Accepted: 23 May 2017

DOI: 10.1111/azo.12202

ORIGINAL ARTICLE

Overlap Indices: Tools to quantify the amount of anatomical overlap among groups of incomplete terminal taxa in phylogenetic analyses

Emanuel Tschopp[1,2,3] | **Friedrich A. Tschopp**[4] | **Octávio Mateus**[2,3]

[1]Dipartimento di Scienze della Terra, Università degli Studi di Torino, Torino, Italy

[2]GeoBioTec, Departamento de Ciências da Terra, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

[3]Museu da Lourinhã, Lourinhã, Portugal

[4]Jona, Switzerland

Correspondence
Emanuel Tschopp, Dipartimento di Scienze della Terra, Università degli Studi di Torino, Torino, Italy.
Email: tschopp.e@gmail.com

Funding information
Fundação para a Ciência e a Tecnologia of the Ministério de Educação e Ciência, Grant/Award Number: SFRH/BD/66209/2009; European Union's Seventh Framework Programme; Marie Skłodowska-Curie, Grant/Award Number: 609402 – 2020

Abstract

Phylogenetic analyses of morphological data are often characterized by missing data due to incomplete operational taxonomic units, as in fossils. This incomplete knowledge derives from various reasons, including—in the case of fossils—the numerous filters an organism has to pass through during taphonomy, fossilization, weathering and collecting. Whereas several methods have been proposed to address issues raised by the inclusion of incomplete terminal taxa, until recently no tool existed to easily quantify the amount of anatomical overlap within a particular clade. The Overlap Indices provide such values and might prove useful for comparative cladistics. We herein describe these new indices and their applications in detail and provide an example file for their calculation. A case study of diplodocid sauropod dinosaurs shows how the Overlap Indices will help to explore and quantify, which one of a number of conflicting tree topologies is supported by more anatomical traits, which skeletal regions are underrepresented in a particular phylogenetic matrix, and which taxon would improve character state score completeness.

KEYWORDS

comparative cladistics, missing data, phylogeny, quantification tool

1 | INTRODUCTION

With the widespread use of phylogenetic analysis, comparative cladistics becomes an increasingly important issue (Peyre de Fabrègues, Allain, & Barriel, 2015; Sereno, 2009; Whitlock & Wilson, 2013). In particular in well-studied taxa, where numerous research groups develop distinct phylogenetic matrices to support their cladograms, conflicting tree topologies are sometimes difficult to resolve. More tools for visualization and quantification of the underlying data and its direct support for the tree topology are therefore needed (Sereno, 2009).

Missing data are a major issue in phenotypic and morphological datasets, in particular in studies of fossil organisms. Even specimens of extant species cannot always be scored completely in morphological, phylogenetic matrices: for instance in vertebrates, study specimens in museums are often either preserved in alcohol or as skeletons, thus inhibiting the ability to score the skeletal or soft tissue characters, respectively. Additionally, dead specimens do not provide information concerning non-morphological phenotypic traits, such as metabolism and behaviour. When preserved as skeletons, lizards in particular are often only represented by skulls in museum collections, due to an alleged uselessness of the postcranial skeleton in morphological studies (Bell & Mead, 2014; ET, unpublished data). Additionally, these skulls are mostly articulated, in some cases, soft tissue still partially covers the bones, and finally bones can fuse during ontogeny, thereby obliterating even more morphological information. CT-scanning can help to discover sutures or visualize the skeleton in specimens preserved with soft tissue (Bell & Mead, 2014), but is not always available and relatively expensive (Smith & Strait, 2008).

Fossil organisms are even more susceptible to loss of morphological information, because they pass through a number of filters after death: taphonomy, preservation, weathering, erosion and collecting (Brusatte, 2012), and even subsequent destruction due to poor preparation, poor curation and accidents. In phylogenetic analyses based on morphological data, missing entries can severely hamper the recovery of a well-resolved cladogram, especially when working at specimen-level (see Tschopp, Mateus, & Benson, 2015). The problem can be circumvented to some extent using more inclusive operational taxonomic units. By doing so, information on species or genus morphology and thus scores for the phylogenetic matrix can be taken from several individual specimens that preserve varying parts of the organism. Thereby, a combined, complete score of the species (or higher-level taxonomic groups) can be used in the phylogenetic matrix. However, this approach is always based on the assumption that earlier referrals of the studied specimens to a particular species are true and might result in a large number of polymorphic characters (Brusatte, 2010). Therefore, in certain cases, specimen-level cladistic analysis has to be preferred over species-level studies (Tschopp et al., 2015).

Several methodological approaches were proposed in the past to minimize the negative effects of missing data on tree topology (e.g., Grillo & Azevedo, 2011; Kearney & Clark, 2003; Norell & Wheeler, 2003; Wilkinson, 1995) and to identify the characters that are more susceptible to missing entries (Pol & Escapa, 2009). However, until recently, no method has been proposed to visualize or quantify the underlying problem of missing data. Missing data hamper direct comparison of OTUs because of lacking anatomical overlap (Tschopp et al., 2015). In cases of extensive missing data, specific OTUs might not have any morphological character for which both could be scored. These fragmentary OTUs can then only be indirectly compared via the inclusion of more complete OTUs that preserve anatomical regions present in the fragmentary ones (Tschopp et al., 2015). Tschopp et al. (2015) therefore introduced a so-called Overlap Index, quantifying the amount of anatomical overlap within a particular clade, and thus indicating how much of the anatomy could actually be compared directly. In the dawn of comparative cladistics (Sereno, 2009), where comparison of basic data underlying phylogenetic hypotheses becomes crucial (Peyre de Fabrègues et al., 2015; Whitlock & Wilson, 2013), indices like the proposed Overlap Index are increasingly important. Herein, we explain in detail how to calculate the two proposed versions of the Overlap Index and explore their potential uses in comparative cladistics.

2 | METHODOLOGY

Two different approaches for calculating an Overlap Index were proposed by Tschopp et al. (2015), the "All Characters Overlap Index" (AOI), and the "Comparable Characters Overlap Index" (COI). The AOI quantifies the number of characters available for analysis within a specific clade, whereas the COI quantifies the number of anatomical overlaps occurring among the characters for which at least two OTUs of the clade under study could be scored (Tschopp et al., 2015). Obviously, the minimum number of OTUs must be two, because a character scored only in a single specimen cannot contribute to any anatomical Overlap Index.

2.1 | How to use the template file

Both indices were calculated in Microsoft Excel®. An updated and simplified template file from the study of Tschopp et al. (2015) is provided as Supporting Information and is suited for matrices comprising up to 200 OTUs and 500 characters. The provided template works with macros, which have to be activated, and the file saved with a different name as a working copy. The basic information for the calculation of the indices are the phylogenetic matrix and the definition of the clades or other groups of OTUs to be compared. The phylogenetic matrix can be pasted into the sheet called "Input Matrix" in the spreadsheet (template provided in the Supporting Information). The file then automatically deletes all entries that are not numbers, and all numbers for character states are substituted by "1." This standardized matrix is saved in a separate, usually hidden, sheet called "MatrixStandardized," which serves for the calculation of the indices and should therefore not be changed. Subsequently, the groups or clades to be compared have to be defined in the sheet "Group Definition." These groups can but do not have to represent the clades recovered in the various trees. In the example of Tschopp et al. (2015), the groups represented the clades found in the strict consensus trees, pruned consensus trees and reduced consensus trees of the analyses performed under equal and implied weights. The group members can be chosen from a drop down menu, or by typing the first few letters of the OTU to be included. The indices are calculated in real time, allowing to assess directly the influence of a specific OTU on the index to be calculated. It is furthermore possible to add group names, which is particularly useful when calculating indices for a number of different groups. With the button "Freeze Group," the indices are automatically saved to the sheet "Results," and the entire spreadsheet will be locked in order to avoid erroneous changes to the matrix. Additional groups can still be defined and "frozen" into the results sheet, but for additional studies with a new matrix, one will have to use the pristine template file again. More detailed results concerning the basic calculations behind the Overlap Indices can be tracked in the sheet "Details." Two hidden sheets ("Group Lists" and "MatrixStandardized") only serve for preparing the calculations and have therefore been hidden to avoid unplanned changes. The entire process is summarized in the flow chart shown in Figure 1.

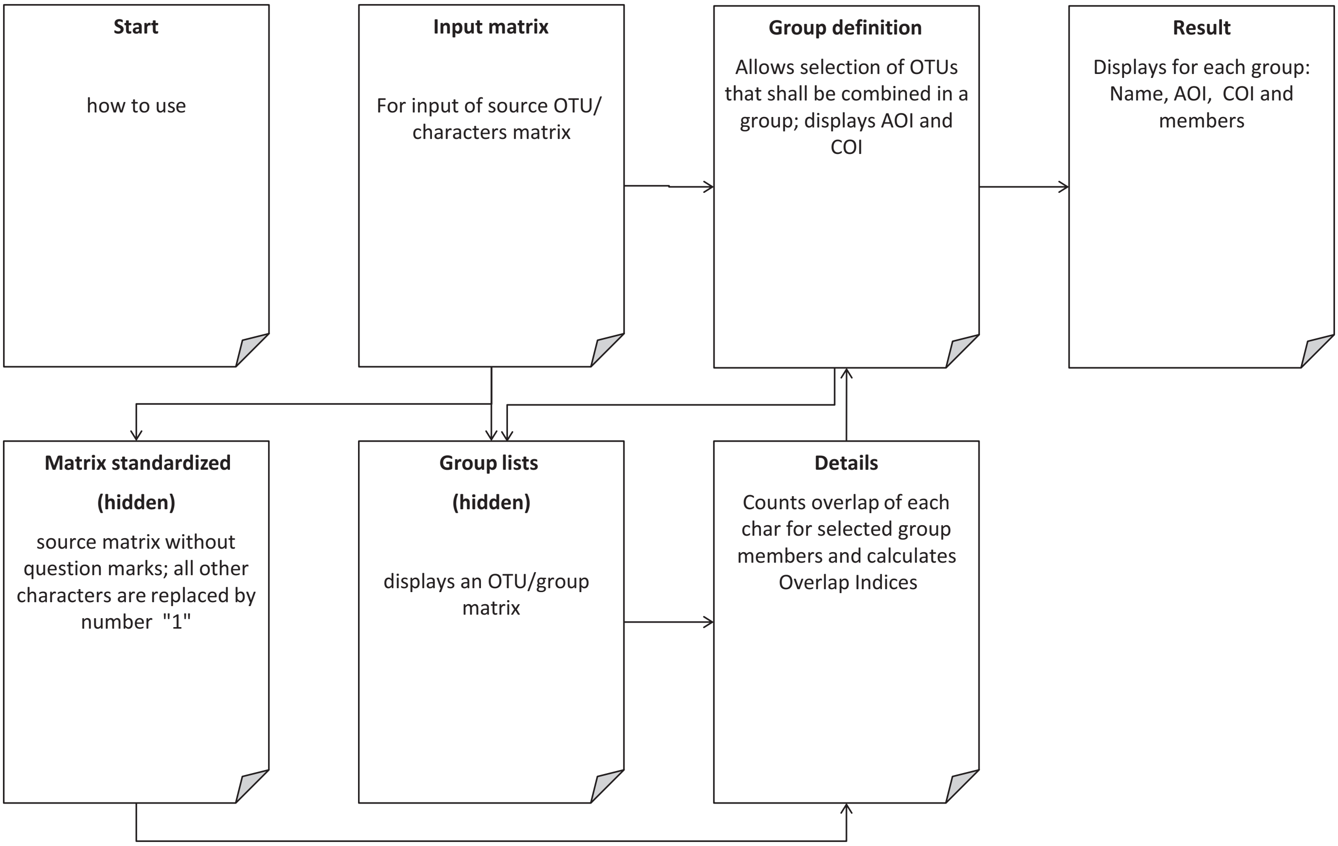

FIGURE 1 Flow chart showing the process of the calculation of the indexes in the template file, and how to use the different sheets

2.2 | Mathematical model

Basic information for both indices are the number of group members (M), the total number of characters (C), and which characters are scored in which group member. From there, the indices calculate the number of possible anatomical overlaps per character ($M-1$), the number of actual overlaps in every single character (Oc), and the number of characters with overlap, that is the ones for which at least two group members could be scored (Co). The number of actual overlaps per character is summed to the total number of overlaps present within a particular group (Om).

The All Characters Overlap Index then calculates the mean amount of overlaps for all characters (Om/C). Finally, it divides the mean number of overlaps per character by the maximum number of possible overlaps per character. The equation for the AOI is therefore the following:

$$\mathrm{AOI} = (\mathrm{Om}/C)/(M-1),$$

where Om is the total number of actual overlaps, C is the total number of characters in the matrix, and M is the number of group members.

The Comparable Character Overlap Index calculates the mean amount of overlaps for the comparable characters only (i.e., the characters with overlap; Om/Co). Finally, it divides the mean number of overlaps per comparable character by the maximum number of possible overlaps per character. The equation for the COI is therefore the following:

$$\mathrm{COI} = (\mathrm{Om}/\mathrm{Co})/(M-1),$$

where Om is the total number of actual overlaps, Co is the number of characters in the matrix with anatomical overlap among group members, and M is the number of group members.

3 | APPLICATIONS

3.1 | All Characters Overlap Index (AOI)

By analysing the number of actually occurring anatomical overlaps within a complete set of anatomical characters in a particular taxonomic group, the AOI quantifies how much of the group's anatomy could be directly compared among the OTUs belonging to the specified group. Tschopp et al. (2015) mention one use of this approach: by comparing the AOI of several clades that could potentially include an unstable OTU, the researcher gets an idea of which phylogenetic position of the questionable OTU is backed up by most direct comparisons of the anatomy. However, because the AOI does not take into account differential character state scoring, which can depend on differential interpretation of morphology (Sereno, 2007, 2009), it cannot be a measure for the support for a particular systematic position. Several indices already exist for the comparison of character state scoring (Sereno, 2009). The AOI has to be seen as a quantification tool for the amount of anatomical overlap, not as a support index for a phylogenetic position.

Recent phylogenetic data sets of different researchers often use varying sets of characters, even when analysing the same or a very similar taxonomic group (Sereno, 2009). Character selection or deletion is rarely explained, and no standard procedures exist for the evaluation of the utility of a character (Jenner, 2004; Poe & Wiens, 2000; Sereno, 2007, 2009). Therefore, comparison of the resulting phylogenetic trees is often difficult, and the recognition of character conflicts that underlie varying systematic positions of certain taxa among different cladograms is laborious and time-consuming (Sereno, 2009). By quantifying the anatomical overlap of specific taxonomic groups, the AOI helps to understand, which of the various positions in the conflicting analyses is backed up by most, directly comparable anatomical data. As such, it serves as a first step for more detailed analyses of character state similarities (Sereno, 2009), as well as for an assessment of the impact of missing data on tree topology.

3.2 | Comparable Characters Overlap Index (COI)

The COI only analyses the characters, where a group shows anatomical overlap, and does not take into account any character for which only one or no OTU could be scored. Because of this, the COI is always equal or higher than the AOI and reaches 100% per definition in groups composed of only two OTUs (Tschopp et al., 2015). The restriction to comparable characters only can be useful to analyse differential taxon sampling among conflicting cladograms. The COI shows which matrix has more complete character scores among the characters comparable within a specific group. By comparing the composition of the group among the conflicting cladograms, the COI can help to identify crucial OTUs for a certain set of characters. Moreover, it indicates where additional, more complete OTUs are needed to decrease the amount of missing data.

In combination with Character Distribution Maps (CDMs; Whitlock & Wilson, 2013), both indices help to identify skeletal regions that would be important for the resolution of the clades under study, but that are underrepresented in the character selection of a particular study. AOI, COI and CDMs are thus promising tools for more transparent and methodic character selection, or to indicate where further or more detailed anatomical studies are needed to resolve conflicting tree topologies.

3.3 | Example

Tschopp et al. (2015) performed two phylogenetic analyses of diplodocid sauropod dinosaurs, which differ in the type of weighting. As reported by them, the two weighting approaches also produced conflicting tree topologies in some cases. One of these conflicting topologies is of particular interest, because it concerns the question of the validity of the popular genus *Brontosaurus*. The validity of this genus depends in part on the phylogenetic position of the holotype of the type species of *Apatosaurus*: *A. ajax* YPM 1860. Under equal weighting, YPM 1860 is found as the sister taxon to a clade composed of four specimens and including the holotype of a second species of *Apatosaurus*, *A. louisae* (Tschopp et al., 2015: fig. 114). Under implied weighting, YPM 1860 was recovered more closely related to the holotype of the type species of *Brontosaurus* (*B. excelsus* YPM 1980) than to *A. louisae*. In addition to the analyses by Tschopp et al. (2015), also Upchurch, Tomida, and Barrett (2004) analysed apatosaurine relationships using specimen-level phylogeny and found YPM 1860 in yet a different position within the tree (Upchurch et al., 2004: fig. 15). The conflicting positions of YPM 1860 in these three analyses are summarized in Figure 2.

By comparing the AOI of the three recovered groupings using the matrix of Tschopp et al. (2015), which has a much larger character sampling compared to Upchurch et al. (2004), we understand which of the three topologies is based on more directly comparable morphological data. Given that

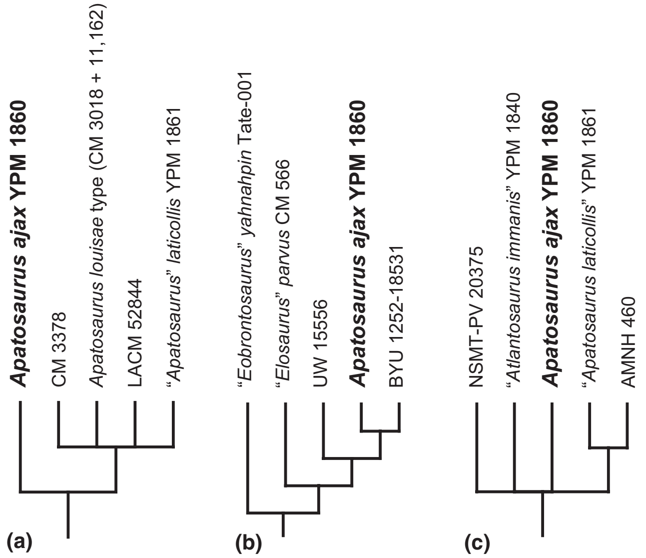

FIGURE 2 Extracts of the trees found by Tschopp et al. (2015) under equal weighting (a) and implied weighting (b), and by Upchurch et al. (2004) (c), showing the different positions in which the holotype of the type species of *Apatosaurus* (*A. ajax* YPM 1860, in bold) was recovered. Note that no OTU other than YPM 1860 is included in all 5-taxon statements compared herein using the AOI

TABLE 1 Results of the case study, as reported in the "Details" sheet on the template file for the calculation of the Overlap Indices

	Group Tested	Number of Group members	Number of characters	Number of possible anatomical overlaps per character	Number of characters with overlap	Total number of overlaps present within group	Mean amount of overlaps for all characters	Standard deviation	Mean amount of overlaps for comparable characters	All Chars Overlap Index	Comparable Chars Overlap Index
Formula		(M)	(C)	(M−1)	(Co)	(Om)	(Om/C)		(Om/Co)	(AOI)	(COI)
AOI example	T15_ew	5	477	4	266	393	0.82	0.73	1.48	**21%**	37%
	T15_iw	5	477	4	262	617	1.29	1.25	2.35	**32%**	59%
	U04 in T15	5	477	4	226	447	0.94	0.99	1.98	**23%**	49%
COI example	COI_U04 in T15	11	477	10	369	1,585	3.32	2.47	4.30	33%	**43%**
	COI_T15	16	477	15	381	2,193	4.60	3.32	5.76	31%	**38%**
	COI_U04	11	32	10	32	164	5.13	1.08	5.13	51%	**51%**

every single OTU influences the Overlap Index of the entire group, in particular if its completeness differs considerably from the average, the AOI is most informative when comparing clades with an equal number of OTUs. Thus, because the clade found under equal weighting includes five OTUs (Tschopp et al., 2015), it makes most sense to compare the AOI for groups of five OTUs including YPM 1860 (Figure 2).

The AOI of the three groups indicates that a sister-taxon relationship of *A. ajax* and the clade including the holotype of *A. louisae* as found by Tschopp et al. (2015) under equal weighting is backed up by the least amount of directly comparable data (21%), followed by a group as recovered by Upchurch et al. (2004) (23%), whereas the clade found under implied weighting ranks highest (32%; Table 1). Even though these values do not provide any direct support for specific groupings (see above), comparing results from different weighting strategies highlights a peculiar issue with implied weighting, correlated with a phenomenon called long-branch attraction. Long-branch attraction occurs when two branches with several convergently acquired features are wrongly grouped together because of lacking intermediate forms, and thus lacking information on the independent accumulation of their similarities (Bergsten, 2005). If large amounts of missing data in a particular OTU concern information on these crucial intermediate steps, the convergently acquired, derived comparable traits might overrule the real phylogenetic signal (Wiens, 1998, 2006). Moreover, given that missing data always also decrease the amount of homoplasies in a specific character and that implied weighting adapts the weight of a character based on its homoplastic rate (Goloboff, 1993), these effects of long-branch attraction due to missing data can be amplified by the use of implied weighting. In the present case, the grouping found under implied weighting has a much higher AOI than the ones under equal weighting, and also the mean amount of overlaps per comparable characters is considerably higher than in the other groupings (Table 1). A significant negative impact of missing data and resulting long-branch attraction under implied weighting can thus probably be excluded in this particular analysis.

The COI is useful to compare data from two different matrices and to explore the effect of the addition of particular OTUs on the anatomical support for the recovered topologies. Therefore, we compare the matrices of Upchurch et al. (2004) and Tschopp et al. (2015) concerning the conflicting topologies within Apatosaurinae, and the effects of a more extensive OTU sampling in Tschopp et al. (2015) compared to Upchurch et al. (2004) (Figure 3). To do so, we calculate the COI for all apatosaur specimens included in Upchurch et al. (2004) using both matrices, and the COI of the entire apatosaurine specimen sampling of the analysis using implied weighting by Tschopp et al. (2015).

The highest score of the COI in this case (51%) was found in the matrix of Upchurch et al. (2004), where all 32

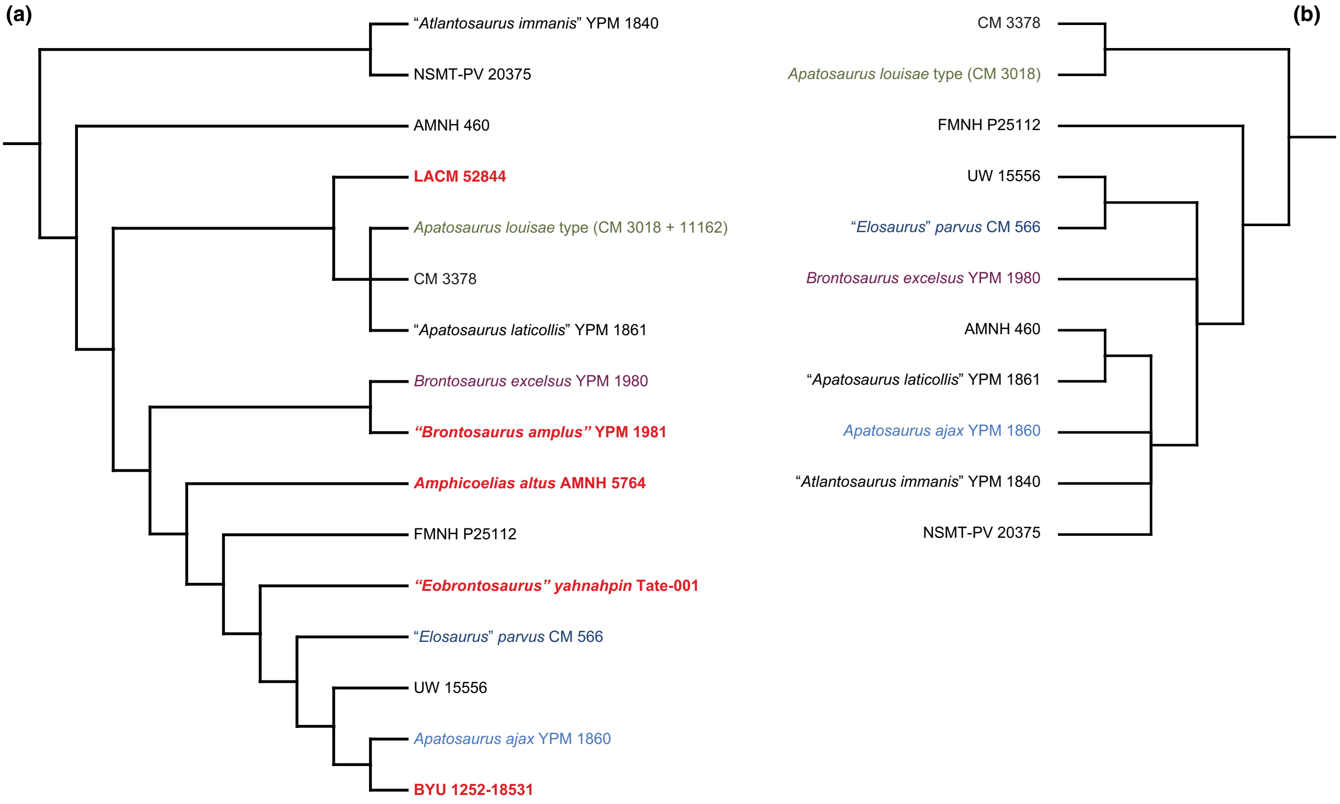

FIGURE 3 Tree topologies as found by (a) Tschopp et al. (2015) under implied weighting and (b) Upchurch et al. (2004). The OTUs highlighted in red/bold were only included in Tschopp et al. (2015). Additional OTUs are highlighted in different colours for comparative purposes (the type specimens of the species *Apatosaurus ajax, A. louisae, Brontosaurus excelsus and B. parvus* (sensu Tschopp et al., 2015). Note that Upchurch et al. (2004) only included the postcranial skeleton of the holotype of *A. louisae* (CM 3018), because their matrix only comprised characters coding for the postcranium. [Colour figure can be viewed at wileyonlinelibrary.com]

characters do show anatomical overlap among the included apatosaurine specimens, followed by the one of Upchurch et al.'s (2004) taxon sampling in the matrix of Tschopp et al. (2015) (43%), and finally the complete sampling of Tschopp et al. (2015) (38%; Table 1). The lower values when using Tschopp et al. (2015) are most probably due to the different taxonomic scope of the two analyses, which is exemplified by a much more restricted taxon and character sampling in the analysis of Upchurch et al. (2004). Being entirely concentrated on resolving relationships within Apatosaurinae, Upchurch et al. (2004) restricted their character sampling such that all the included characters showed anatomical overlap in the specimens they analysed. This also implied that the matrix had to exclude characters from the skull, of which many could only be scored in the type specimen of *Apatosaurus louisae* by Tschopp et al. (2015). The direct comparison of the COI using the same OTU sampling in the two matrices therefore helped to identify the necessity to sample more skull material of apatosaurines in order to significantly increase the amount of comparable characters in Tschopp et al. (2015). Comparing the values of the restricted OTU sampling of Upchurch et al. (2004) and the more complete one of Tschopp et al. (2015) using the matrix of the latter analysis showed that although the mean amount of overlaps per comparable character increased with a larger OTU sampling (5.76 vs 4.30), the COI decreased (Table 1). Subsequent deletion of any of the five added specimens shows that *Amphicoelias altus* AMNH 5764, *Brontosaurus amplus* YPM 1981 and LACM 52844 are responsible for the decreasing COI, whereas the single additions of "*Eobrontosaurus*" *yahnahpin* Tate-001 had no influence on the COI, and the one of BYU 1252-18531 increased the COI from 43% to 44%. Given that Tschopp et al. (2015) scored *B. amplus* YPM 1981 and "*E. yahnahpin*" Tate-001 based on very limited first-hand observations, and none at all, respectively, a rescoring of these two specimens, or the addition of similarly complete individuals as BYU 1252-18531 would be necessary to significantly increase anatomical overlap in Apatosaurinae.

In the case of apatosaurines, the two newly proposed Overlap Indices thus indicated possibly misleading signals due to missing data and helped to understand which specimens should be studied in more detail, in order to increase anatomical overlap. Also, they showed that the addition of new specimens with preserved skull material would be highly beneficial for a better understanding of apatosaurine interrelationships.

4 | CONCLUSION

Indices for meta-analyses across several phylogenetic analyses, comparing the actual morphological data, on which conflicting cladograms are based, are becoming increasingly important. Given the more and more extensive data sets, computed visualization and quantification methods are the only handy tools for a quick assessment of the data underlying differing phylogenetic hypotheses. The recently proposed Overlap Indices provide such a quantification tool for the amount of anatomical overlap in particular clades. Their computational background is herein explained in detail, and a template file is provided that can be easily adapted for any other study.

Two indices were proposed, the All Characters Overlap Index (AOI) and the Comparable Characters Overlap Index (COI). Whereas the AOI includes a measure for completeness of the OTUs, COI does not. As such, AOI is useful for the assessment of the influence of missing data on tree topology, and COI can serve to identify OTUs providing additional anatomical information in otherwise poorly sampled morphological characters. Combined with character distribution maps, both indices help to find underrepresented skeletal regions in particular data sets. They serve as exploratory tools and as precursors of more elaborate and detailed analyses of comparative cladistics concerning for instance character state scoring similarity.

ACKNOWLEDGEMENTS

The Overlap Indices were developed during the PhD of ET, which was funded through the doctoral fellowship from the Fundação para a Ciência e a Tecnologia of the Ministério de Educação e Ciência, Portugal (SFRH/BD/66209/2009). The more detailed assessment of their applications and utility was possible thanks to funding provided to ET from the European Union's Seventh Framework Programme for research and innovation under the Marie Skłodowska-Curie grant agreement No 609402 – 2020 researchers: Train to Move (T2M). We greatly acknowledge the inputs and comment of Claire Peyre de Fabrègues on an earlier version of this manuscript.

REFERENCES

Bell, C. J., & Mead, J. I. (2014). Not enough skeletons in the closet: Collections-based anatomical research in an age of conservation conscience. *The Anatomical Record*, *297*(3), 344–348. https://doi.org/10.1002/ar.22852

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, *21*(2), 163–193. https://doi.org/10.1111/j.1096-0031.2005.00059.x

Brusatte, S. L. (2010). Representing supraspecific taxa in higher-level phylogenetic analyses: Guidelines for palaeontologists. *Palaeontology*, *53*(1), 1–9. https://doi.org/10.1111/j.1475-4983.2009.00918.x

Brusatte, S. L. (2012). *Dinosaur paleobiology* (1st ed.). Hoboken, NJ: Wiley-Blackwell.

Goloboff, P. A. (1993). Estimating character weights during tree search. *Cladistics*, *9*(1), 83–91. https://doi.org/10.1111/j.1096-0031.1993.tb00209.x

Grillo, O. N., & Azevedo, S. A. K. (2011). Recovering missing data: Estimating position and size of caudal vertebrae in *Staurikosaurus pricei* Colbert, 1970. *Anais Da Academia Brasileira de Ciências*, *83*(1), 61–72. https://doi.org/10.1590/s0001-37652011005000003

Jenner, R. A. (2004). The scientific status of metazoan cladistics: Why current research practice must change. *Zoologica Scripta*, *33*(4), 293–310. https://doi.org/10.1111/j.0300-3256.2004.00153.x

Kearney, M., & Clark, J. M. (2003). Problems due to missing data in phylogenetic analyses including fossils: A critical review. *Journal of Vertebrate Paleontology*, *23*(2), 263–274. https://doi.org/10.1671/0272-4634(2003)023[0263:pdtmdi]2.0.co;2

Norell, M. A., & Wheeler, W. C. (2003). Missing entry replacement data analysis: A replacement approach to dealing with missing data in paleontological and total evidence data sets. *Journal of Vertebrate Paleontology*, *23*(2), 275–283. https://doi.org/10.1671/0272-4634(2003)023[0275:merdaa]2.0.co;2

Peyre de Fabrègues, C., Allain, R., & Barriel, V. (2015). Root causes of phylogenetic incongruence observed within basal sauropodomorph interrelationships. *Zoological Journal of the Linnean Society*, *175*(3), 569–586. https://doi.org/10.1111/zoj.12290

Poe, S., & Wiens, J. J. (2000). Character selection and the methodology of morphological phylogenetics. In J. J. Wiens (Ed.), *Phylogenetic analysis of morphological data* (pp. 20–36). Washington, DC: Smithsonian Institution Press. Retrieved from http://www.faculty.biol.ttu.edu/strauss/Phylogenetics/Readings/PoeWiens2000.pdf

Pol, D., & Escapa, I. H. (2009). Unstable taxa in cladistic analysis: Identification and the assessment of relevant characters. *Cladistics*, *25*(5), 515–527. https://doi.org/10.1111/j.1096-0031.2009.00258.x

Sereno, P. C. (2007). Logical basis for morphological characters in phylogenetics. *Cladistics*, *23*(6), 565–587. https://doi.org/10.1111/j.1096-0031.2007.00161.x

Sereno, P. C. (2009). Comparative cladistics. *Cladistics*, *25*(6), 624–659. https://doi.org/10.1111/j.1096-0031.2009.00265.x

Smith, N. E., & Strait, S. G. (2008). PaleoView3D: From specimen to online digital model. *Palaeontologia Electronica*, *11*(2), 1–17.

Tschopp, E., (Unpublished data). Lacertid lizard specimen completeness in various museum collections in Europe.

Tschopp, E., Mateus, O., & Benson, R. B. J. (2015). A specimen-level phylogenetic analysis and taxonomic revision of Diplodocidae (Dinosauria, Sauropoda). *PeerJ*, *3*, e857. https://doi.org/10.7717/peerj.857

Upchurch, P., Tomida, Y., & Barrett, P. M. (2004). A new specimen of *Apatosaurus ajax* (Sauropoda: Diplodocidae) from the Morrison Formation (Upper Jurassic) of Wyoming, USA. *National Science Museum Monographs*, *26*, 1–118.

Whitlock, J. A., & Wilson, J. A. (2013). Character distribution maps: A visualization method for comparative cladistics. *Acta Zoologica*, *94*(4), 490–499. https://doi.org/10.1111/azo.12006

Wiens, J. J. (1998). Does adding characters with missing data increase or decrease phylogenetic accuracy? *Systematic Biology*, *47*(4), 625–640. https://doi.org/10.1080/106351598260635

Wiens, J. J. (2006). Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics*, *39*(1), 34–42. https://doi.org/10.1016/j.jbi.2005.04.001

Wilkinson, M. (1995). Coping with abundant missing entries in phylogenetic inference using parsimony. *Systematic Biology*, *44*(4), 501–514. https://doi.org/10.1093/sysbio/44.4.501

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Tschopp E, Tschopp FA, Mateus O. Overlap Indices: Tools to quantify the amount of anatomical overlap among groups of incomplete terminal taxa in phylogenetic analyses. *Acta Zool*. 2017;00:1–8. https://doi.org/10.1111/azo.12202