

Discrete and morphometric traits reveal contrasting patterns and processes in the macroevolutionary history of a clade of scorpions

N. MONGIARDINO KOCH* , F. S. CECCARELLI†‡ , A. A. OJANGUREN-AFFILASTRO† & M. J. RAMÍREZ†

*Department of Geology & Geophysics, Yale University, New Haven, CT, USA

†Museo Argentino de Ciencias Naturales "Bernardino Rivadavia", Buenos Aires, Argentina

‡Departamento de Biología de la Conservación, Centro de Investigación Científica y de Educación Superior de Ensenada, Ensenada, Baja California, México

Keywords:

disparity;
macroevolution;
morphometrics;
morphospace;
phylogenetic comparative methods.

Abstract

Many palaeontological studies have investigated the evolution of entire body plans, generally relying on discrete character-taxon matrices. In contrast, macroevolutionary studies performed by neontologists have mostly focused on morphometric traits. Although these data types are very different, some studies have suggested that they capture common patterns. Nonetheless, the tests employed to support this claim have not explicitly incorporated a phylogenetic framework and may therefore be susceptible to confounding effects due to the presence of common phylogenetic structure. We address this question using the scorpion genus *Brachistosternus* Pocock 1893 as case study. We make use of a time-calibrated multilocus molecular phylogeny, and compile discrete and traditional morphometric data sets, both capturing the overall morphology of the organisms. We find that morphospaces derived from these matrices are significantly different, and that the degree of discordance cannot be replicated by simulations of random character evolution. Moreover, we find strong support for contrasting modes of evolution, with discrete characters being congruent with an 'early burst' scenario whereas morphometric traits suggest species-specific adaptations to have driven morphological evolution. The inferred macroevolutionary dynamics are therefore contingent on the choice of character type. Finally, we confirm that metrics of correlation fail to detect these profound differences given common phylogenetic structure in both data sets, and that methods incorporating a phylogenetic framework and accounting for expected covariance should be favoured.

Introduction

Ever since the earliest attempts by Linnaeus at classifying the biota, the outstanding diversity of biological form has amazed naturalists and biologists alike, profoundly shaping our approach to most aspects of historical biology. One of the most tantalizing characteristics of morphological diversity, generally referred to as

disparity, is its nonhomogenous distribution. From the differentiated morphotypes of populations to the baupläne of higher level taxonomic categories, disparity shows a highly clustered pattern across the biological hierarchy (Erwin, 2007). Uncovering the causes behind this peculiar pattern has been a central goal of evolutionary biology. In the last few decades, the ever-increasing availability of molecular phylogenies, coupled with the development of rigorous statistical techniques, has resulted in an unprecedented power to tackle these questions through the inference of historical patterns of morphological change (Pagel, 1999), largely contributing to the success of macroevolutionary research (Goswami *et al.*, 2016).

Correspondence: Nicolas Mongiardino Koch, Department of Geology & Geophysics, Yale University, 210 Whitney Ave., New Haven, CT 06511, USA.

Tel.: +1 (203) 432 3114; fax: +1 (203) 432 3134;
e-mail: nicolas.mongiardinokoch@yale.edu

Within this emerging field, evolutionary biologists have generally studied the dynamics of either single characters or of a limited set of traits, for example those characterizing individual morphological modules (Aristide *et al.*, 2016; Garcia-Porta *et al.*, 2016) or hypothesized to play a role in niche exploitation (Mahler *et al.*, 2013). The disparity in size or shape among these characters is captured through the use of either linear measurements, an approach often called traditional morphometrics (Marcus, 1990), or through more elaborate methods such as landmark-based geometric morphometrics (Adams *et al.*, 2004). This focus on the analysis of quantitative morphological variation has led to the development of a wide array of phylogenetic comparative methods (O'Meara, 2012; Pennell & Harmon, 2013), many of which are specifically designed to address the evolutionary history of continuous characters.

On the other hand, palaeontologists have traditionally focused on the origin, assemblage and diversification of entire body plans (Gould, 1990; Wills *et al.*, 1994; Foote, 1997). Given the morphological complexity under study, as well as the taxonomic breadth targeted, identifying homologous measurements or landmarks (a requirement of morphometric-based studies) becomes unfeasible (Wills *et al.*, 1994). Therefore, this kind of research generally relies on matrices built using discrete characters, such as the absence or presence of putatively homologous structures. Although far from new, this approach has recently experienced a resurgence in the literature (Lloyd, 2016).

Morphometric and discrete characters not only differ on the nature of the variation being captured, but are also thought to contrast in aspects such as their rate of evolution, relative degree of independence, the homology of the structures being considered (MacLeod, 2002) and their level of phylogenetic signal and homoplasy (Klingenberg & Gidaszewski, 2010). Therefore, a long-lasting question in disparity-based studies is whether or not the patterns captured by discrete and morphometric variables are congruent. Surprisingly, so far attempts at answering this question have unanimously concluded that the patterns derived from both types of data are compatible (Villier & Eble, 2004; Anderson & Friedman, 2012; Foth *et al.*, 2012; Hetherington *et al.*, 2015). This result is especially intriguing given the fact that palaeontological and neontological approaches seem to offer opposite views on some key evolutionary questions, such as the relative frequency of 'early bursts' of phenotypic evolution (Slater & Pennell, 2013). However, the aforementioned studies exclusively tested for the existence of correlations between either alternative data matrices or results derived from them, such as patterns of morphospace occupation or estimates of disparity. On the other hand, they have rarely incorporated an explicit phylogenetic scaffold into their analyses. As a result, previous studies may have failed to account

for the presence of an expected degree of phylogenetic structure common to different data sets (Anderson & Friedman, 2012), arising from the shared evolutionary history that ultimately generated them. Without an analysis that is performed within an explicit phylogenetic framework, an appropriate null expectation cannot be specified (Lloyd, 2016), and the level to which inferred macroevolutionary patterns and dynamics are contingent on character choice cannot be fully assessed.

We tackled this problem employing the scorpion genus *Brachistosternus* Pocock 1893 as case study. This clade, endemic to South America, is one of the region's best studied genera of scorpions and is composed of medium-sized fossorial species that inhabit arid and semi-arid environments (Kovařík & Ojanguren-Affilastro, 2013). Several subclades have diversified in different biogeographic regions across the southern and eastern regions of the continent (Ojanguren-Affilastro *et al.*, 2016), a process mostly driven by the uplift of the Andean mountain range (Ceccarelli *et al.*, 2016).

Unlike previous efforts at assessing the dependence of macroevolutionary patterns on character choice, we used an independent, time-calibrated molecular phylogeny for the group under study. Using this phylogenetic hypothesis, we studied the evolutionary history of morphological change employing both a discrete and a morphometric character-taxon matrix, applying a variety of phylogenetic comparative methods to address three main questions: Are the patterns of morphological disparity implied by discrete and morphometric variables significantly different from each other? And if so, is this difference within the range expected by simulations of random character evolution? Finally, do these different representations of overall body shape imply that different macroevolutionary processes drove the morphological change in this group of scorpions? By expanding the set of metrics employed to characterize morphological evolution beyond static measures of disparity, as well as incorporating a time-calibrated phylogeny, we are able to provide new insights into the impact of character choice on macroevolutionary inference.

Materials and methods

Phylogeny and data matrices

The phylogeny used is the same as that of a recent biogeographic analysis of the genus (Ceccarelli *et al.*, 2016) and derives from a multispecies coalescent analysis carried out in *BEAST 1.8.2 (Heled & Drummond, 2010) using DNA matrices built from five gene loci sequenced for 120 specimens of *Brachistosternus*. The final topology (Fig. 1) was obtained after pruning outgroups and species for which no morphological information could be gathered. Following previous analysis (Ceccarelli *et al.*, 2016; Ojanguren-Affilastro *et al.*, 2016), the genus was

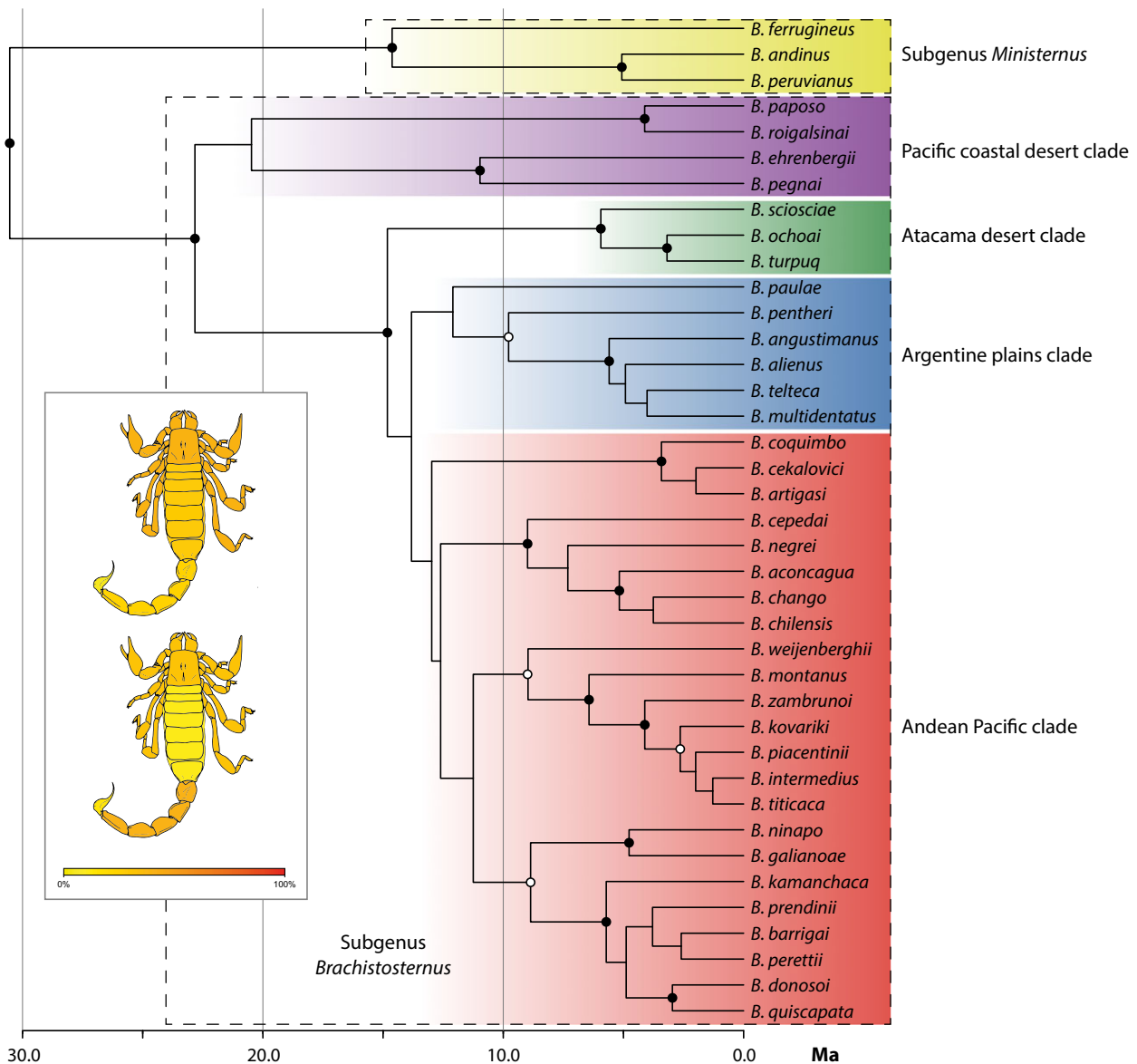


Fig. 1 Phylogeny of the genus *Brachistosternus*. The five subclades used throughout the study are shown in different colours. Black circles represent Bayesian posterior probabilities > 0.95, white circles between 0.9 and 0.95. For details on the methods of phylogenetic inference and time-calibration employed, please refer to Ceccarelli *et al.* (2016). The silhouettes show a dorsal view of *Brachistosternus alienus* Lönnberg 1898 with body regions (herein constituted by prosoma, mesosoma, metasomal segments I–V and telson) colour coded to represent relative character sampling effort between morphological data sets (top, discrete; bottom, morphometric). Sampling is relatively homogenous between body regions, as well as similar for both data sets.

subdivided into five clades that have radiated in different biogeographic regions (colour coded in Fig. 1).

The discrete data set derives from a previous cladistic analysis of the genus (Ojanguren-Affilastro & Ramirez, 2009), which in turn was partially based on characters defined by Prendini (2000, 2003) and Ochoa (2004) to assess phylogenetic relationships among genera of the family Bothriuridae and the superfamily Scorpionoidea.

The data set was reduced to 79 characters by eliminating those that turned invariant under the present taxonomic sampling. The resulting discrete matrix was composed of 46 binary and 33 multistate characters; of the latter, roughly half (45%) were treated as ordered. The use of a taxonomically subsampled matrix, which furthermore consists of characters originally defined to be useful at a higher taxonomic level, led to the

incorporation of a sizable amount of autapomorphies: 22% of characters contain at least one state present in only one species.

On the other hand, a traditional morphometric matrix comprising 68 characters was newly generated for this study. Values were mined from taxonomic descriptions or obtained from museum specimens and comprise mostly log-transformed ratios between linear measurements. This approach was efficient at reducing allometric scaling of variables, with none of the resulting continuous characters showing a significant correlation with total body length (Pearson correlation, all $P > 0.05$ after performing independent Hochberg corrections for male and female data subsets). The relative coverage of body regions was similar across data sets (Fig. 1, inset), allowing us to treat them as competing attempts to represent the overall morphological disparity within the clade. Both data sets were scored for 39 species of *Brachistosternus*, encompassing 90% of all described species of the genus (Kovářík & Ojanguren-Affilastro, 2013). The two data sets are available as Appendices S1 and S2. All subsequent analyses were performed in the R environment (R Core Team, 2015).

Phylomorphospaces and pattern difference

Both morphological data sets were transformed into distance matrices using Gower's coefficient (Gower, 1971a), with function `daisy` of package 'cluster' (Maechler et al., 2015). This distance metric is very common in macroevolutionary studies given its capacity to accommodate missing data (which in our case added up to < 4% of either data set), as well as due to its flexibility to deal with binary, multistate and continuous variables (Lloyd, 2016). Briefly, Gower's coefficient is calculated using the following equation:

$$S_{ij} = \frac{\sum_{k=1}^n s_{ijk} \delta_{ijk}}{\sum_{k=1}^n \delta_{ijk}} \quad (1)$$

where S_{ij} , the overall distance between taxa i and j , is the average score of the individual distances (s_{ijk}) for each character k (with n being the total number of characters) taken over all possible comparisons. Whenever a comparison is possible, $\delta_{ijk} = 1$; whenever there is a missing entry in either one or both of the taxa $\delta_{ijk} = 0$. For binary and unordered multistate characters, $s_{ijk} = 1$ if character states match for both taxa and is 0 otherwise. For ordered and continuous characters, s_{ijk} is the absolute distance between character states, divided by the total range of the variable (Gower, 1971a). In this formulation of Gower's coefficient, S_{ij} will vary between 0 and 1, given that the same is true for all s_{ijk} . In the presence of missing values, Gower's coefficient has been shown to preserve more of the original

distances between taxa than alternative metrics (Lloyd, 2016), as well as to outperform most data estimation methods (Brown et al., 2012). Furthermore, by accommodating all types of characters, Gower's coefficient allowed us to transform both data sets into distance matrices that can be subsequently tackled using the same methodologies.

Principal coordinate (PCO) ordination was performed on the resulting distance matrices (Gower, 1966) using the `cmdscale` function with option `add = TRUE` in order to employ Cailliez (1983) solution to avoid negative eigenvalues. In order to standardize the dimensionality of the study for downstream analyses, Cattell's scree test (Cattell, 1966) was used to determine the retention of the first five dimensions in both cases (as in Hopkins, 2016), accounting for 53% and 67% of total variance for the morphometric and discrete data sets, respectively. Percentage of variance summarized by each PCO, as well as the type of characters that most contribute to the variance captured by them, can be found in Table S1. Morphological disparity, as portrayed by both data sets, was explored using phylomorphospaces (Sidlauskas, 2008) with package 'phytools' (Revell, 2012).

To better evaluate the degree to which the morphospaces derived from discrete and morphometric characters differed from one another, the Procrustes sum of squares (PSS, i.e. the total sum of squared residuals after performing Procrustean superposition, which eliminates differences due to translation, rotation and scaling of the configurations) (Gower, 1971b) was used as a measure of pattern dissimilarity. Procrustean superposition and PSS calculation were performed using package 'vegan' (Oksanen et al., 2016). Furthermore, the obtained PSS value was compared with two null distributions to test different hypothesis. First, we analyzed whether the *a priori* subdivision of characters into discrete and morphometric matrices resulted in significantly different morphospaces. To do this, we applied a permutation test, randomly reassigning characters among matrices. 1000 pseudoreplicated pairs of matrices (of same size as the original ones) were built in such way, and the PSS metric was calculated for each pair after building a five-dimensional morphospace as explained above. The proportion of permuted matrices rendering higher PSS values than the observed one was taken as an estimate of the significance of the test. This procedure is analogous to the one undertaken to evaluate modularity in geometric morphometric data sets (RV coefficient; Klingenberg, 2009) and incongruence among data sources in phylogenetic systematics (ILD test; Farris et al., 1995).

Next, we tested whether the observed degree of pattern difference could be replicated by simulating character evolution in the phylogeny. Once again, 1000 continuous and discrete matrices of same size as the originals were simulated using function `sim.char` in

package ‘geiger’ (Harmon *et al.*, 2008), and compared using the same methodologies (i.e. transformation to distance matrices using Gower’s coefficient, use of PCO ordination and calculation of the PSS on the five-dimensional morphospaces). Parameters for the simulations were obtained from the observed data sets: continuous matrices were evolved under a single Brownian motion model (BM) using the observed evolutionary variance–covariance matrix; discrete traits were modelled using continuous time Markov models, with transition matrices (Q) and maximum-likelihood root states obtained by randomly sampling characters from the discrete data set. Before calculating these parameters, missing values were estimated using ancestral state reconstruction methods with package ‘picante’ (Kembel *et al.*, 2010), an approach that has been shown to be superior to excluding taxa with missing data from morphometric studies (Arbour & Brown, 2014). R code to replicate the permutation and simulation tests is available as Appendix S3, alongside a more detailed discussion of its implementation.

Analysis of disparity

Given that different proxies of disparity often capture different patterns (Ciampaglio *et al.*, 2001), we estimated present-day disparity for the five defined subclades using sum of ranges (SOR), sum of variances (SOV) and mean squared Euclidean distances. To assess the temporal dynamics of disparity, a DTT plot (Harmon *et al.*, 2003) was built for each data matrix. Temporal trajectories of disparity were compared qualitatively with 10 000 replicates of character evolution under BM (Slater *et al.*, 2010). To estimate whether trajectories of disparity through time significantly deviated from the null expectation of a random walk, we calculated the morphological disparity index (MDI) as the area contained between a given trajectory and the median of the BM simulations. A two-tailed test was used to assess the level of significance.

Macroevolutionary model

We performed multivariate maximum-likelihood fits for all dimensions in each partition to four different models of trait evolution, using function transformPhylo.ML of package ‘motmot’ (Thomas & Freckleton, 2012). This approach should alleviate (to some extent) the biases found when fitting individually the first few dimensions of an ordination procedure (Uyeda *et al.*, 2015). Models tested were BM, Ornstein-Uhlenbeck (OU) (Hansen, 1997), and those incorporating the tree stretching parameters δ and κ (Pagel, 1997, 1999; O’Meara, 2012), allowing us to assess a diversity of macroevolutionary scenarios. The fit of a λ model was also explored, but it was excluded from the model selection procedure because it was considered to be redundant in

the context of the other models tested. Models were compared using Akaike information criterion weights (wtAIC).

Correlation metrics and phylogenetic structure

In order to replicate the approach favoured by previous studies, we calculated Spearman’s correlation coefficient for the disparity estimates of the five subclades in the genus according to both data sets, as well as for 10 000 randomly generated groups of 5, 10 and 20 species sampled without replacement. We also employed a Mantel correlation test between both distance matrices and a Procrustes superposition test (PROTEST) between both resulting morphospaces (10^4 permutations in both cases). None of these measures take into account expected levels of covariance due to common phylogenetic history (Anderson & Friedman, 2012; Lloyd, 2016) and may therefore be prone to find significant correlations even when substantial underlying differences are present. We attempted to detect shared phylogenetic structure in both data sets using: (i) an analysis of correlation between phylogenetic independent contrasts (PICs) derived from discrete and morphometric PCOs (calculated using package ‘ape’; Paradis *et al.*, 2004), as well as (ii) between parsimony branch lengths, calculated as the number of unambiguous synapomorphies per branch derived from both partitions, using TNT (Goloboff & Catalano, 2016). Finally, (iii) phylogenetic signal per PCO was calculated using Blomberg’s K (Blomberg *et al.*, 2003), and its significance tested using a randomization test with 1000 replicates.

Results

The patterns of morphological disparity portrayed by the two data sets are strikingly different (Fig. 2). Discrete characters reveal that most main clades within the genus also correspond to morphologically defined clusters, as well as suggesting that a substantial fraction of overall morphological evolution occurred in the branches subtending these clades (although a few prominent cases of long terminal branches can also be seen). On the other hand, very little phylogenetic structure is seen among morphometric characters, with high levels of convergence between members of all groups and a general pattern of very short internal branches and long terminal ones.

A permutation test allowed us to confirm that both morphospaces are the most different characterizations of overall disparity that can be obtained from the scored characters (Fig. 3, $P \leq 0.001$). In fact, a very small fraction of randomly permuted matrices showed PSS values even close to the one observed when comparing the morphospaces derived from morphometric and discrete matrices. To test whether this result was sensitive to

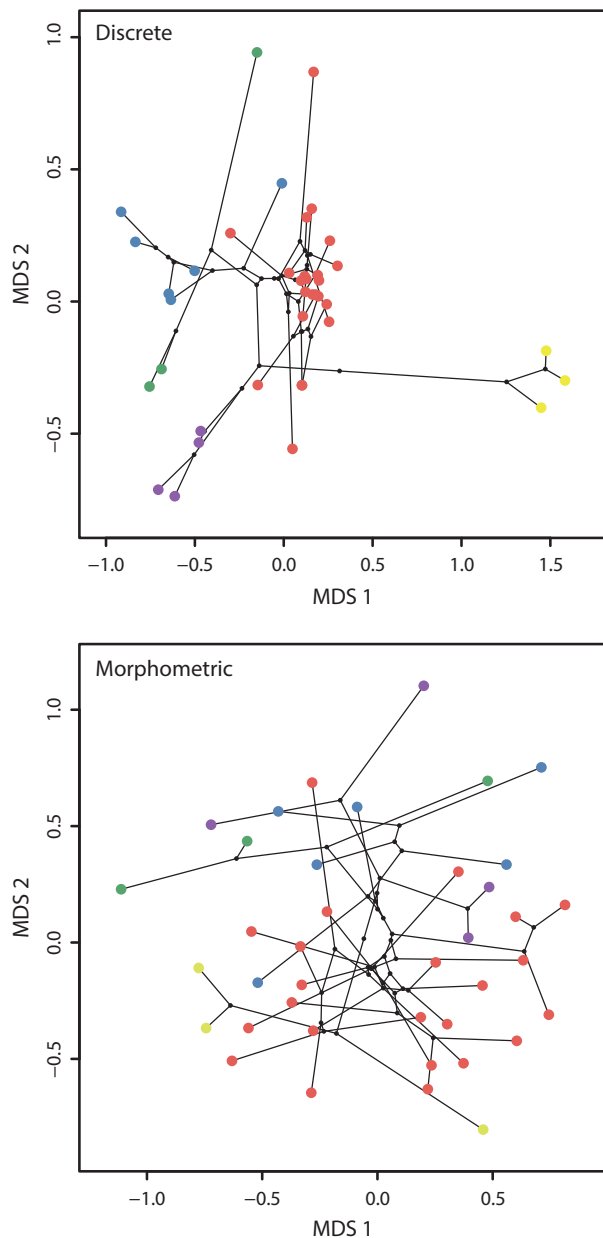


Fig. 2 Morphospaces derived from the discrete (top) and morphometric (bottom) character-taxon matrices. Both graphs are a visual summary of the pattern captured by each data set's five PCOs, obtained using metric multidimensional scaling (MDS). Colours correspond to the five subclades defined in Fig. 1.

the particular way in which Gower's coefficient measures differences among different types of characters, the same procedure was repeated considering all multivariate characters as unordered, treating all as ordered (essentially modifying the way differences are weighted), as well as using a different distance metric, generalized Euclidean distance (see Lloyd, 2016 for

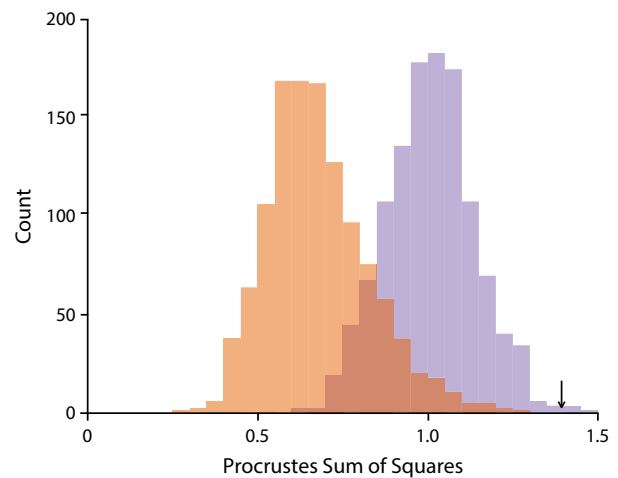


Fig. 3 Difference in overall pattern between the morphometric and discrete morphospaces (arrow), compared with null distributions. Histogram in orange corresponds to the permutation test, histogram in purple to the simulations of random character evolution in the phylogeny.

details). Results were the same in the three cases (see Fig. S1), supporting the conclusion that the subdivision into morphometric and discrete data sets results in patterns of morphospace occupation that are significantly different than those of random subdivisions. Even more, such degree of pattern difference is greater than expected under simulations of random character evolution in the phylogeny ($P = 0.004$). Further tests revealed that this conclusion is robust to whether continuous traits are evolved under independent BMs (i.e. with no covariance), to the treatment of a fraction of simulated multivariate characters as ordered and to the estimation of parameters using only taxa with no missing values (see Fig. S1). Under the null expectation provided by simulating random walks in the phylogeny, the probability of obtaining two morphospaces as contrasting as those in Fig. 2 is negligible. Such strong systematic difference may therefore be the result of different modes of evolution underlying the two data sets.

We estimated several measures of disparity (according to both data sets) for the five previously defined clades. When employing both SOV and SOR, the two partitions showed somewhat similar relative patterns of disparity between clades (Fig. 4a,b). However, an unexpected result is that morphometric traits seem to reveal systematically higher amounts of disparity within clades (one-way ANOVAS, SOV: $F_{1,8} = 9.89$, $P = 0.01$; SOR: $F_{1,8} = 7.77$, $P = 0.02$). In fact, if overall variance is partitioned into a within and between clade sum of squares, opposite patterns emerge from each data set: according to the discrete characters, 73% of total variance occurs between clades, whereas the morphometric

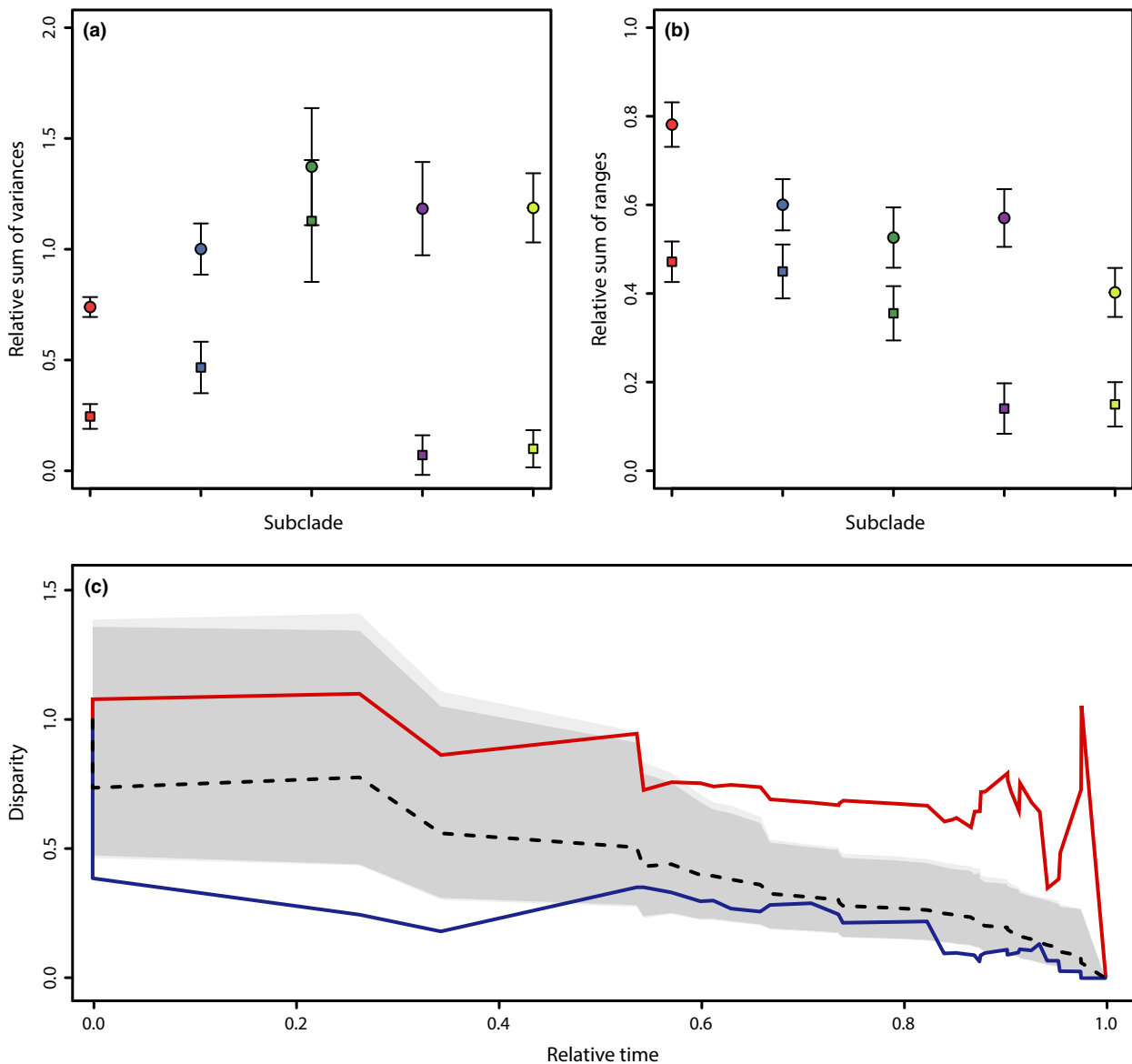


Fig. 4 Morphological disparity among *Brachistosternus* scorpions. Present-day subclade disparity as captured using (a) SOV and (b) SOR. Values are standardized by dividing subclade values by the total disparity of the genus. Error bars were calculated by bootstrapping the original data sets and represent ± 1 standard deviation. Circles correspond to values obtained for the morphometric data set, squares to the ones for the discrete data set. Colours are as in Fig. 1. (c) Disparity through time, showing the trajectory derived from the morphometric (top, red) and discrete (bottom, blue) partitions. The broken line corresponds to the median of 10 000 simulations of character evolution under BM, and grey areas denote the 95% CI of each run.

data set assigns 79% of total variance to differences within clades. As the clades being addressed represent also most of the earliest events of divergence within the genus, it seems likely that the higher amount of disparity between clades according to the morphometric data set – as captured by all three metrics – is actually the effect of partitions showing different patterns of disparity through time. A DTT plot confirmed this: the two data sets offer opposite views on the way

morphological disparity behaved through the evolutionary history of the clade (Fig. 4c). According to the continuous data set, subclade disparity remained higher than expected under a BM model during the entire history of the genus, resulting in an MDI of 0.354. On the other hand, according to the discrete data set, it dropped to very low values early in the radiation of the group, remaining relatively low (< 50% of total clade disparity) until the present (MDI = -0.232). Both

Table 1 Comparison of different models of macroevolution using a simultaneous fit for all dimensions in both data matrices. Models are ranked according to their wtAIC for the morphometric data set.

Model	Morphometric data			Discrete data		
	log-likelihood	AIC	wtAIC	log-likelihood	AIC	wtAIC
δ	204.60	-403.19	0.994	260.72	-515.43	1.000
OU	200.46	-392.92	0.006	241.55	-475.10	0.000
κ	190.49	-374.97	0.000	245.69	-485.37	0.000
BM	186.15	-368.30	0.000	241.55	-479.10	0.000

curves imply MDI values that lie outside of the 95% confidence interval (CI) based on simulated trajectories under BM ($P = 0.012$ and 0.002 for the morphometric and discrete partitions, respectively).

Regarding evolutionary mode, both data sets strongly rejected the hypothesis that trait evolution is compatible with a BM model, while at the same time providing strong support for a δ model (Table 1). This model raises node depths to a power (δ), providing a test of whether there is evidence of changes in the rate of evolution through time (Pagel, 1997). However, the maximum-likelihood estimate for the δ parameter under both partitions suggests contrasting processes of macroevolution. In the case of the continuous data set, we obtained a δ parameter of 8.597 (95% CI = 4.768–18.670), implying accelerating rates of evolution towards the present. This result suggests that the main process governing the morphological evolution of *Brachistoternus* is the acquisition of species-specific adaptations. On the contrary, a δ parameter of 0.086 (95% CI = 0.029–0.209), as supported by the discrete partition, is indicative of a sharp reduction in evolutionary rate, as often is the case with ‘early bursts’ during adaptive radiations.

Despite the profound differences described above, most of the traditional tests used to determine whether different data sets portray similar pictures of disparity detected strong correlations among the morphometric and discrete data sets. Distance matrices were shown to be significantly correlated (Mantel test Pearson’s $r = 0.31$, $P = 0.0004$), as were the relative positions of species in associated morphospaces (Procrustes statistic = 0.45, $P = 0.0003$). Although measures of disparity per clade (Fig. 4a,b) were not significantly correlated (SOV: $t_3 = 0.92$, $P = 0.21$; SOR: $t_3 = 1.72$, $P = 0.09$), the degree of correlation was high in both cases ($r = 0.47$ and 0.70 , respectively), and the lack of significance is most likely a result of sample size. Accordingly, correlation among disparity estimates for random generated subsamples was highly significant (all $P < 10^{-16}$; see Fig. S2 for further detail).

Concerns about the sensitivity of these tests to common phylogenetic structure have been tackled testing

for correlation between PICs derived from both data sets (Anderson & Friedman, 2012). In our case, this test revealed three significant results (12%), although with 25 correlations to perform, none is significant after adjusting for multiple comparisons. Nonetheless, parsimony branch lengths derived from both partitions are significantly correlated ($\rho = 0.39$, $P = 0.0002$), proving that a common phylogenetic structure can indeed be found in both data sets. This result is corroborated by calculating the significance of Blomberg’s K statistic: four discrete and one morphometric PCO showed significant phylogenetic signal (all $P < 0.01$), with other two morphometric dimensions showing marginally nonsignificant results (both $P = 0.06$).

Discussion

Morphological disparity, along with taxonomic diversity, is among the key aspects characterizing biological evolution (Gould, 1990). Understanding their dynamic through time can provide unparalleled insights into evolutionary success and its relationship with ecological opportunity, selective pressures, constraints, biotic interactions and environmental conditions (Gould *et al.*, 1977; Foote, 1992; Wills *et al.*, 1994; Harmon *et al.*, 2003). A lot of effort has therefore been directed at assessing various methodological aspects of their quantification.

We focused on the impact that alternative characterizations of overall morphology, as derived from different character types, have on the inferred macroevolutionary history of a scorpion genus. By employing for the first time an independently derived, time-calibrated phylogeny for the group under study, we were able to tackle this question by incorporating new methods, as well as extending our analyses beyond the comparison of measures of clade disparity.

Where previous studies have systematically found congruence (Villier & Eble, 2004; Anderson & Friedman, 2012; Foth *et al.*, 2012; Hetherington *et al.*, 2015), we found strong evidence of disagreement. Patterns of morphospace occupation, as derived from a discrete and a morphometric data set, were significantly different under a permutation test, thus revealing that the observed variables cannot be partitioned in a way that results in larger differences. Furthermore, the degree of pattern dissimilarity could not be replicated by simulating the evolution of discrete and continuous characters in the phylogeny of the group (Fig. 3), confirming that they are significantly more different than expected by chance alone.

Further analyses revealed the reason for this discrepancy. Both model fitting and DTT plots showed that the disparity captured by discrete characters was divided among subclades early in the history of the group. According to this data set, the genus experienced an ‘early burst’ of morphological diversification, the

signature of an adaptive radiation. The main morphological aspects that it captures, such as differential patterns of setation, evolved during the divergence of the main subclades of the genus, followed by a significant drop in their rate of evolution. On the contrary, the main aspects of morphometric variation in the group, such as the size and shape of the tail segments, underwent an increase in evolutionary rate through time. The resulting pattern is dominated by species-specific adaptations that have overridden most historical signal and resulted in widely convergent morphologies.

The reason behind the difference between our results and those of previous analysis could be ascribed to peculiarities in the evolutionary history of our case study, or to derive from our use of a different methodological approach. We lean towards the second interpretation, mainly because tests of correlation, the methods used so far to detect congruence in patterns derived from different data sets, do not take into account the expected presence of common phylogenetic structure. We confirm this by demonstrating that, in spite of the stark contrast already described, these tests do find high levels of correlation between our data matrices. Furthermore, several lines of evidence confirm the presence of phylogenetic structure among them. We conclude that the methods previously used are inappropriate to assess the congruence of macroevolutionary patterns, being susceptible to finding significant levels of correlation even in the face of strong differences. Methods that incorporate a phylogenetic framework, and are therefore able to account for expected levels of covariance (such as the ones applied above), should be preferred.

The degree to which the patterns and processes captured by both data sets are the result of intrinsic biases in each of them, or otherwise faithfully represent historical processes in this clade of scorpions, is difficult to assess. Discrete data sets, especially those that derive from previous cladistic analyses, are generally biased towards obviating autapomorphic characters, possibly leading to an overestimation of the relative amount of morphological evolution close to the base of the tree. On the other hand, autapomorphies cannot be easily filtered when compiling morphometric data matrices, and we might therefore expect this type of characters to capture relatively more change at the terminal branches of a phylogeny, as is the case in our results. Nonetheless, we believe that there is evidence suggesting that our results also capture biologically relevant phenomena. For example, a significant relationship is found when the morphometric PCO 2 is regressed on a psammophyly index ($F_{1,37} = 15.45$, Bonferroni adjusted $P = 0.002$), representing the degree of adaptation of each species for life on sandy substrates (calculated in Ojanguren-Affilastro *et al.*, 2016; see Prendini, 2001). This increase in specialization has originated multiple independent times, either among individual species or

pairs of species (Ojanguren-Affilastro *et al.*, 2016), and may be one of the reasons why morphometric variables capture an increase in evolutionary rate towards the present. On the other hand, the discrete data set does contain a sizable proportion of autapomorphies (see Materials and methods). For example, 35% of the total number of parsimony steps obtained by optimizing the discrete partition in the molecular tree occur at terminal branches, although this includes a mixture of both autapomorphies and instances of convergence. Likewise, the five longest branches of the tree are all terminal ones (some of which are also captured in the morphospace of Fig. 2). It is in fact likely that the 'early burst' recovered in the analysis of discrete characters, mainly involving an early differentiation among patterns of setae and trichobothria, as well as among configurations of the male reproductive organ, did take place during the early evolution of the group. These traits are not amenable of being coded as continuous variables, and were therefore excluded from the morphometric data matrix, possibly contributing to the lack of signal for early morphological evolution in this data set.

Although some degree of differential character sampling may underlie the differences in the dynamics inferred from different character types, it is also possible that the evidence for 'early bursts' simply has a higher chance of surviving among discrete characters than among morphometric ones. This has strong consequences for the study of macroevolutionary dynamics and yet finds support on a very simple observation: morphological differences among closely related species are generally quantitative in nature (MacLeod & Forey, 2002), but the phylogenetic structure of these characters is more or less rapidly erased by homoplasy (discussed by Klingenberg & Gidaszewski, 2010 and references therein). Although 'early bursts' have been detected with the use of morphometric data (e.g. Mahler *et al.*, 2010; Slater *et al.*, 2010; Weir & Mursleen, 2012; López-Fernández *et al.*, 2013), such studies have generally targeted clades known to have undergone adaptive radiations (*sensu* Schluter, 2000), as well as focused exclusively on traits of established ecological significance. When sought across a broader sample of taxa and morphometric characteristics, evidence for 'early bursts' was found to be surprisingly rare (Harmon *et al.*, 2010). For the vast majority of traits that are not undergoing strong divergent evolution, it seems logical to expect a pattern more consistent with a 'late burst' of morphological disparity.

Likewise, the finding that two equally comprehensive morphological data sets can support opposite macroevolutionary histories for the same group of organisms indicates that we are still far from being able to confidently quantify an 'overall morphological disparity' in the sense of Foote (1993) and Foote (1997). Studies addressing the evolution of entire body plans may

therefore be the ones most susceptible to the effects of alternative character selection, given that the increase in size of the character space from where to sample variables will likely result in less overlap in the variance being captured by different types of characters. We speculate that the discrepancy between the lack of support for early bursts in the comparative literature (Harmou *et al.*, 2010), and its apparent high frequency in the fossil record (Foote, 1994; Lloyd *et al.*, 2012; Hughes *et al.*, 2013) may simply reflect the differences in focus among two fields. Evolutionary biologists use morphometric variables to describe ecologically important characters, whereas palaeontologists use discrete variables to quantify overall morphological disparity.

Finally, it is specifically the presence of a common phylogenetic structure among otherwise widely differing morphological characterizations that likely results in similar patterns of disparity among different subclades, such as shown in Fig. 4a,b. The choice of character type will probably have little effect on studies focusing on this kind of macroevolutionary pattern. All other type of analyses, from those investigating the patterns of morphospace occupation, to those analysing the mode and tempo of evolution, will obtain results that are likely to be as much a result of clade history, as they are of character history.

Acknowledgments

The authors would like to thank Michael J. Donoghue and Pincelli M. Hull for valuable feedback and discussion, as well as Chris Klingenberg, Graham Slater and two anonymous reviewers whose comments significantly improved the manuscript. The Willi Hennig Society is acknowledged for making the software TNT freely available. All authors have seen and agreed on the final version of the manuscript and have no conflict of interest to declare. N.M.K is a graduate student at Yale University, supported by a University Fellowship. This work was funded by research grants PICT 2011-1007 from ANPCyT and PIP 2012-0943 from CONICET, both to M.J.R.

References

- Adams, D.C., Rohlf, F.J. & Slice, D.E. 2004. Geometric morphometrics: ten years of progress following the 'revolution'. *Ital. J. Zool.* **71**: 5–16.
- Anderson, P.S. & Friedman, M. 2012. Using cladistic characters to predict functional variety: experiments using early gnathostomes. *J. Vertebr. Paleontol.* **32**: 1254–1270.
- Arbour, J.H. & Brown, C.M. 2014. Incomplete specimens in geometric morphometric analyses. *Methods Ecol. Evol.* **5**: 16–26.
- Aristide, L., Dos Reis, S.F., Machado, A.C., Lima, I., Lopes, R.T. & Perez, S.I. 2016. Brain shape convergence in the adaptive radiation of New World monkeys. *Proc. Natl. Acad. Sci. USA* **113**: 2158–2163.
- Blomberg, S.P., Garland, T. & Ives, A.R. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**: 717–745.
- Brown, C.L., Arbour, J.H. & Jackson, D.A. 2012. Testing the effect of missing data estimation and distribution in morphometric multivariate data analyses. *Syst. Biol.* **61**: 941–954.
- Cailliez, F. 1983. The analytical solution of the additive constant problem. *Psychometrika* **48**: 305–308.
- Cattell, R.B. 1966. The scree test for the number of factors. *Multivar. Behav. Res.* **1**: 245–276.
- Ceccarelli, F.S., Ojanguren-Affilastro, A.A., Ramírez, M.J., Ochoa, J.A., Mattoni, C.I. & Prendini, L. 2016. Andean uplift drives diversification of the bothriurid scorpion genus *Brachistoternus*. *J. Biogeogr.* **43**: 1942–1954.
- Ciampaglio, C.N., Kemp, M. & McShea, D.W. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. *Paleobiology* **27**: 695–715.
- Erwin, D.H. 2007. Disparity: morphological pattern and developmental context. *Palaentology* **50**: 57–73.
- Farris, J.S., Källersjö, M., Kluge, A.G. & Bult, C. 1995. Constructing a significance test for incongruence. *Syst. Biol.* **44**: 570–572.
- Foote, M. 1992. Paleozoic record of morphological diversity in blastozoan echinoderms. *Proc. Natl. Acad. Sci. USA* **89**: 7325–7329.
- Foote, M. 1993. Contributions of individual taxa to overall morphological disparity. *Palaentology* **19**: 403–419.
- Foote, M. 1994. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. *Paleobiology* **20**: 320–344.
- Foote, M. 1997. The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.* **28**: 129–152.
- Foth, C., Brusatte, S. & Butler, R. 2012. Do different disparity proxies converge on a common signal? Insights from the cranial morphometrics and evolutionary history of Pterosauria (Diapsida: Archosauria). *J. Evol. Biol.* **25**: 904–915.
- García-Porta, J., Šmíd, J., Sol, D., Fasola, M. & Carranza, S. 2016. Testing the island effect on phenotypic diversification: insights from the *Hemidactylus* geckos of the Socotra Archipelago. *Sci. Rep.* **6**: 23729.
- Goloboff, P.A. & Catalano, S.A. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* **32**: 221–238.
- Goswami, A., Mannion, P.D. & Benton, M.J. 2016. Radiation and extinction: investigating clade dynamics in deep time. *Biol. J. Linn. Soc.* **118**: 6–12.
- Gould, S.J. 1990. *Wonderful Life: The Burgess Shale and the Nature of History*. Norton, New York, NY.
- Gould, S.J., Raup, D.M., Sepkoski, J.J., Schopf, T.J. & Simberloff, D.S. 1977. The shape of evolution: a comparison of real and random clades. *Paleobiology* **3**: 23–40.
- Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**: 325–338.
- Gower, J.C. 1971a. A general coefficient of similarity and some of its properties. *Biometrics* **27**: 857–871.
- Gower, J.C. 1971b. Statistical methods of comparing different multivariate analyses of the same data. In: *Mathematics in the Archaeological and Historical Sciences* (F.R. Hodson, D.G. Kendall & P.T. Tautu, eds), pp. 138–149. Edinburgh University Press, Scotland, UK.

- Hansen, T.F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**: 1341–1351.
- Harmon, L.J., Schulte, J.A. II, Larson, A. & Losos, J.B. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. *Science* **301**: 961–964.
- Harmon, L.J., Weir, J.T., Brock, C.D., Glor, R.E. & Challenger, W. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**: 129–131.
- Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W. *et al.* 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* **64**: 2385–2396.
- Heled, J. & Drummond, A.J. 2010. Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**: 570–580.
- Hetherington, A.J., Sherratt, E., Ruta, M., Wilkinson, M., Deline, B. & Donoghue, P.C.J. 2015. Do cladistic and morphometric data capture common patterns of morphological disparity? *Palaeontology* **58**: 393–399.
- Hopkins, M.J. 2016. Magnitude versus direction of change and the contribution of macroevolutionary trends to morphological disparity. *Biol. J. Linn. Soc.* **118**: 116–130.
- Hughes, M., Gerber, S. & Wills, M.A. 2013. Clades reach highest morphological disparity early in their evolution. *Proc. Natl. Acad. Sci. USA* **110**: 13875–13879.
- Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D. *et al.* 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.
- Klingenberg, C.P. 2009. Morphometric integration and modularity in configurations of landmarks: tools for evaluating *a priori* hypotheses. *Evol. Dev.* **11**: 405–421.
- Klingenberg, C.P. & Gidaszewski, N.A. 2010. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Syst. Biol.* **59**: 245–261.
- Kovařík, F. & Ojanguren-Affilastro, A.A. 2013. *Illustrated Catalog of Scorpions. Part II. Bothriuridae; Chaerilidae; Buthidae I, genera Compsobuthus, Hottentotta, Isometrus, Lychas, and Sasaniidotus*. Jakub Rolčík, Prague, Czech Republic.
- Lloyd, G.T. 2016. Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. *Biol. J. Linn. Soc.* **118**: 131–151.
- Lloyd, G.T., Wang, S.C. & Brusatte, S.L. 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi). *Evolution* **66**: 330–348.
- López-Fernández, H., Arbour, J.H., Winemiller, K. & Honeycutt, R.L. 2013. Testing for ancient adaptive radiations in neotropical cichlid fishes. *Evolution* **67**: 1321–1337.
- MacLeod, N. 2002. Phylogenetic signals in morphometric data. In: *Morphology, Shape and Phylogeny* (N. MacLeod & P.L. Forey, eds), pp. 100–138. Taylor & Francis Inc., New York, NY.
- MacLeod, N. & Forey, P.L. 2002. *Morphology, Shape and Phylogeny*. Taylor & Francis Inc., New York, NY.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. 2015. Cluster: cluster analysis basics and extensions. R package version. 2.0.3.
- Mahler, D.L., Revell, L.J., Glor, R.E. & Losos, J.B. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of greater antillean anoles. *Evolution* **64**: 2731–2745.
- Mahler, D.L., Ingram, T., Revell, L.J. & Losos, J.B. 2013. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* **341**: 292–295.
- Marcus, L.F. 1990. Traditional morphometrics. In: *Proceedings of the Michigan Morphometrics Workshop* (F.J. Rohlf & F.L. Bookstein, eds), pp. 77–122. University of Michigan Museum of Zoology, Ann Arbor, MI.
- Ochoa, J.A. 2004. Filogenia del género *Orobothiurus* y descripción de un nuevo género de Bothriuridae (Scorpiones). *Rev. Iber. Aracnol.* **9**: 43–73.
- Ojanguren-Affilastro, A.A. & Ramirez, M.J. 2009. Phylogenetic analysis of the scorpion genus *Brachistosternus* (Arachnida, Scorpiones, Bothriuridae). *Zool. Scr.* **38**: 183–198.
- Ojanguren-Affilastro, A.A., Mattoni, C.I., Ochoa, J.A., Ramírez, M.J., Ceccarelli, F.S. & Prendini, L. 2016. Phylogeny, species delimitation and convergence in the South American bothriurid scorpion genus *Brachistosternus* Pocock 1893: integrating morphology, nuclear and mitochondrial DNA. *Mol. Phylogenet. Evol.* **94**: 159–170.
- Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGinn, D. *et al.* 2016. vegan: community ecology package. R package version 2.4-1. <https://CRAN.R-project.org/package=vegan>.
- O’Meara, B.C. 2012. Evolutionary inferences from phylogenies: a review of methods. *Annu. Rev. Ecol. Evol. Syst.* **43**: 267–285.
- Pagel, M. 1997. Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**: 331–348.
- Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**: 877–884.
- Paradis, E., Claude, J. & Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289–290.
- Pennell, M.W. & Harmon, L.J. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* **1289**: 90–105.
- Prendini, L. 2000. Phylogeny and classification of the superfamily Scorpionoidea Latrielle 1802 (Chelicerata, Scorpiones): an exemplar approach. *Cladistics* **16**: 1–78.
- Prendini, L. 2001. Substratum specialization and speciation in southern African scorpions: the effect hypothesis revisited. Scorpions 2001. In: *Memoriam* (G.A. Polis, V. Fet & P.A. Selden, eds), pp. 113–138. British Arachnological Society, Burnham Beeches, Bucks, UK.
- Prendini, L. 2003. A new genus and species of bothriurid scorpion from the Brandberg Massif, Namibia, with a reanalysis of bothriurid phylogeny and a discussion of the phylogenetic position of *Lisposoma* Lawrence. *Syst. Entomol.* **28**: 149–172.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Revell, L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**: 217–223.
- Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.
- Sidlauskas, B. 2008. Continuous and arrested morphological diversification in sister clades of characiform fishes: a phylogenetic approach. *Evolution* **62**: 3135–3156.

- Slater, G.J. & Pennell, M.W. 2013. Robust regression and posterior predictive simulation increase power to detect early bursts of trait evolution. *Syst. Biol.* **63**: 293–308.
- Slater, G.J., Price, S.A., Santini, F. & Alfaro, M.E. 2010. Diversity versus disparity and the radiation of modern cetaceans. *Proc. Biol. Sci.* **277**: 3097–3104.
- Thomas, G.H. & Freckleton, R.P. 2012. motmot: models of trait macroevolution on trees. *Methods Ecol. Evol.* **3**: 145–151.
- Uyeda, J.C., Caetano, D.S. & Pennell, M.W. 2015. Comparative analysis of principal components can be misleading. *Syst. Biol.* **64**: 677–689.
- Villier, L. & Eble, G.J. 2004. Assessing the robustness of disparity estimates: the impact of morphometric scheme, temporal scale, and taxonomic level in spatangoid echinoids. *Paleobiology* **30**: 652–665.
- Weir, J.T. & Mursleen, S. 2012. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (Aves: Alcidae). *Evolution* **67**: 403–416.
- Wills, M.A., Briggs, D.E.G & Fortey, R.A. 1994. Disparity as an evolutionary index: a comparison of Cambrian and recent arthropods. *Paleobiology* **20**: 93–130.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article: **Figure S1** Sensitivity of permutation test and simulations to alternative settings.

Figure S2 Morphometric and discrete estimates of disparity for random subsamples of taxa.

Table S1 Percentage of variance and type of characters contributing to PCOs.

Appendix S1 Discrete matrix and character description.

Appendix S2 Morphometric matrix and character description.

Appendix S3 R code to replicate the permutation test and simulation analysis.

Received 11 July 2016; revised 5 February 2017; accepted 6 February 2017