

Exploring the effects of phylogenetic uncertainty and consensus trees on stratigraphic consistency scores: a new program and a standardized method

Clint A. Boyd^{a,†,*}, Timothy P. Cleland^a, Nico L. Marrero^b and Julia A. Clarke^{a,†}

^a*Department of Marine, Earth, and Atmospheric Sciences, Campus Box 8208, North Carolina State University, Raleigh, NC 27695, USA;*

^b*New Mexico Institute of Mining and Technology, 801 Leroy Place, Socorro, NM 87801, USA*

Accepted 22 March 2010

Abstract

The stratigraphic record of first appearances provides an independent source of data for evaluating and comparing phylogenetic hypotheses that include taxa with fossil histories. However, no standardized method exists for calculating these metrics for polytomous phylogenies, restricting their applicability. Previously proposed methods insufficiently deal with this problem because they skew or restrict the resulting scores. To resolve this issue, we propose a standardized method for treating polytomies when calculating these metrics: the Comprehensive Polytomy approach (ComPoly). This approach accurately describes how phylogenetic uncertainty, indicated by polytomies, affects stratigraphic consistency scores. We also present a new program suite (Assistance with Stratigraphic Consistency Calculations) that incorporates the ComPoly approach and simplifies the calculation of absolute temporal stratigraphic consistency metrics. This study also demonstrates that stratigraphic consistency scores calculated from strict consensus trees can be overly inclusive and those calculated from less-than-strict consensus trees inaccurately describe the phylogenetic signal present in the source most-parsimonious trees (MPTs). Therefore, stratigraphic consistency scores should be calculated directly from the source MPTs whenever possible to ensure their accuracy. Finally, we offer recommendations for standardizing comparisons between molecular divergence dates and the stratigraphic record of first appearances, a promising new application of these methods.

© The Willi Hennig Society 2010.

Stratigraphic consistency metrics estimate the congruence between the branching pattern of a cladogram and the stratigraphic order of the oldest known records [OKRs *sensu* Walsh (1998); first appearance data of Pol and Norell (2006)] of those taxa with fossil histories. These metrics assume that as our understanding of the fossil record increases, phylogenetic hypotheses should become increasingly congruent with the stratigraphic record. Under this assumption, phylogenies close to the true tree should display a close fit to the fossil record. Over the past two decades, various stratigraphic consistency metrics have been proposed (Gauthier et al., 1988;

Norell and Novacek, 1992; Benton and Storrs, 1994; Huelsenbeck, 1994; Siddall, 1998; Wills, 1999; Marjanovic and Laurin, 2007), investigated for biases and limitations (Norell, 1993; Siddall, 1996, 1997; Hitchin and Benton, 1997a,b; Wills, 1999; Wagner and Sidor, 2000; Pol et al., 2004), and modified to improve their performance (Pol and Norell, 2001, 2006; Wills et al., 2008). Stratigraphic consistency scores have been used as descriptive statistics (Villier et al., 2004; Saucedo et al., 2007; Tetlie and Poschmann, 2008), to compare alternative phylogenetic hypotheses (e.g. Pryer, 1999; O'Leary, 2001; Wilson, 2002; Marivaux et al., 2004; Brusatte and Sereno, 2008), to estimate the completeness of the fossil record for a particular clade (e.g. Kerr and Kim, 2001; Angielczyk and Kurkin, 2003; Jeffery and Emler, 2003), to examine alternative positions of taxa within phylogenies (e.g. Brochu and Norell, 2000; Pol

*Corresponding author:

E-mail address: clintboyd@strafit.org

†Current address: Jackson School of Geosciences, The University of Texas at Austin, 1 University Station C1100, Austin, TX 78712, USA.

and Norell, 2006), and to calculate the incongruence between molecular divergence dates and the fossil record of first appearances for a given topology (Clarke et al., 2007; Marjanovic and Laurin, 2007).

These methods have been applied to a diverse array of taxa, including marsileaceous ferns (Pryer, 1999), harpid gastropods (Merle and Pacaud, 2003), arthropods (Wills, 2001), adelophthalmoid eurypterids (Tetlie and Poschmann, 2008), echinoids (Jeffery and Emlet, 2003; Villier et al., 2004; Saucedo et al., 2007), holothuroids (Kerr and Kim, 2001), lissamphibians (Marjanovic and Laurin, 2007), amniotes (Gauthier et al., 1988), dicynodont therapsids (Angielczyk and Kurkin, 2003), cetaceans (O'Leary, 2001), rodents (Marivaux et al., 2004), and several dinosaurian clades (Brochu and Norell, 2000; Wilson, 2002; Rauhut, 2003; Pol and Norell, 2006; Brusatte and Sereno, 2008; Wills et al., 2008). Despite the widespread use of these metrics, a standardized method for treating polytomies when calculating stratigraphic consistency metrics has yet to emerge, although some authors have commented on this topic. One of the first proposals was to address polytomies by assigning the OKR of the oldest taxon within the polytomy to the uncertain node, effectively disregarding the stratigraphic data provided by the other taxa within the polytomy (Huelsenbeck, 1994). Alternatively, a command within the Manhattan Stratigraphic Measure script (MSM: Siddall, 1998) uses a heuristic search to optimize relationships within polytomies, maximizing stratigraphic congruence. That command was retained in the subsequent modification of the metric by Pol and Norell (2001), termed MSM*.

The most extensive discussion regarding how to calculate stratigraphic consistency metrics for polytomous phylogenies is found in Wills (1999), which introduces the program Ghosts. Ghosts calculates the relative completeness index (RCI; Benton and Storrs, 1994), the stratigraphic consistency index (SCI; Huelsenbeck, 1994), and the gap excess ratio (GER; Wills, 1999) metrics, and allows the user to rearrange taxa within a polytomy such that stratigraphic congruence is either maximized or minimized (i.e. the 'worst case' and 'best case' scenarios of Wills, 1999: supplementary documentation). In the instructions for that program the author suggests the 'worst case' scenario should be employed such that if a polytomous phylogeny is found to be more stratigraphically congruent than an alternative phylogeny it will be because of the order of the resolved nodes and not because of a hypothetical optimal arrangement of taxa within the polytomies.

All of these solutions are limited in that they either trim or rearrange the unresolved branches to produce a dichotomous branching pattern that may not accurately represent the total phylogenetic signal present in the set of most-parsimonious trees (MPTs). The goals of this paper are to document how different approaches to

resolving polytomies affect stratigraphic consistency metrics, to propose a standardized method for calculating stratigraphic consistency measures for polytomous phylogenies, to introduce a new program that simplifies the calculation of stratigraphic consistency metrics, to discuss how molecular divergence dates should be treated when estimating their incongruence with the stratigraphic record of first appearances for a given topology, and to demonstrate that scores calculated from consensus trees imperfectly describe the signal present in the source MPTs. The following discussion is limited to absolute temporal metrics (e.g. MSM* and GER), and their associated range metrics (e.g. MSM* range and GER range; Pol and Norell, 2006), because it was previously demonstrated that these metrics are least influenced by variations in the size, shape, and scale of the tree(s) being analysed (Pol et al., 2004).

Examining methods for treating polytomies

Developing a standardized method for calculating stratigraphic consistency metrics from polytomous phylogenies (i.e. containing at least one polytomy) first requires an understanding of how different methods of treating polytomies affect stratigraphic consistency scores. Five distinct methods for treating polytomies are here recognized and classified as either reducing or restructuring based upon how they deal with taxa within a polytomy. Figure 1 contains a hypothetical pair of alternative phylogenies, one polytomous (Fig. 1a) and the other dichotomous (i.e. fully resolved; Fig. 1b). Both trees contain 13 taxa, designated with the letters A–M, and the OKRs of these taxa were set precisely 1 million years apart starting with taxon A at 1 Myr and increasing in age according to alphabetical order. Examining how each of the five methods for treating polytomies affects the resulting stratigraphic consistency scores of both phylogenies illustrates that the choice of method can influence which tree is found to be more stratigraphically congruent. For all of the methods discussed below, higher-level taxa within polytomies are evaluated relative to each other based upon the oldest OKR present within each taxon, a rule that should be followed whenever calculating stratigraphic consistency metrics for polytomous phylogenies.

Reducing methods

The three reducing methods for treating polytomies are: (i) prune all taxa contained within a polytomy from the tree [e.g. the method employed by Marivaux et al. (2004); hereafter, referred to as the Trimmed Method]; (ii) use only the taxon with the oldest OKR to represent all taxa within the polytomy [i.e. the method proposed by Huelsenbeck (1994); hereafter, referred to as the

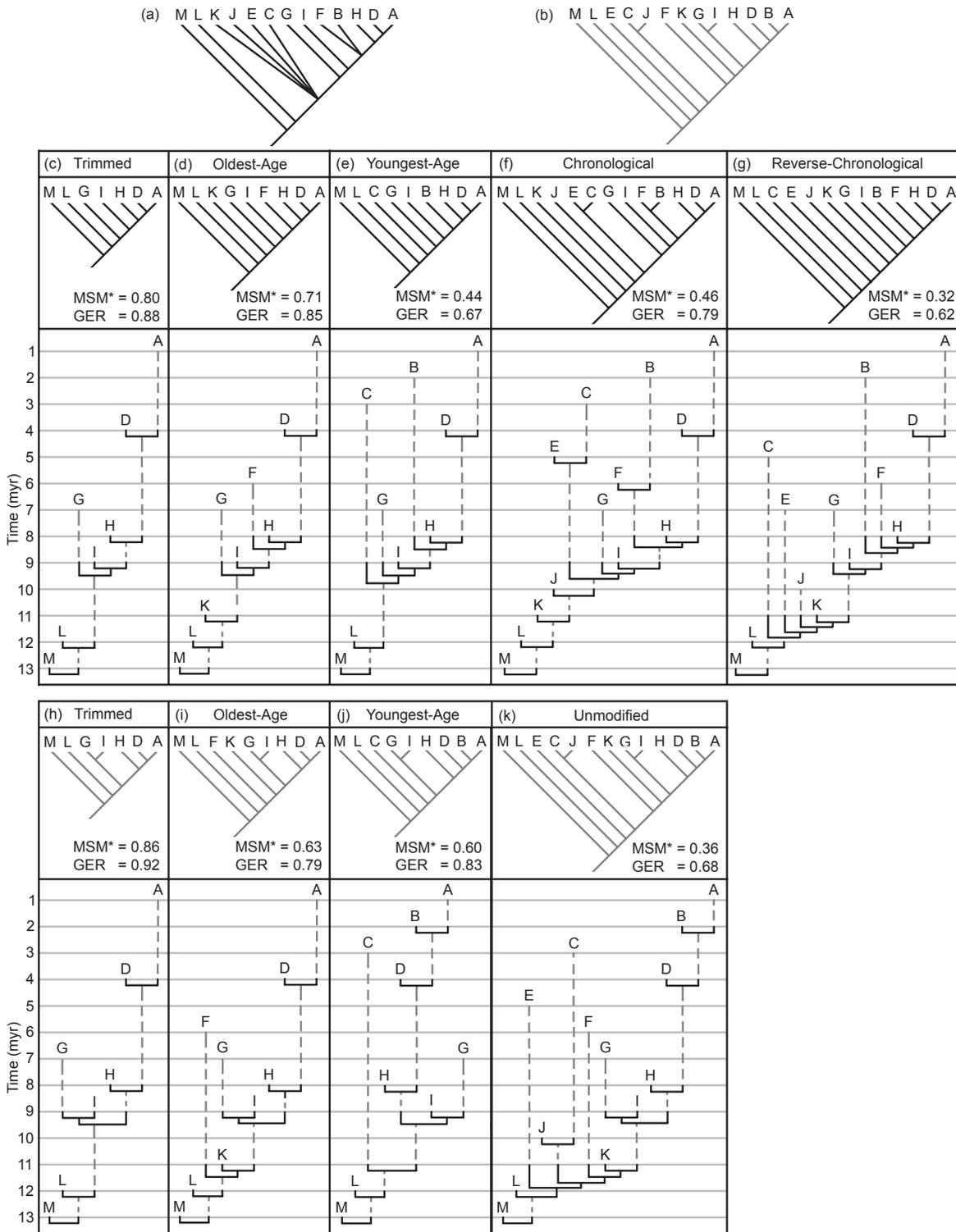


Fig. 1. The effects the five different methods for treating polytomies have on stratigraphic consistency metrics. Two hypothetical phylogenetic hypotheses are compared, one of which contains polytomies (a; drawn in black) while the other is dichotomous (b; drawn in grey). The polytomous phylogeny was modified using the five methods for resolving polytomies: (c) the Trimmed method; (d) the Oldest-Age method; (e) the Youngest-Age method; (f) the Chronological method; and (g) the Reverse-Chronological method. To facilitate accurate comparison of stratigraphic consistency scores, corresponding taxa were trimmed from the dichotomous phylogeny as follows: (h) for comparison with the Trimmed method; (i) for comparison with the Oldest-Age method; (j) for comparison with the Youngest-Age method; and (k) the unmodified tree for comparison with both the Stratigraphic and the Reverse-Stratigraphic methods. The scores for both the MSM* and the GER metrics are provided below each cladogram.

Oldest-Age method]; and (iii) use only the taxon with the youngest OKR to represent all taxa within the polytomy (hereafter, referred to as the Youngest-Age method). All of these methods involve removing taxa from the polytomous tree and ignoring their associated age data (Fig. 1c–e). Therefore, the corresponding taxa must be removed from the dichotomous tree (Fig. 1h–j) prior to calculating stratigraphic consistency scores to facilitate accurate comparisons between these trees (Gauthier et al., 1988; Wills et al., 2008) due to the effect tree size (i.e. number of taxa) has on these metrics (Pol et al., 2004).

Restructuring methods

The two restructuring methods for treating polytomies are: (i) situate taxa in a pectinate arrangement to maximize fit to stratigraphy [e.g. the method implemented by the MSM* script (Siddall, 1998; Pol and Norell, 2001); hereafter referred to as the Chronological method]; and (ii) situate taxa in a pectinate arrangement in reverse chronological order [e.g. the method suggested by Wills (1999); hereafter referred to as the Reverse-Chronological method]. No modifications to the dichotomous tree are necessary when these methods are employed because they do not ignore age data from any taxa, resulting in the retention of identical sets of taxa in each tree. Therefore, scores from both restructured polytomous topologies (Fig. 1f, g) are compared with those from the unmodified dichotomous tree (Fig. 1k).

Results

The example in Fig. 1 illustrates that recognition of the more stratigraphically congruent topology is dependent upon the reducing or restricting method chosen. The Trimmed (Fig. 1c vs. h), Youngest-Age (Fig. 1e vs. j), and Reverse-Chronological (Fig. 1g vs. k) methods result in recovery of relatively lower stratigraphic consistency scores for the modified polytomous tree, while the Oldest-Age (Fig. 1d vs. i) and Chronological (Fig. 1f vs. k) methods result in relatively higher scores for the modified polytomous tree. The scores obtained for the restructured polytomous topologies using the Chronological (Fig. 1f) and Reverse-Chronological (Fig. 1g) methods mark the upper and lower bounds of the total variation that can be produced by restructuring the unresolved branches in the polytomous tree (Wills, 1999), reflecting the maximum amount of variation that could be present in the source MPTs. Both the Trimmed (Fig. 1c) and the Oldest-Age (Fig. 1d) methods produce scores that fall outside of this range because they modify properties that these metrics are sensitive to (e.g. tree size and number of OKRs; Pol et al., 2004).

Similarly, trimming the dichotomous tree to facilitate comparison with the reduced polytomous topologies resulted in stratigraphic consistency scores that are higher than those obtained from the unmodified tree for the same reasons (Fig. 1h–j vs. k). Based on these results, restructuring methods are preferred over reducing methods because they rearrange rather than prune taxa. This allows the full range of variation that phylogenetic uncertainty, indicated by polytomies, imparts on stratigraphic consistency scores to be understood, maximizes the amount of stratigraphic data available to compare alternative phylogenetic hypotheses, and prevents recovery of scores that fall outside of the range of possible variation present in the source MPTs.

ComPoly and the ASCC program suite

The manner in which the uncertainty in the age of the OKRs of taxa with fossil histories is handled has been shown to impact the resulting stratigraphic consistency scores (Pol and Norell, 2006). The MSM* range and GER range metrics were designed so that the effect this uncertainty has on stratigraphic consistency scores is taken into consideration (Pol and Norell, 2006). We propose that the effects of phylogenetic uncertainty, reflected by the presence of polytomies, on these metrics be accounted for in a similar manner: by combining the scores obtained using the Chronological and Reverse-Chronological methods to produce a range of scores that summarize this variation, here termed the Comprehensive Polytoamy approach (ComPoly). When the ComPoly approach is employed when calculating stratigraphic consistency scores, a lower-case p is added to the stratigraphic consistency metric acronym (e.g. MSMp* and GERp), just as the word ‘range’ is added when the uncertainty in the age of OKRs of taxa with fossil histories is taken into consideration (e.g. MSM* range and GER range). If this is done for the polytomous tree from Fig. 1a using the scores from Fig. 1f, g, a range of 0.46–0.32 for MSMp* and 0.79–0.62 for GERp are obtained. The corresponding scores for the unmodified dichotomous tree (Fig. 1k) fall within that range, indicating that neither tree can be selected as more congruent with the stratigraphic record. Implementation of any other approach for treating polytomies would result in a different, inaccurate conclusion. Constructing a range score that accounts for both phylogenetic and age uncertainty requires that the highest score obtained for the Chronologically restructured topology is combined with the lowest score obtained from the Reverse-Chronologically restructured topology. The resulting range score effectively characterizes the full range of variation that age and phylogenetic uncertainty impart on these metrics.

To streamline the process of calculating an array of stratigraphic consistency metrics that account for age and/or phylogenetic uncertainty (e.g. MSM* range and MSMp* range), a collection of files was developed, collectively referred to as the Assistance with Stratigraphic Consistency Calculations (ASCC) program suite (freely available at <http://www.stratfit.org>). A brief introduction to the files included in the program suite is provided below, and more detailed instructions are given in the readme file included in the suite.

Using the ASCC.pl script to construct data files

Previously published scripts that calculated stratigraphic consistency range metrics required manual construction of a set of data files that could be somewhat time-consuming to generate and to troubleshoot when constructed incorrectly. These issues may have discouraged some researchers from calculating these statistics despite their advantages over standard stratigraphic consistency metrics. The ASCC.pl script included in the ASCC program suite simplifies the construction of the agefile, treefile, and data.tnt files required to calculate stratigraphic consistency range metrics by guiding the user through a set of questions about the details of the tree(s) and taxa being analysed. The script can be run in any Perl interpreter (see readme file for suggested programs). Taxa and trees can be read directly from files created by phylogenetic programs (e.g. NEXUS and tree files). Taxon names and trees can also be manually entered in the Newick notation format utilized by the phylogenetic program TNT (i.e. spaces separate terminal taxa, not commas; Goloboff et al., 2008), and terminal taxa should be numbered beginning with 1, because 0 is reserved for a hypothetical root taxon. When using the ComPoly approach, two topologies (Reverse-Chronologically and Chronologically restructured) are entered in that order for each polytymous tree analysed. There is no limit to the size or number of trees that can be entered and analysed at one time, although all trees must include identical sets of taxa to facilitate accurate comparison of scores (Gauthier et al., 1988; Wills et al., 2008). The user assigns each terminal taxon to a specific age bin, which are user-defined age ranges that encompass the uncertainty in the age of each taxon's OKR. When the program is finished, the agefile, treefile, and data.tnt files are created in the same location as the ASCC.pl script.

Using the stratfit.run script to calculate scores

The stratfit.run script is a modified version of the ageuncert.run script created by Pol and Norell (2006) for the phylogenetic program TNT. Whereas the former calculated only the MSM* range metric, the stratfit.run script facilitates the calculation of additional strati-

graphic consistency metrics [i.e. GER range and minimum estimate of stratigraphic gap (MIG; Wills, 1999) range]. Before running stratfit.run, ensure it is in the same folder as the files created by the ASCC.pl script and the program TNT. Next, run TNT and type in 'stratfit.run' and press enter. For each replicate of the analysis (default = 1000), a score for MIG, MSM*, and GER is calculated using a set of OKRs randomly selected for each taxon from the age bins defined using the ASCC.pl script. Three output files are generated (MIG.out, msm.out, and ger.out), which contain a list of the scores calculated for the respective metric during each replicate.

Interpreting the data

A set of Microsoft Excel macro files are included in the ASCC program suite. These files were developed to simplify processing the data generated by TNT. When the scores stored in the output files are imported into these macros, the stratigraphic consistency range scores are automatically reported (see readme.pdf file for detailed instructions). These files also construct histograms of the pair-wise differences between the MSM* and GER scores (or MSMp* and GERp scores when using the ComPoly approach) generated by each replicate of the analysis. Comparison of these output histograms is used to determine if one phylogeny is consistently more stratigraphically congruent (i.e. histogram plots entirely to one side of zero), or if the phylogenies are equally stratigraphically congruent (i.e. histogram crosses zero) when the resulting range scores overlap, as described in Pol and Norell (2006).

Discussion

The restructuring methods used by the ComPoly approach rearrange, rather than trim, taxa placed at unresolved nodes in polytymous phylogenies, thereby maximizing the amount of stratigraphic data available to evaluate congruence between phylogeny and the fossil record. Previously proposed approaches do not describe the full effect that phylogenetic uncertainty has on stratigraphic consistency scores. The ComPoly approach fully describes this variation and can be combined easily with existing stratigraphic consistency metrics (e.g. GER, MIG, and MSM*) using the new ASCC program suite. Here we discuss the benefits of using the ComPoly approach and the ASCC program suite for a new application of stratigraphic consistency metrics and highlight the importance of carefully selecting the type of tree used to calculate stratigraphic consistency scores.

Stratigraphic consistency scores are often used to compare the stratigraphic congruence of alternative

phylogenetic hypotheses for a given clade by holding the stratigraphic data (OKRs) constant while varying the tree topology. Two recent papers have expanded upon that methodology by exploring the ability of these metrics to evaluate the stratigraphic congruence of molecular divergence dates (Clarke et al., 2007; Marjanovic and Laurin, 2007). For example, in the Clarke et al. (2007) analysis, both the OKRs and the tree topology were held constant. Instead, the MIG value obtained by comparing the tree topology solely with the OKRs was contrasted with the value obtained for the same tree topology when molecular divergence dates for certain clades were enforced. Both studies had to account for uncertainty in the OKRs of taxa with fossil histories and the presence of polytomies within the trees analysed. Despite the exhaustive nature of the study by Marjanovic and Laurin (2007), the calculation of stratigraphic consistency metrics was hampered by the fact that the investigation was completed before the publication, and without the benefit, of stratigraphic consistency range metrics (Pol and Norell, 2006), the ComPoly approach, and the ASCC program suite. The study by Clarke et al. (2007) used stratigraphic consistency range metrics (i.e. MIG range), the ComPoly approach, and an early version of the ASCC program suite to address age and phylogenetic uncertainty. Exploring the methodological differences between these studies, with regards to the calculation of stratigraphic consistency scores, clearly demonstrates the advantages of using both the ComPoly approach and the ASCC program suite. It also highlights a few other important methodological details that should be taken into consideration when calculating these metrics.

Marjanovic and Laurin (2007) recognized the high level of uncertainty associated with designating a set date for the OKRs of taxa with fossil histories, specifically citing the difficulty of assessing the position of fossils relative to geochronological boundaries and the uncertainty in the age of geochronological unit boundaries themselves. They chose to address these issues by setting the OKR of each fossil taxon equal to the base of the oldest stage from which it was recovered and assumed that the taxon persisted through the entire stage. However, as demonstrated by Pol and Norell (2006, Figs 1 and 3) fixing the OKRs of taxa with fossil histories to a single age within the range of uncertainty can greatly influence the resulting stratigraphic consistency score. When dealing with topological uncertainty, Marjanovic and Laurin (2007) resolved polytomies using the chronological method, maximizing the fit between stratigraphy and phylogeny. As illustrated by Fig. 1, failing to take into consideration the full range of phylogenetic uncertainty indicated by polytomies can bias the resulting stratigraphic consistency scores. Thus, despite their attempts to thoroughly calculate stratigraphic consistency scores, the resulting scores do not

reflect the full range of age and phylogenetic uncertainty present in the trees analysed. Additionally, both Clarke et al. (2007) and Marjanovic and Laurin (2007) treated molecular divergence estimates as fixed dates. However, because molecular divergence credibility intervals are analogous to the uncertainty in the age of OKRs of taxa with fossil histories when calculating stratigraphic consistency scores (i.e. both represent temporal uncertainty generated by different methods of estimating the timing of cladogenesis), they can, and should, be accounted for using stratigraphic consistency range metrics to define a range of possible dates for the proposed origin of a clade rather than a single date. Despite these issues, which can now be accounted for using the ComPoly approach and the ASCC program suite, the use of stratigraphic consistency metrics to assess the congruence between molecular divergence data and the stratigraphic record of first appearances is a promising new application that deserves further investigation.

Restructuring polytomous strict consensus trees produces multiple dichotomous topologies, some of which may not represent the signal present in the primary trees (Swofford, 1991), resulting in the recovery of less accurate stratigraphic consistency range scores. For example, Fig. 2 displays the two MPTs used to construct the strict consensus tree in Fig. 1a. The ranges calculated from these MPTs using the ASCC program suite (MSM* range = 0.34–0.32; GER range = 0.65–0.62) more accurately represent the signal present in the data than the ranges calculated from the strict consensus tree using the ComPoly approach and the ASCC program suite (MSMp* range = 0.46–0.32; GERp range = 0.79–0.62). Additionally, only the scores calculated from the source MPTs allow the dichotomous tree (Fig. 1k) to be recognized as more congruent with the stratigraphic record. This example clearly demonstrates that stratigraphic consistency scores should be calculated directly from the source MPTs whenever

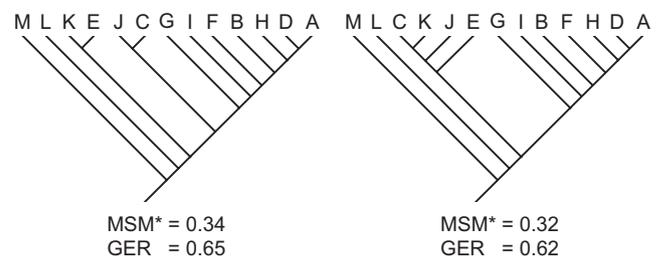


Fig. 2. The two most-parsimonious trees (MPTs) used to construct the polytomous strict consensus tree illustrated in Fig. 1a. MSM* and GER scores for each tree are provided below and to the right of each tree. MSM* and GER range scores constructed by combining the scores from both trees (MSM* range = 0.34–0.32; GER range = 0.65–0.62) are more accurate than those obtained using the ComPoly approach on the polytomous strict consensus tree from Fig. 1a (MSMp* range = 0.46–0.32; GERp range = 0.79–0.62).

possible to ensure the accuracy of the resulting scores. When more than two MPTs are recovered by a given analysis, these metrics are constructed using the ASCC program suite by computing scores for all the MPTs and then combining the highest and lowest recovered scores (see readme file for further instructions). In situations where the topologies of the source MPTs were not reported, reanalysis of the original dataset and calculation of stratigraphic consistency scores from the resulting MPTs is recommended, as the resulting increase in accuracy vastly outweighs the extra time required.

Additionally, computing stratigraphic consistency scores from a less-than-strict consensus tree (e.g. majority-rule tree) is analogous to using the reducing methods for treating polytomies because they do not describe the total signal present in the source MPTs (Nixon and Carpenter, 1996). For example, comparison of two published polytomous phylogenies, the basal ornithischian phylogenetic hypotheses of Butler (2005) and Butler et al. (2008), illustrates the pitfalls of comparing scores calculated from less-than-strict consensus trees. Comparison of the majority-rule consensus trees from these analyses (polytomies treated using the Reverse-Stratigraphic method) resulted in the phylogenetic hypothesis presented by Butler et al. (2008) being identified as more congruent with the stratigraphic record by Wills et al. (2008). However, the results of our analysis of the strict consensus trees from both studies using the ComPoly approach and the ASCC program suite (1000 replicates) does not support this conclusion. Overlapping range scores are recovered for the two phylogenies [for the phylogeny proposed by Butler (2005): MSMp* range = 0.55–0.38, GERp range = 0.87–0.64; for the phylogeny proposed by Butler et al. (2008): MSMp* range = 0.76–0.38, GERp range = 0.95–0.75), requiring that the pair-wise differences between the scores obtained during each replicate be calculated and compiled into a frequency histogram to determine if either phylogeny consistently scores higher than the other regardless of what OKR is randomly assigned to each taxon (Pol and Norell, 2006). In this case, the frequency histograms for both MSM* and GER cross zero, indicating that neither phylogeny is consistently more congruent than the other (Fig. 3). Thus, both phylogenies are equally congruent with the fossil record. As was discussed above and illustrated in Fig. 1, excluding data from the source MPTs when calculating stratigraphic consistency metrics can lead to scores that are artificially higher or lower than those calculated using the full signal present in the source MPTs. This potentially leads to erroneous conclusions regarding the stratigraphic congruence of a phylogeny or a set of alternative phylogenies. Therefore, if circumstances prevent calculation of scores directly from the source MPTs, calculating scores from the strict consensus tree using the ComPoly approach is

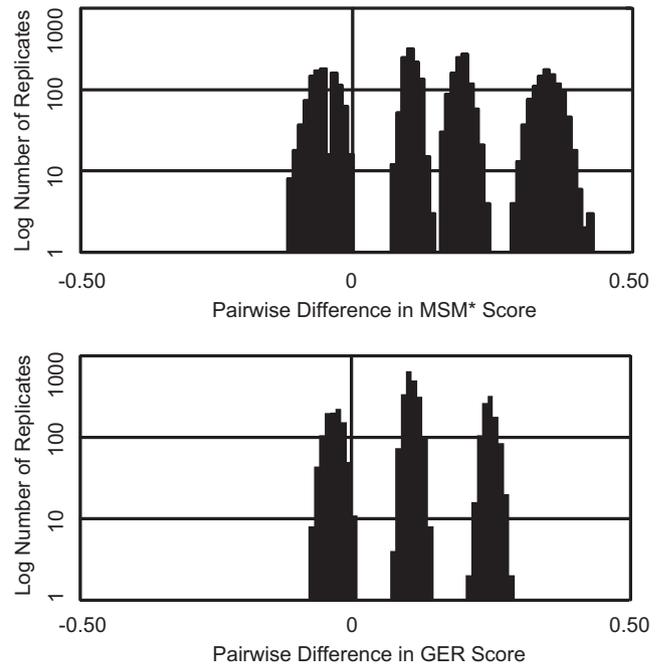


Fig. 3. Histograms of pair-wise differences for the MSMp* and GERp metrics resulting from the comparison of stratigraphic congruence between the strict consensus trees proposed by Butler (2005) and Butler et al. (2008). Because both histograms cross zero, these two phylogenetic hypotheses are considered equally congruent with the fossil record, *contra* Wills et al. (2008).

the preferred solution because this method will always provide a conservative estimate of a phylogeny's stratigraphic congruence and never falsely identify a phylogenetic hypothesis as the most stratigraphically congruent based on spurious resolutions.

Conclusion

Previously, no standardized method existed for calculating stratigraphic consistency metrics for polytomous phylogenies. Researchers chose from a set of imperfect methods for treating polytomies, none of which took into consideration the full range of variation that polytomies impart on stratigraphic consistency scores. We demonstrate that the method by which polytomies are resolved impacts the resulting stratigraphic consistency scores calculated for a given phylogeny. Therefore, it is imperative that the full effects of phylogenetic uncertainty be taken into consideration to ensure accuracy when comparing alternative phylogenies and to prevent erroneous conclusions. The ComPoly approach defines the full range of possible variation polytomies impart on stratigraphic consistency measures. This allows all alternative phylogenies for a given set of taxa to be accurately compared with the stratigraphic record and prevents selection of a subop-

timal phylogenetic hypothesis from a set of alternative hypotheses.

The new ASCC program suite (freely available at <http://www.stratfit.org>) simplifies the calculation of three stratigraphic consistency metrics (GER, MIG, MSM*) and their respective range scores (GER range, MIG range, MSM* range). These metrics can be used as descriptive statistics, or can be used to compare the stratigraphic fit of alternative phylogenetic hypotheses, estimate the completeness of the fossil record for a clade of interest, or examine the stratigraphic congruence of alternative placements of taxa within a phylogeny. Recently implemented methods that facilitate evaluation of the incongruence between the stratigraphic record of first appearances and molecular divergence estimates for a given tree topology highlight a promising new application for these metrics that warrants further investigation (Clarke et al., 2007; Marjanovic and Laurin, 2007). When performing these calculations, the uncertainty in the age of the OKRs of taxa with fossil histories and the credibility intervals for molecular divergence dates should be treated analogously using stratigraphic consistency range metrics that allow these dates to be specified as ranges instead of set points because both represent uncertainty in estimating the timing of cladogenesis. Furthermore, stratigraphic consistency scores should be calculated from the original set of MPTs to ensure the accuracy of the resulting scores. When this is not possible, scores should be calculated from the resulting strict consensus tree so that the total signal present in the source MPTs is represented.

Acknowledgements

We thank A.D.B. Behlke, C. Bell, A. DeBee, D. Eddy, J. Green, E. Johnson, D. Ksepka, D. Pol, T. Rowe, and N.A. Smith for reviewing early drafts of this manuscript; and E.B. Cleland and S.J. Nesbitt for help with troubleshooting the ASCC program suite. We thank D. Pol, an anonymous reviewer, and K. Pickett (editor) for thoughtful comments and suggestions during the review process that improved the clarity of this manuscript.

References

- Angielczyk, K.D., Kurkin, A.A., 2003. Phylogenetic analysis of Russian Permian dicynodonts (Therapsida: Anomodontia): implications for Permian biostratigraphy and Pangaeon biogeography. *Zool. J. Linn. Soc.* 139, 157–212.
- Benton, M.J., Storrs, G.W., 1994. Testing the quality of the fossil record: paleontological knowledge is improving. *Geology* 22, 111–114.
- Brochu, C.A., Norell, M.A., 2000. Temporal congruence and the origin of birds. *J. Vertebr. Paleontol.* 20, 197–200.
- Brusatte, S.L., Sereno, P.C., 2008. Phylogeny of Allosauroidea (Dinosauria: Theropoda): comparative analysis and resolution. *J. Syst. Paleontol.* 6, 155–182.
- Butler, R.J., 2005. The ‘fabrosaurid’ ornithischian dinosaurs of the Upper Elliot Formation (Lower Jurassic) of South Africa and Lesotho. *Zool. J. Linn. Soc.* 145, 175–218.
- Butler, R.J., Upchurch, P., Norman, D.B., 2008. The phylogeny of the ornithischian dinosaurs. *J. Syst. Paleontol.* 6, 1–40.
- Clarke, J.A., Ksepka, D.T., Stucchi, M., Urbina, M., Giannini, N., Bertelli, S., Narvaez, Y., Boyd, C.A., 2007. Paleogene equatorial penguins challenge the proposed relationship between biogeography, diversity, and Cenozoic climate change. *Proc. Natl. Acad. Sci. U S A* 104, 11545–11550.
- Gauthier, J., Kluge, A.G., Rowe, T., 1988. Amniote phylogeny and the importance of fossils. *Cladistics* 4, 105–209.
- Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24, 774–786.
- Hitchin, R., Benton, M.J., 1997a. Congruence between parsimony and stratigraphy: comparisons of three indices. *Paleobiology* 23, 20–32.
- Hitchin, R., Benton, M.J., 1997b. Stratigraphic indices and tree balance. *Syst. Biol.* 46, 563–569.
- Huelsensbeck, J.P., 1994. Comparing the stratigraphic record to estimates of phylogeny. *Paleobiology* 20, 470–483.
- Jeffery, C.H., Emler, R.B., 2003. Macroevolutionary consequences of developmental mode in temnopleurid echinoids from the Tertiary of southern Australia. *Evolution* 57, 1031–1048.
- Kerr, A.M., Kim, J., 2001. Phylogeny of Holothuroidea (Echinodermata) inferred from morphology. *Zool. J. Linn. Soc.* 133, 63–81.
- Marivaux, L., Vianey-Liaud, M., Jaeger, J.-J., 2004. High-level phylogeny of early Tertiary rodents: dental evidence. *Zool. J. Linn. Soc.* 142, 105–134.
- Marjanovic, D., Laurin, M., 2007. Fossils, molecules, divergence times, and the origin of lissamphibians. *Syst. Biol.* 56, 369–388.
- Merle, D., Pacaud, J.-M., 2003. New species of *Eocithara* Fisher, 1883 (Mollusca, Gastropoda, Harpidae) from the Early Paleogene with phylogenetic analysis of the Harpidae. *Geodiversitas* 26, 61–87.
- Nixon, K.C., Carpenter, J.M., 1996. On consensus, collapsibility, and clade concordance. *Cladistics* 12, 305–322.
- Norell, M.A., 1993. Tree-based approaches to understanding history: comments on ranks, rules, and the quality of the fossil record. *Am. J. Sci.* 293, 407–417.
- Norell, M.A., Novacek, M.J., 1992. Congruence between superpositional and phylogenetic patterns: comparing cladistic patterns with fossil records. *Cladistics* 8, 319–337.
- O’Leary, M.A., 2001. The phylogenetic position of cetaceans: further combined data analyses, comparisons with the stratigraphic record and a discussion of character optimization. *Am. Zool.* 41, 487–506.
- Pol, D., Norell, M.A., 2001. Comments on the Manhattan stratigraphic measure. *Cladistics* 17, 285–289.
- Pol, D., Norell, M.A., 2006. Uncertainty in the age of fossils and the stratigraphic fit to phylogenies. *Syst. Biol.* 55, 512–521.
- Pol, D., Norell, M.A., Siddall, M.E., 2004. Measures of stratigraphic fit to phylogeny and their sensitivity to tree size, shape and scale. *Cladistics* 20, 64–75.
- Pryer, K.M., 1999. Phylogeny of marsileaceous ferns and relationships of the fossil *Hydropteris pinnata* reconsidered. *Int. J. Plant Sci.* 160, 931–954.
- Rauhut, O.W.M., 2003. The interrelationships and evolution of basal theropod dinosaurs. *Spec. Pap. Palaeontol.* 69, 1–213.
- Saucede, T., Mooi, R., David, B., 2007. Phylogeny and origin of Jurassic irregular echinoids (Echinodermata: Echinoidea). *Geol. Mag.* 144, 333–359.
- Siddall, M.E., 1996. Stratigraphic consistency and the shape of things. *Syst. Biol.* 45, 111–115.
- Siddall, M.E., 1997. Stratigraphic indices in the balance: a reply to Hitchin and Benton. *Syst. Biol.* 46, 569–573.

- Siddall, M.E., 1998. Stratigraphic fit to phylogenies: a proposed solution. *Cladistics* 14, 201–208.
- Swofford, D.L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? In: Miyamoto, M.M., Cracraft, J. (Eds.), *Phylogenetic Analysis of DNA Sequences*. Oxford University Press, New York, pp. 295–333.
- Tetlie, O.E., Poschmann, M., 2008. Phylogeny and palaeoecology of the Adelophthalmoidea (Arthropoda; Chelicerata; Eurypterida). *J. Syst. Paleontol.* 6, 237–249.
- Villier, L., Neraudeau, D., Clavel, B., Neumann, C., David, B., 2004. Phylogeny of Early Cretaceous spatangoids (Echinodermata: Echinoidea) and taxonomic implications. *Palaeontology* 47, 265–292.
- Wagner, P.J., Sidor, C.A., 2000. Age rank/clade rank metrics—sampling, taxonomy, and the meaning of “stratigraphic consistency”. *Syst. Biol.* 49, 463–479.
- Walsh, S.L., 1998. Fossil datum and paleobiological event terms, paleostratigraphy, chronostratigraphy, and the definition of land mammal “age” boundaries. *J. Vertebr. Paleontol.* 18, 150–179.
- Wills, M.A., 1999. Congruence between stratigraphy and phylogeny: randomization tests and the gap excess ratio. *Syst. Biol.* 48, 559–580.
- Wills, M.A., 2001. How good is the fossil record of arthropods? An assessment using the stratigraphic congruence of cladograms. *Geol. J.* 36, 187–210.
- Wills, M.A., Barrett, P.M., Heathcote, J.F., 2008. The modified gap excess ratio (GER*) and the stratigraphic congruence of dinosaur phylogenies. *Syst. Biol.* 57, 891–904.
- Wilson, J.A., 2002. Sauropod dinosaur phylogeny: critique and cladistic analysis. *Zool. J. Linn. Soc.* 136, 217–276.