

Phylogenetic inference based on landmark data in 41 empirical data sets

SANTIAGO A. CATALANO & AMBROSIO TORRES

Submitted: 20 November 2015
Accepted: 24 March 2016
doi:10.1111/zsc.12186

Catalano, S.A. & Torres, A. (2016). Phylogenetic inference based on landmark data in 41 empirical data sets. —*Zoologica Scripta*, 00: 000–000.

The inference of phylogenetic hypotheses from landmark data has been questioned during the last two decades. Besides theoretical concerns, one of the limitations pointed out for the use of landmark data in phylogenetics is its (supposed) lack of information relevant to the inference of phylogenetic relationships. However, empirical analyses are scarce; there exists no previous study that systematically evaluates the phylogenetic performance of landmark data in a series of data sets. In the present study, we analysed 41 published data sets in order to assess the correspondence between the phylogenetic trees derived from landmark data and those obtained with alternative and independent sources of evidence, and determined the main factors that might affect this inference. The data sets presented a variable number of terminals (5–200) and configurations (1–14), belonging to different taxonomic groups. The results showed that for most of the data sets analysed, the trees derived from landmark data presented a low correspondence with the reference phylogenies. The results were similar irrespective of the phylogenetic method considered. Complementary analyses strongly suggested that the limited amount of evidence included in each data set (one or a few landmark configurations) is the main cause for that low correspondence: the phylogenetic analysis of eight data sets that presented three or more configurations clearly showed that the inclusion of several landmark configurations improves the results. In addition, the analyses indicated that the inclusion of landmark data from different configurations is more important than the inclusion of more landmarks from the same configuration. Based on the results presented here, we consider that the poor results previously obtained in phylogenetic analyses based on landmark data were not caused by methodological limitations, but rather due to the limited amount of evidence included in the data sets.

Corresponding author: *Santiago Andrés Catalano, Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad Ejecutora Lillo (UEL), FML-CONICET, Miguel Lillo 251, 4000 San Miguel de Tucumán, Tucumán, Argentina; Facultad de Ciencias Naturales e IML, Universidad Nacional de Tucumán, S.M. de Tucumán, Argentina, E-mail: sacatalano@gmail.com*

Santiago A. Catalano, Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad Ejecutora Lillo (UEL), FML-CONICET, Miguel Lillo 251, 4000 San Miguel de Tucumán, Tucumán, Argentina. E-mail: sacatalano@gmail.com

Ambrosio Torres, Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad Ejecutora Lillo (UEL), FML-CONICET, Miguel Lillo, 251, 4000 San Miguel de Tucumán, Tucumán, Argentina. E-mail: atorresgalvis@gmail.com

Introduction

Characters that describe the shape of biological structures have been included in phylogenetic studies since the beginning of this discipline (e.g. wing shape in Hennig 1966). The variation in shape is generally translated into a set of discrete states and analysed in combination with other characters. Although this approach has shown to be very useful for incorporating shape information in phylogenetic matrices, it presents some limitations, mainly related to the arbitrary delimitation of discrete states from a variation

that is continuous in nature. In recent times, associated with the publication of new algorithmic implementations (Goloboff *et al.* 2006), the variation in shape has started to be included in morphological and combined matrices as continuous characters (e.g. Clouse *et al.* 2010; Prevosti 2010; Carrizo & Catalano 2015). Nevertheless, the description of shapes as a set of linear distances has additional limitations (Rohlf 2000). With the development of geometric morphometric methods, a new era started in the quantitative analysis of shape variation (Rohlf & Marcus 1993;

Adams *et al.* 2013). These methods are based on a particular definition of shape: all the geometrical information that remains when location, scale and rotation effects are sorted out from an object (Kendall 1977). In recent years, several approaches have been proposed for the phylogenetic treatment of landmark data (Rohlf 2002; Lockwood *et al.* 2004; Caumul & Polly 2005; González-José *et al.* 2008; Catalano *et al.* 2010; Klingenberg & Gidaszewski 2010; Catalano & Goloboff 2012). Although with a degree of mistrust during the first years (e.g. Monteiro 2000), there is now consensus that geometric data, landmark data in particular, can be analysed in a phylogenetic context with the primary aim of analysing morphological evolution (e.g. MacLeod 2002; Klingenberg & Gidaszewski 2010; Adams & Felice 2014). For instance, studies that analyse shape evolution using phylomorphospaces are common nowadays (Sherratt *et al.* 2014; Almécija *et al.* 2015a,b; Püschel & Sellers 2016). However, it is still questioned, on both methodological and empirical grounds, whether landmark data can be used to infer phylogenetic relationships. While in some cases, the empirical results obtained presented some level of agreement with accepted phylogenies, as is the case in studies performed in elephant shrews (Panchetti *et al.* 2008; Scalici & Panchetti 2011), the more general trend has been one of significant incongruence with alternative sources of evidence (e.g. Couette *et al.* 2005; Macholán 2006). The poor results generally obtained were interpreted by some authors (e.g. Klingenberg & Gidaszewski 2010) as indicating that landmark data may not be a suitable source of evidence for phylogenetic analyses. We have recently performed the first phylogenetic analysis based on a high number of configurations (Catalano *et al.* 2015). That analysis was based on nine landmark configurations derived from different skeletal structures. The phylogeny obtained presented a clear concordance with molecular data. In addition, complementary analyses indicated that better results were obtained with the inclusion of a higher number of configurations. Although promising, these results were derived from a single case, and it remains to be tested whether this conclusion can be generalized. In the present study, we analysed 41 previously published data sets in order to evaluate the performance of landmark data to infer phylogenetic relationships. The data sets were analysed with different methodological approaches that were previously employed to analyse landmark data in phylogenetics.

Materials and methods

Data sets

A total of 41 data sets (Table 1, Data S1) were obtained either directly from the authors or downloaded from data repositories (e.g. Dryad). For details of data retrieval

Table 1 Data sets information

Data set	N° species	N° of Conf.	N° of Lmk.	Reference/s
Abe_2012	34	1	37	Abe & Lieberman (2012)
Adams_2014	18	2	11	Adams & Felice (2014)
Almecija_2015	26	1*	26	Almécija <i>et al.</i> (2015a,b)
Alvarez_2013	39	4	90	Álvarez <i>et al.</i> (2013a)
Angielczyk_2010	9	1	9	Angielczyk <i>et al.</i> (2011)
AP_APV	24	2*	56	Álvarez & Perez (2013)/Álvarez <i>et al.</i> (2013b)
Arendt_2010	5	1	5	Arendt (2010)
Aristide_2013	29	1*	29	Aristide <i>et al.</i> (2013)
Astua_2009	32	1	32	Astúa (2009)
Baab_2014	33	1*	33	Baab <i>et al.</i> (2014)
Caumul_2005	12	3	144	Caumul & Polly (2005)
Claverie_2013	25	3	88	Claverie & Patek (2013)
Cruz_2012	5	1	5	Cruz <i>et al.</i> (2012)
de_Freitas_2012	7	3	80	De Freitas <i>et al.</i> (2012)
Foster_2008	7	1	7	Foster <i>et al.</i> (2008)
Foth_2012	31	1	31	Foth <i>et al.</i> (2012)
Franklin_2014	59	1	158	Franklin <i>et al.</i> (2014)
Frederich_2008	8	4	51	Frédérich <i>et al.</i> (2008)
Grieco_2012	6	1	6	Grieco <i>et al.</i> (2012)
GJ_GR	9	14*	476	González-José <i>et al.</i> (2008)/Gómez-Gómez-Robles <i>et al.</i> (2013)
Ivanovic_2011	13	2	26	Ivanović <i>et al.</i> (2012)
Johansson_2009	26	2	24	Johansson <i>et al.</i> (2009)
Klingenberg_2010	9	1	9	Klingenberg & Gidaszewski (2010)
Klingenberg_2013	160	1	160	Klingenberg & Marugán-Lobón (2013)
Klingenberg_2012	14	3	55	Klingenberg <i>et al.</i> (2012)
Markolf_2013	11	1	7	Markolf <i>et al.</i> (2013)
Martin-Serra_2015	46	6*	104	Martín-Serra <i>et al.</i> (2015)
Meloro_2012	14	1	14	Meloro & Jones (2012)
Muñoz_2014	6	1	6	Muñoz-Muñoz <i>et al.</i> (2014)
Neustupa_2007	5	1	5	Neustupa & Škaloud (2007)
Outomuro_2013	32	2	23	Outomuro <i>et al.</i> (2013)
Pierce_2008	24	1	24	Pierce <i>et al.</i> (2008)
Piras_2012	32	1	32	Piras <i>et al.</i> (2012)
Prevosti_2012	151	1	151	Prevosti <i>et al.</i> (2012)
Rohlf_2002	11	1	11	Rohlf (2002)
Santamaria_2013	5	1	5	Santamaria <i>et al.</i> (2013)
Stubbs_2013	53	1	102	Stubbs <i>et al.</i> (2013)
Vera_Candiotti_2009	108	1	108	Vera Candiotti & Altig (2010)
Voyta_2013	8	1	10	Voyta <i>et al.</i> (2013)
Watanabe_2014	10	1*	16	Watanabe & Slice (2014)
Young_2008	17	1*	9	Young (2008)

Lmk., landmark; Conf., configuration; *, 3D datasets.

process, see Data S2. In data sets with two or more configurations, the taxonomic sampling was reduced to those species that presented information for all configurations. Landmark matrices (in TNT format) and trees obtained from those data sets that were available in Dryad were included as Data S3. The rest of the data sets and resulting

trees are available upon request with the express agreement of the original authors.

Data analysis

Assessing methods and data set performance. The evaluation of the performance of different phylogenetic methods or sources of evidence has always been problematic, mainly because the real phylogeny cannot be known. A possible approach is to compare the results obtained in different analyses with phylogenies built considering alternative/independent evidence. This approach is based on a congruence criterion, a criterion that has always been at the base of cladistic analysis (Wheeler 2001). When comparing different methods to infer phylogenetic relationships from landmark data, it is possible to compare the results with phylogenies derived from alternative sources of evidence such as molecules or discrete morphological characters. In the present study, the performance of each data set/method/sampling was evaluated by comparing the results obtained against a reference phylogeny. When possible (87% of the data sets), the reference topology was based on molecular evidence. This was because molecular data offers a more independent source of evidence than traditional morphological characters, since some of these characters may describe the shape of structures that are also analysed as landmark data.

The degree of concordance of the trees derived from landmark data and the reference trees was assessed by calculating a measure of topological similarity based on the number of SPR (Subtree Pruning and Regrafting) moves. This measure is the complement of the number of SPR moves required to convert one tree into the other, divided by $T - 2$, where T is the number of taxa (Goloboff 2007). In addition to the SPR-based measure of topological similarity, Robinson-Foulds distances (Robinson & Foulds 1981) were calculated. The results were similar to those obtained using the SPR-based metric, and are included as supplementary material (Data S4).

Some of the data sets originally included more than one specimen per species while others included only a consensus configuration per species. In those data sets where data from individual specimens were included, all specimens from each species were superimposed using a General Procrustes Analysis (GPA, Gower 1975; Rohlf & Slice 1990). The consensus configurations derived from this step represented the shape of each species. In all cases, the consensus configurations representing each species were used to define a multiple superimposition by means of a new GPA. GPA was conducted using the functions included in the R package geomorph (Adams & Otárola-Castillo 2013). This multiple superimposition was the starting point for all the phylogenetic analyses. In most of the data sets, the

different configurations represent also different structures. Hence, unless stated otherwise, both terms are used here as synonyms.

Phylogenetic searches. The 41 data sets were analysed using ‘Phylogenetic Morphometrics’ (PM), a method developed by Catalano *et al.* (2010) for the analysis of landmark data in phylogenetics. This method is implemented in TNT phylogenetic software (Goloboff *et al.* 2008) and is a direct extension of the parsimony principle (*sensu* Farris 1983) for the analysis of landmark data (for a critical point of view about this method see Adams *et al.* 2013). The tree scores were established using the algorithms described in Goloboff & Catalano (2011) to optimize landmark data on a tree. Since phylogenetic searches for landmark data were not implemented natively in TNT at the time, this study was conducted (the version that analyses landmark data directly has been released in December 2015, Goloboff & Catalano (in press)), the searches were conducted using a script written in TNT macro language. The search strategy consisted in each case of a single RAS (Random Addition Sequence = Wagner trees) followed by rounds of Tree Bisection Reconnection (TBR). All the analyses were run on a cluster of 14 4-core CPUs using the parallel version of TNT. After a standardization step (see below), the multiple superimposition of each structure was incorporated into a matrix as a different character, generating a combined data set. In addition, the 41 data sets were analysed considering five additional phylogenetic approaches that have been previously considered to infer phylogenetic relationships from landmark data. The inclusion of several methods has not been guided by a pluralistic approach. Neither do we consider that all methods are equally good. In fact, many of these approaches have been questioned in the literature, and we agree with most of those criticisms. We included several approaches because we were interested in testing whether the conclusions derived from the analysis of the 41 data sets were consistent irrespective of the method considered.

1. Procrustes distances + Neighbor Joining (PD-NJ). This approach was followed in several studies (e.g. Lockwood *et al.* 2004; Couette *et al.* 2005; Cardini & Elton 2008; Scalici & Panchetti 2011). The processing of the 41 data sets was automated using R scripts (R Development Core Team 2015) that included functions from the following packages: “ape” (Paradis *et al.* 2004), “geomorph” (Adams & Otárola-Castillo 2013), “phangorn” (Schliep 2011) and “shapes” (Dryden 2014). For each structure (configuration), a matrix of Procrustes distances was calculated including all possible pairs of species. Afterwards, a Neighbor-Joining tree (NJ; Saitou & Nei 1987) was inferred. When data sets presented more

- than one configuration, distance matrices were standardized before NJ analysis (see below).
2. Procrustes distances + UPGMA (PD-UPGMA). The procedure followed was identical to the previous approach with the exception that the phylogenetic reconstruction method was UPGMA (Sneath & Sokal 1973). Although this method is phenetic, we included it as a phylogenetic approach because phylogenetic conclusions are generally extracted from the results and because it can be considered as phylogenetic as long as some restricted (and unrealistic) assumptions are held (Felsenstein 2004). This approach was followed in Frédérich *et al.* (2008) and Cardini & Elton (2008) although the authors of those studies clearly stated that the approach was considered only for evaluation of phenetic relationships.
 3. Landmark coordinates as continuous characters + parsimony (LC-P). This analysis was conducted in TNT. A TNT script was written to transform landmark coordinates into continuous data in TNT format. A single RAS + TBR were conducted for each data set. This approach was followed by Gold *et al.* (2014), and criticized, among others, by Catalano *et al.* (2010).
 4. Principal Component (PC) scores as continuous characters + parsimony (PC-P). This approach was proposed by González-José *et al.* (2008). The PC scores were calculated from the GPA superimposition. Each PC was considered as a different continuous character, with scores for each species being the states of the continuous character. Only those PC that represent up to 99% of the cumulative variance explained were included. PCA were done in R using functions included in the “geomorph” package (Adams & Otárola-Castillo 2013). Once the PC scores were obtained, the R script generated a matrix in TNT format including the continuous characters. A single RAS + TBR were run for each data set in TNT software. This approach was criticized by Adams *et al.* (2011).
 5. PC scores as continuous characters + Maximum Likelihood (PC-ML). This approach was followed in Caumul & Polly (2005) and González-José *et al.* (2008). The analysis was conducted following the same procedure as the previous approach with the exception that the ML analysis was run in Phylip (Felsenstein 1989, 2013). The analyses were automated using R scripts and the functions included in the “Rphylip” package (Revell & Chamberlain 2014). This approach has been criticized by Adams *et al.* (2013).

Effect of increasing amount of evidence on phylogenetic analyses of landmark data. To evaluate whether the results obtained in the phylogenetic analysis of landmark data

improve with the inclusion of an increasing number of configurations, a series of analyses were conducted on the eight data sets that have three or more configurations (Table 1). For each data set, a new analysis was conducted considering each possible combination of configurations for each number of configurations. For instance, if the data set had four configurations, the different searches included: (i) the analysis of each configuration independently, (ii) the analysis of each possible pair of configurations and (iii) the analysis of each possible triplet of configurations. Phylogenetic searches were conducted in this case under phylogenetic morphometrics (Catalano *et al.* 2010). The resulting trees were compared with the reference phylogenies using the SPR similarity measure previously described.

Besides sampling different numbers of configurations, we also conducted a series of analyses sampling different numbers of landmarks along all the configurations. To properly compare the results, the number of landmarks sampled in both approaches was the same. As previously indicated, for a given number of configurations, we repeated the analysis considering all the possible combinations of configurations. In the case of sampling different numbers of landmarks, we repeated the analyses considering 10 different random samples of landmarks for each number of landmarks considered. Both analyses were run in TNT and automated using scripts written in TNT macro language.

Standardization. For the simultaneous analysis of multiple landmark configurations, it is necessary to decide how the information of the different configurations is combined. Otherwise, some configurations might have much more influence than the rest in the election of the optimal phylogenetic hypothesis due to uncontrolled and possibly unwanted factors. For instance, if configurations representing different structures have different sizes, the phylogenetic results may be driven by larger configurations (see Catalano *et al.* 2015 for a detailed discussion about this topic). A standardization step is hence essential for the phylogenetic analysis of multiple configurations. Given the nature of the different methodological approaches considered in the present study, different standardization procedures were followed in each case. In the frame of PM analysis, Goloboff & Catalano (2011) proposed a standardization looking for an even contribution of each configuration to the final score. To achieve that goal, the score of each landmark is multiplied by a standardization factor. Alternatively, this factor can be used to modify the size, the configurations and, since the score is the sum of landmark displacements, the same result is obtained following that procedure (Goloboff & Catalano 2011). This was the approach followed in the present study in the analyses performed under phylogenetic morphometrics. For the

analysis of landmark coordinates as independent characters (LC-P), the coordinates were those derived from the standardization procedure previously described. For the analyses that considered PC scores as characters (in both parsimony and ML context), the scores for each PC were standardized between 0 and 1. In distance-based methods, Procrustes distances for each configuration were standardized, so that the minimum interspecific distance was 0 and the maximum interspecific distance was 1.

Results

Characteristic of the data sets

The main characteristics of the data sets analysed in the present study are presented in Table 1, and Data S3. Most of the data sets (approx. 70%) included a single landmark configuration, with only ten per cent including four or more configurations. The number of species was more variable among data sets ranging from five to 160 with Mammals represented two-thirds of the total number of data sets. The data sets included 12 different kinds of structures, the most common being cranium, mandible, wing and general body shape. Most of the structures were digitalized in two-dimensional landmark configurations.

Phylogenetic results

The phylogenetic analysis of the 41 data sets indicated a low congruence between the trees derived from landmark data and the reference phylogenies (Table 2). These results were consistent irrespective of the different phylogenetic methods considered (Fig. 1). The approaches that considered PC scores as characters (either analysed by parsimony or ML) produce the worst results. The approaches that produced the highest level of congruence were PD-UPGMA, PD-NJ and PM. Phylogenetic analyses of landmark data considering different numbers of configurations showed that in six out of eight data sets analysed, there was a clear relationship between the number of configurations and the topological correspondence between landmark-based tree and the reference tree (Fig. 2). Sampling different percentage of landmarks from each configuration instead of sampling complete configurations indicated that, for the same number of landmarks, better results were obtained when landmarks were sampled from different configurations (Fig. 3). The only data set that did not present this pattern was Martín-Serra_2015. In that case, similar values of congruence were obtained sampling landmarks from either a single configuration or from different configurations.

Discussion

Framed in a more general discussion about the utility of morphological characters in phylogenetics (see Giribet

Table 2 Topological correspondence (SPR similarity) between the trees derived from landmark data and the reference phylogeny for the 41 data sets listed in Table 1. The average SPR similarity was 0.51

Data set	SPR similarity
Abe_2012	0.516
Adams_2014	0.667
Almecija_2015	0.565
Alvarez_2013	0.500
Angielczyk_2010	0.667
AP_APV	0.381
Arendt_2010	0.500
Aristide_2013	0.654
Astua_2009	0.448
Baab_2014	0.467
Caumul_2005	0.667
Claverie_2013	0.546
Cruz_2012	1.000
de_Freitas_2012	0.750
Foster_2008	0.250
Foth_2012	0.429
Franklin_2014	0.357
Frederich_2008	0.600
Grieco_2012	0.667
GJ_GR	0.830
Ivanovic_2011	0.300
Johansson_2009	0.348
Klingenberg_2010	0.667
Klingenberg_2013	0.166
Klingenberg_2012	0.364
Markolf_2013	0.500
Meloro_2012	0.455
Muñoz_2014	0.667
Neustupa_2007	1.000
Outomuro_2013	0.379
Pierce_2008	0.381
Piras_2012	0.483
Prevosti_2012	0.270
Rohlf_2002	0.500
Santamaria_2013	0.500
Martín-Serra_2015	0.465
Stubbs_2013	0.340
Vera_Candioti_2009	0.000
Voyta_2013	0.400
Watanabe_2014	0.571
Young_2008	0.600

2015), the case of landmark data is of interest by itself, with considerations that are particular to this kind of characters. On one hand, there have been theoretical concerns about the compatibility of landmark data with the principles of phylogenetics (e.g. Bookstein 1994). On the other hand, there has been an empirical concern: do shape characters (described as landmark configurations) contain historical information as to infer phylogenetic relationships? While most of the discussion found in the literature is related to the first issue (e.g. Bookstein 1994; Monteiro 2000;

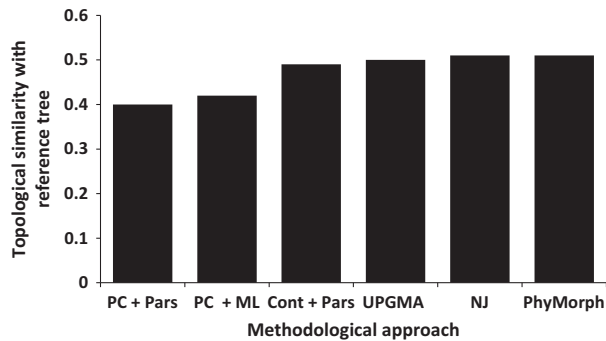


Fig. 1 Comparison of the performance of different phylogenetic approaches to analyse landmark data. The value represents the topological similarity (in terms of SPR moves) between the trees obtained from landmark data and the reference trees for the 41 data sets listed in Table 1. PC, principal component; Pars, parsimony; ML, maximum likelihood; PhyMorph, phylogenetic morphometrics; NJ, Neighbor joining; Cont, continuous. For a detailed description of each approach, see Materials and methods section.

MacLeod 2002; Rohlf 2002; Catalano *et al.* 2010), the empirical evidence considered to assess the performance of landmark data in phylogenetics has only been circumstantial. Studies of individual data sets have shown, in most cases, incongruent results using alternative sources of evidence (e.g. Rohlf 2002; Klingenberg & Gidaszewski 2010). Based on those results, the utility of landmark data has been questioned (e.g. Klingenberg & Gidaszewski 2010). The results of the present study, based on the analysis of 41 data sets strongly suggest that the lack of congruence with the reference phylogenies is associated to the scarce evidence included in the data sets analysed, not to the particular methodological approach considered.

Phylogenetic inference and amount of evidence

The general pattern that emerges from the analysis of the 41 data sets is a low level of congruence between the trees derived from landmark data and the reference phylogenies (Fig. 1, Table 2). The number of SPR moves is on average one-half of the maximum possible values, indicating an important level of topological incongruence. These results were sustained irrespective of the phylogenetic method considered (see below). The low level of congruence found in the present study seems at first sight to support the unsuitability of landmark data for phylogenetic inference (Klingenberg & Gidaszewski 2010). A more detailed evaluation of the results however gives a different perspective, and allows assessment of the main cause for this pattern. A key point for proper interpretation of the results is that approximately 70% of the data sets analysed represented only a single landmark configuration, and an additional 12% represented only two configurations (Table 1). This data by itself suggest that the poor results can result from

the inclusion of insufficient evidence. That scarce evidence is the main limitation of previous phylogenetic analyses on landmark data has been recently suggested by Catalano *et al.* (2015). The present results give support to the finding of Catalano *et al.* (2015) generalizing the pattern previously observed in a particular group of mammals.

Despite the fact that the results improve with additional evidence, the analyses also show that more data are required to properly infer phylogenetic relationships from landmark data. For the data sets analysed, only the GJ-GR data set showed a high congruence with the reference tree. As previously indicated, this data set was composed of data from two different studies (González-José *et al.* 2008 and Gómez-Robles *et al.* 2013), including much more evidence than the rest of the data sets, not only in terms of the absolute number of landmark configurations (14), but also relative to number of terminals included in the analysis (1.5 configurations per species). This higher amount of evidence can be the cause for the high congruence between the landmark data and the reference tree. However, it cannot be discarded that the pattern is caused by both trees being derived from morphological data. In any case, this example still shows that one or a few configurations/structures are insufficient to properly infer phylogenetic relationships from landmark data.

The analyses performed showed that better results were obtained, for the same number of landmarks, when landmarks were sampled from different configurations (Fig. 3). This indicates that the improvement in the results when including a higher number of configurations is not only caused by the inclusion of more landmarks but also by combining information from different structures. There are different reasons why this combination may improve the results. On the one hand, the inclusion of structures with different evolutionary rates would allow resolution of phylogenetic relationships at different levels on the tree. On the other hand, including configurations from different structures can help to reduce the effect of homoplasious characters (Oxnard 2000). Irrespective of the particular cause, it is quite clear that the inclusion of several structures in the same analysis improve the phylogenetic results.

Besides the scarce information included in most of the data sets, there are other factors that might in principle contribute to the incongruence between phylogenetic trees based on molecular data and those derived from landmark data. One of them is related to measurement errors. Measurement errors (mainly related to 2d configurations, Cardini 2014), may introduce “noise” in the data that would in turn worsen the phylogenetic results. A second issue is related to the use of a consensus configuration to represent the shape of each terminal (Cardini *et al.* 2015). Although these two factors may have an important effect when

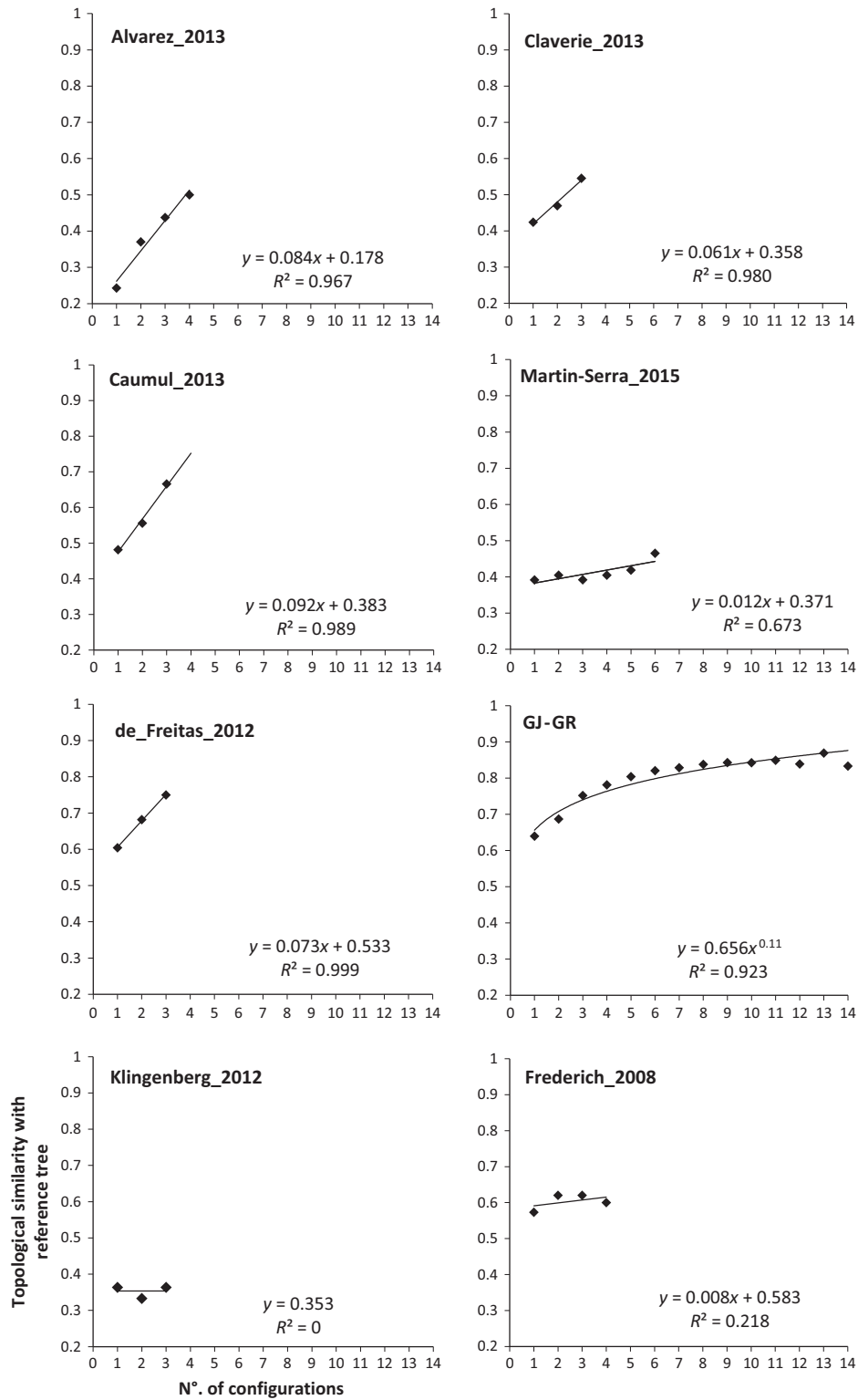


Fig. 2 Correspondence between the trees obtained considering a variable number of landmark configurations and the reference trees measured in terms of SPR similarity. Each point represents the mean value for the trees derived from all possible combinations of configurations (for a given number of configurations). A linear tendency lined is shown except for the case of Martín-Serra_2015 data set where a power trend line is shown.

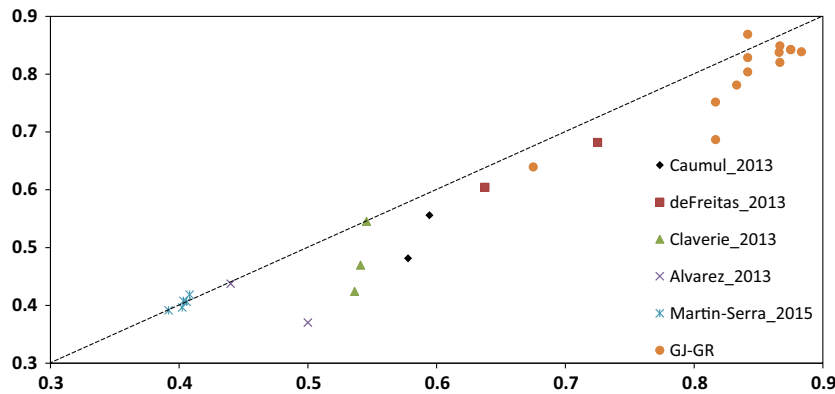


Fig. 3 Comparison of the results obtained by sampling landmarks from different configurations vs. sampling full configurations. Each point in the graph represents the SPR similarity against the reference phylogeny of the trees obtained by sampling a given number of landmarks. The values on the Y axis were obtained by sampling full configurations, while the values in the X axis were obtained by sampling landmarks from all the configurations (See Materials and methods section for details). The line represents equal topological similarity values irrespective of whether the landmarks were sampled from the same configuration or from different configurations. With the exception of the Martin-Serra_2015 data set, higher topological congruence with the reference phylogeny was obtained (for the same number of landmarks), when landmarks were gathered from different configurations. The analysis included only those data sets where results were improved with the inclusion of more configurations (Fig. 2).

working at the population level, we consider that they would have a minor effect in phylogenetic analyses, given that the intraspecific variation is much smaller than the differences among species. In any case, this is just a prediction that should be evaluated in future studies.

Methodological approaches

The decision of which method is selected to perform a certain phylogenetic analysis is based on epistemological (i.e. methods that can be considered as superior in a certain philosophical framework) and practical reasons (e.g. availability of software to perform the analysis). Another relevant issue is the empirical behaviour of the methods in both real and simulated data sets. The present study is the first that compares the empirical performance of the different phylogenetic approaches in a large number of data sets. The different analytical approaches considered to infer phylogenetic relationships from landmark data indicated that PD-NJ, PD-UPGMA and PM approaches gave the best results while PC scores analysed by either parsimony (PC-P) or Maximum Likelihood (PC-L) produced trees with the lowest congruence with the reference phylogenies. Although it is tempting to derive generalizations from these results, some caution should be taken. First because, as previously shown, the evidence included in most of the data sets was scarce, with results that were in general very different from the reference phylogenies. Whether the methods that behaved better under those conditions will also perform better in cases where more evidence is included is an open question. A second point to consider is that the differences in performance among methods are in

general small. For instance, the difference among the three approaches that had better performance can be accounted by only three SPR moves in a single data set. Finally, the results presented here should be taken with caution because each method was run considering a single setting. A detailed comparison of the different approaches would require taking into consideration different settings at different steps of the analyses such as standardization, weighting, superimposition, etc.

Final remarks

Adams *et al.* (2013) indicated that one of the areas of research that has advanced the least in the frame of geometric morphometrics is the analysis of landmark data within a phylogenetic context. These methodological limitations might in principle be a possible explanation for the poor results obtained in empirical studies that intend to infer phylogenetic relationships from landmark data. However, our results strongly suggest that the main limitation of most previous studies was the scarce evidence included in the data sets analysed. If results of future empirical analyses are to be considered to reject the use of landmark data to infer phylogenetics, those analyses should be based on a proper character sampling.

Acknowledgements

The authors thank the Willi Hennig Society for subsidizing the program TNT and making it freely available. We also thank all the authors for sharing their data sets and helping to interpret the data and file formats. Finally, we would like to express our gratitude to Dr. Andrea Cardini

and two anonymous reviewers for their helpful comments. This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP 0260 to SAC) and Agencia Nacional De Promoción Científica y Tecnológica (ANPCyT, PICT 2011-1679, PICT 2014-1930 to SAC).

References

- Abe, F. R. & Lieberman, B. S. (2012). Quantifying morphological change during an evolutionary radiation of Devonian trilobites. *Paleobiology*, *38*, 292–307.
- Adams, D. C. & Felice, R. N. (2014). Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PLoS One*, *9*, 1–8.
- Adams, D. C. & Otárola-Castillo, E. (2013). geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, *4*, 393–399.
- Adams, D. C., Cardini, A., Monteiro, L. R., O’Higgins, P. & Rohlf, F. J. (2011). Morphometrics and phylogenetics: principal components of shape from cranial modules are neither appropriate nor effective cladistic characters. *Journal of Human Evolution*, *60*, 240–243.
- Adams, D. C., Rohlf, F. J. & Slice, D. E. (2013). A field comes of age: geometric morphometrics in the 21st century. *Hystrix, The Italian Journal of Mammalogy*, *24*, 7–14.
- Almécija, S., Orr, C. M., Tocheri, M. W., Patel, B. A. & Jungers, W. L. (2015a). Exploring phylogenetic and functional signals in complex morphologies: the hamate of extant anthropoids as a test-case study. *The Anatomical Record*, *298*, 212–229.
- Almécija, S., Smaers, J. B. & Jungers, W. L. (2015b). The evolution of human and ape hand proportions. *Nature Communications*, *6*, 7717.
- Álvarez, A. & Perez, S. I. (2013). Two- versus three-dimensional morphometric approaches in macroevolution: insight from the mandible of caviomorph rodents. *Evolutionary Biology*, *40*, 150–157.
- Álvarez, A., Ercoli, M. D. & Prevosti, F. J. (2013a). Locomotion in some small to medium-sized mammals: a geometric morphometric analysis of the penultimate lumbar vertebra, pelvis and hindlimbs. *Zoology*, *116*, 356–371.
- Álvarez, A., Perez, S. I. & Verzi, D. H. (2013b). Ecological and phylogenetic dimensions of cranial shape diversification in South American caviomorph rodents (Rodentia: Hystricomorpha). *Biological Journal of the Linnean Society*, *110*, 898–913.
- Angielczyk, K. D., Feldman, C. R. & Miller, G. R. (2011). Adaptive evolution of plastron shape in emydine turtles. *Evolution*, *65*, 377–394.
- Arendt, J. (2010). Morphological correlates of sprint swimming speed in five species of spadefoot toad tadpoles: comparison of morphometric methods. *Journal of Morphology*, *271*, 1044–1052.
- Aristide, L., Rosenberger, A. L., Tejedor, M. F. & Perez, S. I. (2013). Modeling lineage and phenotypic diversification in the New World monkey (Platyrrhini, Primates) radiation. *Molecular Phylogenetics and Evolution*, *82*, 375–385.
- Astúa, D. (2009). Evolution of scapula size and shape in didelphid marsupials (Didelphimorphia: Didelphidae). *Evolution*, *63*, 2438–2456.
- Baab, K. L., Perry, J. M. G., Rohlf, F. J. & Jungers, W. L. (2014). Phylogenetic, ecological, and allometric correlates of cranial shape in Malagasy lemuriforms. *Evolution*, *68*, 1450–1468.
- Bookstein, F. L. (1994). Can biometrical shape be a homologous character? In B. K. Hall (Ed) *Homology: The Hierarchical Basis of Comparative Biology* (pp. 197–227). New York: Academic Press.
- Cardini, A. (2014). Missing the third dimension in geometric morphometrics: how to assess if 2D images really are a good proxy for 3D structures? *Hystrix, The Italian Journal of Mammalogy*, *25*, 73–81.
- Cardini, A. & Elton, S. (2008). Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons *Biological Journal of the Linnean Society*, *93*, 813–834.
- Cardini, A., Seetah, K. & Barker, G. (2015). How many specimens do I need? Sampling error in geometric morphometrics: testing the sensitivity of means and variances in simple randomized selection experiments *Zoomorphology*, *134*, 149–163.
- Carrizo, L. V. & Catalano, S. A. (2015). First phylogenetic analysis of the tribe Phyllotini (Rodentia: Sigmodontinae) combining morphological and molecular data. *Cladistics*, *31*, 593–620.
- Catalano, S. A. & Goloboff, P. A. (2012). Simultaneously mapping and superimposing landmark configurations with parsimony as optimality criterion. *Systematic Biology*, *61*, 392–400.
- Catalano, S. A., Goloboff, P. A. & Gianinni, N. (2010). Phylogenetic morphometrics (I): the use of landmark data in a phylogenetic framework. *Cladistics*, *26*, 539–549.
- Catalano, S. A., Ercoli, M. & Prevosti, F. (2015). The more, the better: the use of multiple landmark configurations to solve the phylogenetic relationships in Musteloids. *Systematic Biology*, *64*, 294–306.
- Caumul, R. & Polly, P. D. (2005). Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (Marmota: Rodentia). *Evolution*, *59*, 2460–2472.
- Claverie, T. & Patek, S. N. (2013). Modularity and rates of evolutionary change in a power-amplified prey capture system. *Evolution*, *67*, 3191–3207.
- Clouse, R. M., de Bivort, B. L. & Giribet, G. (2010). A phylogenetic analysis for the South-east Asian mite harvestman family Stylocellidae (Opiliones: Cyphophthalmi) – a combined analysis using morphometric and molecular data. *Invertebrate Systematics*, *23*, 515–529.
- Couette, S., Escarguel, G. & Montuire, S. (2005). Constructing, bootstrapping, and comparing morphometric and phylogenetic trees: a case study of New World monkeys (Platyrrhini, Primates). *Journal of Mammalogy*, *86*, 773–781.
- Cruz, R. A. L., Pante, M. A. J. R. & Rohlf, F. J. (2012). Geometric morphometric analysis of shell shape variation in *Conus* (Gastropoda: Conidae). *Biological Journal of the Linnean Society*, *165*, 296–310.
- De Freitas, T. R. O., Fernandes, F. A., Fornel, R. & Roratto, P. A. (2012). An endemic new species of tuco-tuco, genus *Ctenomys* (Rodentia: Ctenomyidae), with a restricted geographic distribution in southern Brazil. *Journal of Mammalogy*, *93*, 1355–1367.
- Dryden, I. L. (2014). shapes: Statistical shape analysis. R package version 1.1-10.
- Farris, J. (1983). The logical basis of phylogenetic analysis. In N. I. Platnick & V. A. Funk (Eds) *Advances in Cladistics II* (pp. 7–36). New York: Columbia University Press.

- Felsenstein, J. (1989). PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics*, 5, 164–166.
- Felsenstein, J. (2004). *Inferring Phylogenies*. Sunderland, Massachusetts: Sinauer Associates.
- Felsenstein, J. (2013). PHYLIP – Phylogeny Inference Package. Ver. 3.695. [Computer software and manual]. Seattle, Washington: Department of Genome Sciences, University of Washington, Seattle.
- Foster, D. J., Podos, J. & Hendry, A. P. (2008). A geometric morphometric appraisal of beak shape in Darwin's finches. *Journal of Evolutionary Biology*, 21, 263–275.
- Foth, C., Brusatte, S. L. & Butler, R. J. (2012). Do different disparity proxies converge on a common signal? Insights from the cranial morphometrics and evolutionary history of Pterosauria (Diapsida: Archosauria) *Journal of Evolutionary Biology*, 25, 904–915.
- Franklin, O., Palmer, C. & Dyke, D. (2014). Pectoral fin morphology of Batoid fishes (Chondrichthyes: Batoidea): explaining phylogenetic variation with geometric morphometrics. *Journal of Morphology*, 275, 1173–1186.
- Frédérich, B., Pilet, A., Parmentier, E. & Vandewalle, P. (2008). Comparative trophic morphology in eight species of damselfishes (Pomacentridae). *Journal of Morphology*, 269, 175–188.
- Giribet, G. (2015). Morphology should not be forgotten in the era of genomics—a phylogenetic perspective. *Zoologischer Anzeiger*, 256, 96–103.
- Gold, M. E. L., Brochu, C. A. & Norell, M. A. (2014). An expanded combined evidence approach to the *Gavialis* problem using geometric morphometric data from Crocodylian braincases and Eustachian systems. *PLoS One*, 9, 1–12.
- Goloboff, P. A. & Catalano, S. A. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics*, doi: 10.1111/cla12160.
- Goloboff, P. A. (2007). Calculating SPR distances between trees. *Cladistics*, 24, 591–597.
- Goloboff, P. A. & Catalano, S. A. (2011). Phylogenetic morphometrics (II): algorithms for landmark optimization. *Cladistics*, 27, 42–51.
- Goloboff, P., Mattoni, C. & Quinteros, A. (2006). Continuous characters analysed as such. *Cladistics*, 22, 589–601.
- Goloboff, P. A., Farris, J. S. & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics*, 24, 774–786.
- Gómez-Robles, A., Bermúdez de Castro, J. M., Arsuaga, J. L., Carbonell, E. & Polly, P. D. (2013). No known hominin species matches the expected dental morphology of the last common ancestor of Neanderthals and modern humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 18196–18201.
- González-José, R., Escapa, I., Neves, W. A., Cúneo, R. & Pucciarelli, H. M. (2008). Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. *Nature*, 453, 775–778.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40, 33–51.
- Grieco, T. M., Rizk, O. T. & Hlusko, L. J. (2012). A modular framework characterizes micro- and macroevolution of old world monkey dentitions. *Evolution*, 67, 241–259.
- Hennig, W. (1966). *Phylogenetic Systematics*. Urbana, Illinois: University of Illinois Press.
- Ivanović, A., Dzukić, G. & Kalezić, M. (2012). A phenotypic point of view of the adaptive radiation of crested newts (*Triturus cristatus* Superspecies, Caudata, Amphibia). *International Journal of Evolutionary Biology*, 2012, 1–9.
- Johansson, F., Söderquist, M. & Bokma, F. (2009). Insect wing shape evolution: independent effects of migratory and mate guarding flight on dragonfly wings. *Biological Journal of the Linnean Society*, 97, 362–372.
- Kendall, D. G. (1977). The diffusion of shape. *Advances in Applied Probability*, 9, 428–430.
- Klingenberg, C. P. & Gidaszewski, N. A. (2010). Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Systematic Biology*, 59, 245–261.
- Klingenberg, C. P. & Marugán-Lobón, J. (2013). Evolutionary covariation in geometric morphometric data: analysing integration, modularity, and allometry in a phylogenetic context. *Systematic Biology*, 62, 591–610.
- Klingenberg, C. P., Duttke, S., Whelan, S. & Kim, M. (2012). Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. *Journal of Evolutionary Biology*, 25, 115–129.
- Lockwood, C., Kimbel, W. & Lynch, J. (2004). Morphometrics and hominoid phylogeny: support for a chimpanzee–human clade and differentiation among great ape subspecies. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 4356–4360.
- Macholán, M. (2006). A geometric morphometric analysis of the shape of the first upper molar in mice of the genus *Mus* (Muridae, Rodentia). *Journal of Zoology*, 270, 672–681.
- MacLeod, N. (2002). Phylogenetic signal in morphometric data. In N. MacLeod & P. Forey (Eds) *Morphology, Shape and Phylogeny* (pp. 100–138). London: Taylor & Francis.
- Markolf, M., Rakotonirina, H., Fichtel, C., von Grumbkow, P., Brameier, M. & Kappeler, P.M. (2013). True lemurs... true species-species delimitation using multiple data sources in the brown lemur complex. *BMC Evolutionary Biology*, 13, 233.
- Martín-Serra, A., Figueirido, B., Pérez-Claros, J. A. & Palmqvist, P. (2015). Patterns of morphological integration in the appendicular skeleton of mammalian carnivores. *Evolution*, 69, 321–340.
- Meloro, C. & Jones, M. E. H. (2012). Tooth and cranial disparity in the fossil relatives of *Sphenodon* (Rhynchocephalia) dispute the persistent 'living fossil' label. *Journal of Evolutionary Biology*, 25, 2194–2209.
- Monteiro, L. (2000). Why morphometrics is special: the problem with using partial warps as characters for phylogenetic inference. *Systematic Biology*, 49, 796–800.
- Muñoz-Muñoz, F., Talavera, S., Carpenter, S., Nielsen, S. A., Werner, D. & Pagès, N. (2014). Phenotypic differentiation and phylogenetic signal of wing shape in western European biting midges, *Culicoides* spp., of the subgenus *Avaritia*. *Medical and Veterinary Entomology*, 28, 319–329.
- Neustupa, J. & Škaloud, P. (2007). Geometric morphometrics and qualitative patterns in the morphological variation of five species of *Micrasterias* (Zygnemophyceae, Viridiplantae). *Preslia*, 79, 401–417.
- Outomuro, D., Dijkstra, K. B. & Johansson, F. (2013). Habitat variation and wing coloration affects wing shape evolution in dragonflies. *Journal of Evolutionary Biology*, 26, 1866–1874.

- Oxnard, C. E. (2000). Morphometrics of the primate skeleton and the functional and developmental underpinnings of species diversity. In: P. O'Higgins & M. J. Cohn (Eds) *Development, Growth and Evolution* (pp. 235–263). London: Academic Press.
- Panchetti, F., Scalici, M., Carpaneto, G. M. & Gibertini, G. (2008). Shape and size variations in the cranium of elephant-shrews: a morphometric contribution to a phylogenetic debate. *Zoomorphology*, *127*, 69–82.
- Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290.
- Pierce, S. E., Angielczyk, K. D. & Rayfield, E. J. (2008). Patterns of morphospace occupation and mechanical performance in extant crocodylian skulls: a combined geometric morphometric and finite element modeling approach. *Journal of Morphology*, *269*, 840–864.
- Piras, P., Sansalone, G., Teresi, L., Kotsakis, T., Colangelo, P. & Loy, A. (2012). Testing convergent and parallel adaptations in talpids humeral mechanical performance by means of geometric morphometrics and finite element analysis. *Journal of Morphology*, *273*, 696–711.
- Prevosti, F. J. (2010). Phylogeny of the large extinct South American canids (Mammalia, Carnivora, Canidae) using a “total evidence” approach. *Cladistics*, *26*, 456–481.
- Prevosti, F. J., Turazzini, G. F., Ercoli, M. D. & Hingst-Zaher, E. (2012). Mandible shape in marsupial and placental carnivorous mammals: a morphological comparative study using geometric morphometrics. *Biological Journal of the Linnean Society*, *164*, 836–855.
- Püschel, T. A. & Sellers, W. I. (2016). Standing on the shoulders of apes: analysing the form and function of the hominoid scapula using geometric morphometrics and finite element analysis. *American Journal of Physical Anthropology*, *159*, 25–341.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
- Revell, L. J. & Chamberlain, S. A. (2014). Rphylip: an R interface for PHYLIP. *Methods in Ecology and Evolution*, *5*, 976–981.
- Robinson, D. R. & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, *53*, 131–147.
- Rohlf, F. J. (2000). Statistical power comparisons among alternative morphometric methods. *American Journal of Physical Anthropology*, *111*, 463–478.
- Rohlf, F. J. (2002). Geometric morphometrics and phylogeny. In N. Macleod & P. Forey (Eds) *Morphology, Shape, and Phylogeny* (pp. 175–193). London, UK: Taylor & Francis.
- Rohlf, F. J. & Marcus, L. F. (1993). A revolution in morphometrics. *Trends in Ecology and Evolution*, *8*, 129–132.
- Rohlf, F. J. & Slice, D. E. (1990). Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, *39*, 40–59.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.
- Santamaria, C. A., Mateos, M., Taiti, S., DeWitt, T. J. & Hurtado, L. A. (2013). A complex evolutionary history in a remote archipelago: phylogeography and morphometrics of the hawaiian endemic *Ligia* Isopods. *PLoS One*, *8*, 1–14.
- Scalici, M. & Panchetti, F. (2011). Morphological cranial diversity contributes to phylogeny in soft-furred sengis (Afrotheria, Macroscelidea). *Zoology*, *114*, 85–94.
- Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*, 592–593.
- Sherratt, E., Gower, D. J., Klingenberg, C. P. & Wilkinson, M. (2014). Evolution of cranial shape in caecilians (Amphibia: Gymnophiona). *Evolutionary Biology*, *41*, 528–545.
- Sneth, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy*. San Francisco: Freeman.
- Stubbs, T. L., Pierce, S. E., Rayfield, E. J. & Anderson, P. S. L. (2013). Morphological and biomechanical disparity of crocodyline archosaurs following the end-Triassic extinction. *Proceedings of the Royal Society B*, *280*, 1–10.
- Vera-Candioti, M. F. & Altig, R. (2010). A survey of shape variation in keratinized labial teeth of anuran larvae as related to phylogeny and ecology. *Biological Journal of the Linnean Society*, *101*, 609–625.
- Voyta, L. L., Golenishchev, F. N. & Tiunov, M. P. (2013). Analysis of shape and size variation of the first lower molar in Far-Eastern grey voles of genus *Alexandromys* (Rodentia: Cricetidae) from Russian fauna using geometric morphometrics. *Russian Journal of Theriology*, *13*, 19–32.
- Watanabe, A. & Slice, D. E. (2014). The utility of cranial ontogeny for phylogenetic inference: a case study in crocodylians using geometric morphometrics. *Journal of Evolutionary Biology*, *27*, 1078–1092.
- Wheeler, W. (2001). Homology and the optimization of DNA sequence data. *Cladistics*, *17*, S3–S11.
- Young, N.M. (2008). A comparison of the ontogeny of shape variation in the anthropoid scapula: functional and phylogenetic signal. *American Journal of physical anthropology*, *136*, 247–264.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Characteristics of each dataset.

Data S2. Data gathering.

Data S3. Data sets and resulting trees.

Data S4. Topological correspondence between the trees derived from landmark data and the reference phylogenies for the 41 datasets listed in Table 1.