

# Polymorphic characters in the reconstruction of the phylogeny of geoemydid turtles

RAFAELLA C. GARBIN<sup>1\*</sup>, EDUARDO ASCARRUNZ<sup>1†</sup> and WALTER G. JOYCE<sup>1</sup>

<sup>1</sup>Department of Geosciences, University of Fribourg, Chemin du Musée 6, 1700 Fribourg, Switzerland

Received 10 March 2017; revised 27 October 2017; accepted for publication 22 November 2017

Several attempts to resolve the phylogeny of turtles in the clade Geoemydidae using morphology have been unsuccessful, in part because of unusually high levels of polymorphism. This has hindered the integration of the geoemydid fossil record into a phylogenetic framework. Many methods, shown to improve phylogenetic inference, allow the incorporation of different amounts of state frequency information from polymorphic characters into a phylogenetic analysis. Here, we present a new character matrix for the shell of geoemydids and assess the performance of polymorphism coding methods ('majority', 'generalized frequency coding', 'polymorphic' and 'missing') in a phylogenetic analysis by comparing the result topology of each method with a reference molecular phylogeny. The four coding methods failed to recover trees that were both well resolved and highly congruent with the reference phylogeny. Moreover, contrary to previous studies, the coding methods that made more use of character states frequencies did not perform better. However, a leave-one-out subsampling analysis suggested that despite these problems, the new matrix can still be used to place fossils in the geoemydid phylogeny with some reliability. Finally, we provide a list of characters that diagnose the major clades in our molecular reference tree.

ADDITIONAL KEYWORDS: morphology – palaeontology – systematics – Testudines.

## INTRODUCTION

Geoemydidae is a diverse clade of turtles with ~71 extant species currently distributed in the tropical to temperate regions of Asia, Europe, North Africa and the Americas (Ernst & Barbour, 1989; Turtle Taxonomy Working Group, 2017). The group has a particularly rich fossil record from the Tertiary period in the Northern Hemisphere (Hay, 1908; Mlynarski, 1976; Lapparent de Broin, 2001; Claude & Tong, 2004; Hervet, 2004; Danilov, 2005; Hutchinson, 2006), but little is known about the group's evolutionary history, as the phylogenetic relationships of most fossils have not been established with any confidence.

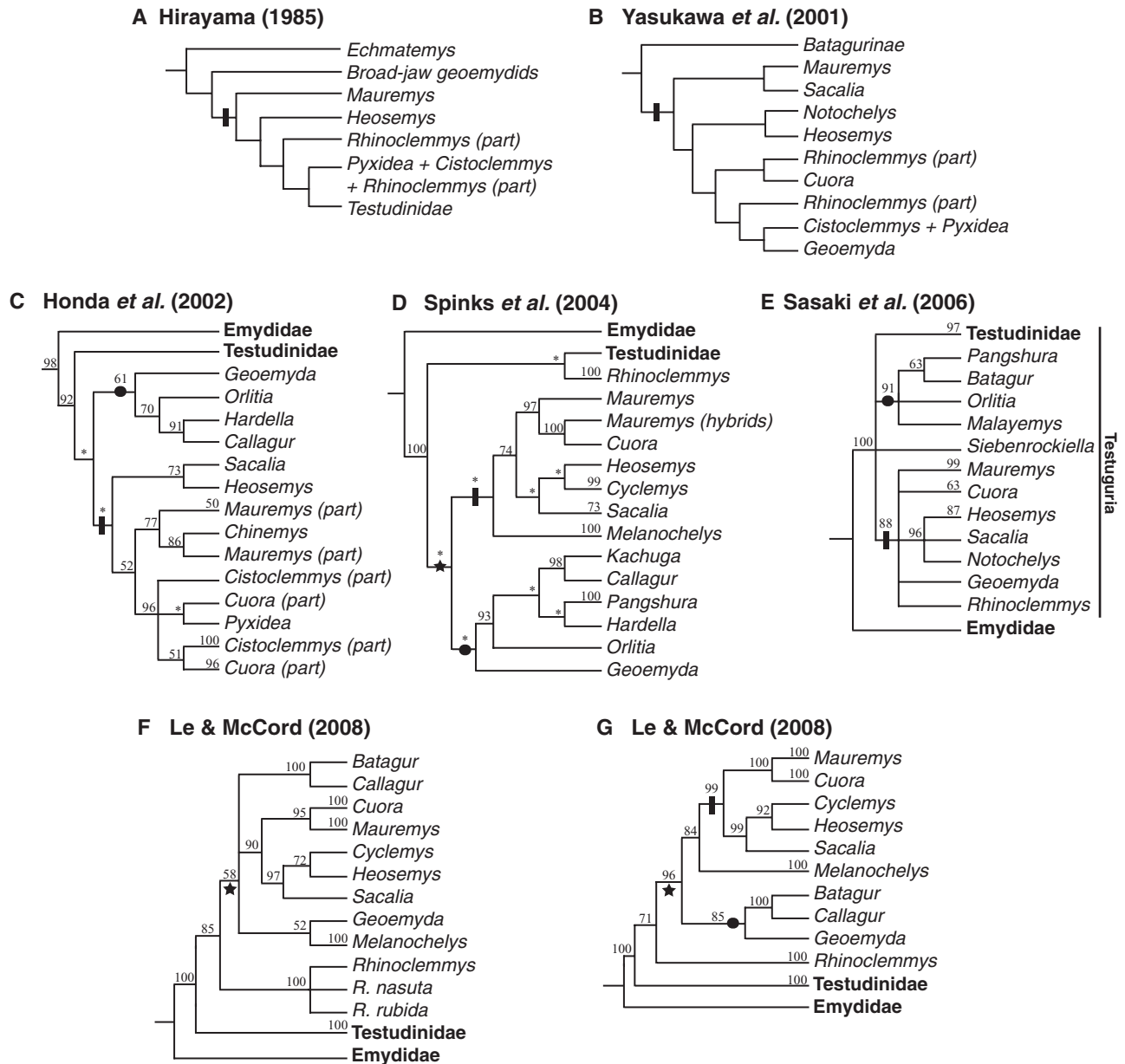
The phylogeny of extant geoemydids has been studied using morphological and, more recently, molecular data (Hirayama, 1985; Yasukawa, Hirayama & Hikida, 2001; Honda *et al.*, 2002; Barth *et al.*, 2004; Spinks *et al.*, 2004; Diesmos *et al.*, 2005; Sasaki *et al.*, 2006; Le & McCord, 2008). Although molecular studies agree overall (Honda *et al.*, 2002; Barth *et al.*, 2004;

Spinks *et al.*, 2004; Diesmos *et al.*, 2005; Sasaki *et al.*, 2006; Le & McCord, 2008), morphological studies do not (Hirayama, 1985; Yasukawa *et al.*, 2001; Claude & Tong, 2004), conflicting both with each other and with the molecular data (Fig. 1).

The topological disagreement between molecular and morphological phylogenies (Fig. 1) has been interpreted as indicative of high morphological homoplasy (Sasaki *et al.*, 2006; Claude *et al.*, 2012). For instance, the morphology-based phylogeny by Hirayama (1985) suggested that various geoemydids with a secondary palate form a monophyletic group, as they share many traits related to the expansion of the triturating surfaces, but molecular phylogenies (e.g. Spinks *et al.*, 2004) suggest that geoemydids evolved secondary palates multiple times, a conclusion that also makes morphological sense, as differences in secondary palate type can be distinguished among the groups recovered in molecular analyses. Likewise, whereas the phylogeny by Hirayama (1985) united all geoemydids with a 'reversed' neural pattern, rendering the clades *Cuora* and *Rhinoclemmys* polyphyletic, the topology by Spinks *et al.* (2004) concluded that these taxa are monophyletic and that 'reversed' neural patterns were

\*Corresponding author. E-mail: [rafaella.decarvalhogarbin@unifr.ch](mailto:rafaella.decarvalhogarbin@unifr.ch)

†These authors contributed equally to this work.



**Figure 1.** A summary of previously proposed phylogenetic relationships for Geoemydidae. Cladograms were shortened from the originals for simplicity. A, Hirayama's (1985) manual cladistic analysis, with 82 morphological characters and four chromosomal. B, maximum parsimony (MP) tree from Yasukawa et al. (2001), using 35 morphological characters. C, molecular phylogeny from Honda et al. (2002), using 12S and 16S, MP optimization. D, one of the topologies retrieved from Spinks et al. (2004) using 12S, cytochrome *b* and R35 maximum likelihood (ML). E, Sasaki et al. (2006) topology from short interspersed nuclear elements (SINE) insertion into 49 loci. F, strict consensus of MP topology from Le & McCord (2008), using 12S, 16S, Rag1, C-mos and cytochrome *b*. G, ML and Bayesian analysis topology from Le & McCord (2008) for the same genes as in F. Length of branches does not correspond to rate of substitution per site in ML analyses. Bootstrap values > 50% are shown over the branches when applicable. Star indicates a clade consisting of the Old-World Geoemydidae (all species except *Rhinoclemmys*). Small bar indicates a clade corresponding to the 'narrow-jaw' geoemydids of Hirayama (1985). Circle indicates a clade corresponding to the 'broad-jaw' geoemydids that was retrieved in different studies. Taxonomic changes: after Honda et al. (2002), the genera *Pyxidea* and *Cistoclemmys* have been synonymized with *Cuora*, as well as *Kachuga* and *Callagur* with *Batagur* (after Le et al., 2007).

acquired at least twice, a conclusion otherwise supported by biogeographical data.

Focusing on solving disagreements attributed to morphological characters, [Joyce & Bell \(2004\)](#) reviewed morphological characters used in testudinoid systematics (e.g. [Hirayama, 1985](#)) and produced a new character matrix. However, the authors decided not to publish a phylogenetic analysis because they did not believe the maximum parsimony phylogeny reconstructed from their morphological matrix, which was incongruent with the molecular phylogenies and traditional turtle taxonomy. [Joyce & Bell \(2004\)](#) examined multiple specimens for several species, enabling them to attest to the presence of polymorphism in many characters. They concluded that a more detailed study of polymorphism, and possibly new methods, were necessary to recover the phylogenetic signal from morphological data. In the present study, we address the magnitude and effects of polymorphism with a revised data set.

In general, the effects of polymorphism depend on the phylogenetic informativeness of the state frequencies within the characters and the coding method, i.e. the strategy used to represent the observation of multiple states of a character in a single terminal taxon. When coding methods that do not capture information about state frequency are used, every occurrence of polymorphism will decrease the explanatory power of a character under the criterion of maximum parsimony and the most common implementations of maximum-likelihood inference packages. This is because cells with multiple states are compatible with more bipartitions than cells coded for a single state. In matrices with a modest character/species ratio, the depressed explanatory power of the characters can have a significant impact on the resolution of the tree. Furthermore, the same phenomenon allows characters that are more informative, but misleading, to have a greater influence on the topology. Yet, if state frequencies can be estimated reliably and are phylogenetically informative, frequency-aware methods of polymorphic coding would, in theory, enhance phylogenetic performance ([Wiens, 1999, 2001](#)). With the aim of exploiting the latter possibility, a set of methods have been developed that allow the use of state frequencies of polymorphic characters ([Wiens, 1999](#)), but they have not been used extensively. We constructed a morphological dataset based on new and revised definitions of shell characters (mainly from [Hirayama, 1985](#); [Claude & Tong, 2004](#); [Joyce & Bell, 2004](#)), with the goal of laying the foundation for the future integration of fossil data into geoemydid phylogeny. We recorded data in a master specimen-level matrix that allows us to document the frequency of intraspecific character variation. We applied four coding methods to deal with polymorphism and assessed the performance of

the methods through comparison with a reference molecular phylogeny. Additionally, we discuss possible synapomorphies for select geoemydid subclades.

## MATERIAL AND METHODS

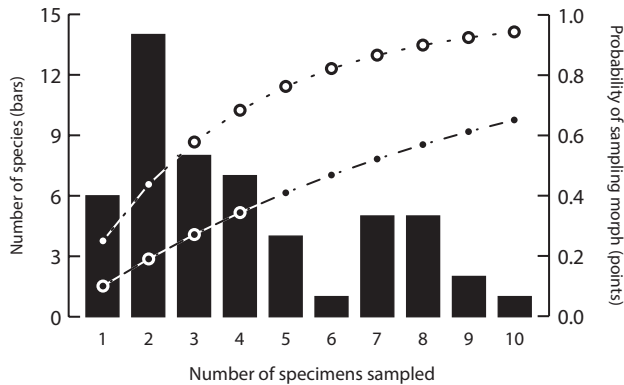
### TAXON SAMPLING

To explore the impact of polymorphic data on the phylogeny of geoemydid turtles, we examined and scored the skeletal morphology of 225 adult or subadult specimens lacking obvious deformities, representing 51 of the 71 extant geoemydid species currently accepted ([Turtle Taxonomy Working Group, 2017](#)), with two extant species of testudinids [*Stigmochelys pardalis* (Bell, 1828) and *Gopherus polyphemus* (Daudin, 1801)] and three extant species of emydids [*Emys blandingii* Duméril, 1805, *Chrysemys picta* (Schneider, 1783) and *Clemmys guttata* (Schneider, 1792)] serving as the outgroups. We conducted our analyses at the species level, according to the updated alpha taxonomy presented by the [Turtle Taxonomy Working Group \(2017\)](#). Although we attempted to include as many specimens as possible per species, we were able to include an average of only 3.88 specimens per species ([Fig. 2](#)), as well-prepared skeletal specimens of wild-caught geoemydids are rare in museum collections and because of time constraints.

All examined specimens are housed in the following institutions: American Museum of Natural History, New York, NY, USA (AMNH); Field Museum of Natural History, Chicago, IL, USA (FMNH); Muséum d'Histoire Naturelle de la Ville de Genève, Geneva, Switzerland (MHNG); Museum für Tierkunde Senckenberg, Dresden, Germany (MTD); Chelonian Research Institute, Oviedo, FL, USA (PCHP); National Museum of Natural History, Washington, DC, USA (USNM); Naturhistorisches Museum Basel, Basel, Switzerland (NMB); and Naturhistorisches Museum Wien, Vienna, Austria (NMW). The complete list of specimens is provided in the Supporting Information (Appendix S1).

### CHARACTER SAMPLING

A total of 80 shell characters were scored for each specimen: 55 from the carapace and 25 from the plastron. These characters were modified from previous character lists ([Hirayama, 1985](#); [Gaffney & Meylan, 1988](#); [Meylan & Sterrer, 2000](#); [Yasukawa \*et al.\*, 2001](#); [Stephens & Wiens, 2003](#); [Joyce & Bell, 2004](#); [Claude & Tong, 2004](#); [Hervet, 2006](#); [Joyce & Lyson, 2010](#)), newly extracted from descriptive studies ([Boulenger, 1889](#); [Ernst & Barbour, 1989](#)), or created based on personal observations. As most fossils from this group are only shell material, we restricted character selection



**Figure 2.** Distribution of specimen sample sizes for the species studied. The probability of sampling a morph that has a frequency of 10% (closed points) or 25% (open points) in the population, as calculated using equation 1 from Wiens & Servedio (2000), is also shown to highlight whether sampling is sufficient.

to this region. A complete list of character definitions and discussions is given in the Supporting Information (Appendix S2). For pictures of most characters, see the Supporting Information (Appendix S8, Figs S7–S74).

#### CODING METHODS

From the coding methods available to address polymorphic characters, we selected four, referred to here as majority (MJ), missing (MI), generalized frequency coding (GFC) and polymorphic (PL). These methods were chosen because they represent a wide range of strategies for incorporating state frequency information and can be applied to complex character matrices with multistate characters. Furthermore, for these methods the number of terminal analytical units that can be used is unlimited, in contrast to step-matrix frequency coding, which can only be used with up to 32 terminals with PAUP\* (Wiens, 2001; Lawing, Meik & Schargel, 2008). We created four separate species-level character matrices from our specimen-level character matrix using the four selected coding schemes and protocols described in the following subsections. All matrices can be found in the Supporting Information (Appendix S3).

##### Majority (MJ)

The MJ method was originally developed to code allelic polymorphism (Johnson, Zink & Marten, 1988). Each character is simply coded for its most frequent (modal) state. If there is a tie between the most common states, those states are left in the corresponding cell (e.g. if the most common states are zero and two, the cell is coded as '(0 2)' in the NEXUS file). Whenever a specimen shows left–right variation for a character (thereby creating a polymorphic cell), we considered

only the most common state for that species. Coding was expedited through the use of a script in R (R Core Team, 2016; Supporting Information, Appendix S4).

##### Missing (MI)

For the MI method (Pimentel & Riggins, 1987), any polymorphic cell in the matrix is coded as missing data (i.e. '?'). All information pertaining to polymorphism is thereby lost. Therefore, if two species share the presence of a derived state, but this state is not fixed in both species, the derived state will not support the grouping of the two species (Wiens, 1995, 1999). PL and MI produce identical parsimony scores for binary characters, as noted by Wiens (1995).

##### Generalized frequency coding (GFC)

The GFC method (Smith & Gutberlet, 2001) is similar to other frequency-coding methods (e.g. frequency-bins, step-matrix frequency coding), as it uses the precise frequency of traits within a species when coding polymorphism. In this method, each character state of the original matrix is rephrased as a new character ('subcharacter' of Smith & Gutberlet, 2001) in a new, expanded matrix. The cumulative frequencies of states in each character are calculated and attributed to new subcharacters. Then, the values of the subcharacters are coded based on the frequency-bin method (Wiens, 1993, 1995). After deleting the non-informative subcharacters, a maximal weight is given to all subcharacters and then adjusted accordingly. Unequal subcharacter weighting (USW) gives a different weight to each subcharacter (from the same group) according to the number of steps and number of informative subcharacters in the group. We used FastMorphologyGFC v.1.0 (Chang & Smith, 2001) to generate a GFC-coded species-level matrix including frequencies, cumulative frequencies and different weights. For more details, refer to Smith & Gutberlet (2001).

##### Polymorphic (PL)

The PL method (Campbell & Frost, 1993) leaves all observed states coded in the cell without any changes (e.g. '0&1'). As most of the characters in this study are multistate, some species presented more than two states when scored (Fig. 3). This method was implemented by merging specimen data into their respective species in Mesquite v.3.04 build 725 (Maddison & Maddison, 2016).

#### PHYLOGENETIC ANALYSES

##### Reference phylogeny

To build a molecular reference tree that combines different types of molecular evidence, we performed a maximum likelihood (ML) analysis of a concatenated matrix of three mitochondrial (12S, cytochrome *c* oxidase I and cytochrome *b*) and four nuclear loci (R35

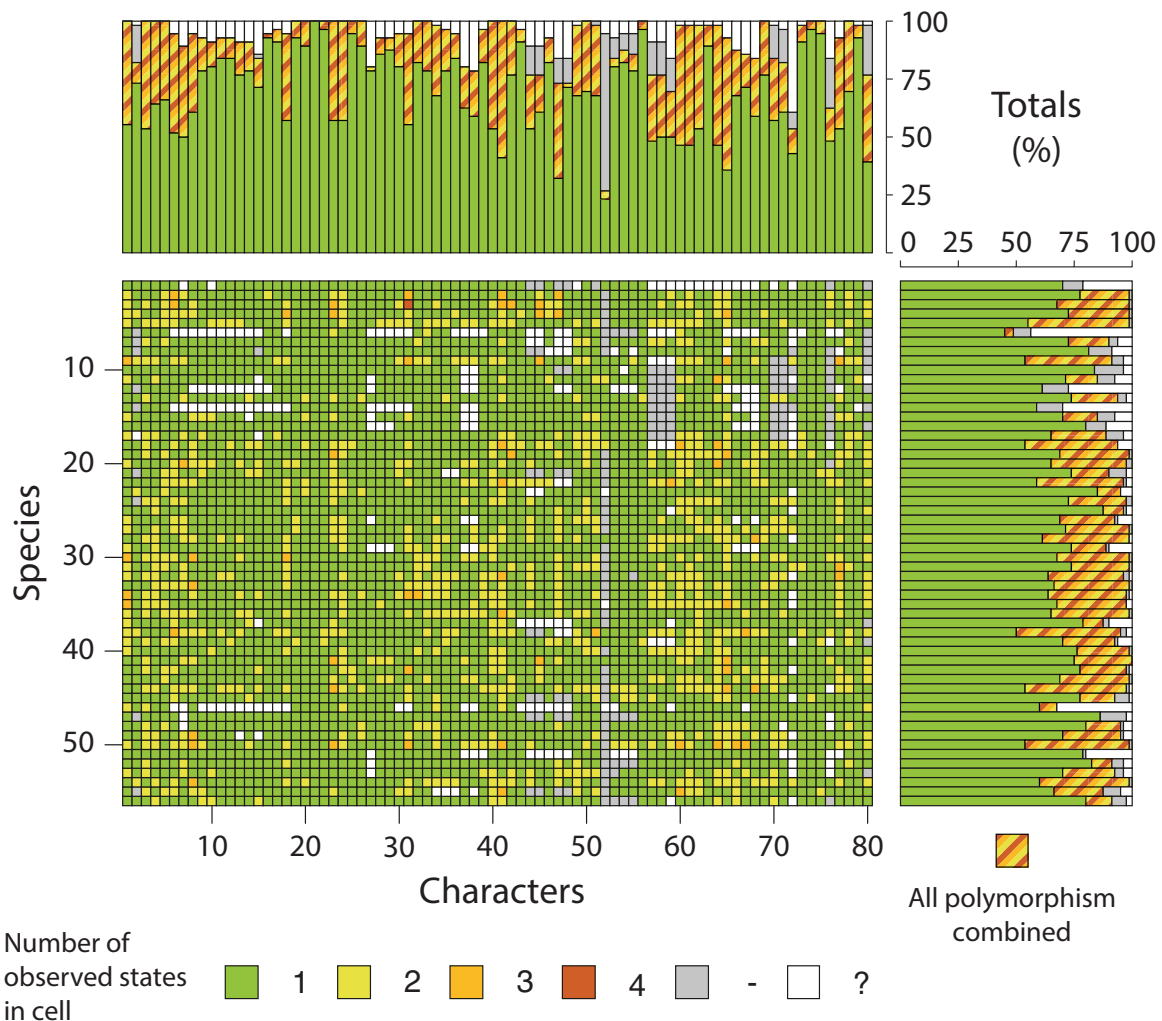


intron, c-mos, Rag1 and Rag2). Sequence sampling prioritized the maximal inclusion of geoemydid species and used only sequences previously published by Honda *et al.* (2002), Spinks *et al.* (2004) and Le & McCord (2008) (see Supporting Information, Appendix S5 for GenBank accession numbers), or available on the BOLD workbench databases in the case of cytochrome c oxidase I (Ratnasingham & Herbert, 2007). The minimal level of sampling and analysis was the species. Sequences belonging to different subspecies were therefore combined to maximize locus sampling for each species. In total, we sampled the molecular sequences of 60 geoemydid

species, with five testudinid species and one emydid as outgroups.

Sequence alignment, when required, was first performed in MUSCLE v. 3.8.31 (Edgar, 2004), then visually inspected and manually adjusted as needed. We concatenated all sequences into a single matrix of 4299 sites, but locus sampling was patchy (Supporting Information, Appendix S6).

Selection of partition scheme and model was performed using the greedy merging strategy (Lanfear *et al.*, 2014) implemented in IQTree v.1.5.4 (Nguyen *et al.*, 2015; Chernomor, von Haeseler & Minh, 2016). Sequence composition was homogeneous within each



**Figure 3.** Distribution of polymorphism in the species-level morphological character matrix (with polymorphic coding method). The graphs surrounding the central image show the proportion of fixed, polymorphic, missing and inapplicable cells by character (upper graph) and by species (right graph). Out of the 4480 data cells, 963 display polymorphism (22%), and 48 out of 56 surveyed species (86%) have at least one character coded as polymorphic. All 80 characters have polymorphism for at least one species, except character 21, which is also uninformative. Character 21 relates to the presence or absence of the cervical scute. The absence is autopomorphic for *Stigmochelys pardalis*, but reportedly shared with other testudinids absent in our sample (Joyce & Bell, 2004).

locus, according to IQTree's  $\chi^2$  composition test. For the ML tree estimation, we performed 15 heuristic searches in IQTree. Branch support was assessed with 5000 ultra-fast bootstrap replicates (Minh, Nguyen & von Haeseler, 2013) and branches with < 95% ultra-fast bootstrap support were collapsed. The data were also analysed under maximum parsimony (MP) using TNT v. 1.5 (Goloboff & Catalano, 2016), following the same procedure described for the morphological data.

We based our reference tree on the maximum-likelihood topology. As we did not have sequences of the testudinid *S. pardalis*, we manually placed it as sister to the other testudinid, *G. polyphemus*, in the reference tree. For the final reference tree, we also pruned all species not represented in the morphological matrix. All files used to produce the reference tree are provided in the Supporting Information (Appendix S6).

#### *Analysis of morphological matrices*

The four species-level character matrices were analysed under the parsimony optimality criterion with TNT. The choice of parsimony for the phylogenetic analyses was to accommodate the four coding methods, as GFC is not compatible with Bayesian and ML analysis. All multistate characters were considered ordered, except for character 8 (or the equivalent subcharacters in the GFC matrix). Tree searching was performed with 5000 rounds of random addition sequence with TBR optimization holding up to 10 trees, followed by TBR branch swapping of the trees retained in the tree buffer. Bootstrap support values were calculated with 1000 replications of symmetrical subsampling, each with 100 rounds of the random addition sequence with TBR optimization. Branches without support or ambiguous support were collapsed (Coddington–Schraff's rule 3; Coddington & Schraff, 1994). If more than one most parsimonious tree (MPT) was found, a strict consensus was calculated. For bootstrap values, the majority consensus was preferred. In all analyses, the emydid species *Emys blandingii* was set as the outgroup.

As a post-analysis, we used the iterPCR method of Pol & Escapa (2009) to identify possible 'rogue' terminals or clades that could be having a disproportionate negative effect on topological resolution. Dropping 'rogue' terminals from the MPTs can significantly improve the resolution of the strict consensus.

In addition to the analyses of morphology alone, we conducted a total evidence analysis (TEA) using the same search strategy described above on a concatenated matrix that combined the molecular data and the PL matrix.

#### COMPARISON OF METHODS AND PERFORMANCE CRITERIA

##### *Accuracy and resolution*

In this paper, we consider topological 'accuracy' to mean the degree to which the resolved relationships of a given tree are compatible with those of a reference tree. In other words, accuracy does not take into account relationships that may be resolved in one tree but unresolved in the other. This conception follows the 'soft' interpretation of polytomies (Maddison, 1989), where internal nodes of a degree greater than three represent a failure to discriminate among the possible dichotomous resolutions, rather than true polytomous speciation events. For the quantification of topological accuracy, we used the contradiction difference (CD) of Bapst, Schreiber & Carlson (2017). The CD is defined as the total number of conflicting splits across two topologies with equal numbers of terminals and can be scaled by dividing that quantity by  $2 \times (\text{number of shared terminals} - 2)$ , corresponding to the maximal number of splits that could be in conflict between the two phylogenies. We use the scaled CD to measure the percentage of topological accuracy of the tree, defined as  $100 \times (1 - \text{CD}_{\text{scaled}})$ , where 0% accuracy means total split conflict between the two trees, and 100% means total lack of conflict (agreement between relationships that are resolved in both trees). Therefore, our measure of accuracy is never depressed by the presence of polytomies.

We define topological resolution as the degree to which a tree resolves relationships, correctly or incorrectly. This is measured as the ratio between the number of internal nodes that are resolved (i.e. of degree 3) in the tree and the number of internal nodes in a fully binary tree with the same number of terminals.

##### *Leave-one-out subsampling analysis of phylogenetic placement performance*

If the phylogeny inferred from molecular data is deemed more reliable than the results of morphological analyses alone, it is advantageous to use the molecular signal to find the position of the fossils. This position can be determined by a total-evidence analysis, where both the molecular and the morphological partitions are considered together for the computation of the tree scores during tree search, or by a backbone analysis that constrains the topology of all extant representatives using the molecular tree ('phylogenetic fossil placement' analysis of Berger & Stamatakis, 2010).

We assessed the potential performance of our morphological data in backbone analyses under the different coding schemes (i.e. MJ, PL, GFC and MI). We used a leave-one-out subsampling procedure, implemented with a TNT script (Supporting Information,

Appendix S7), that simulates the behaviour of extant taxa with ‘known’ phylogenetic relationships as if they were fossils being reinserted into an otherwise constrained tree. The script loops through the list of ingroup species. For each species, the reference tree is set up as a topological constraint, and only the query species is allowed to affect its position. A parsimony search for all constrained most parsimonious trees (CMPTs) is then performed. The analysis is guaranteed to find all CMPTs because the backbone constraint reduces the tree search space to  $2N - 5$ , where  $N$  is the number of species in the tree, allowing us to use the branch-and-bound tree-search algorithm (‘implicit enumeration’ in TNT). We used the strict consensus of the CMPTs of each species placement to compute its accuracy and resolution. This procedure gives us an approximation of the possible position of a single extinct species in a best-case scenario where the character and specimen sampling of the fossil species is as good as the sampling of the extant species.

#### IMPACT OF SPECIMEN UNDER-SAMPLING

To assess the impact of specimen under-sampling, we performed a polymorphic entry replacement data analysis (PERDA; Watanabe, 2016). PERDA takes a species-level matrix with polymorphic coding (PL) and generates multiple subsampled replicates of the matrix, randomly selecting a single state observed in the polymorphic cells. This recreates coding a matrix from a single observation per species. Then, phylogenetic analyses of the original polymorphic matrix and of all of its subsampled replicates are performed. Finally, all trees generated from the subsampled matrices are compared against the tree from the original polymorphic matrix, and a strict consensus of the trees of the subsampled matrices (‘metaconsensus’) is computed. Topological congruence, or lack thereof, between the subsampled matrix trees and the PL matrix tree indicates the strength of the phylogenetic signal, despite intra-specific variation.

We ran PERDA with 4000 subsamples on TNT, using the TNT script from the original implementation (Watanabe, 2016). Parsimony searches were performed with 1000 random addition sequences with TBR optimization holding up to 10 trees, followed by TBR branch swapping of the trees in memory. By default, the PERDA script uses a custom topological distance based on groupings. To facilitate comparisons with other PERDA analyses, we report our results using PERDA’s topological distance, instead of the contradiction difference measure used elsewhere in this paper.

## RESULTS

### REFERENCE PHYLOGENY

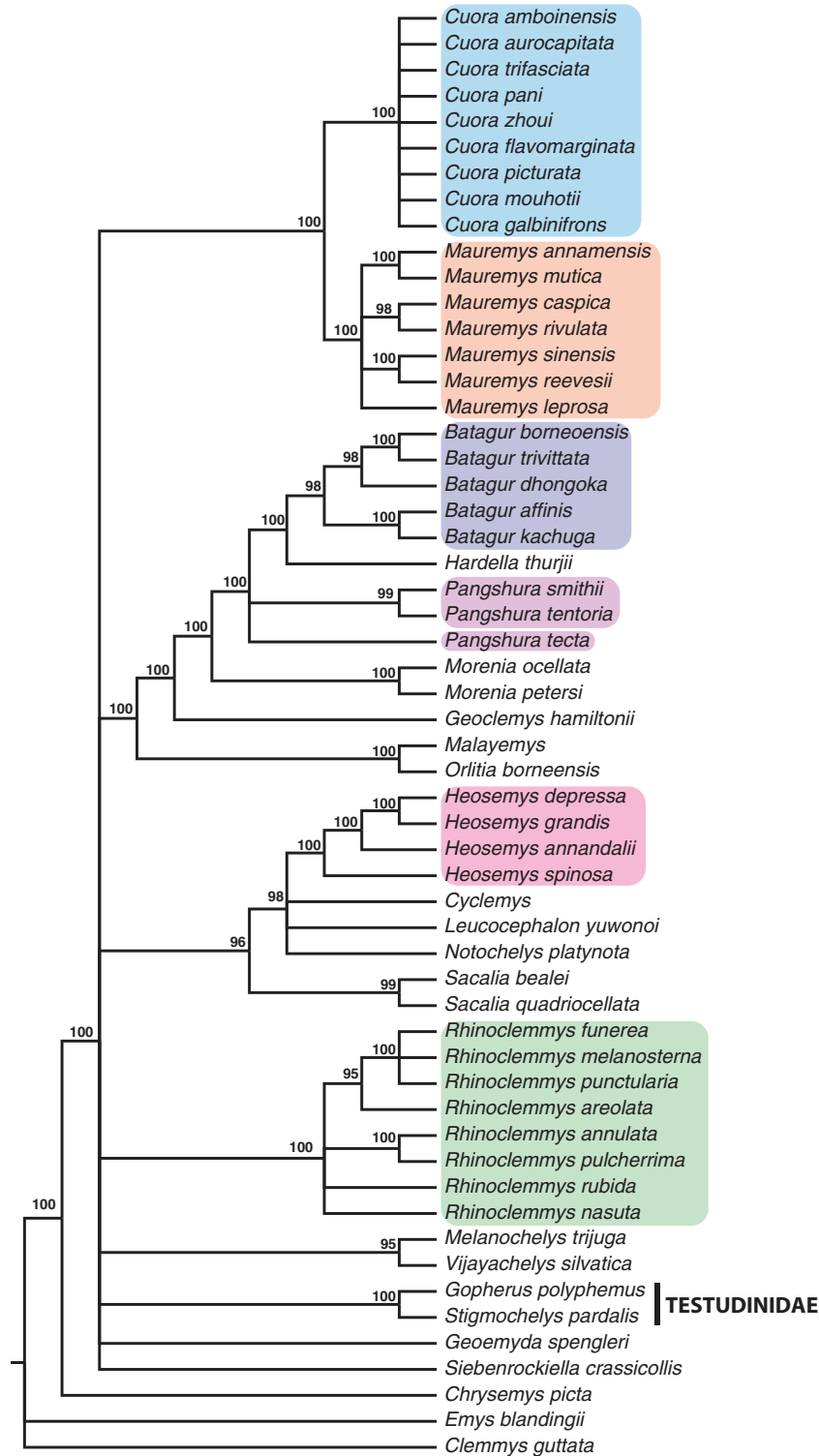
The tree from our ML molecular phylogenetic analysis is given in the Supporting Information (Appendix S8, Fig. S1). The reference tree (Fig. 4) is in broad agreement with the most recent global molecular sequence analyses of geoemydids (Barth *et al.*, 2004; Spinks *et al.*, 2004; Diesmos *et al.*, 2005; Prashag *et al.*, 2006; Le & McCord, 2008), as all of the currently recognized ‘genera’ (see Turtle Taxonomy Working Group, 2017) are recovered as monophyletic. A peculiar feature of the ML tree (Supporting Information, Appendix S8, Fig. S1) is that *Geoemyda* + *Siebenrockiella* were recovered as sister to all other geoemydids, instead of *Rhinoclemmys* (Spinks *et al.*, 2004; Diesmos *et al.*, 2005). However, the branches implied in those relationships, and other basal relationships within Geoemydidae, had poor ultra-fast bootstrap support (< 95%). We collapsed those poorly supported branches in our reference tree to evaluate only the performance of the morphological data against relationships for which there is strong molecular evidence. The collapsed reference tree (Fig. 4) retained 60% of the nodes resolved.

The maximum parsimony analysis of the same molecular data yielded 24 MPTs, whose strict consensus is broadly similar to the ML tree (Supporting Information, Appendix S8, Fig. S2). There are some differences in the arrangement of rootward nodes and other slight differences in the resolution of the most recent clades. However, there is no topological conflict between the MP tree and the ML tree if one does not take into account branches with low support (< 70% bootstrap support in the MP tree, < 95% ultra-fast bootstrap support in the ML tree).

### MORPHOLOGICAL PHYLOGENETIC ANALYSES

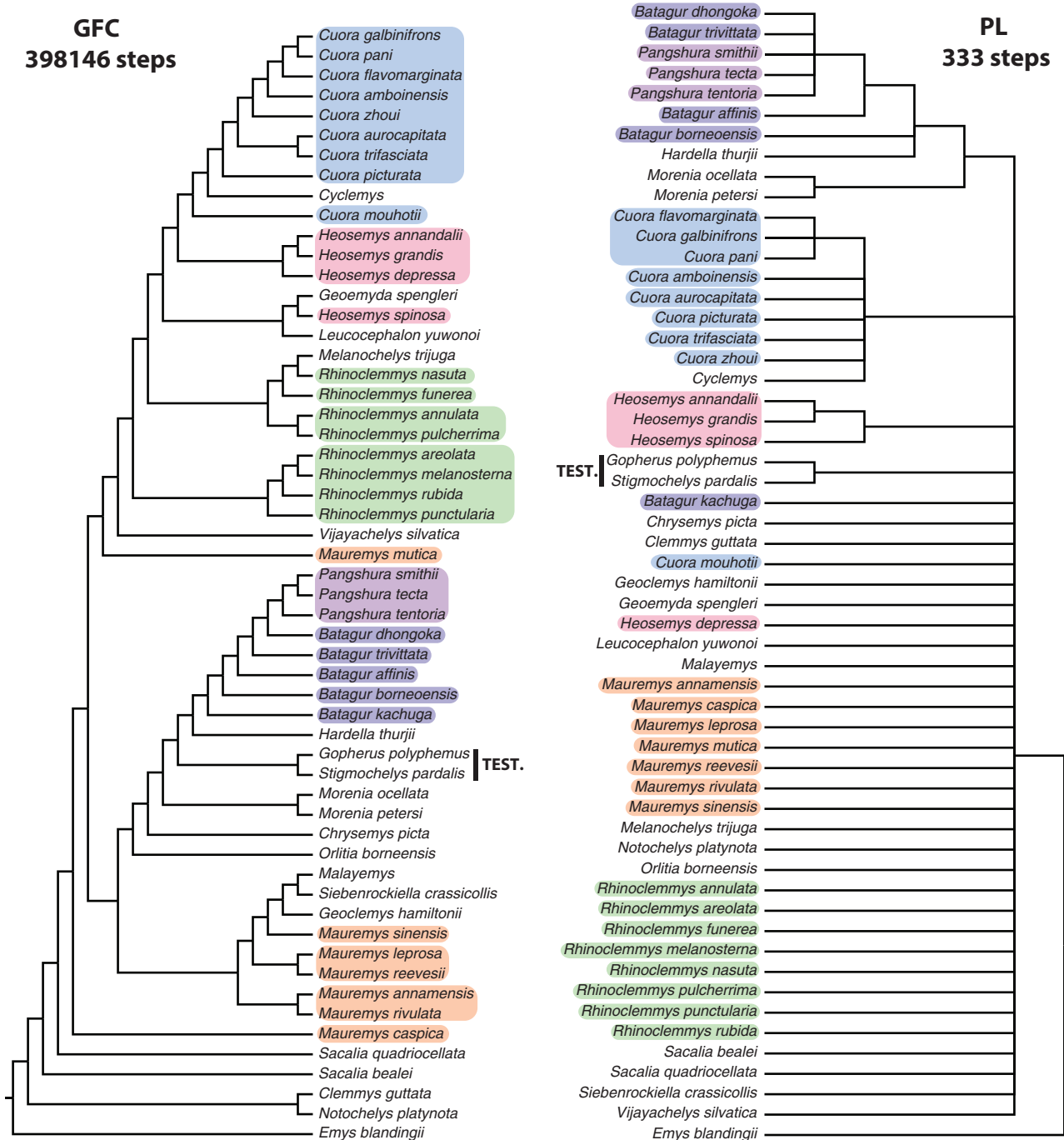
The trees resulting from each coding method are shown in Figures 5 and 6 (for bootstrap support values, see Supporting Information, Appendix S8, Fig. S3). The performance of the four methods is summarized in Table 1. No correlation is apparent between the amount of frequency information used by each method and the performance of the analysis.

The PL method yielded a total of 3237 MPTs, with 450 hits out of 5000 replications and a best score of 333 (Fig. 5). A traditional branch swapping (TBR) was run, finding 89 067 trees and the same best score. The GFC method yielded only one MPT hit 19 times, with a best score of 398 146. The MI method yielded 430 MPTs, with 44 hits and a best score of 312; > 100 000 trees were found after TBR, and the score remained



**Figure 4.** Reference tree derived from molecular data, used for comparisons with the results of the morphological trees. This molecular phylogeny of Geoemydidae was derived from sequences of three mitochondrial genes and four nuclear loci of previous studies (Spinks *et al.*, 2004; Le *et al.*, 2007). Phylogenetic analysis was performed with maximum likelihood. The topology is mostly in agreement with the results of the studies published in the last 15 years.

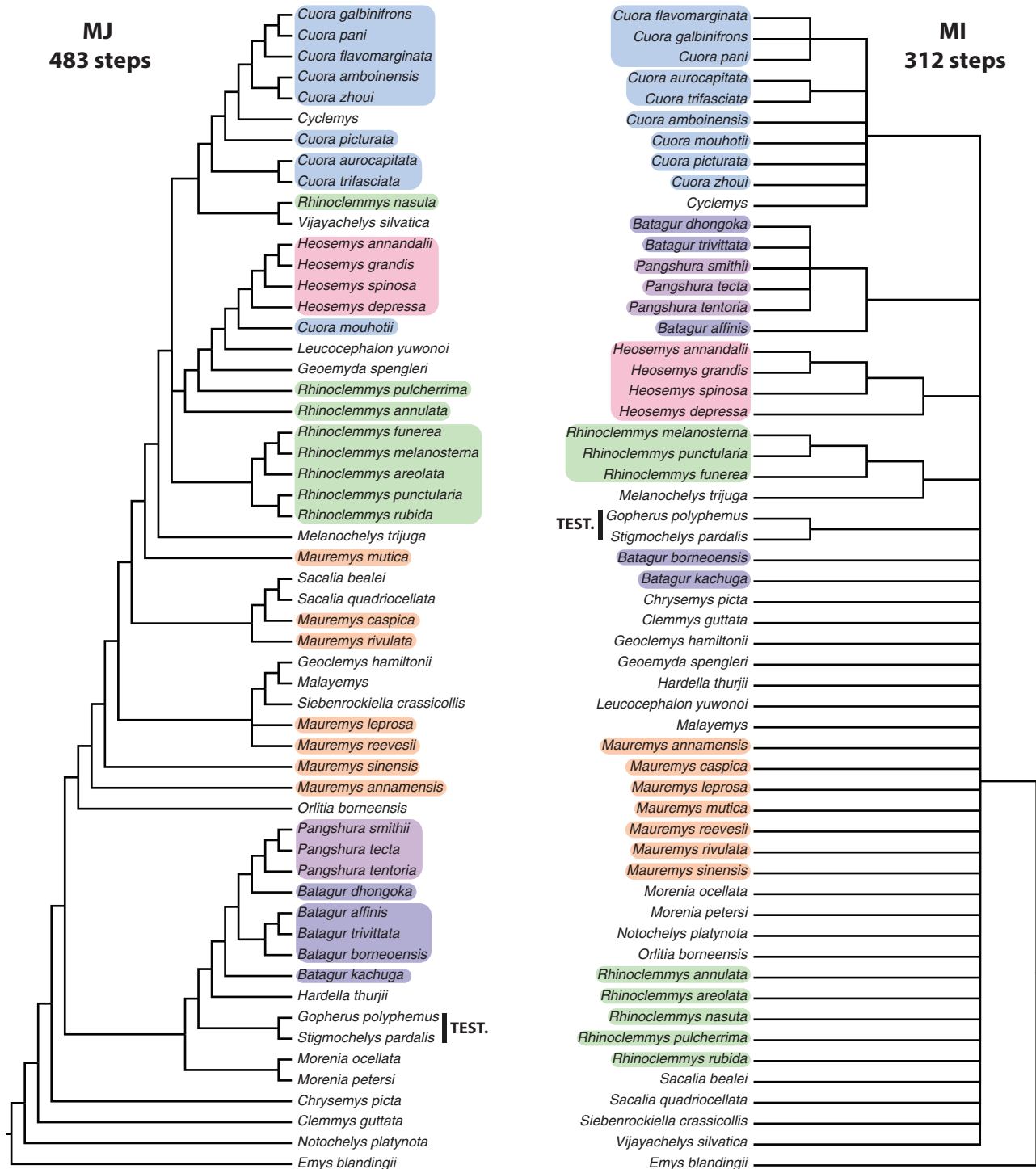




**Figure 5.** Most parsimonious tree for the generalized frequency coding (GFC) method (left) and strict consensus tree of the polymorphic (PL) method (right). Both trees are unconstrained. Colours refer to those groups retrieved in the reference phylogeny (Fig. 4).

the same. The MJ method yielded four MP trees, hit 48 times, with a best score of 483; neither the number of trees nor the score changed after branch swapping. A strict consensus was calculated for all MPTs for each method (except GFC) with no additional tree search.

Of the four coding methods evaluated, MI yielded a strict consensus tree that was the most accurate (82%) but not very well resolved (25% of node resolution). PL tied with MI for accuracy, but managed to resolve two fewer nodes (22% of node resolution). MJ and GFC yielded trees that were far better



**Figure 6.** Strict consensus of the unconstrained trees of the majority (MJ) method (left) and the missing (MI) method (right). Colour groupings refer to the reference tree clades.

resolved, with 100% node resolution for GFC (only one MPT was found) and 96% for MJ, but were also much less accurate, with scores < 40%. Collapsing internal branches with low (< 70%) bootstrap support increased the accuracy of all trees, but decreased

resolution severely, to no more than 20%, and as little as 4% for MI and PL.

Although the PL consensus tree had only 12 internal nodes (22% of the total possible internal nodes), it was able to recover monophyly of *Morenia* and *Pangshura*,

**Table 1.** Resolution and accuracy of the strict consensus of the most parsimonious trees obtained with the four coding methods evaluated

Coding method	Resolution [supported branches only] (%)	Accuracy [supported branches only] (%)
GFC	100 [20]	38 [94]
PL	22 [4]	82 [100]
MJ	96 [15]	40 [85]
MI	26 [4]	82 [100]

Branches are considered supported when they have a bootstrap frequency of  $\geq 70\%$ .

suggesting a close affinity between most species of *Batagur*. Meanwhile, the GFC tree did not recover the clades of *Batagur*, *Rhinoclemmys*, *Mauremys* and *Heosemys* as monophyletic. Among more speciose 'genera', only *Pangshura* was recovered as monophyletic. In general, most methods were able to retrieve a close affinity between species of *Pangshura* and some species of *Batagur*, *Heosemys* and the two species of Testudinidae.

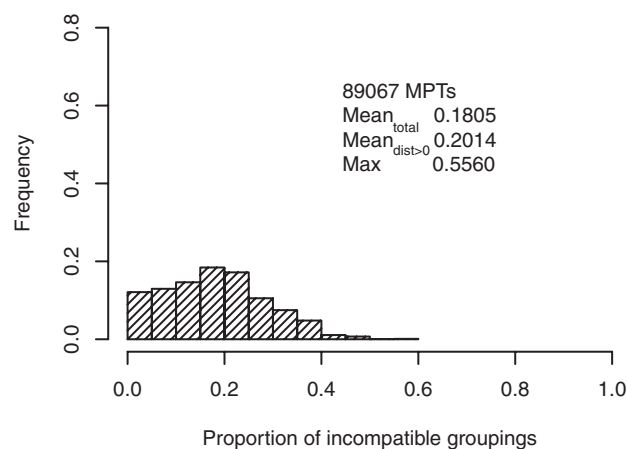
IterPCR did not identify sets of specimens or clades whose exclusion could improve the resolution of the strict consensus of MI or PL by  $> 9\%$ . Therefore, we conclude that rogue species (or clades) are not one of the major causes of the lack of strict consensus resolution.

The total-evidence analysis combining the molecular data and the PL matrix recovered a strict consensus tree (Supporting Information, Appendix S8, Fig. S4) that was well resolved (95%) and highly accurate (97%). Collapsing internal branches with low bootstrap support depressed the resolution to 63%, comparable to the resolution of the reference tree (60%; Fig. 4). The accuracy also increased to 100%, highlighting that the total evidence tree is dominated by the molecular signal.

#### POLYMORPHIC ENTRY REPLACEMENT DATA ANALYSIS

The tree distance histogram produced by the PERDA (Fig. 7) showed that the great majority (90%) of the trees generated from PERDA-subsampled matrices have groupings incompatible with the PL strict consensus tree. The dispersion of tree distances is slightly greater than the most extreme results of Watanabe (2016) and shows that topological estimation is highly sensitive to intraspecific sampling.

Likewise, the PERDA metaconsensus tree was completely unresolved, save for the sister-group relationship between the testudinids *G. polyphemus* and *S. pardalis*. Individual PERDA trees were typically better resolved than the strict consensus of the PL

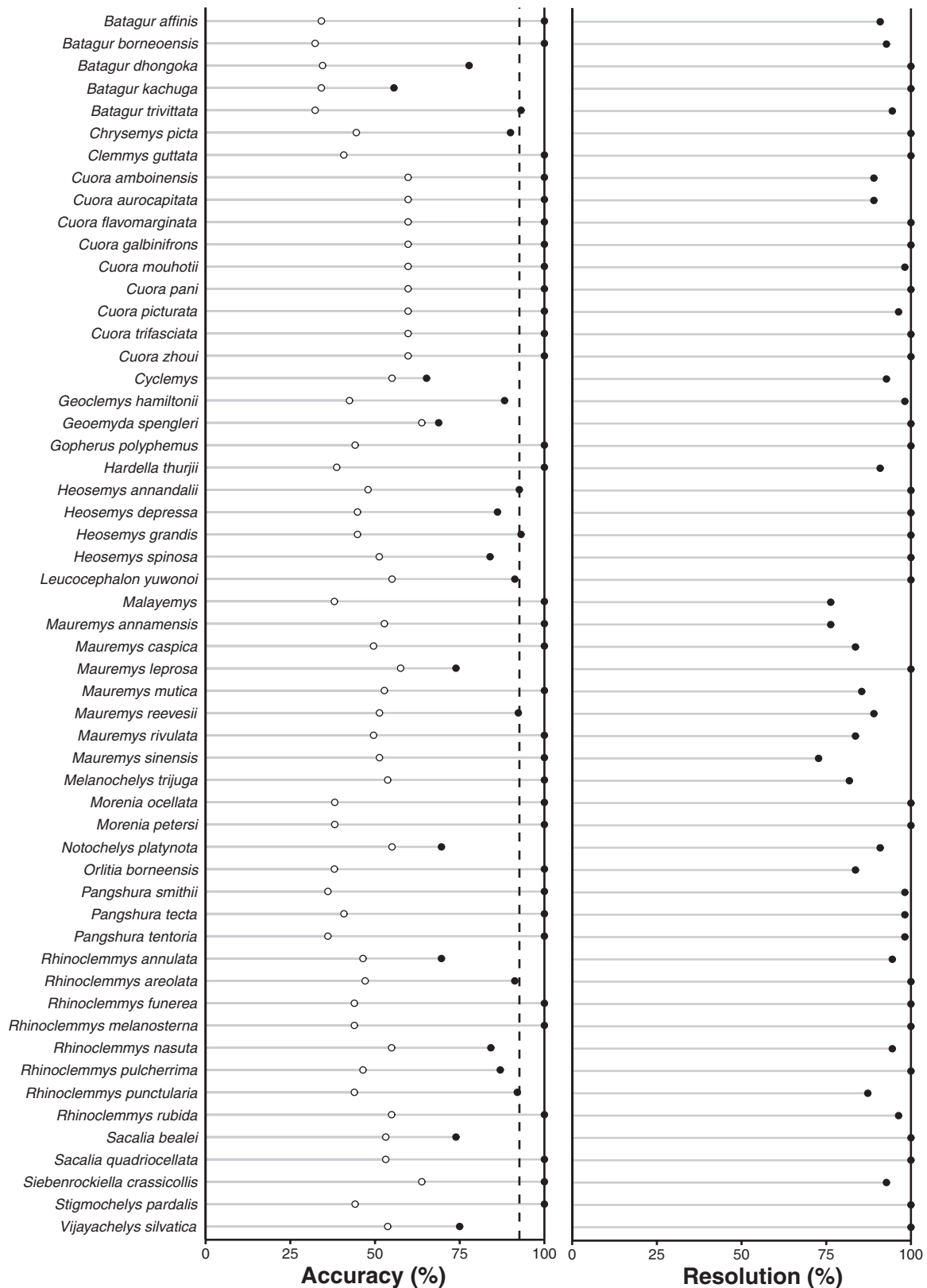
**Figure 7.** Histogram of tree distances (as proportion of incompatible groupings) between the tree obtained from the analysis of the polymorphic (PL) matrix and the trees obtained from polymorphic entry replacement data analysis (PERDA). PERDA generates matrices by randomly sampling a single state from each cell in the PL matrix. Thus, the trees estimated from PERDA matrices represent an estimation of alternative results that could have been obtained from a single-specimen sampling strategy.

matrix (12 nodes in the strict consensus vs.  $> 20$  nodes in 75% of the PERDA trees), with a few (0.008%) even attaining maximal node resolution. This indicates that the lack of resolution of the metaconsensus comes from topological conflict between the individual PERDA trees and not their lack of resolution. Therefore, none of the resolved nodes in the geoemydid ingroup is robust to random morph sampling under a single-specimen sampling strategy.

#### LEAVE-ONE-OUT SUBSAMPLING ANALYSIS OF PHYLOGENETIC PLACEMENT PERFORMANCE

Our leave-one-out subsampling analysis (Fig. 8; Supporting Information, Appendix S8, Figs S5, S6) shows that the phylogenetic placement performance was similar among all methods; the best was MI (at the expense of some resolving power; Fig. 8), and the worst was GFC (Table 2). However, the species placements under MI, PL and MJ do not differ greatly in mean accuracy and resolution (up to 4% difference in accuracy, up to 3% difference in resolution; Table 2). A closer examination of the placement performance of individual species suggests that certain species are more susceptible to misplacement than others. Therefore, caution is needed when placing a fossil that seems closely related to one of these species or has similar character combinations. For example, *Notochelys platynota* (Gray, 1834) had, with all the methods, accuracy  $< 75\%$  and was even lower than the average of

## Leave-one-out subsampling analysis for MI method





**Table 2.** Mean resolution and accuracy attained in species placement with the leave-one-out subsampling analysis, using the four methods for coding intraspecific variability

Coding method	Resolution (%)	Accuracy (%)
GFC	100	80
PL	95	90
MJ	98	88
MI	95	93

all possible placements in the trees when using PL, MJ or GFC (Supporting Information, Appendix S8, Fig. S5). Other terminals that could not be placed with high accuracy were *Malayemys* sp. (with GFC and PL), *Orlitia borneensis* Gray, 1873 (GFC and MJ), *Vijayachelys silvatica* (Henderson, 1912) and *Cyclemys* sp. (with GFC and PL).

## DISCUSSION

### MORPHOLOGICAL PHYLOGENETIC ANALYSES AND POLYMORPHISM CODING METHODS

Many studies have attempted to resolve the phylogeny of geoemydid turtles using morphological data (Hirayama, 1985; Yasukawa *et al.*, 2001; Claude & Tong, 2004; Joyce & Bell, 2004; Joyce & Lyson, 2010), but the results have been disparate and often at odds with the molecular data and traditional taxonomy. Additionally, the scarcity of convincing synapomorphies and the high levels of polymorphism and homoplasy (deduced from mapping characters onto molecular phylogenies) led to a lack of confidence in the ability of morphological characters to resolve overall geoemydid relationships (Joyce & Bell, 2004; Claude *et al.*, 2012). In parallel, new methods were developed to optimize the coding of polymorphic characters for phylogenetic analysis (Campbell & Frost, 1993; Wiens, 1995; Johnson *et al.*, 1988; Smith & Gutberlet, 2001). Given that morphological characters are essential for integrating fossils into geoemydid phylogeny, and polymorphism is an unavoidable feature of the skeletal morphology of geoemydids, we explored the performance of polymorphism coding methods in geoemydid phylogeny.

Wiens (1995, 1998) previously concluded that frequency-based coding methods (i.e. frequency coding, step-matrix frequency coding, majority) usually perform better than other methods when using polymorphic characters, as these methods use the

maximal amount of information stored in the matrix. Furthermore, Swofford & Berlocher (1987) and Wiens (1995, 1998) noted that these methods are less sensitive to sampling error than others because they use the actual frequency of each trait.

Smith & Gutberlet (2001) developed generalized frequency coding (GFC) as a frequency-based coding method suitable for multistate characters that allows the relative frequencies of character states to be coded directly into the character matrix. In previous studies, when compared with other multistate character coding methods (i.e. step-matrix and gap-weighting), GFC performed as well as or better than others, because this method uses all intraspecific variation in the coding (Smith & Gutberlet, 2001; Lawing, Meik & Schargel, 2008) and allows an arbitrary number of terminals in most software implementations.

In the present study, we compared the performance of GFC and three other polymorphism coding methods: polymorphic (PL), majority (MJ) and missing (MI). We used a reference tree based on molecular data (Fig. 4) to measure the performance of these methods in terms of resolution (the number of internal nodes in the tree) and accuracy (topological compatibility with the reference tree). We used a phylogeny inferred from molecular data as our reference tree because of the lack of contradiction in the supported relationships between the total evidence tree and the molecular trees inferred with ML and MP, indicating that the molecular data contain the stronger signal.

When compared with previous morphological phylogenetic analyses of testudinoid taxa that did not assess polymorphism in detail (Yasukawa *et al.*, 2001; Joyce & Bell, 2004), the relationships found in some of our analyses are more accurate with respect to the reference tree (Table 3). However, this is not indicative of good performance. None of the coding methods yielded topologies with both resolution and accuracy > 50%, regardless of how much state frequency information these methods introduced into the analyses. These results strongly suggest that our information about the frequency of character states carries little phylogenetic signal in geoemydids. Furthermore, GFC, which makes the most use of frequency information, gave the most misleading results, with a fully resolved but very inaccurate tree (38% accuracy; Table 1).

The poor performance of GFC was probably caused by a phenomenon described by Bardin *et al.* (2014): the increase in the number of states of characters in pure-noise matrices (character matrices generated without reference to a phylogeny) induces highly resolved but

**Figure 8.** Placement accuracy (left) and resolution (right) for each species from our leave-one-out subsampling analysis of the morphological matrix using the missing (MI) method (black circles). Dashed line shows the mean for all species. Open circles represent the mean accuracy score for all the possible placements of one species.

**Table 3.** Comparison of morphological matrices of Geoemydidae from previous studies and the present study

Morphological matrix	Total number of species	Number of Geoemydidae species	Species in common with reference tree	Geoemydidae species in common with reference tree	Total number of characters	Number of shell characters	Accuracy (%)	Resolution (%)
Joyce & Bell (2004)	53	25	30	25	70	27	42	73
Yasukawa <i>et al.</i> (2001)	28	28	28	28	35	9	48	100
Present study (PL)	56	51	56	51	80	80	82	22

This study retrieved the most accurate result when compared with other morphological matrices, but the improved sampling of polymorphism also resulted in a significant loss of resolution. Joyce & Bell (2004) coded their matrix with polymorphic (PL), and Yasukawa *et al.* (2001) used a mixed strategy for polymorphic coding. The strict consensus trees from the Joyce & Bell (2004) and Yasukawa *et al.* (2001) matrices were estimated using the same search strategy used in the analyses of our data.

inaccurate trees in parsimony analysis. GFC introduces many subcharacters for each original character in the matrix. Each of those subcharacters has 32 states, and most of them are ordered in our matrix, resulting in a high granularity in the parsimony scores. Confounding factors, such as high amount of polymorphism and homoplasy, amplified by the granularity of the parsimony score, are likely to be overwhelming the small phylogenetic signal that GFC extracts more efficiently from state frequency information.

Bardin *et al.* (2014) suggested that bootstrapping could mitigate the effects of the artefactual resolution induced by increasing the number of states. Indeed, a bootstrap analysis of the GFC matrix yields a more accurate topology that is not completely compatible with the MP tree of the original matrix. However, the number of correct nodes with bootstrap support is very low, once again negating the advantage of GFC's greater ability to make use of frequency information. Although GFC can be useful, future studies should be cautious of the effects that we observed with our data.

Although state frequencies may not be phylogenetically informative, the assessment of polymorphism remains crucial. As the PERDA shows (Fig. 7), the high amount of polymorphic cells in our matrix makes tree inference sensible when a single-specimen sampling strategy is used.

#### MAKING USE OF MOLECULAR DATA

As mentioned before, phylogenetic placement is one of the most common ways to incorporate fossil and molecular data into a comprehensive phylogeny. The other dominant approach is total-evidence analysis. The following discussion addresses phylogenetic placement, but many points are valid for both approaches.

As in our standard phylogenetic analyses (Figs 5, 6), our leave-one-out subsampling analysis of the performance of phylogenetic placement (Fig. 8; Supporting Information, Appendix S8, Figs S5, S6) supports that

state frequency information is not phylogenetically informative for our data. Placement accuracy seems to be inversely proportional to the amount of state frequency information conveyed by the coding method (Table 2). However, the overall performance of all coding methods was much better for phylogenetic placement, as all methods have average accuracy and resolution measures of 80% or greater (Table 2). Therefore, we are cautiously optimistic about the use of our matrix in future studies placing fossil geoemydids in the current molecular phylogenetic framework. Special caution should be taken regarding four general points:

1. Phylogenetic placement can only be as good as the molecular phylogeny skeleton. Despite extensive work on geoemydid phylogeny in the past decade, molecular studies have not converged entirely at the global or local level. For instance, there are disagreements in the arrangement of *Geoemyda* and *Rhinoclemmys*, depending on the type of molecular data analysed (i.e. Spinks *et al.*, 2004 with sequence data, and Sasaki *et al.*, 2006 with SINE insertions).
2. Our leave-one-out subsampling procedure does not cover cases in which two or more species are placed in the reference phylogeny. We do not know how well our matrix will perform when placing multiple, closely related fossils.
3. We currently have no strong synapomorphies for Geoemydidae. Therefore, discrimination of early crown or late stem geoemydid species will be challenging.
4. Our character matrix has low amounts of missing entries, typical of neontological data, not palaeontological. Therefore, the quality of actual fossil placements is likely to be lower than suggested by the leave-one-out subsampling analysis.

These shortcomings can be minimized, however. Uncertainty in the molecular phylogeny can be reduced by performing the phylogenetic placement for several estimated topologies and averaging the

results. Alternatively, one could collapse poorly supported nodes in the reference phylogeny and accept lower performance in placement resolution.

Furthermore, examination of fossil material will help to identify new characters to add to our matrix and thus clarify our understanding and improve the coding of the current specimens. Increasing the number of characters is probably the most crucial improvement that can be made to our present data, as it would improve the accuracy and resolution of our inferences.

More methodological improvements include the use of weights, additional sources of data, and possibly, the use of probabilistic phylogenetic inference methods. Recent developments on the dynamic estimation of weights based on character self-consistency have become relatively popular (Goloboff, 1993, 2014). However, in a phylogenetic placement framework, character weights should be estimated in relationship to the reference phylogeny. This approach has been put forward for maximum likelihood by Berger & Stamatakis (2010). Additional sources of data include fossil age estimates and biogeographical distribution, which can be incorporated into a coherent framework for total-evidence phylogenetic analysis (not simple phylogenetic placement), owing to recent work on Bayesian methods (e.g. Pyron, 2011; Ronquist *et al.*, 2012; Gavryushkina *et al.*, 2017).

Recently, there has been a debate on the relative performance of maximum parsimony and statistical methods of phylogenetic inference, applicable to this study in the development of new methods for phylogenetic inference of morphological data (Wright & Hillis, 2014; O'Reilly *et al.*, 2016; Goloboff, Torres & Arias, 2017; Puttick *et al.*, 2017). Maximum parsimony analysis also offers another alternative: one can attempt to represent the same data from a shell morphology in a different manner. Recent studies have used continuous characters (Goloboff *et al.*, 2006) and even three-dimensional landmark coordinates (Catalano *et al.*, 2010) in parsimony analysis. From our examination of geoemydid shells, there is an amount of morphological variation that does not lend itself to easy discretization or characterization as a conventional suite of characters; for instance, the overall shape of the carapacial dome. These new approaches using additional and alternative sources remain very recent innovations, so future implementations should assess their behaviour carefully.

Finally, although MI showed the greatest mean accuracy with relatively low loss in resolution, the difference in phylogenetic placement performance between the different methods is not large, and the relative performance of the methods could be highly contingent on the character sampling, given the noisy nature of our data. Therefore, future studies could benefit from

a leave-one-out subsampling analysis to inform the choice of coding method for each dataset. Our leave-one-out subsampling procedure can easily be adapted to maximum likelihood analyses, and possibly, to Bayesian inference.

#### SYNAPOMORPHIES OF GEOEMYDID CLADES

To clarify whether morphological characters are available to diagnose the various clades identified using molecular data, we mapped the distribution of all character states of our PL character matrix onto the reference tree (Fig. 4) using the maximum parsimony ancestral state optimization in Mesquite. In the following subsections, we discuss possible synapomorphies for select geoemydid clades and the historical importance of some characters. We consider a synapomorphy to be a derived character state shared between descendants of the same ancestor but not necessarily fixed or present in all of them (i.e. 'polymorphic synapomorphy' from Hillis, 1987).

*Geoemydidae*: A literal interpretation of our results implies that a median carapacial keel is a synapomorphy of *Geoemydidae*, as this feature is generally absent in the outgroups. Within the clade, these keels may be reduced or lost in late adult stages of larger species, such as in *Batagur* species and *Heosemys depressa* (Anderson, 1875), or in species with a domed carapace, such as *Cuora galbinifrons* Bourret, 1940. We are hesitant, however, to suggest using this character to diagnose the group, because a number of emydids and testudinids not sampled herein have keels as juveniles or adults (Joyce & Bell, 2004).

Claude & Tong (2004) considered the presence of three carapacial keels, at least in juveniles, to be a character that unites all geoemydids except for *Rhinoclemmys* and the extinct *Echmatemys*. This study also includes a character pertaining to the presence of these lateral keels (in adults only), but it is not diagnostic for any particular clade. However, given that we did not systematically survey juveniles or include a character for juveniles only, our observations are not sufficient to contradict the original assertion of Claude & Tong (2004). We urge further study, but given that juvenile turtles can rarely be identified with confidence in the fossil record, we are sceptical that further investigation would yield a useful character for palaeontologists.

A literal interpretation of our results confirms the assertions of Hirayama (1985) and Yasukawa *et al.* (2001) that paired anterior and posterior musk duct foramina are a synapomorphy of *Geoemydidae*. Within the ingroup, these foramina are absent, or at least, were not observed by us in *Batagur dhongoka* (Gray, 1832), *Batagur borneoensis* (Schlegel & Müller, 1845), *Cuora*



*aurocapitata* Luo & Zong, 1988, *Cuora flavomarginata* (Gray, 1863; only the posterior foramina were absent for the last three), *Cyclemys*, *Notochelys platynota*, *Morenia petersi* Anderson, 1879, *Rhinoclemmys areolata* (Duméril & Bibron, 1851) and *Rhinoclemmys melanosterna* (Gray, 1861). Although Yasukawa *et al.* (2001) already noted that musk duct foramina might be absent in *Morenia*, Le & McCord (2008) stated that they were able to see small musk ducts in *Morenia ocellata* (Duméril & Bibron, 1835), and we agree. Given that musk glands and musk duct foramina are known to occur in many emysternids (Waagen, 1972; Joyce & Bell, 2004) and that rigorous analysis that would confirm the absence of musk duct foramina in the testudinoid stem lineage is still outstanding, we are reluctant to endorse the usage of this character to diagnose fossils as geoemydids.

In most geoemydids, the fifth and sixth marginal scutes overlap onto the hyoplastra and hypoplastra by crossing over the osseous bridge. This character is inapplicable for hinged species because they do not have a bridge. A reversion is apparent in the *Heosemys* clade and *Geoemyda spengleri* (Gmelin, 1789). Although this character is not apparent in the outgroup we used, a survey of additional taxa (e.g. *Trachemys scripta*) revealed the occasional presence of these characteristics in some extant emydids. We therefore cannot endorse this character as a unique synapomorphy for the group.

McDowell (1964) noted that geoemydids (his Bataguridae) could be distinguished from emydids by the presence of a fully divided suprapygal. We here confirm the general presence of this characteristic in geoemydids with the exception of most representatives of *Rhinoclemmys*, *Geoemyda spengleri*, *Notochelys platynota*, *Leucocephalon* and some specimens of the *Mauremys* clade. As most of these exceptions are located near the base of the tree, however, it remains unclear whether this character can be optimized to represent the plesiomorphic condition for the clade.

*Melanochelys-Vijayachelys clade*: The phylogenetic position of *Melanochelys trijuga* (Schweigger, 1812) has remained uncertain. This species was retrieved as sister to all other 'narrow-jawed' geoemydids (Spinks *et al.*, 2004), the *Heosemys-Sacalia* clade (Sasaki *et al.*, 2006) or *Geoemyda* (Diesmos *et al.*, 2005, Le & McCord, 2008). Prashag *et al.* (2006) were the first to propose the phylogenetic position of *M. trijuga* as sister to *V. silvatica* at the base of the tree. Here, we observe that in two species from the Indian subcontinent, the anterior margin of the entoplastron is intersected by the gularohumeral sulcus and that the inguinal buttress only contacts the fifth costal bone (shared with *Rhinoclemmys*). Most importantly, both *M. trijuga* and *V. silvatica* can retain three carapacial keels as adults, a great exception among geoemydids (see Geoemydidae discussion above).

*Rhinoclemmys clade*: All species of *Rhinoclemmys* occur in tropical Central and South America and are therefore united by a biogeographical signal, but it has been difficult to retrieve morphological characters that unite them as a group. Hirayama (1985) found *Rhinoclemmys* to be paraphyletic, because *Rhinoclemmys rubida* (Cope, 1870) and *Rhinoclemmys annulata* (Gray, 1860) shared more similarities with some *Cuora* species than with other *Rhinoclemmys* species. Yasukawa *et al.* (2001) retrieved a similar topology with the neighbor-joining method, also using morphological characters. Molecular phylogenies, however, have retrieved a monophyletic *Rhinoclemmys*, even though Spinks *et al.* (2004; e.g. combined analysis) and Diesmos *et al.* (2005) grouped this clade with Testudinidae, thereby rendering Geoemydidae paraphyletic. The phylogeny of Sasaki *et al.* (2006), using SINE insertion, recovered *Rhinoclemmys* species inside Geoemydidae, but not the monophyly. Le & McCord (2008) again supported the monophyly of *Rhinoclemmys* based on two nuclear and three mitochondrial genes.

One of the reasons it is difficult to find characters that diagnose this New World clade is because representatives of *Rhinoclemmys* lack unusual morphological adaptations, such as kinetic shells or secondary palates, and therefore symplesiomorphically resemble many geoemydids in the Old World (McDowell, 1964; Pritchard & Trebbau, 1984; Carr, 1991). Pritchard & Trebbau (1984) diagnosed *Rhinoclemmys* species by the absence of lateral keels on the carapace, hexagonal neurals with short posterior sides, and absence of a plastral hinge, but these characters occur broadly among other geoemydids as well.

We conclude that representatives of *Rhinoclemmys* share the presence of a carapace gutter formed by the peripherals (absent in *R. rubida* and *R. areolata*), incomplete subdivision of the pygal by intermarginal the sulcus (except in *Rhinoclemmys nasuta* Boulenger, 1902; shared with Emydidae), and placement of the posterior musk duct foramina at the suture between the peripheral and inguinal buttress, when present. Unlike Claude & Tong (2004), we cannot consider the absence of lateral keels exclusive to *Rhinoclemmys*, as it occurs in the majority of adult geoemydids, and we did not analyse juvenile specimens (see 'Geoemydidae clade' above).

*Heosemys-Sacalia clade*: This clade includes all species currently classified in *Heosemys* and *Sacalia*, as well as *Leucocephalon yuwonoi* (McCord, Iverson & Boeadi, 1995), *Notochelys platynota*, and all specimens of *Cyclemys* united under *Cyclemys* sp. The close relationship of these species has been retrieved by many phylogenetic analyses, with both molecular and



morphological data (Yasukawa *et al.*, 2001; Spinks *et al.*, 2004; Diesmos *et al.*, 2005; Sasaki *et al.*, 2006; Le & McCord, 2008). Following our analysis, these species have an entoplastron that is longer posteriorly than anteriorly (with reversion in *Sacalia bealei* Gray, 1831), usually lack axillary and inguinal scutes (polymorphic in most species), and possess inguinal buttresses that reach the costal sutures with the bridge. The last characteristic is present in all species of the clade (except *L. yuwonoi*) and is polymorphic in *Sacalia quadriocellata* (Siebenrock, 1903), *Cyclemys*, *N. platynota* and most *Heosemys*.

**Sacalia clade:** The two species of *Sacalia* present an oval first neural followed by a series of anteriorly short-sided neurals. In some specimens of *S. quadriocellata*, the second neural is quadrangular, probably owing to minor abnormalities in the region of contact. In both *Sacalia* species, the sulcus between vertebral I and pleural I (seam A, *sensu* Tinkle, 1962) contacts the second marginal and not the first marginal, as in most geoemydids.

**Cyclemys–Notochelys–Leucocephalon–Heosemys clade:** This is a subclade of *Heosemys–Sacalia*. The species in this group share the presence of significant serration on the posterior peripherals (seventh to 11th), a derived condition in Geoemydidae (Yasukawa *et al.*, 2001). This condition was not observed in *H. depressa*, probably because our sample for this species consisted of particularly large adults, which also tend to have a smoother carapace. In all species of this group, with the exception of *N. platynota*, the posterior margin of first vertebral is wider than the anterior margin in at least one specimen. Although frequently observed, the presence of an anterolateral step on the first vertebral is polymorphic for this group and shared with the *Batagur–Orlitia* clade.

**Heosemys clade:** The four species of *Heosemys*, including *Heosemys annandalii* (Boulenger, 1903), share the presence of posteriorly short-sided third to sixth neural bones, a round seventh neural, and an anteriorly short-sided eighth neural (except for *Heosemys spinosa* (Gray, 1830), which has a posterior short-sided one). The second neural is posteriorly short sided in *H. spinosa* and *H. depressa*, whereas it is quadrangular in *H. annandalii* and *Heosemys grandis* (Gray, 1860).

The fifth marginal scute does not overlap the hyoplastron or hypoplastron in *H. annandalii*, *H. grandis* and *H. spinosa*, a homoplasy with Testudinidae. Within the group, this character is secondarily lost in *H. depressa*, thereby displaying the same condition as seen in most other geoemydids (with exception of *Mauremys leprosa* Schweigger, 1812).

Although not a focus of this paper, all four *Heosemys* share a secondary loss of the quadratojugal bone and the temporal arc on the skull. This condition is also present in *Siebenrockiella leytensis* (Taylor, 1920), *Cuora picturata* Lehr, Fritz & Obst, 1998, *Cu. flavomarginata* and *L. yuwonoi* (McCord *et al.*, 2000; Diesmos *et al.*, 2005).

**Batagur–Malayemys clade:** This clade includes all species currently classified as *Batagur*, *Morenia* and *Pangshura*, as well as *Geoclemys hamiltonii* (Gray, 1830), *Hardella thurjii* (Gray, 1831), *Malayemys* sp. and *Orlitia borneensis*. McDowell (1964) first established a group reminiscent of this clade as the ‘batagurines with secondary palate’, although he included *Mauremys sinensis* (Gray, 1834) and *Mauremys reevesii* (Gray, 1831), as these taxa also have extensive secondary palates. Later, Hirayama (1985) recognized this grouping as a distinct phylogenetic clade, which Gaffney & Meylan (1988) named Batagurinae. Joyce & Lyson (2010) provided the name Palatochelydia for the monophyletic extraction recognized by molecular results from this grouping (minus *Malayemys* and *Orlitia*), but referred to the most inclusive clade containing these species, instead of the least inclusive one. Representatives of the *Batagur–Malayemys* clade share the following combination of characters: second to sixth neural bones anteriorly short sided, strong inguinal buttresses in clear contact with both fifth and sixth costals, and the universal presence of a triangular or rounded anal notch.

**Batagur–Hardella–Pangshura–Morenia clade:** This is a well-supported subclade of the *Batagur–Malayemys* clade retrieved by molecular phylogenies in the past (Spinks *et al.*, 2004; Diesmos *et al.*, 2005; Le, McCord & Iverson, 2007). These broad-jaw geoemydids have a sulcus between pleural I and II (seam B of Tinkle, 1962), pleural II and III (seam C of Tinkle, 1962), and pleural III and IV (seam D of Tinkle, 1962), reaching the fourth, sixth and eighth marginals, respectively, but *Batagur kachuga* (Gray, 1831), *Batagur trivittata* (Duméril & Bibron, 1835), *B. borneensis*, *Pangshura tentoria* (Gray, 1834) and *M. petersi* show polymorphism for these characters. In addition, species of this clade usually have a first vertebral that is as long as it is wide and an entoplastron that is never intersected by the humero-pectoral sulcus, the latter character state being shared with Testudinidae.

**Batagur–Hardella–Pangshura clade:** This subclade has broad support in the literature (Spinks *et al.*, 2004; Diesmos *et al.*, 2005; Le, McCord & Iverson, 2007; Joyce & Lyson, 2010). Representatives of this clade can be diagnosed by the presence of strong axillary buttresses that directly contact the first costal

rib and the overlap of the fourth and sixth marginal onto the neighbouring costals. [Le et al. \(2007\)](#) noted that male representatives of *Batagur* exhibit costal fontanelles as adults, but we found open fontanelles in all representatives of this clade, with the exception of *B. trivittata*, and the consistent presence of fontanelles in females, at least as labelled in museum collections.

**Pangshura clade:** Although not present as a monophyletic clade in our reference tree ([Fig. 4](#)) owing to low branch support, here we discuss the *Pangshura* clade as a whole (with exception of *Pangshura sylhetensis* Jerdon, 1870, which was not analysed in this study), including the fossil *Pangshura tatrotia* Joyce & Lyson, 2010. All mentioned species share the presence of an octagonal fourth neural (i.e. with anterior and posterior short sides), a fourth vertebral scute that is longer than it is wide and that always covers the fourth neural anteriorly, a long contact between the last suprapygal and tenth peripheral bone, and contact of the sulcus between the fourth pleural and fifth vertebral scutes (seam E of [Tinkle, 1962](#)) with the tenth marginal scute. Furthermore, these taxa share the presence of an anteromedial projection on the interpleural I–II sulcus and on the interpleural II–III sulcus; whereas the former sulcus consistently laps onto the second costal, the latter does not always cross the anterior margin of fourth costal.

**Batagur clade:** Representatives of the recently established *Batagur* clade ([Le et al., 2007](#)), including the recently resurrected *Batagur affinis* (Cantor, 1847; [Praschag et al., 2007, 2008](#)), have flared eighth to 12th marginal scutes and a straight anterior plastron margin without spikes or lateral tuberosities. Within this clade, *B. dhongoka*, *B. trivittata* and *B. borneensis* share a fourth vertebral scute that is longer than it is wide (also present in *Pangshura*). The subclade consisting of by *B. kachuga* and *B. affinis* lacks a notch on the pygal bone (also as in *B. trivittata*), whereas *B. affinis* uniquely exhibits a medial contact of the seventh costal bones.

**Cuora–Mauremys clade:** The species of *Mauremys* and *Cuora* form a clade in many molecular phylogenies ([Honda et al., 2002](#); [Barth et al., 2004](#); [Spinks et al., 2004](#); [Diesmos et al., 2005](#)) and are known to hybridize between and within these clades (e.g. ‘*Mauremys iversoni*’, ‘*Ocadia philippeni*’, ‘*Ocadia glyphistoma*’, ‘*Mauremys pritchardi*’, ‘*Cuora serrata*’; [Parham et al., 2001](#); [Wink et al., 2001](#); [Spinks et al., 2004](#); [Stuart & Parham, 2004, 2007](#)). [Claude et al. \(2012\)](#) mentioned that weak plastral buttresses, intersection of the entoplastron by the humeropectoral sulcus, and the presence of three carapacial keels until adulthood diagnose representatives of this clade, but our analysis does not support any of those synapomorphies, even

though we included them in our matrix. We are therefore unaware of any morphological support for this clade from the shell.

**Mauremys clade:** [Barth et al. \(2004\)](#) and [Spinks et al. \(2004\)](#) found the *Mauremys* group to be paraphyletic with respect to turtles historically classified as *Chinemys* and *Ocadia* based mostly on mitochondrial DNA, but no morphological characters have been proposed for the expanded definition of *Mauremys* that is now pervasive in the literature. In this study, we found *Mauremys* species to share the following combination of morphological characters: second to sixth neural bones anterior short sided (third and fourth are polymorphic in *Mauremys mutica* Cantor, 1842), sulcus between first vertebral and first pleural (seam A) contacts second marginal scute or the sulcus between first and second marginal (with exception of *M. sinensis*), sulcus between second and third pleural scutes (seam C) contacts seventh marginal or the sulcus between sixth and seventh marginal, and axillary buttress clearly contacts first costal bone (polymorphic in *M. leprosa*).

**Cuora clade:** The highly specialized shell of the Asian box turtles shows characters related to the presence of a plastral hinge, such as a pivoting process on the fifth peripheral bone, carapace and plastron attached by connective tissue, a hyo-hyoplastron suture completely confluent with the pectoro-abdominal sulcus, and small axillary and inguinal buttresses that do not contact the costal bones. Also, these turtles have a posteriorly short-sided sixth neural and commonly do not show lateral spikes on the anterior plastron margin, except for *Cuora trifasciata* (Bell, 1825) and *Cuora mouhotii* (Gray, 1862).

Within Geoemydidae, a true plastral hinge was observed in the species of *Cuora*, *Cyclemys* and *N. platynota* (i.e. absent bony bridge, presence of medially directed pivotal process on fifth peripheral), although the plastral lobes are more kinetic in *Cuora* than the two others, and some juveniles of *N. platynota* and *Cyclemys* may not have a hinge ([Yasukawa et al., 2001](#)).

This study supports the conclusion by [Honda et al. \(2002\)](#) that the plastral hinge emerged independently two times in Geoemydidae: in the common ancestor of the *Cuora* clade and in that of *Cyclemys* and *Notochelys*.

#### POLYMORPHISM IN GEOEMYDIDAE

Our analyses suggest that the polymorphic state frequencies in our dataset carry only a small phylogenetic signal. This contradicts previous studies of empirical and simulated data, which found that coding methods that incorporate more frequency information into the analysis generally outperform coding methods that ignore it, even with small sample sizes ([Fig. 2](#); [Wiens,](#)

1995; Wiens & Servedio, 1997). From those studies and the arguments presented by Wiens (1999), we are convinced that polymorphic state frequencies can contribute to significant phylogenetic signal. Therefore, we wonder what particular conditions in the intra-specific phenotypic variation of geoemydids render it uninformative. A major possible factor is that a significant part of the observed phenotypic variation is the result of random variations in developmental conditions and phenotypic plasticity, thereby not reflecting underlying genetic differences. Indeed, the occasional presence of left–right differences within the specimens we sampled supports the idea that many of the alternative phenotypes encoded by our characters can be produced by the same genotypes in certain conditions.

Furthermore, Fritz *et al.* (2007) and Mikulíček *et al.* (2013) reported that the extensive morphological differentiation of populations of *Testudo graeca* Linnaeus 1758 does not correlate with their phylogenetic and taxonomic structure as inferred from mitochondrial and nuclear AFLP data (amplified fragment length polymorphism), suggesting that their morphological differences are a plastic response to local environments. Similar disagreement between morphology and well-supported genetic differentiation has been observed in *Graptemys* sp. (Praschag *et al.*, 2017). Those studies, taken together with the observations presented here and by Joyce & Bell (2004), suggest that phenotypically plastic morphological variation could be a widespread phenomenon among testudinoids. Unfortunately, we have not found a practical approach to distinguish the genetic from the non-genetic variation in our data, especially as it is likely that the causes of polymorphism should be assessed not only per character, but also per species. Nevertheless, even if much of the observed polymorphism does not have a genetic basis and thus is not even strictly homoplastic, this does not negate the usefulness of the characters affected.

Although the polymorphic cells appear to contribute little, our leave-one-out subsampling analysis showed that the fixed-state cells still allow phylogenetic placement of a single species somewhat reliability. This gives further grounds to the call by Wiens (1999) to include polymorphic characters in phylogenetic analysis.

Finally, although we have emphasized the problem of recovering the phylogenetic signal from polymorphic cells, other factors could also be hindering the performance of our phylogenetic analyses. Many of our characters seem phylogenetically labile, and we do not assess in detail the extent and impact of homoplasy in this paper. Another important issue is that most shell characters refer to tightly integrated structures that may co-vary because of developmental and mechanical constraints and may therefore be co-dependent. Different modes of cladogenesis and morphological

differentiation do not guarantee the evolution of morphological synapomorphies to allow resolving and supporting every node of a tree. Bapst (2013) recently demonstrated that conditions that produce unresolved clades may be common, so we may have to lower our expectations of tree resolution.

## OUTLOOK AND CONCLUSION

We attempted to assemble a comprehensive morphological dataset for the shell of geoemydid turtles to help resolve the shell-dominated fossil record of the group. We focused on documenting polymorphism, as geoemydids are known to exhibit intraspecific variation in their shells. Although the majority of our informative specimens came from institutions known for their osteological turtle collections, our sampling is still limited, with an average of 3.88 specimens per species, which highlights that fully prepared turtle shells are relatively rare in research collections. Although more data would certainly improve the dataset, we are sceptical that much additional sampling across the tree is possible using fully prepared specimens without prohibitive amounts of travel.

A number of methods are available to codify state frequency information from polymorphic characters. We investigated the resolution and accuracy of four methods for our data set [majority (MJ), missing (MI), generalized frequency coding (GFC) and polymorphic (PL)] by comparing the results they retrieve within a parsimony framework to a molecular reference tree built using previously published sequence data. We used a molecular reference tree, because molecular data provide a signal that is likely to be fully independent from our morphological data and produce a result that is both non-random and consistent with other, external data, such as biogeography. All methods failed to produce trees that were both accurate and well resolved. We therefore cannot recommend any particular coding method for the present dataset.

Given the unsatisfactory results achieved with morphological data alone, we investigated whether our dataset could be used in combination with a molecular backbone constraint to obtain meaningful results with a leave-one-out subsampling analysis, for which we reinserted extant species with ‘known’ relationships into our molecular reference tree using the morphological data. The majority of species were reinserted back into the tree close to their correct position.

We hope that the new character matrix produced in this study and our dataset will be useful for future studies on the phylogenetic relationships between extinct geoemydids and their living relatives.



## ACKNOWLEDGEMENTS

We would like to thank the following people for providing access to the material held at their institutions: David Kizirian and Lauren Vonnahme (American Museum of Natural History), Alan Resetar (Field Museum of Natural History), Andreas Schmitz (Muséum d'Histoire Naturelle de la Ville de Genève), Uwe Fritz, Raffael Ernst and Markus Auer (Museum für Tierkunde Dresden), Peter Pritchard and Zach Burke (Chelonian Research Institute), Addison Wynn (National Museum of Natural History), Loïc Costeur (Naturhistorisches Museum Basel) and Silke Schweiger and Georg Gassner (Naturhistorisches Museum Wien). We thank the editor Louise Allcock, Uwe Fritz and two anonymous reviewers for their comments and suggestions on this manuscript. We acknowledge the Willi Hennig Society for making possible the use of TNT free of charge. This work was supported by the Swiss National Science Foundation Grant 20021\_153502 to W.G.J. We hereby state that we have no conflicts of interests to disclose.

## REFERENCES

- Bapst DW. 2013.** When can clades be potentially resolved with morphology? *PLoS ONE* **8**: e62312.
- Bapst DW, Schreiber HA, Carlson SJ. 2017.** Combined analysis of extant Rhynchonellida (Brachiopoda) using morphological and molecular data. *Systematic Biology*. doi:10.1093/sysbio/syx049.
- Bardin J, Rouget I, Yacobucci MM, Cecca F. 2014.** Increasing the number of discrete character states for continuous characters generates well-resolved trees that do not reflect phylogeny. *Integrative Zoology* **9**: 531–541.
- Barth D, Bernhard D, Fritzsche G, Fritz U. 2004.** The freshwater turtle genus *Mauremys* (Testudines, Geoemydidae) — a textbook example of an east–west disjunction or a taxonomic misconception? *Zoologica Scripta* **33**: 213–221.
- Berger SA, Stamatakis A. 2010.** Accuracy of morphology-based phylogenetic fossil placement under maximum likelihood. In: *ACS/IEEE International Conference on Computer Systems and Applications - AICCSA 2010*, Hammamet, Tunisia, 1–8. doi:10.1109/AICCSA.2010.5586939.
- Boulenger GA. 1889.** *Catalogue of the chelonians, rhynchocephalians, and crocodiles in the British Museum (Natural History)*. London, UK: Trustees of the Museum of Natural History.
- Campbell JA, Frost DR. 1993.** Anguid lizards of the genus *Abronia*: revisionary notes, descriptions of four new species, a phylogenetic analysis, and key. *Bulletin of the American Museum of Natural History* **216**: 1–122.
- Carr JL. 1991.** Phylogenetic analysis of the neotropical turtle genus *Rhinoclemmys* Fitzinger (Testudines: Emydidae). D. Phil. Thesis. Carbondale, IL: Southern Illinois University.
- Catalano SA, Goloboff PA, Giannini NP. 2010.** Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework. *Cladistics* **26**: 539–549.
- Chang V, Smith EN. 2001.** *FastMorphologyGFC. Version 1.0.* Available at: <http://www3.uta.edu/faculty/ensmith>
- Chernomor O, von Haeseler A, Minh BQ. 2016.** Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65**: 997–1008.
- Claude J, Tong H. 2004.** Early Eocene testudinoid turtles from Saint-Papoul, France, with comments on the early evolution of modern Testudinoidea. *Oryctos* **5**: 3–45.
- Claude J, Zhang J-Y, Li J-J, Mo J-Y, Kuang X-W, Tong H. 2012.** Geoemydid turtles from the Late Eocene Maoming basin, southern China. *Bulletin de la Société Géologique de France* **183**: 641–651.
- Coddington J, Scharff N. 1994.** Problems with zero-length branches. *Cladistics* **10**: 415–423.
- Danilov IG. 2005.** Die fossilen Schildkröten Europas. In: Böhme W, ed. *Handbuch der Reptilien und Amphibien Europas*. Wiebelsheim, Germany: Aula-Verlag, 329–441.
- Diesmos AC, Parham JF, Stuart BL, Brown RM. 2005.** The phylogenetic position of the recently rediscovered Philippine forest turtle (Bataguridae: *Heosemys leytensis*). *Proceedings of California Academy of Sciences* **56**: 31–41.
- Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Ernst CH, Barbour RW. 1989.** *Turtles of the world*. Washington, DC: Smithsonian Institution press.
- Fritz U, Hundsdoerfer AK, Široký P, Auer M, Kami H, Lehmann J, Mazanaeva LF, Türkozan O, Wink M. 2007.** Phenotypic plasticity leads to incongruence between morphology-based taxonomy and genetic differentiation in western Palaearctic tortoises (*Testudo graeca* complex; Testudines, Testudinidae). *Amphibia-Reptilia* **28**: 97–121.
- Fritz U, Guicking D, Auer M, Sommer RS, Wink M, Hundsdoerfer AK. 2008.** Diversity of the Southeast Asian leaf turtle genus *Cyclemys*: how many leaves on its tree of life? *Zoologica Scripta* **37**: 367–390.
- Gaffney ES, Meylan PA. 1988.** A phylogeny of turtles. In: Benton MJ, ed. *The phylogeny and classification of the tetrapods, volume 1 – amphibians, reptiles and birds*. Oxford: Oxford University Press, 157–219.
- Gavryushkina A, Heath TA, Ksepka DT, Stadler T, Welch D, Drummond AJ. 2017.** Bayesian total-evidence dating reveals the recent crown radiation of penguins. *Systematic Biology* **66**: 57–73.
- Goloboff PA. 1993.** Estimating character weights during tree search. *Cladistics* **9**: 83–91.
- Goloboff PA. 2014.** Extended implied weighting. *Cladistics* **30**: 260–272.
- Goloboff PA, Catalano SA. 2016.** TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**: 221–238.
- Goloboff PA, Mattoni CI, Quinteros AS. 2006.** Continuous characters analyzed as such. *Cladistics* **22**: 589–601.



- Goloboff PA, Torres A, Arias JS. 2017.** Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. *Cladistics*. doi:10.1111/cla.12205
- Hay OP. 1908.** *The fossil turtles of North America*. Washington, DC: Carnegie Institution of Washington.
- Hervet S. 2004.** Systematic of the “Palaeochelys sensu lato – Mauremys” group (Chelonii, Testudinoidea) from the Tertiary of Western Europe: principal results. *Annales de Paléontologie* **90**: 13–78.
- Hervet S. 2006.** The oldest European ptychogasterid turtle (Testudinoidea) from the lowermost Eocene amber locality of Le Quesnoy (France, Ypresian, MP7). *Journal of Vertebrate Paleontology* **26**: 839–848.
- Hillis DM. 1987.** Molecular versus morphological approaches to systematics. *Annual review of Ecology and Systematics* **18**: 23–42.
- Hirayama R. 1985.** Cladistic analysis of batagurine turtles. *Studia Palaeocheloniologica* **1**: 140–157.
- Honda M, Yasukawa Y, Hirayama R, Ota H. 2002.** Phylogenetic relationships of the Asian box turtles of the genus *Cuora* sensu lato (Reptilia: Bataguridae) inferred from mitochondrial DNA sequences. *Zoological Science* **19**: 1305–1312.
- Hutchinson JH. 2006.** *Bridgeremys* (Geoemydidae: Testudines), a new genus from the middle Eocene of North America. *Russian Journal of Herpetology* **13**(Suppl.): 63–83.
- Johnson NK, Zink RM, Marten JA. 1988.** Genetic evidence for relationships in the avian family Vireonidae. *The Condor* **90**: 428–445.
- Joyce WG, Bell CJ. 2004.** A review of the comparative morphology of extant testudinoid turtles (Reptilia: Testudines). *Asiatic Herpetological Research* **10**: 53–109.
- Joyce WG, Lyson TR. 2010.** *Pangshura tatrotia*, a new species of pond turtle (Testudinoidea) from the Pliocene Siwaliks of Pakistan. *Journal of Systematic Paleontology* **8**: 449–458.
- Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. 2014.** Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* **14**: 82.
- Lapparent de Broin F. 2001.** The European turtle fauna. *Dumerilia* **4**: 155–217.
- Lawing AM, Meik JM, Schargel WE. 2008.** Coding meristic characters for phylogenetic analysis: a comparison of step-matrix gap-weighting and generalized frequency coding. *Systematic Biology* **57**: 167–173.
- Le M, McCord WP. 2008.** Phylogenetic relationships and biogeographical history of the genus *Rhinoclemmys* Fitzinger, 1835 and the monophyly of the turtle Geoemydidae (Testudines: Testudinoidea). *Zoological Journal of the Linnean Society* **153**: 751–767.
- Le M, McCord WP, Iverson JB. 2007.** On the paraphyly of the genus *Kachuga* (Testudines: Geoemydidae). *Molecular Phylogenetics and Evolution* **45**: 398–404.
- Maddison W. 1989.** Reconstructing character evolution on polytomous cladograms. *Cladistics* **5**: 365–377.
- Maddison WP, Maddison DR. 2016.** *Mesquite: a modular system for evolutionary analysis. Version 3.04*. Available at: <http://mesquiteproject.org>
- McCord WP, Iverson JB, Spinks PQ, Shaffer HB. 2000.** A new genus of geoemydid turtle from Asia. *Hamadryad* **25**: 20–24.
- McDowell SB. 1964.** Partition of the genus *Clemmys* and related problems in the taxonomy of the aquatic Testudinidae. *Proceedings of the Zoological Society of London B: Biological Sciences* **143**: 239–278.
- Meylan PA, Sterrer W. 2000.** *Hesperotestudo* (Testudines: Testudinidae) from the Pleistocene of Bermuda, with comments on the phylogenetic position of the genus. *Zoological Journal of the Linnean Society* **128**: 51–76.
- Minh BQ, Nguyen MAT, von Haeseler A. 2013.** Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30**: 1188–1195.
- Mikulíček P, Jandzik D, Fritz U, Schneider C, Šíroký P. 2013.** AFLP analysis shows high incongruence between genetic differentiation and morphology-based taxonomy in a widely distributed tortoise. *Biological Journal of the Linnean Society* **108**: 151–160.
- Mlynarski M. 1976.** Testudines. In: Kuhn O, ed. *Encyclopedia of paleoherpetology*. Stuttgart, Germany: Gustav Fischer Verlag, 1–130.
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015.** IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**: 268–274.
- O'Reilly JE, Puttick MN, Parry L, Tanner AR, Tarver JE, Fleming J, Pisani D, Donoghue PC. 2016.** Bayesian methods outperform parsimony but at the expense of precision in the estimation of phylogeny from discrete morphological data. *Biology letters* **12**: 20160081.
- Parham JF, Simison WB, Kozak KH, Feldman CR, Shi H. 2001.** New Chinese turtles: endangered or invalid? A reassessment of two species using mitochondrial DNA, allozyme electrophoresis and known-locality specimens. *Animal Conservation* **4**: 357–367.
- Pimentel RA, Riggins R. 1987.** The nature of cladistic data. *Cladistics* **3**: 201–209.
- Pol D, Escapa IH. 2009.** Unstable taxa in cladistic analysis: identification and the assessment of relevant characters. *Cladistics* **25**: 515–527.
- Praschag P, Hundsdoerfer AK, Fritz U. 2007.** Phylogeny and taxonomy of endangered South and South-east Asian freshwater turtles elucidated by mtDNA sequence variation (Testudines: Geoemydidae: Batagur, Callagur, Hardella, Kachuga, Pangshura). *Zoologica Scripta* **36**: 429–442.
- Praschag P, Ihlow F, Flecks M, Vamberger M, Fritz U. 2017.** Diversity of North American map and sawback turtles (Testudines: Emydidae: *Graptemys*). *Zoologica Scripta* **46**: 675–682.
- Praschag P, Schmidt C, Fritzsche G, Müller A, Gemel R, Fritz U. 2006.** *Geoemyda silvatica*, an enigmatic turtle of the Geoemydidae (Reptilia: Testudines), represents a distinct genus. *Organisms Diversity & Evolution* **6**: 151–162.
- Praschag P, Sommer RS, McCarthy C, Gemel R. 2008.** Naming one of the world's rarest chelonians, the southern *Batagur*. *Zootaxa* **1758**: 61–68.
- Pritchard PCH, Trebbau P. 1984.** *The turtles of Venezuela*. Oxford, OH: Society for the Study of Amphibians and Reptiles.

- Puttick MN, O'Reilly JE, Tanner AR, Fleming JF, Clark J, Holloway L, Lozano-Fernandez J, Parry LA, Tarver JE, Pisani D, Donoghue PC. 2017.** Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proceedings of the Royal Society of London B: Biological Sciences* **1846**: 20162290.
- Pyron RA. 2011.** Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60**: 466–481.
- Ratnasingham S, Hebert PDN. 2007.** The barcode of life data system (<http://www.barcodinglife.org>). *Molecular Ecology Notes* **7**: 355–364.
- R Core Team. 2016.** *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>
- Ronquist F, Klopstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. 2012.** A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61**: 973–999.
- Sasaki T, Yasukawa Y, Takahashi K, Miura S, Shedlock AM, Okada N. 2006.** Extensive morphological convergence and rapid radiation in the evolutionary history of the family Geoemydidae (old world pond turtles) revealed by SINE insertion analysis. *Systematic Biology* **55**: 912–927.
- Smith EN, Gutberlet RL Jr. 2001.** Generalized frequency coding: a method of preparing polymorphic multistate characters for phylogenetic analysis. *Systematic Biology* **50**: 156–169.
- Spinks PQ, Shaffer HB, Iverson JB, McCord WP. 2004.** Phylogenetic hypotheses for the turtle family Geoemydidae. *Molecular Phylogenetics and Evolution* **32**: 164–182.
- Stephens PR, Wiens JJ. 2003.** Ecological diversification and phylogeny of emydid turtles. *Biological Journal of the Linnean Society* **79**: 577–610.
- Stuart BL, Parham JF. 2004.** Molecular phylogeny of the critically endangered Indochinese box turtle (*Cuora galbinifrons*). *Molecular Phylogenetics and Evolution* **31**: 164–177.
- Stuart BL, Parham JF. 2007.** Recent hybrid origin of three rare Chinese turtles. *Conservation Genetics* **8**: 169–175.
- Swofford DL, Olse H. 1990.** Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. *Systematic Biology* **39**: 411–426.
- Tinkle DW. 1962.** Variation in shell morphology of North American turtles I. The carapacial seam arrangements. *Tulane Studies in Zoology* **9**: 331–349.
- Turtle Taxonomy Working Group. 2017.** Turtles of the world: annotated checklist and atlas of taxonomy, synonymy, distribution and conservation status (8<sup>th</sup> edition). *Chelonian Research Monographs* **7**: 1–292.
- Waagen GN. 1972.** *Musk glands in recent turtles*. Masters dissertation. Salt Lake City, UT: University of Utah.
- Watanabe A. 2016.** The impact of poor sampling of polymorphism on cladistic analysis. *Cladistics* **32**: 317–334.
- Wiens JJ. 1993.** Phylogenetic systematics of the tree lizards (genus *Urosaurus*). *Herpetologica* **49**: 399–420.
- Wiens JJ. 1995.** Polymorphic characters in phylogenetic systematics. *Systematic Biology* **44**: 482–500.
- Wiens JJ. 1998.** Testing phylogenetic methods with tree congruence: phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. *Systematic Biology* **47**: 427–444.
- Wiens JJ. 1999.** Polymorphism in systematics and comparative biology. *Annual Review of Ecology and Systematics* **30**: 327–362.
- Wiens JJ. 2001.** Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology* **50**: 689–699.
- Wiens JJ, Kuczyński CA, Stephens PR. 2010.** Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biological Journal of the Linnean Society* **99**: 445–461.
- Wiens JJ, Servedio MR. 1997.** Accuracy of phylogenetic analysis including and excluding polymorphic characters. *Systematic Biology* **46**: 332–345.
- Wiens JJ, Servedio MR. 2000.** Species delimitation in systematics: inferring diagnostic differences between species. *Proceedings of the Royal Society of London B: Biological Sciences* **267**: 631–636.
- Wink M, Guicking D, Fritz U. 2001.** Molecular evidence for hybrid origin of *Mauremys iversoni* Pritchard & McCord, 1991, and *Mauremys pritchardi* McCord, 1997. *Zoologische Abhandlungen* **51**: 41–49.
- Wright AM, Hillis DM. 2014.** Bayesian analysis using a simple likelihood model outperforms parsimony for estimation of phylogeny from discrete morphological data. *PLoS ONE* **9**: e109210.
- Yasukawa Y, Hirayama R, Hikida T. 2001.** Phylogenetic relationships of geoemydine turtles (Reptilia: Bataguridae). *Current Herpetology* **20**: 105–133.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix S1.** Specimens examined.

**Appendix S2.** Characters examined, with comments on their use and state distributions.

**Appendix S3.** TNT matrices for the four coding methods.

**Appendix S4.** Python script for majority coding method in R.

**Appendix S5.** GenBank accession numbers used for the reference phylogeny.

**Appendix S6.** Reference tree sequence alignment files.

**Appendix S7.** TNT script for our leave-one-out subsampling analysis.

**Appendix S8.** Supplementary figures.

**Figure S1.** Our molecular reference tree using maximum likelihood approach, uncollapsed, with bootstrap values over the branches. For modifications made over this tree, see Material and methods section, 'Reference phylogeny'.

**Figure S2.** Strict consensus of the most parsimonious trees for the molecular reference phylogeny. Bootstrap values > 70% are given over the branches.

**Figure S3.** Majority consensi of the most parsimonious trees estimated with each method, with bootstrap values over the branches.

**Figure S4.** The strict consensus tree of our total-evidence analysis combining molecular data and the polymorphic coding matrix.

**Figure S5.** Leave-one-out subsampling analysis accuracy results. Higher scores reflect higher accuracy. Dashed line shows the mean for all species. White circles represent the mean accuracy score for all the possible placements of one species.

**Figure S6.** Leave-one-out subsampling analysis resolution results. Higher scores reflect higher resolution.

**Figures S7–S74.** Illustration of character state definitions.