

This article was downloaded by:[National University of Ireland, Maynooth]
[National University of Ireland, Maynooth]

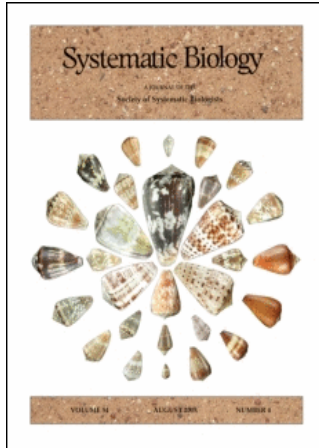
On: 14 March 2007

Access Details: [subscription number 769592964]

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954

Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Systematic Biology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713658732>

Investigating Stagnation in Morphological Phylogenetics Using Consensus Data

To link to this article: DOI: 10.1080/10635150601115624

URL: <http://dx.doi.org/10.1080/10635150601115624>

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

© Taylor and Francis 2007

Points of View

Syst. Biol. 56(1):125–129, 2007
 Copyright © Society of Systematic Biologists
 ISSN: 1063-5157 print / 1076-836X online
 DOI: 10.1080/10635150601115624

Investigating Stagnation in Morphological Phylogenetics Using Consensus Data

SIMON R. HARRIS,^{1,2,3} DAVIDE PISANI,^{1,4} DAVID J. GOWER,¹ AND MARK WILKINSON¹

¹Department of Zoology, The Natural History Museum, London, SW7 5BD, UK;
 E-mail: simon.harris@ncl.ac.uk (S.R.H.)

²School of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK

³School of Biology and Psychology, Division of Biology, University of Newcastle-upon-Tyne, NE1 7RU, UK

⁴Bioinformatics Laboratory, The National University of Ireland, Maynooth, Co. Kildare, Ireland

“Such debates are often a result of large gaps in the fossil record, rapid diversification within a lineage or highly derived morphologies within extant lineages that make phylogeny reconstruction difficult” (Meyer and Zardoya, 2003).

Rieppel and Kearney (2002) have recently called attention to the lack of progress in resolving several high-profile disagreements in vertebrate phylogeny that, because of the crucial importance of fossil taxa, depend upon the interpretation of morphological data. Perhaps the best-known example concerns the phylogenetic placement of Testudines (turtles) within Amniota. Most textbooks assert that turtles lie outside a clade including all other extant reptiles because they lack temporal fenestrae in their skull (the presumed plesiomorphic, anapsid condition). However, there has long been dissent based on the possibility that, like all other extant reptiles, turtles are diapsids, and that they have secondarily lost the two fenestrae for which this group is named (e.g., Goodrich, 1916). The controversy stems from the highly divergent morphology of the group and the lack of intermediate forms. The earliest known turtle, *Proganochelys quenstedtii* Baur, 1887, already has many of the features of modern turtles. This makes assessments of homology with potential close relatives difficult and, as evidenced by the literature, open to conflicting interpretations (homology assessments) that cannot all be correct.

It might have been hoped or expected that the application of modern numerical phylogenetic techniques would resolve the phylogenetic position of turtles and their extinct relatives. Molecular data provide some support for the diapsid hypothesis (Iwabe et al., 2005, and references therein), but recent numerical phylogenetic studies of morphology have largely recapitulated rather than resolved the controversy (Fig. 1), with osteological data interpreted repeatedly as supporting either anapsid (e.g., Lee, 1997, 2001) or diapsid (e.g., Rieppel and deBraga, 1996; Rieppel and Reisz, 1999) affinities of turtles. Consensus objects (e.g., trees, sequences) are widely used in biology (Day and McMorris, 2003). Here we use data pertaining to the relationships of turtles to illustrate

the use of a consensus approach for investigating conflict in morphological phylogenetics. We use this approach to identify and investigate agreements and disagreements between recent osteological data sets employed in the debate over diapsid and anapsid hypotheses and to investigate their importance in the stagnation of the debate on turtle origins.

Most systematic biologists have some familiarity with the use of consensus techniques to provide representations of and/or inferences from, for example, sets of trees. Consensus objects can be viewed as representations of the input objects and/or as inferences from the input objects. In general, a consensus method is a mapping of a set of input objects (e.g., trees) onto a set of one or more output objects of the same type (e.g., trees). The mapping corresponds to the application of some consensus rule, the commonest of which (e.g., strict, majority-rule) relate to the extent of agreement among the input objects. Because many objects are complex structures that may agree in some respects while disagreeing in others, construction of consensus objects usually involves the decomposition of complex objects into component parts to which the consensus rule is applied. For example, in constructing a strict component consensus tree, the input trees are decomposed into sets of components (full splits, clades in rooted trees) and unanimous agreement upon individual components (i.e., their presence in every input tree) is the condition for inclusion of a component in the consensus tree.

Where the objects are data matrices, the decomposition is into individual entries in each data matrix (i.e., the scoring of a taxon for a character) and it is the corresponding entries in different data matrices (i.e., the alternative scorings of a taxon for a character) that are subjected to a consensus rule. For example, unanimous agreement across the input data matrices is required for a datum to be included in strict consensus data. Where there is disagreement over the scoring of a character for a taxon, then at least one of the alternative scorings is incorrect and the conflict is represented by a missing

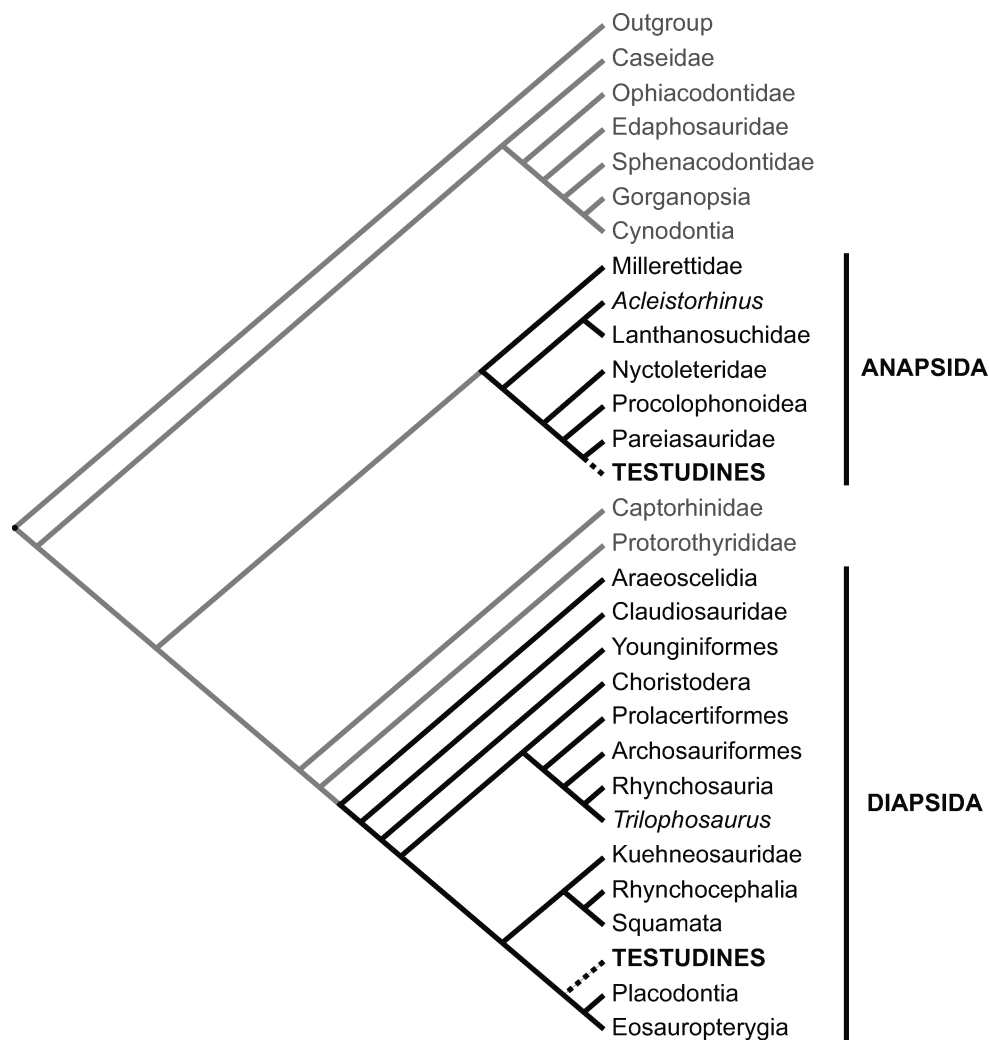


FIGURE 1. Phylogenetic tree showing with dashed lines the diapsid and anapsid placements of Testudines (turtles) favored by Rieppel and Reisz (1999) and Lee (2001), respectively, relative to most parsimonious relationships (solid lines) upon which these authors and the consensus data agree. Taxa in grey are neither diapsid nor anapsid.

entry in the strict consensus data. Varieties of consensus data, corresponding to different consensus rules, are considered in the appendix.

Given that alternative data sets for turtles have been analyzed using the same methodology and software, parsimony analysis as implemented in PAUP* 4.0b10 (Swofford, 2003), different results must, heuristics aside, be attributable to differences in the data. In fact, the most recent osteological data matrices of the leading protagonists are remarkably similar in size and scope with that of Lee (2001) having been developed from Rieppel and Reisz's (1999) compilation of 168 osteological characters for 34 taxa. However, minor differences in the ways in which Lee (2001) and Rieppel and Reisz (1999) divided or aggregated overlapping characters and taxa required that these data matrices be slightly modified in order to facilitate the comparison of alternative scorings of the same characters/taxa. Lee (2001) divided Rieppel and

Reisz's (1999) character 51 into two characters (51 + 169) using additive binary coding. These contrasting character constructions are analytically equivalent, and we arbitrarily chose to recombine Lee's (2001) characters 51 and 169 in a modified version of his matrix. Rieppel and Reisz (1999) included three exemplar pareiasaurs, *Anthodon*, *Bradysaurus*, and *Scutosaurus*, and two placodonts, *Placodus* and *Cyamodus*, that Lee (2001) combined into the Pareiasauridae and Placodontia, respectively. We combined the exemplar taxa in a modified version of Rieppel and Reisz's (1999) data, so that the new inclusive taxa were scored as possessing all states that were present in one or more of the exemplars. With the exception of the differences in included taxa, most parsimonious trees for the modified data sets are identical to those for the corresponding original input matrices. Thus our modifications do not alter the principal signals (*sensu* Pisani and Wilkinson, 2002) of the original data matrices.

TABLE 1. Five characters that contained significantly more scoring disputes than expected by random at the 5% level. Character numbers (Char) and descriptions from Rieppel and Reisz (1999). Taxa = taxa over which scoring is disputed; RR = scoring for the taxa in Rieppel and Reisz (1999); Lee = scoring in Lee (2001).

Char	Description	Taxa	RR	Lee
41	Quadrate anterior process: long, extending forward along its sutural contact with the quadrate process of the pterygoid to nearly reach the level of the transverse flange (0); short, not extending anteriorly beyond 55% the length of the quadrate process of the pterygoid (1).	Acleistorhinidae, Placodontia, Eosauropterygia	1	?
63	Basioccipital/basisphenoid relationship: floor of braincase with gap between both elements (0), elements fused to floor of brain cavity (1).	Ophiacodontidae, Millerettidae, Nycteroleteridae, Younginiformes	?	0
		Rhynchocephalia	?	1
		Edaphosauridae, Sphenacodontidae	1	0
		Squamata	0	1
69	Occipital flange: absent (0), present (1).	Outgroups, Caseidae, Ophiacodontidae, Edaphosauridae, Sphenacodontidae, Gorganopsia, Cynodontia, Captorhinidae, Protorothyrididae, Millerettidae, Acleistorhinidae, Lanthanosuchidae, Nycteroleteridae, Araeoscelididae, Claudiosauridae, Younginiformes	1	0
74	Suborbital fenestra: absent (0), present but with contribution from either maxilla or jugal along lateral border (1), present but with both maxilla and jugal excluded from lateral border (2).	Captorhinidae, Millerettidae, Procolophonoidea, Nycteroleteridae	1 ?	0 0
140	Femoral fourth trochanter: present (0), absent (1).	Pareiasauridae Placodontia Eosauropterygia	0 1 1	1 0 0 and 1

We constructed a strict consensus data matrix from these two modified matrices to assert only the observations and interpretations upon which Rieppel and Reisz (1999) and Lee (2001) explicitly agree. Consistent with our strict consensus rule, all characters (Lee's 170 to 176) and taxa (Rieppel and Reisz's Seymouridae) not present in both modified matrices were scored entirely with missing data in the consensus matrix and excluded from our further analyses. Across the 29 taxa, only 77 scoring differences in 38 of the 168 characters (<2% of cells) were represented with missing entries in the analyzed consensus matrix. That the disparate results of the previous studies turn on such a small number of disagreements over the scoring of a few morphological characters suggests that neither diapsid nor anapsid affinities for turtles can be considered robust (see also Wilkinson et al., 1997). Randomly assigning 77 "disagreements" to cells in a 29 × 168 matrix (10,000 replicates), we find that five characters (Table 1) have significantly more ($P < 0.05$) disagreements than expected by chance alone. Thus we can reject the null hypothesis that the scorings over which Rieppel and Reisz (1999) and Lee (2001) disagree are not concentrated in any particular characters. These are obvious targets for further scrutiny and possible reassessment.

Three taxa (Eosauropterygia, Pareiasauridae, and Testudines) also have significantly more ($P < 0.05$) conflicting scorings than expected by chance alone. The Testudines are the subject of most scoring conflicts (14), followed by Pareiasauridae (10) and Eosauropterygia (8). It is noteworthy that the latter two extinct taxa are those inferred by Rieppel and Reisz (1999) and Lee (2001),

respectively, to be the closest relatives of turtles. This concentration of disagreements is consistent with at least two explanations: that these taxa are the most intrinsically difficult to score, or that they have merely attracted greater attention and scrutiny. Whereas turtles are universally considered to have a disparate morphology over which disagreements might well be expected, the positioning of Eosauropterygia and Pareiasauridae as diapsids and anapsids, respectively, are less controversial. This suggests that the putative close relationship of these taxa to turtles has prompted greater scrutiny of their morphologies. It also raises the possibility that different assumptions, made consciously or otherwise, regarding the closeness of relationship of these taxa to turtles may have impacted upon how their morphologies have been interpreted by different experts. Thus, in a thought experiment where turtles are unknown, we might expect there to be more agreement in the scorings of the morphologies of the Eosauropterygia and Pareiasauridae.

In addition to locating and quantifying conflict, the strict consensus approach provides us with a matrix from which all scoring conflicts have been removed. Consensus data can be analyzed to investigate what phylogenetic inferences are supported by the less contentious and seemingly more certain data upon which the different experts agree. We can be certain that some mistakes (scoring errors) are excluded from the consensus data but at the cost of also excluding data that might well be free of error. It might be hoped that the strict consensus data contain a residual signal that favors the anapsid or diapsid hypothesis, demonstrating that one of the alternative hypotheses is more dependent upon

the least certain and more contentious data, and perhaps suggesting a resolution. Any residual signal in consensus data might alternatively support some other relationship of turtles or the removal of relevant data could simply destabilize their relationships. In fact, none of these results is obtained. Parsimony analysis of our strict consensus data yields three equally parsimonious trees, the two equally optimal diapsid trees of Rieppel and Reisz (1999) and the single anapsid solution of Lee (2001). This surprising result demonstrates that the data upon which the experts do not disagree still contain conflicting signals supporting both the diapsid and anapsid affinities of turtles. Disagreement over the relationships of turtles is not contingent upon, and cannot be simply attributed to, the disagreements and errors in the subjective interpretation of morphology. We interpret this as underscoring the real challenge that turtle morphology provides for homology assessment.

Rieppel and Kearney (2002:78) note "at least some debates in systematics devolve into what appear to be irresolvable arguments over character interpretations" and that resolution will only be achieved by more rigorous character analysis employing classical tests of homology. They note also (p. 59) that there is "a threat that preconceived notions of phylogeny influence character analysis," something considered also by Lee (1995) and Rieppel (1995). Our analysis suggests that this threat may have been realized in the debate over the affinities of turtles. Although rigorous character analysis can only help, we suspect that morphological interpretations are often made in the light of presumed understanding of the interrelationships of taxa and that phylogenetic assumptions may be difficult to avoid (see also Gower and Weber, 1998:405). In practice, which taxa are allowed to shed the most light equates to assumptions about their relative phylogenetic relevance, differences in which may contribute to the stagnation of morphological phylogenetic controversies. Perhaps the best we can hope for is that alternative assumptions be recognized, explored, and subject to extensive critical discussion.

In our view, the current lack of resolution of arguments over character interpretations in turtle phylogenetics reflects the inability of experts to identify or agree upon objective criteria for resolving their seemingly irreconcilable differences in how best to score particular taxa for particular characters: manifestations of differences in philosophies of character scoring, phylogenetic assumptions, or both. This is surprising and disappointing given the fundamental importance of such criteria and the age of the discipline of comparative morphology. It is to be hoped that modern phylogenetic comparative morphology can make more rapid progress in addressing the fundamental issues that must underpin differences in scoring. In the case of turtles, we believe the best hope for resolving entrenched disagreements would be provided by the injection of a healthy dose of independent critical scrutiny of the osteological data and the generation of a more than two-sided debate. Surprisingly, perhaps, although the debate over turtle affinities has achieved some notoriety, the recent morphological

data compilations have not yet received extensive independent scrutiny.

The controversy might also be resolved by the discovery of extinct taxa displaying intermediate morphologies, incorporating additional taxa and/or morphological characters into analyses (e.g., Hill, 2005), or more probably by increased molecular data, all of which would then allow some assessment of the relative performance of experts' seemingly private or subjective criteria of homology assessment and character construction. In this study we have used a strict consensus rule to illustrate the use of consensus data, but obvious alternatives, such as the majority rule, may prove useful where there are more than two expert opinions. In addition to identifying disagreement, consensus rules could provide an objective basis for choosing among hypotheses if there really is no other way of resolving disagreements. We hope that consensus data might prove a useful tool for exploring controversies in morphological phylogenetics, and that there would be some basis other than a consensus rule for resolving any disagreements.

ACKNOWLEDGEMENTS

This work was supported by NERC PhD studentship NER/S/A/2000/03255, an NHM MRF award, BBSRC grant 40/G18385, a Marie Curie Intra European Individual Fellowship (Contract Number MEIF-CT-2005-010022), and by the Michigan Center for Theoretical Physics. We thank Michael Benton, James Cotton, Samantha Mohun, Hendrik Müller, Buck McMorris, and David Polly for helpful discussions and Mike Lee and Olivier Rieppel for critical reviews.

REFERENCES

- Baur, G. 1887. Ueber den Ursprung der Extremitäten der Ichthyopterygia. Jahresberichte und Mitteilungen des Oberrheinischen Geologischen Vereines 20:17–20.
- Day, W. H. E., and F. R. McMorris. 1992. Critical comparison of consensus methods for molecular sequences. *Nucleic Acids Res.* 20:1093–1099.
- Day, W. H. E., and F. R. McMorris. 2003. Axiomatic consensus theory in group choice and bioinformatics. *Frontiers in Applied Mathematics* Volume 39. Society for Industrial and Applied Mathematics, xvi + 155.
- Goodrich, E. S. 1916. On the classification of the Reptilia. *Proc. R. Soc. Lond. B.* 89:261–276.
- Gower, D. J., and E. Weber. 1998. The braincase of *Euparkeria*, and the evolutionary relationships of birds and crocodylians. *Biol. Rev.* 73:367–411.
- Hill, R. V. 2005. Integration of morphological data sets for phylogenetic analysis of Amniota: The importance of integumentary characters and increased taxon sampling. *Syst. Biol.* 54:530–547.
- Iwabe, N., Y. Hara, Y. Kumazawa, K. Shibamoto, Y. Saito, T. Miyata, and K. Katoh. 2005. Sister group relationships of turtles to the bird-crocodylian clade revealed by nuclear DNA-coded proteins. *Mol. Biol. Evol.* 22:810–813.
- Lee, M. S. Y. 1995. Historical burden in systematics and the interrelationships of "parareptiles." *Biol. Rev.* 70:459–547.
- Lee, M. S. Y. 1997. Reptile relationships turn turtle. *Nature* 389:245–246.
- Lee, M. S. Y. 2001. Molecules, morphology, and the monophyly of diapsid reptiles. *Contrib. Zool.* 70:1–22.
- Meyer, A., and R. Zardoya. 2003. Recent advances in the (molecular) phylogeny of vertebrates. *Ann. Rev. Ecol. Evol. Syst.* 34:311–338.
- Pisani, D., and M. Wilkinson. 2002. MRP, taxonomic congruence and total evidence. *Syst. Biol.* 51:151–155.
- Rieppel, O. 1995. Studies on skeleton formation in reptiles: Implications for turtles relationships. *Zoology* 98:298–308.
- Rieppel, O., and M. deBraga. 1996. Turtles as diapsid reptiles. *Nature* 384:453–455.

- Rieppel, O., and M. Kearney. 2002. Similarity. *Biol. J. Linn. Soc.* 75: 59–82.
- Rieppel, O., and R. R. Reisz. 1999. The origin and early evolution of turtles. *Ann. Rev. Ecol. Syst.* 30:1–22.
- Swofford, D. L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, Massachusetts.
- Wilkinson, M. 1994. Common cladistic information and its consensus representation: Reduced Adams and reduced cladistic consensus trees and profiles. *Syst. Biol.* 43:343–368.
- Wilkinson, M., J. Thorley, and M. J. Benton. 1997. Uncertain turtle relationships. *Nature* 387:466.

First submitted 13 October 2005; reviews returned 21 February 2006;

final acceptance 6 June 2006

Associate Editor: Michael Lee

APPENDIX

VARIETIES OF CONSENSUS DATA

There are a variety of rules that could be used to produce consensus data. We have used a simple, strict consensus rule to illustrate the potential for consensus investigations to yield useful insights. Here we further characterize this rule and two others, while noting the potential for variants that that may prove more or less useful. Consensus concepts have probably been used informally in phylogenetics in constructing data sets that incorporate previous work, and consensus sequences (Day and McMorris, 1992) are ubiquitous in molecular biology.

We take as input a (multi)set of data matrices, each of which comprises scores for exactly the same characters and taxa. For each character of each taxon we compare the corresponding scores across the data matrices and apply a rule that determines the corresponding score in the consensus data as a function of the scores in the input matrices.

Strict Consensus

Strict consensus data include all and only those scores that are identical across all input matrices. Any departure from unanimity leads to a corresponding missing entry (a null score) in the consensus data. Thus the strict consensus data are those upon which all sources explicitly agree.

Missing entries, “polymorphic” scores, and ambiguity codes (or any score for that matter) might stand for a number of things. For simplicity, any score in an input matrix is interpreted as the assertion (by the source) that the available data are best represented (for whatever reason) by that score (whatever its interpretation); i.e., that the score is correct. Note that a missing entry in strict consensus data indicates either disagreement among the input matrices or that all have a corresponding missing entry.

Majority-Rule Consensus

Majority-rule consensus data include all and only those scores that are in a majority of the input matrices, those upon which a majority of sources agree. Lack of a majority leads to a corresponding missing entry in the consensus data. The interpretation of scores, including missing entries and polymorphisms, is as above. Thus, the majority-rule naturally generalizes to stricter (up to 100%, the strict) rules.

Semistrict Consensus

Semistrict consensus data include all and only those scores that are uncontradicted. Lack of an uncontradicted score leads to a corresponding missing entry in the consensus data. The semistrict consensus data are those upon which the sources do not explicitly disagree.

If we interpret all scores, including missing entries and polymorphic scores, as the correct scores according to the source, as above, then scores contradict (over which is the correct score) if they are different. In that case, there is no difference between strict and semistrict consensus data. The semistrict method becomes distinct when we interpret input scores differently, as the source’s lesser assertion that their score does not contradict the correct score.

Polymorphic scores (and ambiguity codes), e.g., $\{X/Y\} = X$ or Y , can represent either real polymorphism or uncertainty as to which score obtains, analogous to the hard and soft interpretations of polytomies in trees, respectively. Only with the latter is noncontradiction of different scores a possibility. Adopting this interpretation, we consider that a score J does not contradict another score K if and only if K entails J . For example, Y entails X or Y , and thus $\{X/Y\}$ does not contradict Y . In contrast, $\{X/Y\}$ does not entail Y , and thus Y contradicts $\{X/Y\}$. In general, the semistrict rule returns the intersection of the scores (e.g., $\{X/Y\} \cap \{X\} = \{X\}$), with polymorphic scores interpreted as sets of scores and missing entries interpreted as the set of all possible scores.

Other Possibilities

Consider a character with the three alternative scores X , Y , and Z . It might be argued that the polymorphic score $\{X/Y\}$ and the alternative score Y agree that Z is an incorrect score, and that this agreement should be represented in the strict consensus, using the score $\{X/Y\} = \text{not } Z$, rather than a missing entry. Wilkinson (1994) called such negative agreements disqualifiers. The same logic extends to multistate characters in the absence of polymorphism; i.e., the scores X and Y in input matrices lead to the polymorphic score $\{X/Y\}$ (the disqualifier, not Z) in the consensus. This does not define a strict consensus method, because a score in the consensus data may not be in every (or in any) input matrix. Nonetheless, adding this additional agreement to the strict (or semistrict) consensus data might yield additional insights. In the present case, the corresponding disqualifier consensus leads to a nonsignificant difference between the anapsid and diapsid trees, with the anapsid tree one step shorter.

The majority rule, as defined, can be interpreted like the strict in terms of the sources asserting that their scores are correct. Thus the polymorphic score $\{X/Y\}$ would be in the majority-rule consensus data, as defined, only if it is present in a majority of the input trees, and a number of variants can be envisaged that might usefully employ “softer” interpretations of the input data. Most importantly, a score that is present in a majority of the input matrices that cast a vote may not be a majority of all the matrices when some contain missing entries. In that case, it seems sensible to only count the votes cast and to ignore missing entries.

Beyond consensus, data matrices may include different characters and taxa. The semistrict approach generalizes readily to the construction of supermatrices. Conceptually, the supermatrix problem is converted into a consensus problem by embedding each input matrix in a supermatrix otherwise composed only of missing entries. Majority-rule also generalizes if missing entries are not counted.