalzielii
Parthenocissus_laetevirens
Parthenocissus_feddei
Parthenocissus_himalayana
Cissus_penninervis
Parthenocissus_quinquefolia
Parthenocissus_vitacea
Parthenocissus_heptaphylla
Parthenocissus_clarnensis_0.96
463

Figure S1. Vitaceae fossil placements inferred under the compound dirichlet branch length prior. Fossil taxa and branches are highlighted in red. Values following fossil tip labels indicate posterior support for placement. Topology is summarized from the posterior using the set of maximally credible clades (MCC). The placements are depicted only in the subclade containing all 6 fossils.

References

Amson, E. and de Muizon, C. 2014. A new durophagous phocid (mammalia: Carnivora) from the late neogene of peru and considerations on monachine seals phylogeny. *Journal of Systematic Palaeontology*, 12(5): 523–548.

Berger, S. A. and Stamatakis, A. 2010. Accuracy of morphology-based phylogenetic fossil placement under maximum likelihood. pages 1–9.

Brazeau, M. D. 2011. Problematic character coding methods in morphology and their effects. *Biological Journal of the Linnean Society*, 104(3): 489–498.

Butler, M. A. and King, A. A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.*, 164(6): 683–695.

Catalano, S. A., Goloboff, P. A., and Giannini, N. P. 2010. Phylogenetic morphometrics (i): the use of landmark data in a phylogenetic framework. *Cladistics*, 26(5): 539–549.

Cavalli-Sforza, L. L. and Edwards, A. W. 1967. Phylogenetic analysis. models and estimation procedures. *Am. J. Hum. Genet.*, 19(3): 233.

Chen, I. 2009. *History of Vitaceae inferred from morphology-based phylogeny and the fossil record of seeds*. University of Florida.

Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.*, 25(5): 471.

Felsenstein, J. 1981. Evolutionary trees from dna sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17(6): 368–376.

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics*, 19(1): 445–471.

Felsenstein, J. 2002. Quantitative characters, phylogenies, and morphometrics. *Morphology, shape and phylogeny*, pages 27–44.

González-José, R., Escapa, I., Neves, W. A., Cúneo, R., and Pucciarelli, H. M. 2008. Cladistic analysis of continuous modularized traits provides phylogenetic signals in homo evolution. *Nature*, 453(7196): 775.

Guindon, S. 2018. Accounting for calibration uncertainty: Bayesian molecular dating as a "doubly intractable" problem. *Syst. Biol.*

Hastings, W. K. 1970. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1): 97–109.

Hauser, D. L. and Presch, W. 1991. The effect of ordered characters on phylogenetic reconstruction. *Cladistics*, 7(3): 243–265.

Hawkins, J. A., Hughes, C. E., and Scotland, R. W. 1997. Primary homology assessment, characters and character states. *Cladistics*, 13(3): 275–283.

Heath, T. A., Huelsenbeck, J. P., and Stadler, T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences*, 111(29): E2957–E2966.

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., and Ronquist, F. 2016. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.*, 65(4): 726–736.

Jones, K. E., Smaers, J. B., and Goswami, A. 2015. Impact of the terrestrial-aquatic transition on disparity and rates of evolution in the carnivoran skull. *BMC evolutionary biology*, 15(1): 8.

Klingenberg, C. P. 2011. Morphoj: an integrated software package for geometric morphometrics. *Molecular ecology resources*, 11(2): 353–357.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.*, 50(6): 913–925.

MacLeod, N. 2002. Phylogenetic signals in morphometric data. *Morphology, shape and phylogeny*, 100: 138.

O'Meara, B. C., Ané, C., Sanderson, M. J., and Wainwright, P. C. 2006. Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5): 922–933.

Parins-Fukuchi, C. 2017. Use of continuous traits can improve morphological phylogenetics. *Syst. Biol.*

Pleijel, F. 1995. On character coding for phylogeny reconstruction. *Cladistics*, 11(3): 309–315.

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of lissamphibia. *Syst. Biol.*, 60(4): 466–481.

R Core Team 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rannala, B., Zhu, T., and Yang, Z. 2011. Tail paradox, partial identifiability, and influential priors in bayesian branch length inference. *Molecular biology and evolution*, 29(1): 325–335.

Revell, L. J. 2012. phytools: an r package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 3(2): 217–223.

Ronquist, F., Klopfstein, S., Vilhelmsen, L., Schulmeister, S., Murray, D. L., and Rasnitsyn, A. P. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the hymenoptera. *Syst. Biol.*, 61(6): 973–999.

Ronquist, F., Lartillot, N., and Phillips, M. J. 2016. Closing the gap between rocks and clocks using total-evidence dating. *Phil. Trans. R. Soc. B*, 371(1699): 20150136.

Scotland, R. and Pennington, R. T. 2000. *Homology and systematics: coding characters for*

phylogenetic analysis. CRC Press.

Scotland, R. W., Olmstead, R. G., and Bennett, J. R. 2003. Phylogeny reconstruction: the role of morphology. *Syst. Biol.*, 52(4): 539–548.

Simões, T. R., Caldwell, M. W., Palci, A., and Nydam, R. L. 2017. Giant taxon-character matrices: quality of character constructions remains critical regardless of size. *Cladistics*, 33(2): 198–219.

Slater, G. J., Harmon, L. J., and Alfaro, M. E. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution*, 66(12): 3931–3944.

Smith, U. E. and Hendricks, J. R. 2013. Geometric morphometric character suites as phylogenetic data: extracting phylogenetic signal from gastropod shells. *Syst. Biol.*, 62: 366–385.

Soltis, D. E., Smith, S. A., Cellinese, N., Wurdack, K. J., Tank, D. C., Brockington, S. F., Refulio-Rodriguez, N. F., Walker, J. B., Moore, M. J., Carlsward, B. S., *et al.* 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany*, 98(4): 704–730.

Sukumaran, J. and Holder, M. T. 2010. Dendropy: a python library for phylogenetic computing. *Bioinformatics*, 26(12): 1569–1571.

Wiens, J. J. 2009. Paleontology, genomics, and combined-data phylogenetics: can molecular data improve phylogeny estimation for fossil taxa? *Systematic Biology*, 58(1): 87–99.

Wiens, J. J., Kuczynski, C. A., Townsend, T., Reeder, T. W., Mulcahy, D. G., and Sites, J. W. 2010. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. *Syst. Biol.*, 59(6): 674–688.

Wilkinson, M. 1995. A comparison of two methods of character construction. *Cladistics*, 11(3): 297–308.

Wright, A. M. and Hillis, D. M. 2014. Bayesian analysis using a simple likelihood model

outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS One*, 9(10): e109210.

Yang, Z. 2014. *Molecular evolution: a statistical approach*. Oxford University Press.