CSIRO PUBLISHING
www.publish.csiro.au/journals/is *Invertebrate Systematics*, 2009, **23**, 515–529

A phylogenetic analysis for the South-east Asian mite harvestman family Stylocellidae (Opiliones : Cyphophthalmi) – a combined analysis using morphometric and molecular data

Ronald M. Clouse[A,C], *Benjamin L. de Bivort*[A,B] *and Gonzalo Giribet*[A]

[A]Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA.
[B]Rowland Institute at Harvard, 100 Edwin Land Boulevard, Cambridge, MA 02142, USA.
[C]Corresponding author. Email: clouse@fas.harvard.edu

Abstract. In an effort to place type specimens lacking molecular data into a phylogenetic framework ahead of a taxonomic revision, we used morphometric data, both alone and in combination with a molecular dataset, to generate phylogenetic hypotheses under the parsimony criterion for 107 members of the South-east Asian mite harvestman family Stylocellidae (Arachnida: Opiliones: Cyphophthalmi). For the morphometric analyses, we used undiscretised characters, analysed for independence and collapsed by principal components analysis (PCA) when dependent. Two challenges not previously encountered in the use of this method were (a) handling terminals with missing data, necessitated by the inclusion of old and damaged type specimens, and (b) controlling for extreme variation in size. Custom scripts for independence analysis were modified to accommodate missing data whereby placeholder numbers were used during PCA for missing measurements. Size was controlled in four ways: choosing characters that avoided misleading size information and were easily scaled; using only locally scaled measurements; adjusting ratios by *y*-intercepts; and collapsing dependent characters into one. These steps removed enough size information that miniaturised and large species, suspected from molecular and discrete morphological studies to be closely related, were closely placed using morphometric data alone. Both morphometric and combined analyses generated relationships that positioned type specimens in agreement with taxonomic expectations and our knowledge of the family from prior studies. The hypotheses generated here provide new direction in linking molecular analyses with established taxonomy in this large group of South-east Asian arachnids.

Introduction

The South-east Asian mite harvestmen comprise a diverse family, Stylocellidae, that extends from the eastern Himalayas in China and India to western New Guinea (Juberthie 1988; Rambla 1991; Shear 1993;Giribet 2000; Boyer *et al*. 2007; Clouse and Giribet 2007; Giribet *et al*. 2007). One of the largest groups in the opilionid suborder Cyphophthalmi, Stylocellidae is in urgent need of a taxonomic revision. A recent molecular phylogeny of the family has revealed the main clades (Clouse and Giribet 2010), but the positions of all type species for each postulated genus within that phylogeny remain pending. For this reason, we sought to conduct a combined molecular and morphological analysis, allowing the placement of specimens for which we have no molecular data. Such specimens include the type for what is currently the largest genus in the family, *Stylocellus sumatranus* Westwood, 1874.

A significant problem in the family is a large clade detected by molecular data composed of species from the Indo-Malay Archipelago. This clade lacks diagnostic morphological features, and discrete morphological characters have shown little promise for resolution. In molecular studies, the clade also lacks resolution among species groups found on various islands, perhaps owing to their peculiar history of rapid diversification on a large peninsula that has a dynamic geologic history (Clouse and Giribet 2010). For this reason, we turned to undiscretised morphometric data, which we have shown previously to contain phylogenetic information in Cyphophthalmi (de Bivort *et al*. 2010).

There has been significant debate about combining different data partitions in phylogenetic analyses (e.g. Kluge 1989; Bull *et al*. 1993; Miyamoto and Fitch 1995; Huelsenbeck *et al*. 1996*a*, 1996*b*), but such methods have been demonstrated as useful in Opiliones, including molecular–morphological combinations (Giribet *et al*. 1999, 2002; Boyer and Giribet 2007). Combining undiscretised (uncoded) morphometric and molecular data is far more novel. The only such analysis of which we are aware is that of Hardy *et al*. (2008), who combined 15 undiscretised morphometric characters with 138 discrete morphological characters and 6831 base pairs of aligned molecular data in a phylogenetic analysis of Cape reeds (Restionaceae). Their motivation was partly the same as ours here – the placement of species for which sequence data are lacking – and they found that the combined analyses increased resolution and support over trees generated from molecular or discrete morphological data alone. Also, analyses including taxa with only morphological data

10.1071/IS09044 1445-5226/09/060515

placed them reasonably. However, important methodological concerns need to be addressed when using morphometric data, especially when undiscretised, and here we also explore methods for detecting independence among characters, managing missing values, controlling for size information and differentially weighting data partitions.

Materials and methods

Independence analysis

We used a method developed earlier (de Bivort *et al*. 2010) whereby normalised character values were compared in a pairwise fashion ('independence analysis', IA) (Fig. 1), and sets of characters that were correlated without outliers (i.e. containing the same information) were collapsed by principal component analysis (PCA) and replaced with their first principal component (PC1). The degree to which outliers would be tolerated in pairwise comparisons was optimised. The resulting undiscretised datasets were analysed under the parsimony criterion in the program TNT (Goloboff *et al*. 2006, 2008), and the optimal dataset was determined by evaluating its success at retrieving suspected clades and tree-wide bootstrap support.

Missing data

Our first methodological concern in this study was handling missing data. Not only were most of our important type specimens missing all molecular data, but they also could not provide a complete set of morphological assessments. Some were badly damaged, such as the only known specimen of *Stylocellus sumatranus* (Fig. 2*A*, *B*). Others were known only from females, like *S. pocockii* Hansen & Sørensen, 1904. Cyphophthalmi are sexually dimorphic, and their taxonomy is based largely on male specimens, which have several important diagnostic features related to the adenostyle, gonostome, exocrine glands and anal region (Giribet and Boyer 2002). Likewise, many characters studied under the scanning electron microscope could not be examined for certain types. Thus we modified the methods described in de Bivort *et al*. (2010) to accommodate missing data, as described below.

When collapsing dependent character networks that contained missing data, we inserted a placeholder number for each missing value, which was the mean of the values of the other characters undergoing collapse. As an average value, such a placeholder will not skew the PCA. If more than 50% of all character values undergoing collapse were missing for a particular terminal, it was eliminated from the PCA, and its PC1 value was replaced with '?'

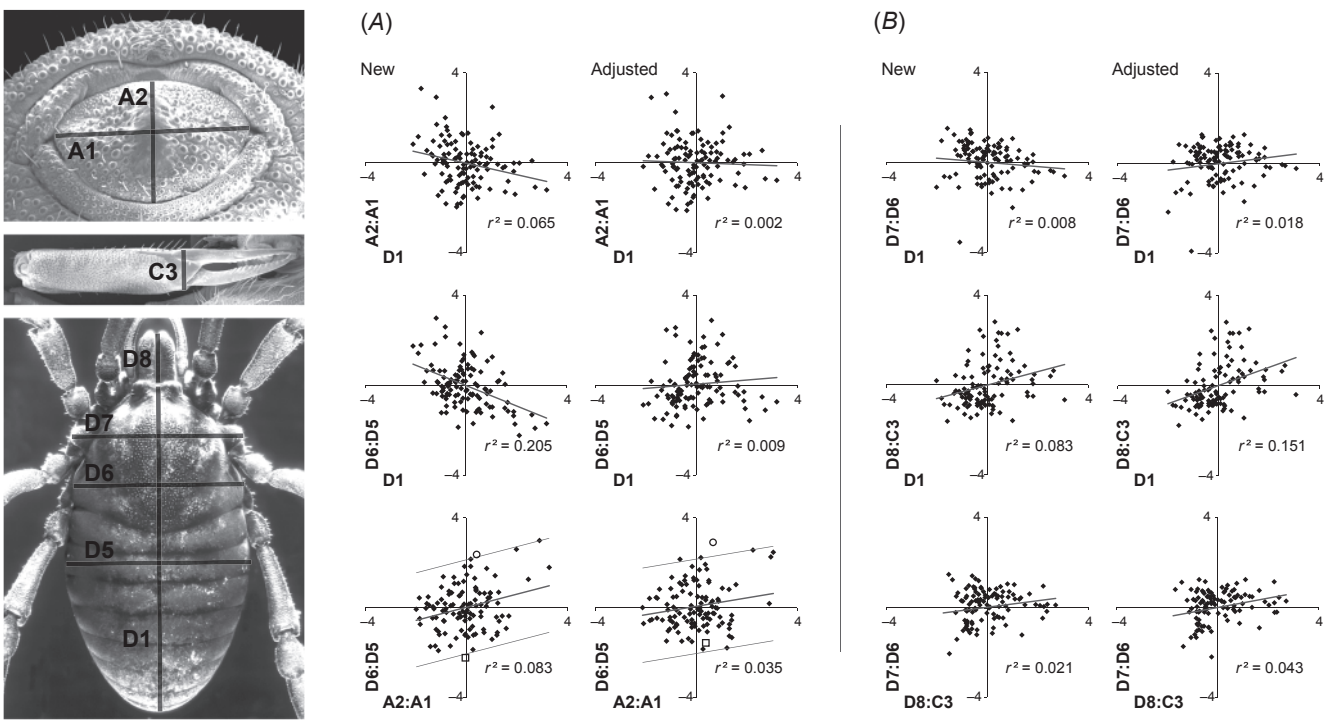

Fig. 1. The relationships between two body measurement ratios to body size (*A* and *B*, upper two rows) and to each other (bottom row), for the 'New' and 'Adjusted' datasets (left and right, respectively), for two different sets. Measurements are *z*-score-normalised and are illustrated on specimen photographs of the anal region (left, top), chelicerae (middle) and dorsal view (bottom). Both A2 : A1 and D6 : D5 are less correlated with body size (D1) in the 'Adjusted' dataset (*A*, upper two rows). For the 'Adjusted' dataset, this results in the two ratios being treated as independent across a larger range of cutoff values, for the specimen represented by an open circle moves farther from the trendline (*A*, bottom row). Also the most distant outlier, the specimen represented by an open square in the 'New' dataset, changes (*A*, bottom row). The thin lines in the bottom row are the all the same distance from the trendlines. For the ratio sets D7 : D6 and D8 : C3 (*B*), adjusting the data by the *y*-intercept strengthens the relationship between each ratio and body size as well as to each other. (The bottom row of *B* is missing the highly negative value for D7 : D6 seen in the top row, because that specimen was missing a value for D8 : C3.)

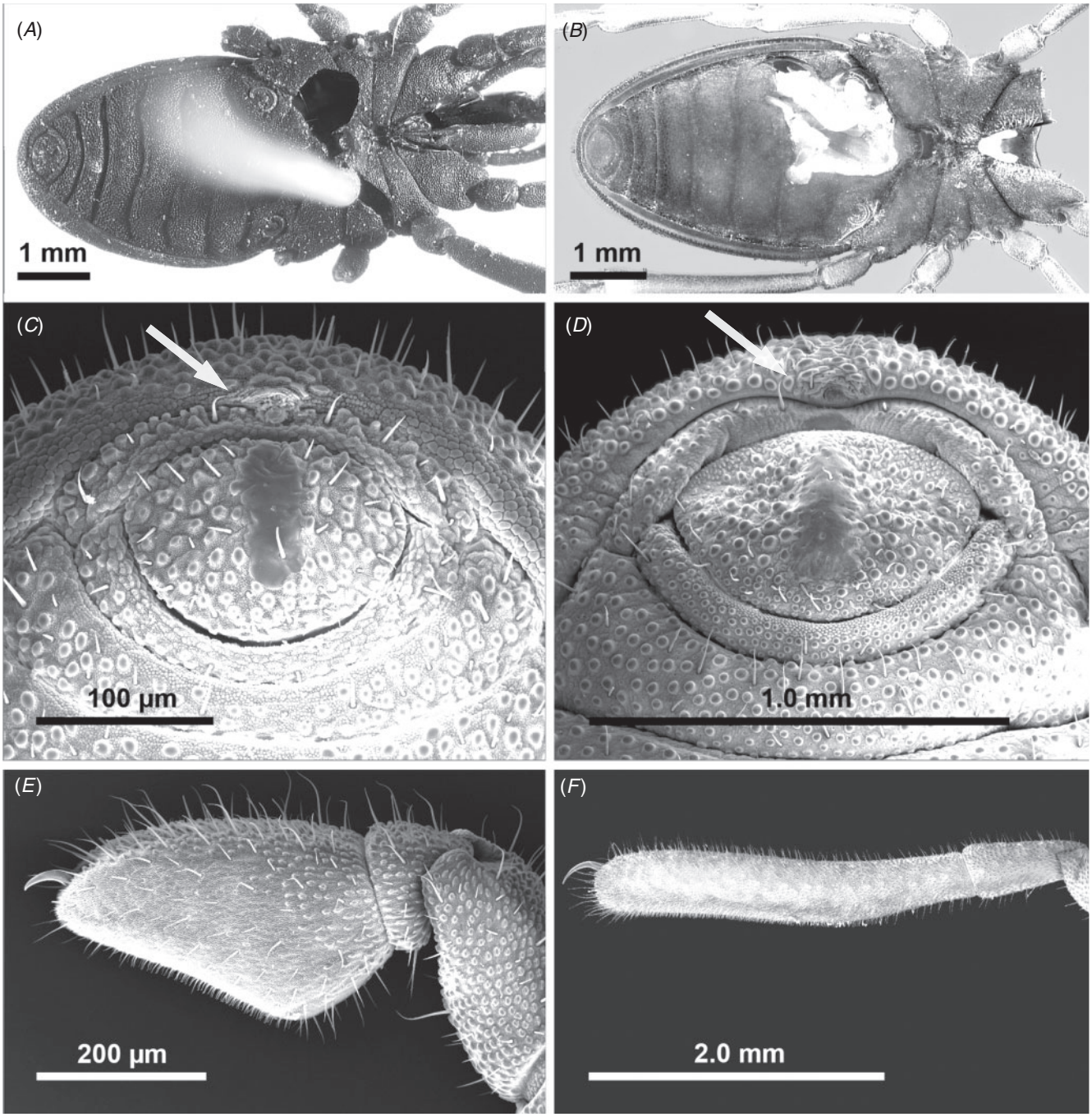

Fig. 2. Dorsal views of (*A*) the pinned holotype of *Stylocellus sumatranus* and (*B*) *S. gryllospecus*. The anal region of the small Sumatra sp. 13 and the large Borneo sp. 13 (*C* and *D*, respectively), showing the anal gland pore (arrows) and the smooth region on the anal plate. Tarsi I for the same species (*E* and *F*).

in the final dataset to be used in TNT. Before IA, the data were normalised by conversion into *z*-scores. They were also made positive by subtracting the smallest value in the dataset, since TNT cannot use negative character values; thus, the placeholder mean was not zero, as would be expected with normalised data. After IA, the data were normalised and made positive again, since PCs are not necessarily *z*-scores when calculated.

Size corrections and the construction of morphometric datasets

Our second analytic challenge was the influence of size variability across terminals. Stylocellidae contains the largest Cyphophthalmi known as well as several miniaturised species, especially in a Bornean-centred clade known from molecular

studies (called the 'Borneo+*Miopsalis*' clade in Clouse and Giribet 2007 and 'Clade B' in Clouse and Giribet 2010). Large and small species in this clade (with a 5-fold length difference) do share important morphological synapomorphies (Fig. 2*C*, *D*), but body size has clearly not been conserved and appears to correlate with particular morphometric differences (Fig. 2*E*, *F*). Using raw measures, then, especially if poorly chosen, could lead to a mass repetition of misleading size information throughout our dataset, resulting in phylogenetic hypotheses that were no more than sortings of terminals by size (Rae 2002). Nonetheless, size can carry phylogenetic information (Bookstein *et al*. 1985), and we successfully used raw measures in our previous analysis of the relatively evenly sized family Pettalidae (de Bivort *et al*. 2010). Here we attempted to control for size in four ways: selecting characters known to be free of spurious size information (such as those affected by troglomorphism) and easily scaled[1] by orthogonal or parallel measurements, using only ratios composed of nearby measurements (what we call here 'local scaling'), adjusting ratios to control for linearity and non-zero *y*-intercepts (Albrecht *et al*. 1993) and collapsing redundant characters into one through IA.

Correcting for body size started with the choice of raw measurements. We chose measurements that aligned with body axes and could thus be scaled by parallel measurements (illustrations of all measurements are available in an Accessory Publication on the *Invertebrate Systematics* website). For example, the width of the gonostome could be divided by the width of the prosoma, and the distance between sternite sutures by the length of the opisthosoma. In all, 93 raw measurements were taken for each terminal by examining dorsal, lateral, ventral and appendage photographs and scanning electron micrographs from our database. Twenty-four terminals had two specimens photographed, one terminal had three specimens and two terminals had four specimens; in any cases where multiple specimens were available for any measurement, all were taken and averaged for that terminal. Specimens for the species *Stylocellus sedgwicki* Shear, 1979, *S. thorellii* Hansen & Sørensen, 1904 and *S. weberii* Hansen & Sørensen, 1904, are currently unavailable, so they could only be measured from illustrations published with their original descriptions. *Miopsalis pulicaria* Thorell, 1890, could not be included owing to a lack of information and illustrations. Although the types for *M. pulicaria* are supposedly deposited in the Museo Civico di Storia Naturale (Genova), we have not been able to examine them. *Stylocellus spinifrons* Roewer, 1942 was not included because the description illustrations lack detail and a scale bar. Females were included when no males were available, although they are slightly larger than males in most species; their measurements for sexually dimorphic features, e.g. the genital opening (gonostome), were input as missing data.

Three morphometric datasets were created to pass through IA, PCA and tree-searching. The first dataset, 'Original', consisted of our initial collection of simple ratios that appeared to adequately correct for size without any adjustments. The second dataset, 'New', added the species *Stylocellus tarumpitao* Shear, 1993 from Palawan Island, Philippines, eliminated 13 ratios that appeared to add only noise to the analysis (i.e. were inferred to be highly homoplastic), added four new ratios from the chelicerae and gonostome and added a measure of the Rambla's organ for species in the genus *Meghalaya*. Rambla's organ is a patch of modified cuticle on the last article of the fourth walking leg of males (Schwendinger and Giribet 2005) that has been proposed to be homologous with the large depression in the same region in males of *Meghalaya* (Giribet *et al*. 2007); so for species that had this depression, a rough measure of it was made by taking half the height and length of the article. The third dataset, 'Adjusted', consisted of the 'New' dataset adjusted as per Albrecht *et al*. (1993); the relationship between the numerator and denominator of each ratio character was inspected, and all were found to have a positive relationship. For no ratio characters were component measurements better fit by a curvilinear model than a linear one, so only a *y*-intercept adjustment was made for each ratio by subtracting the *y*-intercept from the numerator. The final counts of morphometric characters and terminals for the 'Original' dataset were 60 characters for 117 terminals (106 from Stylocellidae), and for the 'New' and 'Adjusted' datasets were 51 characters for 118 terminals (107 from Stylocellidae). All morphometric data, in raw, normalised and collapsed form, are available in the Accessory Publication on the *Invertebrate Systematics* website.

Combination with molecular data and phylogenetic searches

For each of the three morphometric datasets, tree-searches were conducted by themselves and in combination with 6571 molecular characters. The molecular dataset was an implied alignment from Clouse and Giribet (2010) output by the phylogenetic program POY4, build 2885 (Varón *et al*. 2008), and it came from six molecular markers: two nuclear ribosomal (18S rRNA, 28S rRNA), two nuclear protein-encoding (histones H3 and H4), one mitochondrial ribosomal (16S rRNA) and one mitochondrial protein-encoding (cytochrome *c* oxidase subunit I). For all molecular analyses, the costs of transversions and insertions–deletions (indels) were set at two, with transitions at one, as was found earlier to be the optimal cost set for that dataset (Clouse and Giribet 2010). When the molecular data were added, so were an additional 21 outgroups and 18 Stylocellidae taxa that had molecular but no morphometric data, making the final terminal counts for the combined analyses 156 for the 'Original' dataset and 157 for the 'New' and 'Adjusted' datasets.

Tree searches using only morphometric data usually settled on the shortest tree quickly after starting with 100 Wagner trees and employing the sectorial, ratchet, drift and fuse functions as implemented in TNT (Goloboff *et al*. 2008), but searches on combined data were more involved. We started with 100 Wagner trees and used all searching options, fused the shortest trees from those searches with ones found using other datasets ('Original', 'New' or 'Adjusted') and different molecular transformation costs, and did another round of fusing with Wagner and random trees improved only with tree bisection and reconnection (TBR) branch swapping.

[1]Goloboff *et al.* (2006) used the term 'scaling' for a different type of correction.

We also did tree searches where the morphometric data were weighted so as to provide an approximately equal number of steps to the final tree, thereby having a comparable impact on the tree as the thousands of molecular characters. This weighting factor was calculated by finding the optimal tree when the morphometric data were not weighted and taking the ratio of the tree length with all the data optimised on the tree to the tree length with only the morphometric data optimised. Other variations consisted of excluding females and using a dataset where all raw measures were made into *y*-intercept-adjusted ratios over each terminal's measure of total body length.

Evaluating trees

The actual number of characters used in any tree search depended on the cutoff in the independence analysis, and the 'Original' morphometric dataset was used to explore the full range of cutoff values and their effect on the final datasets and resulting trees. Trees were evaluated in two ways: (1) by summing the bootstrap support for all nodes receiving support >50%; and (2) by averaging the average pairwise distances (measured as the number of separating branches) between terminals for certain clades and relationships found in previous phylogenetic analyses (Boyer *et al*. 2007; Clouse and Giribet 2007; Clouse and Giribet 2010) – a measure of how close to monophyletic these groups were. The latter was done by reformatting parenthetical trees from TNT such that all branch lengths were set to 1, importing them into APE (Analysis of Phylogenetics and Evolution; Paradis *et al*. 2004), and generating a matrix of pairwise distances using the cophenetic.phylo command. From this, average distances were calculated among members of the following groups:

(1) Each cyphophthalmid family with more than one representative (Pettalidae, Sironidae, Neogoveidae, and Stylocellidae);
(2) The stylocellid genera *Fangensis* and *Meghalaya* (Clade B);
(3) The large, non-*Fangensis* species from Borneo or species with anal gland pores;
(4) Species from Sulawesi plus *Stylocellus novaguinea* Clouse & Giribet, 2007;
(5) The miniaturised species on Sumatra with an anal gland pore (Sumatra sp. 13) plus the large, non-*Fangensis* species with anal gland pores;
(6) Each *Fangensis* species plus the outgroup species; and
(7) The residents of each large Indomalay island for which we had no prior evidence of multiple lineages: Java, Borneo, and Sulawesi.

The nature of the best IA cutoff value from this analysis was used to more quickly narrow the cutoff when analysing the 'New' and 'Adjusted' datasets.

Resampling support

Bootstrap and jackknife resampling supports were calculated for both morphometric and combined morphometric and molecular trees for all three datasets. Bootstrap support was calculated during the process of choosing an optimal IA cutoff point as well, but it was found during that process that support increased with a more thorough search of each pseudoreplicate, even when the number of pseudoreplicates was very low. Thus, when choosing the optimal IA cutoff, bootstrap support was calculated using 100 pseudoreplicates with five Wagner starting trees each, and employing the sectorial, ratchet, drift and fuse functions as implemented in TNT (Goloboff *et al*. 2008). After the optimal tree was found, support was calculated with an emphasis on searching each pseudoreplicate: for both the morphometric dataset and the combined morphometric and molecular datasets, resampling was calculated using 20 pseudoreplicates, 10 Wagner starting trees each and the default settings of sectorial, ratchet, drift and fuse functions. This search required ~12 h on a desktop computer for each combined dataset and for each type of resampling measure (bootstrap and jackknife). For bootstrap calculations, the probability of character selection was 50%, and for jackknifing, the probability of character removal was 36% (Farris *et al*. 1996).

As discussed in de Bivort *et al*. (2010), clades found using the morphometric data alone were not expected to have high resampling support values, since IA removes redundant information from morphometric datasets by design. Conversely, the molecular data are known to provide high support for the major clades as shown in the inset tree in Fig. 3; nodes receiving less than 50% bootstrap support (thin branches) are all with Clade C (Clouse and Giribet 2010). However, despite the high support using the molecular data alone, and despite its prominence in each dataset (several thousand more characters and about seven times the number of steps contributed to each tree over the morphometric data), we (Clouse and Giribet 2010) and Hardy *et al*. (2008) have found that the inclusion of terminals with a very small proportion of the data can considerably lower resampling support of otherwise reasonable phylogenies. The analyses in this study focus mainly on placing species that have no molecular data, and they constitute around one-fourth of our terminals. Thus, we would expect our morphometric-only phylogenies to have low support due to the removal of redundant characters, and the combined morphometric and molecular phylogenies to have low support due to the inclusion of terminals with significant amounts of missing data.

Results

In our analysis of the 'Original' dataset, the IA cutoff where we had the best recovery of suspected clades was on either side of a range where there was a convergent collapse of characters and an increase in bootstrap support, as we found in our study of the Pettalidae (de Bivort *et al*. 2010). Changes in character collapse, bootstrap support and clade recovery are illustrated as a function of IA cutoff in the Accessory Publication. The trees on either side of this range were very similar, with those from IA cutoff 1.48 (which resulted in 53 characters) finding more suspected relationships. For the 'New' and 'Adjusted' datasets, the optimal cutoff was chosen by first looking for a similar range of cutoff values and then tree-searching datasets from those cutoff values. For both datasets, the best trees were found around this range; if there was no tree that was clearly superior at recovering suspected relationships, the cutoff where the resulting tree had the highest bootstrap support was chosen.

The shortest trees found from each dataset were of the following lengths and number. From the 'Original' dataset,

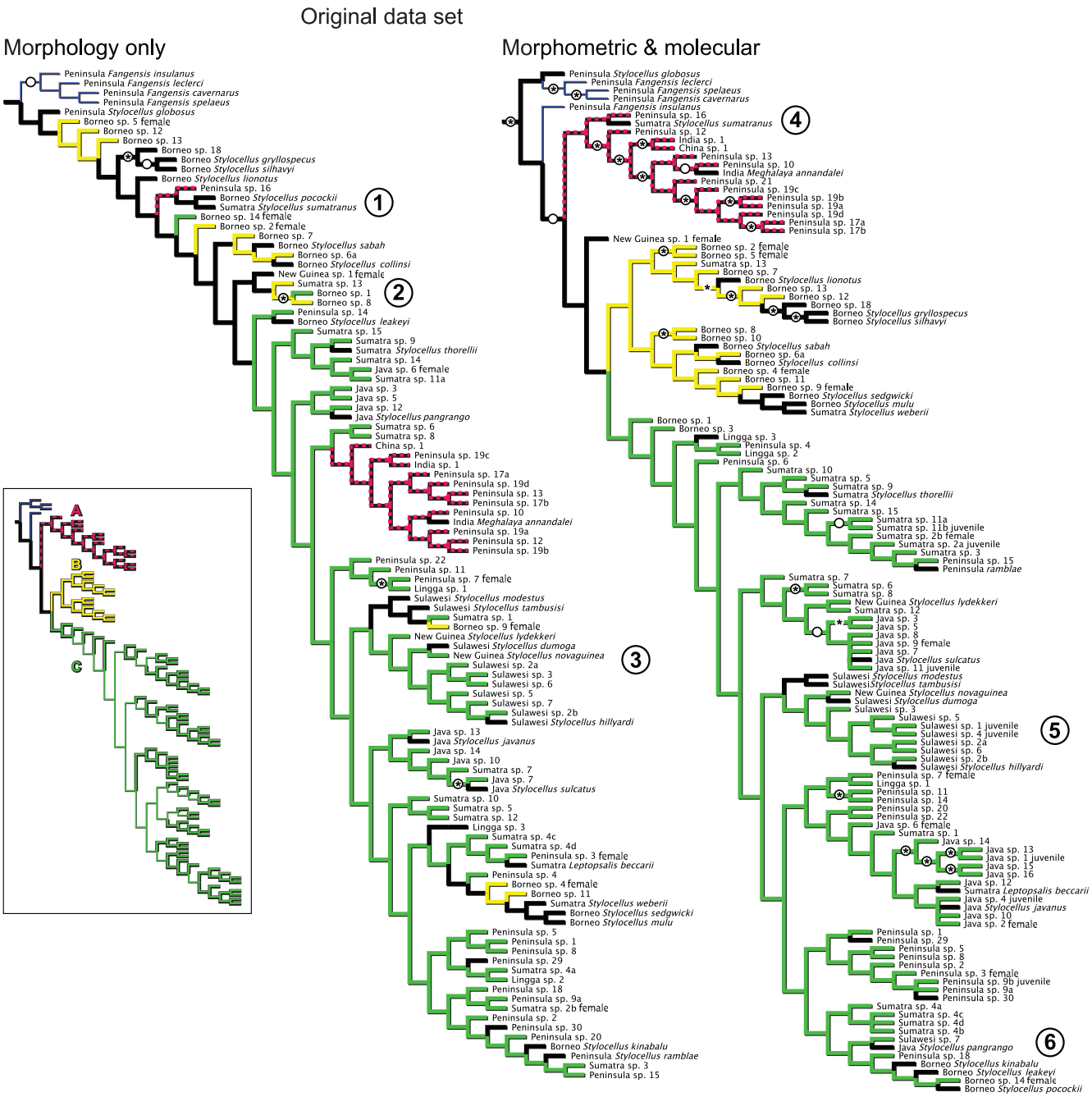

Fig. 3. Strict consensuses of the shortest trees found using the 'Original' morphometric dataset, showing Stylocellidae only (outgroups in Fig. 6). The analysis of only the morphometric data is on the left, and a combination of those data with the molecular data is on the right. Marked branches (blue, red, yellow and green in colour versions; thin, dashed, light grey and dark grey in greyscale versions) correspond to clades found in Clouse and Giribet (2010) using only molecular data (inset); thin branches in Clade C of the inset tree received bootstrap support less than 50%. Symbols on branches indicate resampling support above 50%, an asterisk for bootstrap and a circle for jackknife. Black branches are terminals lacking molecular data. Circled numbers indicate zones of interest discussed in the text.

there was a single morphometric tree of length 4247.653, and the combined analysis settled on a consensus tree of length 28 599.566. The 'Original' dataset's combined analysis shortest-length trees were found after extensive searching while the number of trees in the memory was constrained to 100; we did two subsequent searches starting from those trees but allowing more trees in the memory, and those searches found 1595 and 2501 trees of the same length and strict consensus topologies as the original 100 (Fig. 3). From the 'New' dataset, we found one morphometric tree of length 3489.020 and 61 combined-analysis trees of length 27 836.500 (Fig. 4). From the 'Adjusted' dataset, there was one morphometric tree of length 3270.596 and four shortest combined-analysis trees of length 27 569.183 (Fig. 5). The outgroups for resulting trees are

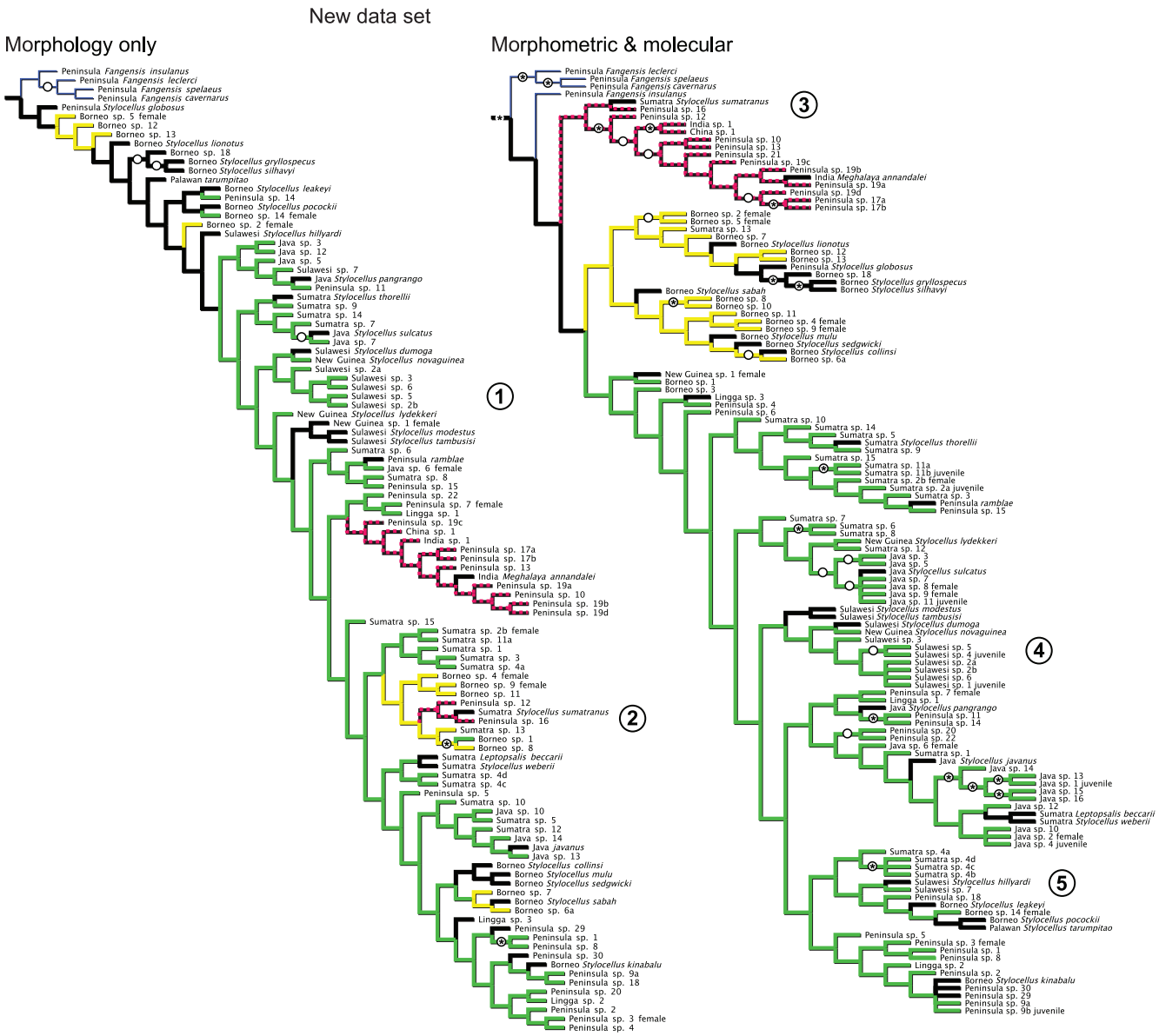

Fig. 4. Strict consensuses of the shortest trees found using the 'New' morphometric dataset, showing Stylocellidae only (outgroups in Fig. 6). Colours and symbols are described in the legend of Fig. 2.

shown in Fig. 6, and branches of the Stylocellidae-only trees are coloured to match the topology of the molecular phylogeny found by Clouse and Giribet (2010) (Fig. 3, inset). The genus *Fangensis*, paraphyletic at the first nodes in the tree, has thin, blue branches; Clade A, which is essentially *Meghalaya* plus Peninsula sp. 16, has dashed, red branches; Clade B, which was identified earlier as the 'Borneo+*Miopsalis*' clade (Clouse and Giribet 2007) and contains both large, troglomorphic species and the miniaturised Sumatra sp. 13, is in yellow (light grey); and Clade C, which contains most of the species on Sumatra and the Thai–Malay Peninsula, a few on Borneo and all the species on Java, Sulawesi and New Guinea, is in green (dark grey).

The best trees from all three morphometric datasets analysed alone recovered several relationships suspected from molecular studies and the biogeography of the region, discussed in more detail below. When morphometric data were combined with molecular data, trees consisted of the well supported clades already recovered using the molecular data alone (Clouse and Giribet 2010). Clade C was recovered but had different rearrangements of its constituent lineages with the addition of new terminals and the different sets of morphometric data. This was also expected, because relationships among the various island lineages in Clade C have been shown to have low support and be highly sensitive to searching parameters and terminal selection (Clouse and Giribet 2010) (Fig. 3, inset).

The dataset comprised of *y*-intercept-adjusted ratios for all raw measurements divided by just the body length recovered many suspected relationships but generally failed to recover

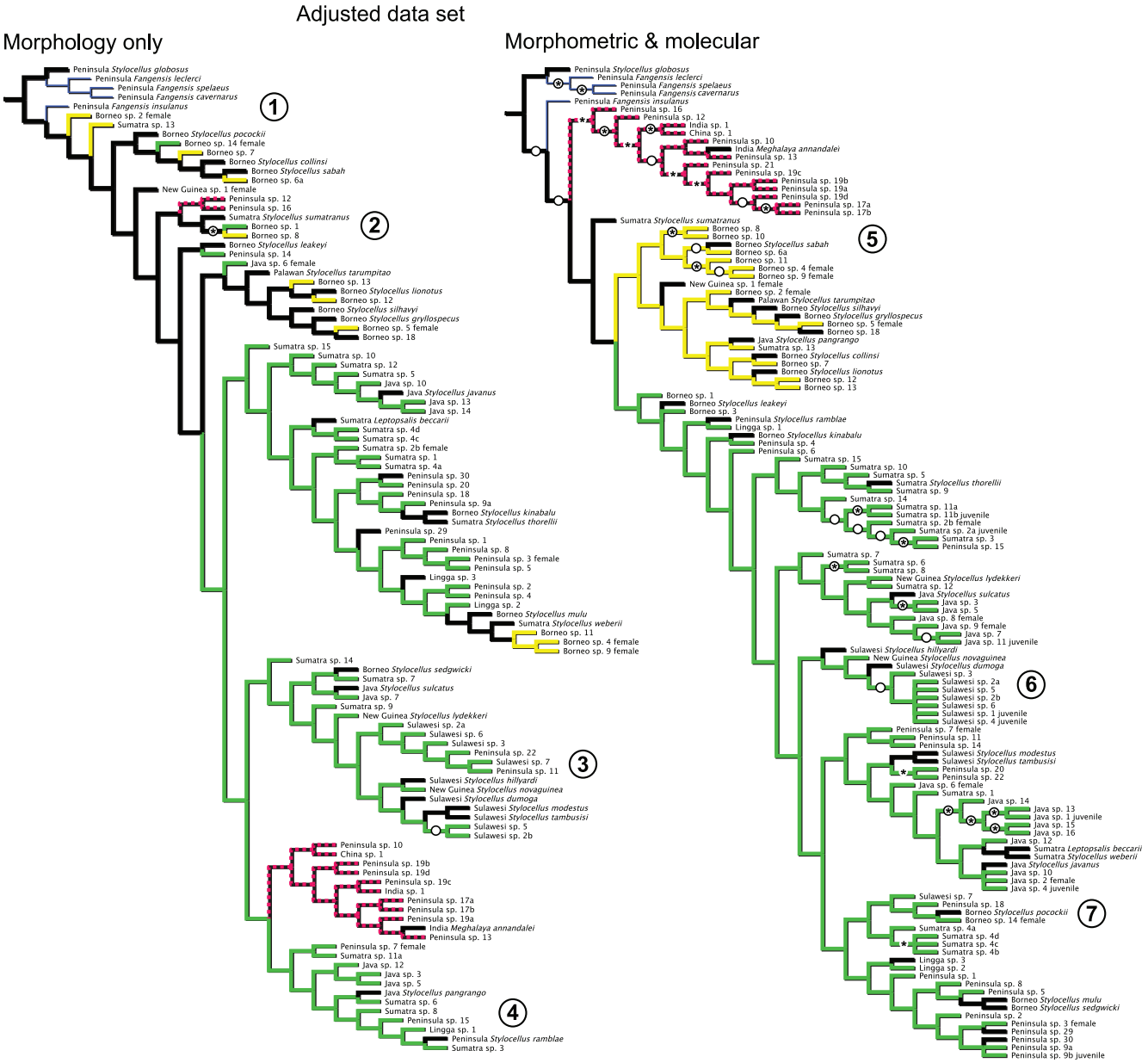

Fig. 5. Strict consensuses of the shortest trees found using the 'Adjusted' morphometric dataset, showing Stylocellidae only (outgroups in Fig. 6). Colours and symbols are described in the legend of Fig. 2.

relationships among the families (not shown). Removing females from analyses had little if any effect on the relationships of remaining terminals (not shown). Weighting morphometric data such that they contributed as many steps to the final tree as the molecular data resulted in trees that generally resembled the molecular phylogeny with two exceptions: Clade A (*Meghalaya*, in red) was consistently recovered as sister to Clade C, not Clade B; and Borneo sp. 1 placed inside Clade B with most of the other Bornean species, not in Clade C (not shown). This was not surprising: members of Clades A and C have strongly convergent morphologies (loss of the anal gland pore, moderate size, similar overall shape, and large eye lenses), and Borneo sp. 1 is one of only three species in Clade C that have anal gland pores, which is a common character in Clade B.

As expected, clades with high resampling support were generally few across all trees. Within Stylocellidae, support tended to concentrate in Clade A (*Meghalaya*, in red), in *Fangensis* and the other early splits in the family, and among some groups of very closely related species in Clade C (in green).

Discussion

The Cyphophthalmi present key opportunities to evaluate morphometric phylogenies because they have been the subject

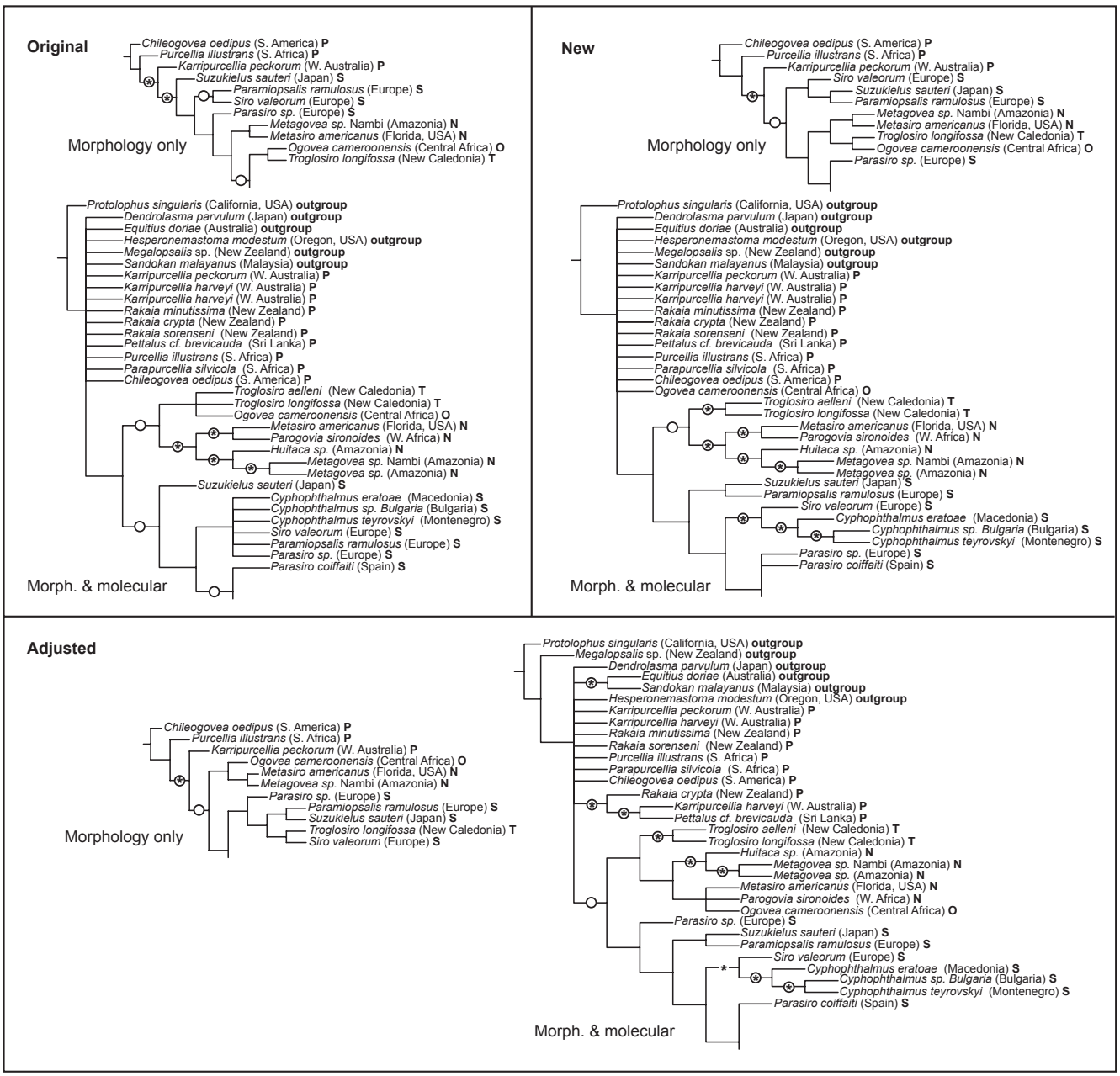

Fig. 6. Outgroup relationships using morphometric and morphological combined with molecular data, for the 'Original', 'New' and 'Adjusted' datasets. Letters after each terminal name are first initials of the current familial placement: Pettalidae, Sironidae, Neogoveidae, Ogoveidae and Troglosironidae.

of several molecular phylogenetic studies and these historical hypotheses have been shown to closely match proposed movements of its native landmasses (Boyer *et al*. 2007). This is especially true of Stylocellidae, the molecular phylogeny of which is consistent with the origin of the family on the Sibumasu terrane ~300 Mya (Hall 2002; Metcalfe 2002; Clouse and Giribet 2010). This terrane rifted from Gondwana and moved north to accrete to Eurasia, today underlying most of the Thai–Malay Peninsula and northern Sumatra. Stylocellidae appears to have diversified out of this peninsula north to the eastern Himalayas, then, with the arrival of Borneo in the Cretaceous, south into an expanded southern peninsula and Borneo. Their present distribution on western Sulawesi and Java is consistent with those areas being part of the southern peninsula before its breakup into the Malay Archipelago during the Cenozoic. The presence of Stylocellidae in northern Sulawesi is consistent with its attachment to the rest of the island 15 Mya, but the three species on New Guinea appear to have arrived by transoceanic dispersal, the only cases known in the suborder (Clouse and Giribet 2007; 2010).

Phylogenetic signal in morphometric data

Five key relationships found using morphometric data alone were recovered consistently and in agreement with molecular studies, indicating that the morphometric data contained significant phylogenetic information. First is the monophyly of Stylocellidae itself and monophyly or paraphyly among other family members (Fig. 6). Second, among the families, the close relationship between Troglosironidae and Neogoveidae, as found in previous molecular studies (Boyer *et al*. 2007; Sharma and Giribet 2009*a*), was recovered in the 'Original' and 'New' morphometric datasets – this despite not using an important character that is now interpreted as synapomorphic for the clade: the serrated claw of leg II. Third, with all three morphometric datasets we recovered the suspected relationship of the West African species *Ogovea cameroonensis* Giribet & Prieto, 2003, from the family Ogoveidae, as sister to either the Neogoveidae or Troglosironidae (which are themselves often recovered as sister clades in molecular studies). Unpublished molecular data from *O. cameroonensis* also suggest affinities to both families, and the distribution of Ogoveidae overlaps that of West African Neogoveidae (Giribet 2000; Boyer *et al*. 2007). In the combined analysis of the 'Adjusted' dataset, *O. cameroonensis* was recovered in a polytomy with the West African *Parogovia sironoides* Hansen, 1921, and the North American *Metasiro americanus* (Davis, 1933), both in Neogoveidae.

Fourth, within Stylocellidae we recovered *Fangensis* as the first splits (Figs 3–5, blue branches), which matches one of the most consistent results of molecular studies of the family (Schwendinger *et al*. 2004; Clouse and Giribet 2007; Clouse and Giribet 2010). Moreover, with the 'Adjusted' dataset we recovered *Fangensis* as paraphyletic, with *F. insulanus* Schwendinger & Giribet, 2005 as sister to the rest of the family (Fig. 5, zone 1), a common molecular finding that may justify the designation of a new generic status for this species.

Finally, with the morphometric data we recovered Peninsula spp. 12 and 16 in close association with each other and with Clade A. Most of the species in Clade A have several morphological synapomorphies defined in the description of the genus *Meghalaya* (Giribet *et al*. 2007), but Peninsula sp. 12, although generally resembling *Meghalaya*, also has an anal gland pore, absent in *Meghalaya*. Peninsula sp. 16 has an anal gland pore, too, but its overall morphology bears no obvious resemblance to *Meghalaya*. Thus it was expected that morphometric data alone would not find a close relationship between Peninsula sp. 16 and the remainder of Clade A, but we were interested in the placement of Peninsula sp. 12 and its possible influence on the whole of Clade A. The 'Original' morphometric dataset placed Peninsula sp. 12 as a derived member of Clade A, which conflicted with molecular results, showing it as splitting earlier from the rest of the clade, and the 'New' and 'Adjusted' datasets placed Peninsula sp. 12 well outside Clade A, among Bornean species with anal gland pores (mostly from Clade B, in yellow, Figs 4, 5, zone 2). However, among those Bornean species was also Clade A's Peninsula sp. 16, and using the 'New' morphometric dataset, it and Peninsula sp. 12 formed a monophyletic group with *Stylocellus sumatranus*; moreover, Peninsula spp. 12 and 16 were sister to each other using the 'Adjusted' morphometric dataset. This association between Peninsula spp. 12 and 16 is the first morphological support for the unintuitive molecular finding that Peninsula sp. 16 and *Meghalaya* have a close relationship.

Effects of missing data and Stylocellus sumatranus

The most striking effect of missing data in our analysis was the lowering of resampling support across all trees despite the recovery of clades suspected from molecular studies, current taxonomy and biogeographic evidence. This echoes the findings of Hardy *et al*. (2008), who found that the inclusion of 20 termimals (out of a total of 317) that had only morphological data in a combined morphological and molecular analysis lowered resampling support in spite of reasonably placing most terminals at the genus level. It is also consistent with our findings in a purely molecular analysis (Clouse and Giribet 2010), in which the inclusion of 15 terminals (out of 137) with very few data caused support values to drop significantly despite the recovery of a reasonable, resolved phylogeny. Here we included 29–30 terminals with no molecular data (black branches) and 39 terminals with no morphometric data (out of 117–118), so this effect was especially pronounced in these analyses.

Stylocellus sumatranus and some other terminals without molecular data were also missing a sizeable portion of the morphometric data, but their consistent recovery in certain positions suggests relationships meriting further investigation. In all analyses, there was a close association between *Stylocellus sumatranus* and Peninsula sp. 16, even when the molecular data separated Peninsula sp. 16 from the Clade-B species that *S. sumatranus* superficially resembles (Fig. 3, zone 3; Fig. 4, zone 3; Fig. 5, zone 5). We have known that *S. sumatranus* probably does not bear a close relationship to many of the species currently classified in the genus *Stylocellus*, because its anal gland pore and heavily sculptured chelicerae are not common among other *Stylocellus* species. On the whole, we have been inclined to place *S. sumatranus* in Clade B, where we find other large species with these character states, even if the only known Sumatran member of the clade is highly miniaturised (Sumatra sp. 13) (Clouse and Giribet 2007). This would have conveniently allowed us to use the four established genera for the family (*Miopsalis* too poorly defined to be useful and the type species not available for examination) for all the major clades: *Fangensis*, although often paraphyletic, at the base; *Meghalaya* for Clade A; *Stylocellus* for Clade B; and *Leptopsalis* (given the placement here of *L. beccarii* Thorell, 1882/83) for Clade C. However, with *S. sumatranus* never being found inside Clade B nor found in close association with Clade B members among their more scattered recovery in the morphometric-only analyses, consistently being found in close relationship with Peninsula sp. 16, and having a biogeography more reasonably aligned with Peninsula sp. 16 (found on a small island off the west coast of Thailand) than Bornean species, the proper generic revision of the family may be more complicated than originally thought. This is exacerbated by the fact that finding fresh material of *S. sumatranus* seems implausible due to the lack of specific collecting data (the type locality is 'Sumatra') and the large

amount of suitable habitat present on that island. However, it seems clear that *Leptopsalis* is not in synonomy with *Stylocellus* as previously thought (see, for example, Giribet 2000).

We asked if the close association between *S. sumatranus* and Peninsula sp. 16 was due to a general similarity in all characters that made this more of a phenetic association than a parsimonious optimisation. For the 'New' dataset, we examined the differences in normalised character values between *S. sumatranus* and the other terminals, summing their absolute values; *S. sumatranus* is less similar to Peninsula sp. 16 than to 16 other terminals in our analysis. Given that *S. sumatranus* was missing 35% of the morphometric data – the type is missing both fourth walking legs, has a damaged gonostome and a deformed opisthosoma – we expected it to be highly unstable in our analyses or to find a relationship with other terminals missing large amounts of data, such as *S. thorellii*, missing 41%, and *S. pangrango* Shear, 1993, missing 33%; thus, the stability of its relationship to Peninsula sp. 16 is surprising.

Biogeography of Sulawesi and New Guinea

We were particularly interested in the species from Sulawesi and New Guinea for five reasons. First, a counterintuitive finding of molecular studies has been a close association between *S. novaguinea* and the species from Northern Sulawesi but the lack of a close relationship among the species from New Guinea or between the species from Western and Northern Sulawesi (Clouse and Giribet 2007, 2010). Second, all of these species have, like many species in Clade C, a distinct lack of morphological synapomorphies (no Rambla's organ, no anal gland pore, no cheliceral sculpturing, etc.) but perhaps share subtle shape similarities, so they were an important test of whether such perceived similarities were meaningful. Third, among the Sulawesi species is the large *S. hillyardi* Shear, 1993, which falls into a class of vague, large species, like the female Borneo sp. 14 and Palawan's *S. tarumpitao*. Is *S. hillyardi* just a large member of a Sulawesi clade, or does it represent a different lineage that also invaded Palawan? Fourth, the geologic history of Sulawesi is conducive to the species there forming a clade. Western Sulawesi accreted on to the ancient Sundaland Peninsula for only about ten million years around the K–T boundary (Hall 2002). Only after it separated did Eastern and Northern Sulawesi attach, allowing clades derived from the Western Sulawesi inhabitants to invade those regions. Thus, monophyly of the island's inhabitants is strongly suspected, but to date only the Northern Sulawesi species have been recovered as monophyletic. Finally, as with other groups, we wanted to place described species from the island for which we had no modern specimens. Besides *S. hillyardi* and the undescribed species we collected from northern and western Sulawesi, there are three other described species, one from each region: *S. dumoga* Shear, 1993 from northern Sulawesi; *S. modestus* Hansen & Sørensen, 1904 from western Sulawesi; and *S. tambusisi* Shear, 1993 from eastern Sulawesi.

All analyses of just morphological data, especially the 'Original' and 'Adjusted' datasets, found a close relationship among all Sulawesi and the described New Guinean species (Fig. 6, zone 3; Fig. 4, zone 1; Fig. 5, zone 3). The addition of molecular data tended to separate out Sulawesi sp. 7 and *S. lydekkeri* Clouse & Giribet, 2007, as expected (Clouse and Giribet 2007, 2010) (Fig. 6, zones 5 and 6; Fig. 4, zones 4 and 5; Fig. 5, zones 6 and 7). With the 'New' dataset, *S. hillyardi* was never recovered particularly close to the other Sulawesi species, but when molecular data pushed Sulawesi sp. 7 away from the other Sulawesi species, it then became sister to *S. hillyardi*. With the 'Adjusted' dataset, molecular data caused *S. modestus* and *S. tambusisi* to move away from other Sulawesi species, but they remained sister to each other. New Guinea sp. 1, an especially vague species known only from females and yielding little molecular data (not included here), tended to be unstable in analyses, finding placement among various members of Clade B with the 'Original' and 'Adjusted' datasets, which seems unlikely; however, with the 'New' dataset, it was recovered as sister to *S. modestus* and *S. tambusisi*. These analyses suggest a close relationship among all the Sulawesi species, and perhaps between them and all the New Guinean ones; a relationship, which, for Sulawesi sp. 7 and *S. lydekkeri*, is in conflict with our previous molecular results but which we will be exploring further. The presence of Stylocellidae on New Guinea has been difficult to explain (Clouse and Giribet 2007), especially three different species from different locations. Likewise, different origins for the Sulawesi fauna would be difficult to explain, given the island's brief attachment to Borneo in the early Cenozoic. Thus, pulling Sulawesi sp. 7 and *S. lydekkeri* (both of which have been unstable in molecular analyses) into a clade with the very stable and well supported relationship among the other Sulawesi species and *S. novaguinea* requires invoking fewer dispersal events.

Size information and phylogenetic signal

A major challenge at the start of this analysis was controlling for size. Size surely contains some phylogenetic information, since certain species groups tend to be similar in size, but it also appears to be a labile trait in this family. The largest Cyphophthalmi known (Borneo sp. 13, 7.5 mm long) is a member of Clade B with the smallest Stylocellidae (Sumatra sp. 13, 1.4 mm long), and large size appears to be part of the suite of troglomorphic features we needed to avoid in the cave-dwelling species anyway (*Fangensis* spp., *Stylocellus globosus* Schwendinger & Giribet, 2004 and several Bornean members of Clade B). We used the closeness of Borneo sp. 13 and Sumatra sp. 13 (Fig. 6, zone 2; Fig. 4, zone 2; Fig. 5, zone 1) in our trees as the first evaluation of our efforts to eliminate size information from our characters, and our preliminary analyses often placed clusters of small species far from clusters of large species, suggesting that size information was still masking phylogenetic signal – at least that suggested by the molecular data.

The problem of extracting phylogenetic information from morphometric data in a clade with a wide range of body sizes was addressed specifically by Gilbert and Rossie (2007) in their study of the monkey tribe Papionini, which includes 50 kg mandrills and 3 kg mangabeys. They developed a technique whereby large and small species were coded separately, and characters known to be influenced by allometry were assessed within the context of the particular size group. Thus, a large and a small species could both have 'short' snouts, even though in absolute terms the snout of the former was larger than that of the latter. The phylogeny resulting from this 'narrow allometric' method was congruent with molecular phylogenies. Nonetheless,

even after size corrections, otherwise reasonable morphometric phylogenies may have especially large or small species misplaced (Macholan 2008), and approaches like the narrow allometric method are best used when there exists a natural binning of species into large and small categories.

The misplaced species in Macholan's (2008) study are interesting in that they occur despite the use of Burnaby's (1966) multiple-group principal components analysis (MGPCA), a method endorsed by Rohlf and Bookstein (1987) over Humphries *et al*.'s (1981) sheared principal components, which was found to be ineffective in removing all size information; apparently even MGPCA can leave behind size information, and the most effective and appropriate scaling procedure is still a matter of debate. A common method has been to take regression residuals (Rohlf 1990; Singleton 2000; Stephens and Wiens 2003; Poe 2004; Bergmann and Russell 2007), but this method has been criticised by Rae (2002) for its reliance on a correction factor derived from the entire terminal set. That is, the scaling denominator applied to any one terminal is dependent on the selection of terminals used in the analysis, much like Jungers *et al*.'s (1995) division of individual character values by the geometric mean of all. Rae (2002) questioned the independence of data that depend on other taxa in the analysis and recommended using the individual's own measurements as its scaling factor, as he had done with hominid data (Rae 1997). For example, measurements could be expressed as ratios to each other (e.g. O'Grady and May 2003) or as ratios to a single individual size measure (e.g. O'Neill and Dobson 2008, who used body mass × bone length).

Still, the use of simple ratios has itself been the subject of debate, as they have their drawbacks. Atchley *et al*. (1975, 1976) argued against the use of simple ratios in morphometrics due to their alteration of data distributions, introduction of new relationships and failure to completely eliminate size information. Anderson and Lydic (1977) provided further evidence from biological examples, and later it was shown that the Kluge–Kerfoot phenomenon (a positive correlation of character variances within versus between populations; Kluge and Kerfoot 1973) was a statistical artefact of a size correction (Bookstein *et al*. 1985). However, Corruccini (1977) argued that adjusting by the *y*-intercept in the numerator was an adequate correction already in the literature, Dodson (1978) argued that the failure of ratios to completely eliminate size was long-known and Hills (1978) argued that much of the problem with ratios was known to be corrected by log-adjusting them. There were rebuttals to the negative responses (Atchley 1978; Atchley and Anderson 1978), and later Albrecht *et al*. (1993) offered a useful accounting of proper ratio adjustments to maximise the elimination of size (momentarily setting aside problems with their influence on statistical distributions). Albrecht *et al*. (1993) drew attention to whether the relationship between the numerator and demoninator of a simple ratio was linear with the *y*-intercept equal to zero. Only in that circumstance could a simple ratio eliminate size information, and it was argued that this is rarely the case with morphometric data. It is easy to see how a curvilinear relationship between numerator and denominator could retain size information by imagining an appendage that reaches an upper size limit while the body as a whole continues to grow; continuing to divide the appendage length by body size in large specimens would result in ever smaller fractions. The problem of the *y*-intercept can be shown by an example of two measurements, A and B, drawn from two specimens, I and II : I(A,B) = (1,3) and II(A,B) = (2,4). Taking the ratio B/A for both, we would obtain different fractions, 3 and 2. Thus, we would still be able to distinguish the small from the large specimen, i.e. indicators of overall size have not been removed from the data. However, this can be corrected by first finding the relationship between these two characters, with A as the independent variable. Then, subtracting the *y*-intercept, 2, from the numerator before calculating the ratio for I and II would result in the same fraction for both (1).

Ultimately, we decided to control for size in four ways. First, we avoided measurements diagonal to body axes that would retain size information when scaled by orthogonal or parallel measurements, and we favoured measurements around regions shown from earlier studies to be informative and not heavily affected by troglomorphism. Second, we used only locally scaled measurements.

Third, we adjusted ratios by the *y*-intercept. Adjusting the 'New' dataset by the *y*-intercept caused Sumatra sp. 13 to move, appropriately, from a derived clade of mixed affinities to a position near *Fangensis* and the large Borneo sp. 2 (Fig. 5,

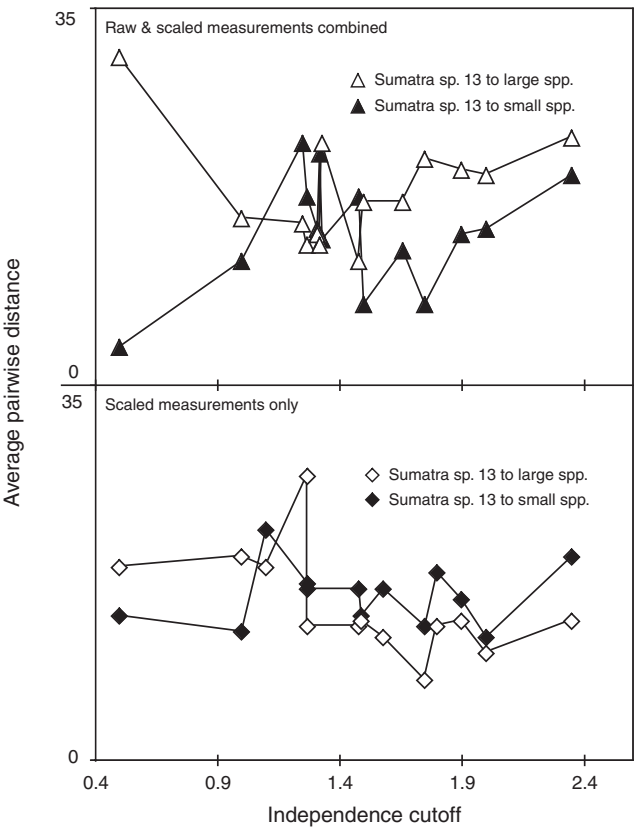

Fig. 7. Average pairwise distances from Sumatra sp. 13 to two other small species, *Stylocellus ramblae* and Lingga sp. 1 (black symbols), and to the two large species, *S. globosus* and Borneo sp. 13 (white symbols) using the 'Original' morpometric dataset. The upper graph shows the distance at different cutoff values of the independence analysis using raw and scaled data combined, and the bottom graph shows the same for only the scaled data.

zone 1). Also, large *S. hillyardi* moved further into the Sulawesi and New Guinea clade, becoming sister to the smaller *S. novaguinea* (Fig. 5, zone 3); likewise, small Peninsula sp. 12 and large Peninsula sp. 16 became sister to each other (Fig. 5, zone 2). The effects of adjusting the data by the *y*-intercept are illustrated with examples in Fig. 1. For some body measurement ratios, subtracting the *y*-intercept of their trendline from the numerator before taking the ratio rendered the ratio effectively uncorrelated with body size (Fig. 1*A*). The removal of spurious size information from all ratios led to a weakening of correlations between those ratios that show little or no true allometry. However, the same adjustment on other ratios revealed a slightly stronger correlation between each ratio and body size; consequently, there was a stronger correlation between the ratios (Fig. 1*B*). Thus, the adjustment can also reveal true allometry and help collapse such ratios into a single character.

Fourth, we collapsed characters that were correlated with no outliers. Many of these characters were united by their large content of size information, and their collapse into one PC eliminated repetitions of size in the analysis. Using the 'Original' dataset, we examined the distance from Sumatra sp. 13 to two closely related, large species (*S. globosus* and Borneo sp. 13) versus two distantly related, small species (*S. ramblae* Giribet, 2002, and Lingga sp. 1) in trees from different IA cutoffs (Fig. 7). As expected, increasing the IA cutoff (and thus the number of collapsed characters) caused Sumatra sp. 13 to approach the large species and avoid the small ones, especially when raw measurements were included in the data.

Conclusions

Virtually every aspect of our study – the phylogenetic analysis of undiscretised data, use of measurements as characters, calculation of ratios, discovery of character dependence, collapse of characters into principal components – touches on major controversies of morphometrics in phylogenetics. We have discussed many of these in our original description of the independence analysis technique (de Bivort *et al*. 2010), and here we have mostly focused on the issue of using ratios to control for size. Major objections invoke the complications (or impossibility) of determining homology in measurements and shape descriptors (Bookstein *et al*. 1985; Bookstein 1994; Zelditch *et al*. 1995; Rohlf 1998), but there has been recognition that homology can be approximated (Naylor 1996), and that homology errors with shape descriptors can be inconsequential at certain scales (Bookstein 1994). Likewise, although we have no disagreement with demonstrations that eliminating size information is both difficult and can create statistical effects that make the data useless in certain kinds of analyses (Atchley *et al*. 1976; Anderson and Lydic 1977; Atchley 1978; Atchley and Anderson 1978; Phillips 1983), we also feel there may be an acceptable level of approximation that allows efficient use of the limited information at hand.

Sorting the enormous clade of indistinct species spread over the Indo–Malay Archipelago and placing types with missing body parts were goals that needed to be accomplished before attempting a taxonomic revision. Morphology is the only way certain species can be included in phylogenies (e.g. Wheeler 1992; Hillis and Wiens 2000), and when species lack easily discretised features, morphometrics may be the only way to examine morphology. Morphometric data – undiscretised, combined in shape descriptors and analysed in a parsimony framework as we have done here – have been argued to contain considerable phylogenetic signal (González-José *et al*. 2008), and morphometric data have regularly been searched for new sources of characters in systematics (e.g. Fink and Zelditch 1995; Bookstein 2002; Pelser *et al*. 2004; Dessein *et al*. 2005; Abdala 2007; Domínguez and Roig-Juñent 2008; Moon *et al*. 2008). We consider the phylogenetic hypotheses found here from our morphometric data reasonable enough to provide direction for many species, and we anticipate referring to these hypotheses during the course of a taxonomic revision of the cyphophthalmid family Stylocellidae.

Acknowledgements

Peter J. Schwendinger, Louis Deharveng, Alex Riedl and Cahyo Rahmadi generously shared specimens used for study. Janet Beccaloni (BMNH), Norman I. Platnick and Lorenzo Prendini (AMNH), Nikolaj Scharff (ZMUC), Peter J. Schwendinger (MHNG), Petra Sierwald (FMNH), and James Hogan (Oxford Museum) sent type specimens for study. We are grateful for the generosity of Jeremy A. Miller, David H. Kavanaugh and Charles E. Griswold (CAS) for making the Chinese material available; it was collected on an expedition funded by National Science Foundation award BSI-0103795, 'Biotic Survey of the Gaoligongshan, a Biodiversity Hotspot in Western Yunnan, China' to CEG. This material is based in part on work supported by the National Science Foundation under Grant No. 0236871. Putnam Expedition Grants (MCZ) funded travel to Indonesia, and Cahyo Rahmadi assisted with permits and field work. Richard Schalek and Ellen Hodges (Center for Nanoscale Systems, Harvard University) provided s.e.m. support. The Willi Hennig Society is acknowledged for the sponsored version of TNT. Diane Sheridan, Joey Pakes, Jessica Baker, Alejandra Guzmán, and Jamie Sodikoff provided laboratory support; Laura Leibensperger and Adam Baldinger of the MCZ Department of Invertebrate Zoology provided essential collections support.

References

Abdala, C. S. (2007). Phylogeny of the *boulengeri* group (Iguania: Liolaemidae, *Liolaemus*) based on morphological and molecular characters. *Zootaxa* **1538**, 3–84.

Albrecht, G. H., Gelvin, B. R., and Hartman, S. E. (1993). Ratios as a size adjustment in morphometrics. *American Journal of Physical Anthropology* **91**, 441–468. doi:10.1002/ajpa.1330910404

Anderson, D. E., and Lydic, R. (1977). On the effect of using ratios in the analysis of variance. *Biobehavioral Reviews* **1**, 225–229. doi:10.1016/0147-7552(77)90024-9

Atchley, W. R. (1978). Ratios, regression intercepts, and the scaling of data. *Systematic Zoology* **27**, 78–83. doi:10.2307/2412817

Atchley, W. R., and Anderson, D. (1978). Ratios and the statistical analysis of biological data. *Systematic Zoology* **27**, 71–78. doi:10.2307/2412816

Atchley, W. R., Gaskins, C. T., and Anderson, D. (1975). Statistical properties of ratios in biological data. *American Zoologist* **15**, 829.

Atchley, W. R., Gaskins, C. T., and Anderson, D. (1976). Statistical properties of ratios. I. Empirical results. *Systematic Zoology* **25**, 137–148. doi:10.2307/2412740

Bergmann, P. J., and Russell, A. P. (2007). Systematics and biogeography of the widespread Neotropical gekkonid genus *Thecadactylus* (Squamata), with the description of a new cryptic species. *Zoological Journal of the Linnean Society* **149**, 339–370.

Bookstein, F. L. (1994). Can biometrical shape be a homologous character? In 'Homology: The Hierarchical Basis of Comparative Biology'. (Ed. B. K. Hall.) pp. 197–227. (Academic Press: New York.)

Bookstein, F. L. (2002). Creases as morphometric characters. In 'Morphology, Shape and Phylogeny'. (Eds N. Macleod and P. L. Forey.) pp. 139–174. (Taylor and Francis: London.)

Bookstein, F. L., Chernoff, B., Elder, R. L., Humphries, J. M. Jr, Smith, G. R., and Strauss, R. E. (1985). 'Morphometrics in Evolutionary Biology: The Geometry of Size and Shape Change, With Examples from Fishes.' (Academy of Natural Sciences: Philadelphia, PA.)

Boyer, S. L., and Giribet, G. (2007). A new model Gondwanan taxon: systematics and biogeography of the harvestman family Pettalidae (Arachnida, Opiliones, Cyphophthalmi), with a taxonomic revision of genera from Australia and New Zealand. *Cladistics* **23**, 337–361. doi:10.1111/j.1096-0031.2007.00149.x

Boyer, S. L., Clouse, R. M., Benavides, L. R., Sharma, P., Schwendinger, P. J., Karunarathna, I., and Giribet, G. (2007). Biogeography of the world: a case study from cyphophthalmid Opiliones, a globally distributed group of arachnids. *Journal of Biogeography* **34**, 2070–2085. doi:10.1111/j.1365-2699.2007.01755.x

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L., and Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Systematic Biology* **42**, 384–397.

Burnaby, T. P. (1966). Growth-invariant discriminant functions and generalized distances. *Biometrics* **22**, 96–110. doi:10.2307/2528217

Clouse, R. M., and Giribet, G. (2007). Across Lydekker's Line – first report of mite harvestmen (Opiliones : Cyphophthalmi : Stylocellidae) from New Guinea. *Invertebrate Systematics* **21**, 207–227. doi:10.1071/IS06046

Clouse, R. M., and Giribet, G. (2010). When Thailand was an island—the phylogeny and biogeography of mite harvestmen (Opiliones, Cyphophthalmi, Stylocellidae) in Southeast Asia. *Journal of Biogeography.* doi:10.1111/j.1365-2699.2010.02274.x

Corruccini, R. S. (1977). Correlation properties of morphometric ratios. *Systematic Zoology* **26**, 211–214. doi:10.2307/2412844

de Bivort, B. L., Clouse, R., and Giribet, G. (2010). A morphometrics-based phylogeny of the temperate Gondwanan mite harvestmen (Opiliones, Cyphophthalmi, Pettalidae). *Journal of Zoological Systematics and Evolutionary Research.* doi:10.1111/j.1439-0469.2009.00562.x

Dessein, S., Ochoterena, H., De Block, P., Lens, F., Robbrecht, E., Schols, P., Smets, E., Vinckier, S., and Huysmans, S. (2005). Palynological characters and their phylogenetic signal in Rubiaceae. *Botanical Review* **71**, 354–414. doi:10.1663/0006-8101(2005)071[0354:PCATPS]2.0.CO;2

Dodson, P. (1978). On the use of ratios in growth studies. *Systematic Zoology* **27**, 62–67. doi:10.2307/2412814

Domínguez, M. C., and Roig-Juñent, S. A. (2008). A phylogeny of the family Fanniidae Schnabl (Insecta: Diptera: Calyptratae) based on adult morphological characters, with special reference to the Austral species of the genus *Fannia*. *Invertebrate Systematics* **22**, 563–587. doi:10.1071/IS08003

Farris, J. S., Albert, V. A., Kallersjo, M., Lipscomb, D., and Kluge, A. G. (1996). Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**, 99–124. doi:10.1111/j.1096-0031.1996.tb00196.x

Fink, W. L., and Zelditch, M. L. (1995). Phylogenetic analysis of ontogenetic shape transformations: A reassessment of the piranha genus *Pygocentrus* (Teleostei). *Systematic Biology* **44**, 343–360.

Gilbert, C. C., and Rossie, J. B. (2007). Congruence of molecules and morphology using a narrow allometric approach. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11910–11914. doi:10.1073/pnas.0702174104

Giribet, G. (2000). Catalogue of the Cyphophthalmi of the world (Arachnida, Opiliones). *Revista Ibérica de Aracnología* **2**, 49–76.

Giribet, G., and Boyer, S. L. (2002). A cladistic analysis of the cyphophthalmid genera (Opiliones, Cyphophthalmi). *The Journal of Arachnology* **30**, 110–128. doi:10.1636/0161-8202(2002)030[0110:ACAOTC]2.0.CO;2

Giribet, G., Rambla, M., Carranza, S., Baguñà, J., Riutort, M., and Ribera, C. (1999). Phylogeny of the arachnid order Opiliones (Arthropoda) inferred from a combined approach of complete 18S and partial 28S ribosomal DNA sequences and morphology. *Molecular Phylogenetics and Evolution* **11**, 296–307. doi:10.1006/mpev.1998.0583

Giribet, G., Edgecombe, G. D., Wheeler, W. C., and Babbitt, C. (2002). Phylogeny and systematic position of Opiliones: a combined analysis of chelicerate relationships using morphological and molecular data. *Cladistics* **18**, 5–70.

Giribet, G., Sharma, P. P., and Bastawade, D. B. (2007). A new genus and species of Cyphophthalmi (Arachnida: Opiliones) from the north-eastern states of India. *Zoological Journal of the Linnean Society* **151**, 663–670.

Goloboff, P. A., Mattoni, C. I., and Quinteros, A. S. (2006). Continuous characters analyzed as such. *Cladistics* **22**, 589–601. doi:10.1111/j.1096-0031.2006.00122.x

Goloboff, P. A., Farris, J. S., and Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786. doi:10.1111/j.1096-0031.2008.00217.x

González-José, R., Escapa, I., Neves, W. A., Cúneo, R., and Pucciarelli, H. M. (2008). Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. *Nature* **453**, 775–778. doi:10.1038/nature06891

Hall, R. (2002). Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences* **20**, 353–431. doi:10.1016/S1367-9120(01)00069-4

Hardy, C. R., Moline, P., and Linder, H. P. (2008). A phylogeny for the African Restionaceae and new perspectives on morphology's role in generating complete species phylogenies for large clades. *International Journal of Plant Sciences* **169**, 377–390. doi:10.1086/526467

Hillis, D. M., and Wiens, J. J. (2000). Molecules versus morphology in systematics: conflicts, artifacts, and misconceptions. In 'Phylogenetic Analysis of Morphological Data'. (Ed. J. J. Weins.) pp. 1–19. (Smithsonian Institution Press: Washington, DC.)

Hills, M. (1978). On ratios—a response to Atchley, Gaskins, and Anderson. *Systematic Zoology* **27**, 61–62. doi:10.2307/2412813

Huelsenbeck, J. P., Bull, J. J., and Cunningham, C. W. (1996*a*). Combining data in phylogenetic analysis. *Trends in Ecology & Evolution* **11**, 152–158. doi:10.1016/0169-5347(96)10006-9

Huelsenbeck, J. P., Bull, J. J., and Cunningham, C. W. (1996*b*). Combining data in phylogenetic analysis – reply. *Trends in Ecology & Evolution* **11**, 335. doi:10.1016/S0169-5347(96)91643-2

Humphries, J. M., Bookstein, F. L., Chernoff, B., Smith, G. R., Elder, R. L., and Poss, S. G. (1981). Multivariate discrimination by shape in relation to size. *Systematic Zoology* **30**, 291–308. doi:10.2307/2413251

Juberthie, C. (1988). Les Opilions Cyphophthalmes: biogéographie, vitesse d'évolution, périodes de colonisation du milieu souterrain. *TUB-Dokumentation Kongresse und Tagungen, Berlin* **38**, 303–308.

Jungers, W. L., Falsetti, A. B., and Wall, C. E. (1995). Shape, relative size, and size-adjustments in morphometrics. *Yearbook of Physical Anthropology* **38**, 137–161. doi:10.1002/ajpa.1330380608

Kluge, A. G. (1989). A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* **38**, 7–25. doi:10.2307/2992432

Kluge, A. G., and Kerfoot, W. C. (1973). The predictability and regularity of character divergence. *American Naturalist* **107**, 426–442. doi:10.1086/282844

Macholan, M. (2008). The mouse skull as a source of morphometric data for phylogeny inference. *Zoologischer Anzeiger* **247**, 315–327. doi:10.1016/j.jcz.2008.06.001

Metcalfe, I. (2002). Permian tectonic framework and palaeogeography of SE Asia. *Journal of Asian Earth Sciences* **20**, 551–566. doi:10.1016/S1367-9120(02)00022-6

Miyamoto, M. M., and Fitch, W. M. (1995). Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* **44**, 64–76.

Moon, H. K., Vinckier, S., Walker, J. B., Smets, E., and Huysmans, S. (2008). A search for phylogenetically informative pollen characters in the subtribe Salviinae (Mentheae: Lamiaceae). *International Journal of Plant Sciences* **169**, 455–471. doi:10.1086/526463

Naylor, G. J. P. (1996). Can partial warp scores be used as cladistic characters? In 'Advances in Morphometrics'. (Eds L. F. Marcus, M. Corti, A. Loy, G. P. L. Naylor and D. E. Slice.) pp. 519–530. (Plenum Press: New York.)

O'Grady, E. W., and May, M. L. (2003). A phylogenetic reassessment of the subfamilies of Coenagrionidae (Odonata: Zygoptera). *Journal of Natural History* **37**, 2807–2834. doi:10.1080/0022293021000007453

O'Neill, M. C., and Dobson, S. D. (2008). The degree and pattern of phylogenetic signal in primate long-bone structure. *Journal of Human Evolution* **54**, 309–322. doi:10.1016/j.jhevol.2007.08.008

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics (Oxford, England)* **20**, 289–290. doi:10.1093/bioinformatics/btg412

Pelser, P. B., van den Hof, K., Gravendeel, B., and van der Meijden, R. (2004). The systematic value of morphological characters in *Senecio* sect. *Jacobaea* (Asteraceae) as compared to DNA sequences. *Systematic Botany* **29**, 790–805. doi:10.1600/0363644041744482

Phillips, R. B. (1983). Shape characters in numerical taxonomy and problems with ratios. *Taxon* **32**, 535–544. doi:10.2307/1221721

Poe, S. (2004). Phylogeny of anoles. *Herpetological Monograph* **18**, 37–89. doi:10.1655/0733-1347(2004)018[0037:POA]2.0.CO;2

Rae, T. C. (1997). The early evolution of the hominoid face. In 'Function, Phylogeny, and Fossils: Miocene Hominoid Evolution and Adaptations'. (Eds D. Begun, C. Ward and M. Rose.) pp. 59–77. (Plenum: New York.)

Rae, T. C. (2002). Scaling, polymorphism, and cladistic analysis. In 'Morphology, Shape and Phylogeny'. (Eds N. Macleod and P. L. Forey.) pp. 45–52. (Taylor and Francis: London.)

Rambla, M. (1991). A new *Stylocellus* from some caves of Borneo, Malaysia (Opiliones, Cyphophthalmi, Stylocellidae). *Memoires de Biospeologie* **18**, 227–232.

Rohlf, F. J. (1990). Morphometrics. *Annual Review of Ecology and Systematics* **21**, 299–316. doi:10.1146/annurev.es.21.110190.001503

Rohlf, F. J. (1998). On applications of geometric morphometrics to studies of ontogeny and phylogeny. *Systematic Biology* **47**, 147–158. doi:10.1080/106351598261094

Rohlf, F. J., and Bookstein, F. L. (1987). A comment on shearing as a method for "size correction". *Systematic Zoology* **36**, 356–367. doi:10.2307/2413400

Schwendinger, P. J., and Giribet, G. (2005). The systematics of the south-east Asian genus *Fangensis* Rambla (Opiliones: Cyphophthalmi: Stylocellidae). *Invertebrate Systematics* **19**, 297–323. doi:10.1071/IS05023

Schwendinger, P. J., Giribet, G., and Steiner, H. (2004). A remarkable new cave-dwelling *Stylocellus* (Opiliones, Cyphophthalmi) from peninsular Malaysia, with a discussion on taxonomic characters in the family Stylocellidae. *Journal of Natural History* **38**, 1421–1435.

Sharma, P., and Giribet, G. (2009*a*). A relict in New Caledonia: phylogenetic relationships of the family Troglosironidae (Opiliones: Cyphophthalmi). *Cladistics* **25**, 279–294. doi:10.1111/j.1096-0031.2009.00252.x

Sharma, P., and Giribet, G. (2009*b*). Sandokanid phylogeny based on eight molecular markers—the evolution of a southeast Asian endemic family of Laniatores (Arachnida, Opiliones). *Molecular Phylogenetics and Evolution* **52**, 432–447. doi:10.1016/j.ympev.2009.03.013

Shear, W. A. (1993). New species in the oilionid genus *Stylocellus* from Malaysia, Indonesia and the Philippines (Opiliones, Chyphophthalmi, Stylocellidae). *Bulletin of the British Arachnological Society* **9**, 174–188.

Singleton, M. (2000). The phylogenetic affinities of *Otavipithecus namibiensis*. *Journal of Human Evolution* **38**, 537–573. doi:10.1006/jhev.1999.0369

Stephens, P. R., and Wiens, J. J. (2003). Ecological diversification and phylogeny of emydid turtles. *Biological Journal of the Linnean Society. Linnean Society of London* **79**, 577–610. doi:10.1046/j.1095-8312.2003.00211.x

Varón, A., Vinh, L. S., Bomash, I., and Wheeler, C. W. (2008). 'POY version 4.0.2885.' (American Museum of Natural History: New York.)

Wheeler, W. C. (1992). Extinction, sampling, and molecular phylogenetics. In 'Extinction and Cladistic Analysis'. (Eds M. Novacek and Q. D. Wheeler.) pp. 205–215. (Columbia University Press: New York.)

Zelditch, M. L., Fink, W. L., and Swiderski, D. L. (1995). Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. *Systematic Biology* **44**, 179–189.

Manuscript received 22 October 2009, accepted 5 January 2010