When discrete characters are wanting: Continuous character integration under the phylospecies concept informs the revision of the Australian land snail *Thersites* (Eupulmonata, Camaenidae)

Guoyi Zhang[1,2,*], Gerasimos Cassis[1], Frank Köhler[2]

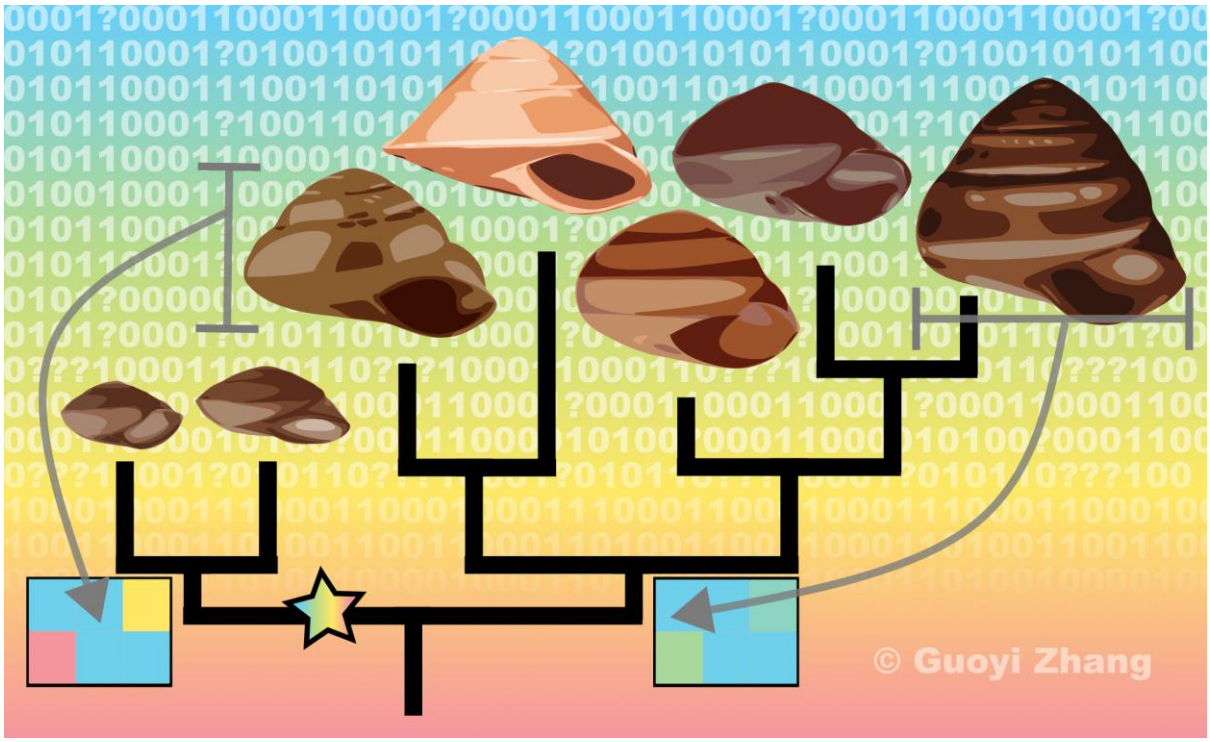
[1] School of Biological, Earth & Environmental Sciences, University of New South Wales, High St, Kensington, Sydney NSW 2052, Australia

[2] Australian Museum Research Institute, 1 William Street, Sydney NSW 2010, Australia

* guoyi.zhang@unsw.edu.au; starsareintherose@gmail.com

Running title: Continuous characters integration under phylospecies

Abstract



Species that are predominantly characterized by continuous instead of discrete morphological characters pose a challenge to species delimitation. Under the phylospecies concept, species are delimited by apomorphies, which are difficult to establish when characters are not discrete. In the present study, we address these challenges in the Australian land snail genus *Thersites* (family Camaenidae), where accepted species exhibit a scarcity of discrete distinguishing characters and differ in a limited range of continuous characters according to prior taxonomic studies. We integrate analyses of genome-scale molecular data with evaluations of several continuous, qualitative, and discrete morphological characters derived from landmarks to delineate species through detecting apomorphies. We found that dimensions of shell and genitals overlapped considerably among the currently accepted species. These overlaps may indicate morphospace saturation, which may affect species delineation through the gen-morph species concept. Statistical methods, such as Dunn's test, failed to consistently delineate monophyletic taxa. Additionally, we could not derive apomorphies from discretized landmarks for the prevalence of outlier specimens. However, applying the Kruskal-Wallis test at certain nodes of the tree revealed significant differences in certain continuous characters. We propose that these inferred differences represent apomorphies. Ultimately, our results suggest that of the four species currently accepted, both *T. mitchellae* and *T. novaehollandiae* should be synonymized as *T. novaehollandiae* is paraphyletic with respect to *T. mitchellae*. In addition, *T. darlingtoni* nests within the *T. richmondiana* lineage, and both taxa are also considered synonyms. Distinct character states of the umbilicus support the existence of two independent lineages. One with an open (*T.* sp1 + *T.* sp2) and one with a closed umbilicus (*T. novaehollandiae* + *T. richmondiana*). The Kruskal-Wallis tests

support the recognition of four distinct species, *T. novaehollandiae* and *T. richmondiana* plus two new species.

Introduction

Species delimitation is fundamental to biological and evolutionary studies (Simpson, 1951). Morphological characters are commonly employed for this purpose (Zapata & Jiménez, 2012), and traditionally taxonomists have exclusively relied on morphological traits to define species for a long time indeed (Blackwelder, 1967). Morphological characters can broadly be categorized as either discrete or continuous. When discrete characters are available, qualitative assessments are typically used in taxonomic inferences (Farris, 1998; Swofford, 2003; Goloboff and Morales, 2023). In contrast, when characters are continuous, statistical methods are often applied to distinguish character states. However, the statistical approach has proven to be of limited value because the homology of characters is not considered (Sneath and Sokal, 1973).

The challenge of delimiting species is further compounded by the lack of standardized approaches for processing the different types of characters mentioned above. Moreover, various species concepts differ in how they select, process, estimate, and apply characters. Wilkins (2018) catalogued 28 species concepts, among which the phylospecies concept (Hennig, 1950) is one of the more significant ones (Wilkins, 2018; Odenbaugh, 2022). This concept defined species as the smallest diagnosable monophyletic biological entity identifiable by apomorphies (Hennig, 1950; Wilkins, 2018). However, challenges arise for groups lacking discrete morphological apomorphies even when significant molecular differences may be evident. This issue is exacerbated by the prevalent preference for discrete characters as apomorphies over continuous ones (Rae, 1998). Although efforts have been made to incorporate continuous data into phylogenetics (Thiele, 1993; Goloboff et al., 2006) and even landmark datasets (Zelditch et al., 1995), these approaches have largely fallen short of establishing reliable apomorphies that are derived from continuous characters.

Integrative species concepts promise to overcome the limitation of artificial species boundaries by combining multiple criteria. For example, Hong (2020) proposed the gen-morph species concept, which builds upon de Queiroz's (1998) general lineage species concept and integrates several existing approaches to enhance operability. A key requirement of this concept is that a species must exhibit at least two distinct, statistically discontinuous apomorphies. However, when species occupy broad ecological niches with high niche variability while character variation is constrained, a continuous overlap in the morphospace (Budd, 2021) of taxa may occur (Foote, 1994; Paup, 1966; Erwin, 2006). Therefore, the gen-morph species concept is challenged by saturation in morphospace. Moreover, Mayden's (1997) and Wiley and Lieberman (2011) argued that operational species concepts or species concepts requiring additional criteria ultimately capture only a subset of biodiversity.

The Australian land snail *Thersites* Pfeiffer, 1855 (Camaenidae) exemplifies the challenges of delimiting species under these constraints. The members of this group exhibit predominantly continuous differences in morphological characters while they may also face morphospace saturation due to ecological, functional, and genetic constraints. Currently, the genus comprises four accepted species: *T.*

richmondiana (Reeve, 1852), *T. novaehollandiae* (Gray, 1834), *T. mitchellae* (Cox, 1864), and *T. darlingtoni* Clench & Archer, 1938 (Stanisic et al., 2010). Bishop (1978) reviewed the genus based on comparative morphology using continuous characters derived from the shell and reproductive organs. In this revision, Bishop (1978) treated *T. darlingtoni* as a junior synonym of *T. richmondiana*. However, subsequently Stanisic et al. (2010) treated *T. darlingtoni* as an accepted species for being 'distinguishable by smaller size and a more rounded body whorl'. As noted above, the lack of discrete diagnostic features hampers the identification of unambiguous apomorphies rendering the current classification phenetic. To complicate matters, a recent molecular phylogenetic study revealed that *Thersites novaehollandiae*, a species characterized by a variable morphology and a wide distribution, is made up of multiple mitochondrial lineages, and is paraphyletic with respect to *T. mitchellae* (Hugall & Stanisic, 2011). This finding contradicts the earlier taxonomic classification of Bishop (1978) who accepted *T. novaehollandiae* and *T. mitchellae* as two distinct species. The inconsistency between mitochondrial phylogenetics and morphology-based taxon delineation by which two morphologically distinguishable forms are not mutually monophyletic in a phylogenetic tree underscore the broader challenges of delineating species by emphasizing continuous traits.

Thersites species inhabit a variety of wet to deciduous forest types throughout eastern New South Wales and south-eastern Queensland. However, recent surveys in more xeric landscapes west of the Great Dividing Range have recorded presumedly undescribed species of *Thersites*, which implies that this genus may not be limited solely to the mesic forests in central-eastern Australia (Murphy and Shea, 2015, 2023). These potentially new species, *Thersites* sp. "MtKaputar" (Murphy and Shea, 2015) and *Thersites* sp. "Coolah Tops" (Murphy and Shea, 2023), have been discovered in drier sclerophyll woodlands in central-eastern New South Wales. However, the taxonomic status of these two candidate taxa remains to be clarified.

The evolutionary history of *Thersites* is poorly understood, partly because previous studies have not sampled all taxa comprehensively and relied exclusively on mitochondrial DNA. In this study, we employe a genome-wide SNP dataset covering multiple populations across all accepted taxa to test hypotheses concerning the species-level classification as well as the evolutionary history of this group. Given the lack of clearly distinguishing, especially categorical, characters, we aimed to objectively quantify various allegedly diagnostic traits to derive additional categorical characters that could serve as potential apomorphies in species delimitation. These characters include Voronoi distances used to assess soft body features and landmark datasets used to assess shell shapes.

Materials and Methods

Our methodology pipeline diagram is shown as Figure 1.

Taxonomic classification

The latest taxonomic classification of *Thersites* by Stanisic et al. (2010) serves as our null hypothesis in species delineation. In addition, our null hypothesis includes the two supposedly undescribed species recognised by Murphy and Shea (2015, 2023).

Materials examined

Vouchers studied here, including types, are kept in the collections of the Australian Museum, Sydney (AM), Queensland Museum, Brisbane (QM), Museum Victoria, Melbourne (NMV), Natural History Museum, London (NHMUK), and the Museum of Comparative Zoology, Cambridge, Mass. (MCZ).

Data repository and analytical scripts used

All used scripts used in this manuscript and related instructions have been deposited in the GitHub repository (http://github.com/starsareintherose/Thersites-Data). Datasets are available from Dryad (https://datadryad.org/stash/dataset/doi: 10.5061/dryad.2rbnzs81c). Operations were taken under BioArchLinux environment (Zhang et al., 2025)

Species occurrences

We used QGIS to plot all available specimen records on a map, categorized by originally assigned taxon names. Occurrences of specimens included in the molecular analyses are marked with a star symbol.

Shell measurements

We analyzed the shell dimensions of 323 adult specimens to test if the shell width–height ratio is suitable to distinguish species. We applied focus stacking to generate high-resolution images. Subsequently, the shells were extracted from the background and resized to a uniform pixel scale using Adobe Photoshop 2024 or GIMP 2.10 with G'MIC-Qt plugin. Next, we converted all images to a white-shell and black-background format using ImageMagick. Finally, we utilized a custom FIJI macro script to measure shell width and height, which were then visualized as scatter plots using matplotlib implemented in Python.

After excluding broken shells unsuitable for morphometric analysis, we conducted a landmark-based morphometric analysis on 321 aperture view photographs to assess shell shape variation. Fixed landmarks and two semi-landmark curves were digitized using tpsDig 2.32 (Rohlf, 2021a). The points along each curve were resampled into 50 equidistant points by length, and both semi-landmarks and landmarks were treated equally in the analysis. Specifically, we digitized two curves. One traces the left

outline of the body whorl and the other delineates the aperture shape and seven landmarks, whose positions are shown in Fig. 2. We then analyzed three sets of landmarks or semi-landmarks separately: the two curves as distinct groups and the seven landmarks as a single group. Given that *Thersites richmondiana* exhibits significantly different shell shapes, these specimens were treated as a negative control. For each individual group and the combined dataset, we performed the morphometric analyses independently. In each analysis, we applied a Procrustes fit by aligning the principal axes to mitigate potential manual rotation errors. Covariant distance matrices and principal component analysis (PCA) were then computed using MorphoJ 1.08.02 (Klingenberg, 2011), and PCA scatter plots with convex hulls were visualized. Finally, we calculated the average observations for initial hypothesis and visualized them using the thin plate spline network method in MorphoJ. Following these separate analyses, we examined the scatter plots to ensure that the negative control specimens were not completely intermixed with the other groups. After filtering out the negative control mixing group, we combined all points and repeated the analysis.

Genital system measurements

Bishop (1978) stated that the flagellum is shorter than 1/4 of the length of the proximal epiphallus in *T. novaehollandiae*, about half as long in *T. richmondiana*, and about as long as the proximal epiphallus in *T. mitchellae*. We measured the genital systems of 89 specimens using FIJI (Schindelin et al., 2012), separately quantifying the lengths of phallus, epiphallus, and flagellum. Epiphallus1 and epiphallus2 are delimited by the junction with the penial retractor muscle (Fig. 3). Epiphallus1 is the part of the epiphallus proximal and epiphallus2 distal of that junction (Fig. 3). Measurements of genital organs are visualized using histograms. We performed Shapiro-Wilk's test to test for normal distributions. The proximal epiphallus to the flagellum ratio was then visualized in violin maps and scatter plots. We tested Bishop's (1978) hypothesis that species differ in the relative lengths of their genital organs by analysing the ratios between the lengths of the proximal ephiphallus, whole phallus, epiphallus1, phallus + epiphallus1, phallus + epiphallus, respectively. Additionally, we plotted the epihallus1 – epiphallus ratio in a violin map and scatter plot. We performed a Kruskal–Walli's test to assess significant differences among the groups, followed by Dunn's test as a post-hoc analysis to determine the specific groupings based on the Kruskal–Wallis results.

We evaluated the flagellum shape using width measurements and landmark-based morphometrics. After discarding incomplete or broken flagellum unsuitable for analysis, candidate photographs were standardized to a uniform pixel scale and converted to a white flagellum on a black background using ImageMagick. Voronoi distances were computed using FIJI and the width profile tools, and the resulting flagellum shape metrics for 81 specimens were analyzed. After collecting the Voronoi distance data, we used a custom Python script to resample the distance points to 100 points per specimen. We retained

the 50% and top 70% of points from the flagellum tip separately, focusing on tip variation. The datasets corresponding to 100%, 70%, and 50% of the flagellum lengths were analyzed using the following methods. We visualized the flagellum shape curves using matplotlib in Python and converted the resampled points to TPS files via a custom Python script. Subsequently, all landmarks were analyzed in MorphoJ, where we performed a Procrustes fit, computed covariant distance matrices, and conducted PCA. We also calculated the average observations by group. Finally, PCA results were visualized using scatter plots and convex hulls, and group average observations were displayed using the thin plate spline network.

SNP generation and analyses

DNA extracts from 80 individuals were used to perform DArTseq high-density genotyping (DArT Pty Ltd., Canberra, Australia), a SNP-based population genetic method. The workflow of this commercial application includes a double restriction enzyme-mediated genome complexity reduction method to target multiple loci that are randomly distributed across the genome and uses next-generation sequencing platforms to sequence the resulting DNA fragments (Melville et al., 2017). For our current dataset, the PstI/HpaII enzyme combination was employed to generate short DNA fragments, which were sequenced as single-ended reads on an Illumina Hiseq 2500 platform. These reads were processed using a proprietary DArT analytical pipeline (Jaccoud et al. 2001) to obtain a SNP matrix for subsequent analyses.

We used the dartR package for the R statistical environment (R Core Team, 2025) to process and analyse the SNP matrix. Initially, we filtered the data as follows: exclude loci with a read depth outside the 5–30, removal of markers with reproducibility scores below 100%, exclude loci with coverage of below 80% of all individuals, and excluding individuals with coverage of below 50% of all loci, removal of all monomorphic markers. We also carried out quality control by filtering out SNPs with minor allele frequencies below 0.01 and randomly removing secondary markers to mitigate potential linkage. We calculated the Principal Coordinates Analysis (PCoA) to visualize the genetic structure using ggplot2. Finally, we converted the SNP dataset to FASTA format files for subsequent phylogenetic analyses.

Molecular phylogenetic analyses

We employed the SNP data to reconstruct the phylogenetic relationships among the sequenced specimens. We reconstructed Maximum Parsimony (MP) phylogenies using TNT 1.6 (Goloboff and Morales, 2023) with guoyi.run (Zhang, 2025) under three weighting schemes: Equal weighting, implied weighting, and extended implied weighting. For implied weighting and extended implied weighting,

we set the K value to 12 following the recommendations of Goloboff et al. (2017). The script performed 1,000 iterations of TBR branch swapping to generate Maximum Parsimony Trees (MPTs), which were subsequently summarized using the strict consensus method to produce a strict tree. Group supports were estimated using 1,000 jackknifing iterations and 1,000 symmetric resampling replicates. Additionally, we calculated the consistency index (CI), retention index (RI), and tree length (TL) via guoyi.run.

For model-based phylogenetic analyses, we employed modeltest-ng (Darriba et al., 2020) to identify the most suitable substitution model base on Bayesian Information Criterion (BIC), incorporating 16 gamma categories. The appropriate model was implemented in a Maximum Likelihood (ML) analysis performed by using raxml-ng. We utilized 10 parsimony starting trees and 10 random starting trees and applied the SPR branch swapping to obtain the best Maximum Likelihood tree. Group support was estimated by performing 1,000 Felsenstein bootstrap replicates. We reconstructed a Bayesian Inference (BI) phylogeny by using MrBayes 3.2.7 (Ronquist et al., 2012) by running two independent runs for 2,000,000 generations, sampling every 2,000 generations, with a chain temperature of 0.05. The final consensus tree and posterior probabilities were determined from the last 50% of all sampled trees (burn-in = 50%).

ML bootstrap (BS), MP jackknifing (JK), symmetric resampling (SR) values above 90% and Bayesian posterior probabilities (BPP) above 0.9 are considered as indicating strong support; values of BS/JK/SR >70% or BPP >0.8 indicate good support; lower values are regarded as insufficient support.

Morphological character optimization analysis

We used phytools (Revell, 2024) to visualize continuous measurements along the phylogenetic tree. Specifically, we mapped continuous character states relating to shell width, shell height, lengths of phallus, epiphallus1, epiphallus2, and flagellum, onto the molecular phylogeny. We imputed missing data using a lineage-specific approach. We applied a Kruskal-Wallis test to determine whether continuous dimensions significantly differed among taxa, and the results were visualized using heatmaps. Our analysis aggregated data from descendant taxa at internal nodes and calculated p-values for each variable. If any variable has an empty group (i.e., no valid data), the test is skipped, and the p-value is assigned as n/a (not applicable).

We used the nts files generated from tpsRelw (Rohlf, 2021b) alongside raw landmark data to categorize the landmarks as discrete characters. Silhouette scores were first computed for the variation matrix to determine the optimal number of clusters (best K value). We then applied K-means clustering with this optimal K to assign specimens into two clusters. This analysis was performed separately for four groups

of landmarks: shell landmarks, body whorl semi-landmarks, aperture semi-landmarks, and half-flagellum landmarks. For the raw landmarks, we additionally applied Generalized Procrustes Analysis (GPA) before generating a matrix to maintain consistency with the analyses conducted in MorphoJ.

Finally, after categorizing both the landmark and absence or presence of umbilicus, we mapped these characters onto the phylogeny to identify apomorphic and homoplastic character states. This was achieved using three character-optimization methods, fast optimization (ACCTRAN), unambiguous optimization, and slow optimization (DELTRAN), as implemented in WinClada (Nixon, 2021).

Results

Species distributions

Thersites species are primarily found in the forests of New South Wales and Queensland in eastern Australia (Fig. 4). *Thersites novaehollandiae* is the most widely distributed species, ranging from Newcastle to Coffs Harbour along the coast. Its range extends inland to include the Gibraltar and Main Ranges. *Thersites richmondiana* occurs in coastal areas from Lismore to the Sunshine Coast, while *T. mitchellae* inhabits the coastal rainforests from Lismore to the Gold Coast. *Thersites darlingtoni* is restricted to parts of the Border Ranges, *T.* sp1 is confined to Mount Kaputar, and *T.* sp2 to the Coolah Tops. The sequenced samples cover the known distributions of all species well except *T. richmondiana*.

Shell dimensions

Shell widths of all measured specimens ranged from 2.25 to 8.20 cm, with an average of 4.90 ± 0.91 cm. Shell heights varied from 1.58 to 6.41 cm, averaging 4.00 ± 0.67 cm. The width-to-height ratio spanned from 0.83 to 1.18, with an average of 0.83 ± 0.12. *Thersites* sp1 and sp2 were notably smaller than the currently recognized species (see Dryad support file). The two thresholds in height-to-width ratio of 0.7 and 0.8 did not delineate the morphospace of any taxon (Fig. S1).

The Shapiro-Wilk test suggested that shell height and width measurements did not follow normal distributions; however, the height-to-width ratio did (Fig. 5A-C). Dunn's test (Fig. 5D-F) indicated that *T. novaehollandiae*, *T.* sp1, and *T.* sp2 consistently clustered together based on these dimensions. When grouped by shell height, *T. darlingtoni* and *T. richmondiana* formed one group while *T. mitchellae* formed a separate group. When grouped by shell width, *T. mitchellae* and *T. darlingtoni* clustered together while *T. richmondiana* stood alone. Finally, based on the shell height-to-width ratio, *T.* sp1, *T. darlingtoni*, and *T. richmondiana* grouped together, *T.* sp2 and *T. novaehollandiae* formed second group, and *T. mitchellae* a third.

Shell landmark-based morphometrics analyses

In the overall landmarks PCA (Fig. 6A), PC1 accounted for 55.17% of the variation and PC2 for 20.06%. The scatter plots showed that *T. richmondiana* overlapped markedly with *T. mitchellae* and *T. darlingtoni*, and partially with *T. novaehollandiae*. Both, *T*. sp1 and *T*. sp2 fell within the range of *T. novaehollandiae*.

In the PCA of body whorl curve landmarks (Fig. 6B), PC1 explained 74.22% of the variation and PC2 accounted for 14.73%. The scatter plots indicated that *T. richmondiana* could clearly be distinguished from any other species, except for some overlap with *T. darlingtoni*. *Thersites mitchellae* and *T. novaehollandiae* largely overlapped, while *T. darlingtoni* overlap with both *T. novaehollandiae* to some extent and in part with *T. richmondiana*. Additionally, *T.* sp1 fell entirely within the range of *T. novaehollandiae*, whereas *T*. sp2 was similar to *T. darlingtoni* and *T. richmondiana*.

For the aperture curve landmarks PCA (Fig. 7A), PC1 explained 69.47% of the variation and PC2 21.04%. The scatter plots revealed that *T. richmondiana* could not be distinguished from any other species, as all species overlapped substantially with one another.

In the combined landmarks PCA (Fig. 8), PC1 explained 23.57% of the variation and PC2 explained 24.43%. Here, *T. richmondiana* overlapped almost entirely with *T. darlingtoni* and in parts with *T. novaehollandiae* and *T. mitchellae*. *Theristes mitchellae* and *T. novaehollandiae* overlapped almost completely with each other, while *T.* sp1 fell within their range. Notably, *T.* sp2 could be distinguished from the other species.

After removing the aperture curve landmarks PCA (Fig. 7B), which showed negative control mixing with other groups, the PCA yielded PC1 explaining 44.53% of the variation and PC2 explained 38.21%. The scatter plots indicated that *T. richmondiana* partially overlapped with *T. darlingtoni*. *Thersites mitchellae* and *T. novaehollandiae* showed small overlap, whereas *T. darlingtoni* exhibited partial overlap with *T. novaehollandiae*. *Thersites* sp1 fell within the range of *T. novaehollandiae*, and *T.* sp2 within the range of *T. richmondiana*.

Thin plate spline network maps (Figs 6-8) revealed that *T. richmondiana* and *T. darlingtoni* tended to exhibit a sharper keel on the body whorl, a more depressed body whorl, a narrower aperture opening, and an overall higher shell spire. By contrast, the maps for *T. mitchellae* and *T. novaehollandiae* indicated a rounder body whorl and a larger aperture. However, *T. mitchellae* tended to have a more elevated spire, while *T. novaehollandiae* typically had a lower spire. Furthermore, *T.* sp1 showed a trend towards a larger aperture, a rounder body whorl, and a lower spire, whereas *T.* sp2 tended to have a narrower aperture, a sharper body whorl, and a lower spire.

Genital system measurements

Histograms (Fig. 9) and the Shapiro-Wilk test both indicated that the lengths of various genital organs were normally distributed, except for the flagellum. Violin plots combined with Dunn's test results (Fig. 10) revealed significant differences in genital system dimensions among the groups. Based on flagellum length, *T. mitchellae*, *T. darlingtoni*, and *T. richmondiana* formed one group, while *T. novaehollandiae*, *T.* sp1, and *T.* sp2 formed another. Phallus measurements differentiated two distinct clusters, one consisting of *T.* sp1, *T.* sp2, and *T. darlingtoni*, and the other of all remaining species. Similarly, length of epiphallus1 clearly separated *T.* sp1 and *T.* sp2 from all other species. Moreover, both the lengths of ephallus2 and ephiphallus distinguished three groups: *T.* sp1 and *T*. sp2, *T. novaehollandiae*, and all remaining species. The analyses of the epiphallus and phallus data suggested that *T. novaehollandiae* grouped with *T. darlingtoni*, *T. mitchellae* clustered with *T. richmondiana*, and *T*. sp1 and *T*. sp2 consistently formed a separate group.

In contrast, ratio-based analyses (Fig. S2) were less discriminative than absolute measurements. The ratios of flagellum/epiphallus displayed a pattern, grouping *T. novaehollandiae*, *T*. sp1, and *T*. sp2 together. Likewise, the ratios of flagellum/epiphallus1, and flagellum/(phallus+epiphallus1), flagellum/epiphallus, flagellum/(phallus+epiphallus), and epiphallus1/epiphallus indicated that *T.* sp2 and *T. novaehollandiae* clustered together, with the other species forming a separate group.

Flagellum shape variation analysis

No clear pattern has been observed for the curves based on 100% and 70% of flagellum length (Figs S3-S4). However, the curve for the distal half of the flagellum (50% length) indicated that both *Thersites michellae* and *T. novaehollandiae* exhibited a similar trend: the flagellum was slimmer and thinner toward the tip. *Thersites* sp1 and *T.* sp2 displayed a similar pattern, while *T. darlingtoni* and *T. richmondiana* showed an initial thickening of the flagellum followed by a tapering tip.

Principal component analysis (PCA) (Fig. S3) based on the 100% flagellum length revealed that PC1 accounted for 38.505% of the variation and PC2 explained 24.703%. *Thersites novaehollandiae* encompassed nearly all the variation in *Thersites* flagellum shape, clustering tightly together. Similar patterns have been observed in PCA based on 70% flagellum length (Fig. S4; PC1: 39.539%, PC2: 30.199%) and 50% flagellum length (Fig. 11; PC1: 62.218%, PC2: 10.816%). Most *T. novaehollandiae* samples clustered with *T. michellae*, *T.* sp1, and *T.* sp2, including one sample from *T. darlingtoni*. Meanwhile, *T. richmondiana* and the majority of *T. darlingtoni* samples grouped together, even incorporating some *T. novaehollandiae* specimens.

SNP analysis

The initial SNP dataset included 80 genotypes and had a total length of 103,926 SNPs. After filtering,

73 genotypes containing 765 SNPs remained. The underlying multiple sequence alignment contained 52,408 bp, including 10 % of missing data. In the PCoA scatter plot (Fig. 12), PC1 accounted for 33.3% of the variation and PC2 for 12.2%. *Thersites novaehollandiae*, *T. mitchellae*, *T. darlingtoni*, and *T. richmondiana* clustered closely along PC1, whereas *T.* sp1 and *T.* sp2 were more distinctly separated on this axis. Notably, some individuals of *T. darlingtoni* grouped with *T. richmondiana*. *Thersites mitchellae* and *T. novaehollandiae* were positioned near each other with *T. mitchellae* exhibiting considerable variation along PC2.

Molecular phylogenetic analysis

The topologies produced by extended implied weighting (EIW) and implied weighting (IW) under Maximum Parsimony (MP) were identical (Fig. 13A). However, the equal weighting (EW) method tended to resolve some lineages as polytomies (Fig. S5). However, this did not affect the phylogenetic position of any nominal taxon. Moreover, compared with the topologies obtained by employing Bayesian Inference (BI) and Maximum Likelihood (ML), the EIW/IW approach under MP more frequently resulted in unresolved nodes (Fig. 13B). Notably, the topologies inferred by BI and ML were identical.

All analyses revealed two principal clades. The first clade comprised *Thersites richmondiana*, *T. darlingtoni*, *T. novaehollandiae*, and *T. mitchellae*, whereas the second consisted of *T.* sp1 and *T.* sp2. Within the first clade, *T. richmondiana* and *T. darlingtoni* formed a subclade, and *T. novaehollandiae* and *T. mitchellae* another.

The clade containing *Thersites richmondiana*, *T. darlingtoni*, *T. novaehollandiae*, and *T. mitchellae* was maximally supported in all analyses (jackknifing, JK = 100; symmetric resampling, SR = 100; Bayesian posterior probability, BPP = 1.00; bootstrap, BS = 100). Within this clade, *T. novaehollandiae* has been recovered as paraphyletic because *T. mitchellae* occupied a crown group position nested among *T. novaehollandiae* lineages. The monophyly of *T. mitchellae* was maximally supported by jackknifing (JK = 100) and symmetric resampling (SR = 100) in the MP trees, and by Bayesian posterior probability (BPP = 1.00) and ML bootstrap (BS = 100) in BI and ML analyses. Furthermore, *T. mitchellae* and *T. novaehollandiae* together consistently formed a strongly supported monophyletic group (JK = 95, BPP = 1.00, SR = 89, BS = 86). In the MP analyses, *T. richmondiana* and *T. darlingtoni* clustered together with maximal support (JK = 100; SR = 100, BPP = 1.00; BS = 100) in all analyses. However, *T. darlingtoni* occupied the crown position, rendering *T. richmondiana* paraphyletic. The monophyly of *T. darlingtoni* is strongly supported by BI (BPP=0.98) but weakly by ML (BS=74).

The second main clade comprising *T.* sp1 and *T.* sp2 is maximally supported across all analyses (JK = 100; SR = 100; BPP = 1.00; BS = 100), with both taxa forming a maximally supported sister group (JK = 100; SR = 100; BPP = 1.00; BS = 100).

Morphological character optimization analysis

The heatmap-coloured phylogenies (Fig. 14) indicated that continuous measurements varied substantially among species. We found not clear differences between the inter- and intraspecific ranges of variation. We found that *T*. sp1 and *T*. sp2 were consistently smaller than all other species. *Thersites mitchellae* and *T. novaehollandiae* were found to be highly variable while *T. darlingtoni* differed little from *T. richmondiana*. The Kruskal-Wallis test heatmaps plotted onto the phylogeny (Fig. 16) revealing statistically significant differences between the species pair *T*. sp1 + *T*. sp2 and the group of *T. darlingtoni* , *T. richmondiana*, *T. mitchellae*, and *T. novaehollandiae* in shell width and height as well as lengths of phallus, epiphallus1, epiphallus2, and flagellum. In addition, significant differences have been observed between *T*. sp1 and *T*. sp2 in shell width and height as well as in the lengths of epiphallus2 and flagellum. Comparisons of *T. darlingtoni* and *T. richmondiana* with *T. mitchellae* and *T. novaehollandiae* revealed significant differences in shell width and height, epiphallus2 and flagellum. The differences between *T. mitchellae* and the crown group of *T. novaehollandiae* were significant for shell width and the lengths of phallus and flagellum. Other significant differences have also been detected among various *T. novaehollandiae* lineages.

Silhouette scores for all landmarks (Fig. 15) indicated that the optimal clustering was achieved with two clusters. Based on flagellum shape, *T. novaehollandiae*, *T. darlingtoni*, and *T. richmondiana* are distributed across both clusters, whereas *T.* sp1, *T.* sp2, and *T. mitchellae* occurred exclusively in cluster 1. When considering all landmarks, all species were represented in both clusters except *T.* sp1 and *T.* sp2. Analysis of body whorl semi-landmarks showed that *T.* sp2 was present in both clusters, nearly all *T. richmondiana* were confined to cluster 1, and almost all *T. darlingtoni* were found in cluster 0, with the remaining species occurring in a single cluster. Regarding aperture shape, *T.* sp2 appeared only in one cluster, nearly all *T. darlingtoni* grouped in a single cluster, while the other species were distributed across both clusters.

Under both fast and slow optimization (Fig. 16), the only apomorphy that has been identified supports the sister group relationship between *T.* sp1 and *T.* sp2. This apomorphy is an open umbilicus. By contrast, we identified multiple homoplasies across various nodes in the tree.

Discussion

Taxonomic significance of shell and genital morphometrics

In the latest taxonomic revision, Bishop (1978) employed simple ratios in shell and genital dimensions to distinguish *Thersites* species claiming that they can reliably be distinguished by means of two independent, continuous character states: Height/width ratio of the shell and the ratio between flagellum

and proximal epiphallus lengths. Specifically, Bishop (1978) stated that for height/width ratio, *T. mitchellae* exhibits a height–width ratio greater than 0.8, whereas *T. richmondiana*, *T. darlingtoni* and *T. novaehollandiae* show ratios smaller than 0.7. For ratio between flagellum and proximal epiphallus, *T. novaehollandiae* exhabits the ratio lower than 0.25, *T. richmondiana* and *T. darlingtoni* are about 0.5, *T. mitchellae*'s approach 1. However, none of these delimitation boundaries worked, we also found no boundary within any other ratios (Figs S1, S2). Even when constraining the range of the numerator or denominator, overlap persists between species.

While Bishop (1978) synonymized *T. darlingtoni* with *T. richmondiana*, subsequently Stanisic et al. (2010) treated the former as an accepted species for being distinguished from *T. richmondiana* by shell size and smaller size and in having rounded instead of flat whorls. However, this statement was not backed up by an analysis of any data (Fig. 6).

We tested the classification proposed by Stanisic et al. (2010) by rigorously analysing the supposedly diagnostic character states in a statistical framework as well as conducting genomic analyses. Our comparative analyses of morphological traits revealed that the four currently accepted *Thersites* species cannot be reliably distinguished using shell shape characteristics nor by using reproductive traits, such as flagellum length as incorrectly stated by Bishop (1978). In fact, based on the character distributions alone, we cannot recognize distinguishable groups. The Shapiro–Wilk test indicated that measurements, including shell width, shell height, and flagellum length, were not normally distributed. However, histograms failed to reveal distinct states in any of these characters. That means that taxa could not be reliably distinguished by means of nay of these characters. Moreover, the measurements of genital organs and shells revealed no clear separation between the amounts of intraspecific and interspecific variation in taxa as previously delineated (Figs 5, 9, 10).

Species monophyly

As the current morphology-based classification has not withstood statistical scrutiny, but was also not rejected outright, we also employed a molecular phylogeny to put the initial taxonomic hypothesis to a test. The phylospecies concept defines species as the smallest monophyletic group that is distinguished from other such groups by one or more apomorphies (Wilkins, 2018). Consequently, it is also required that species are monophyletic in phylogenetic trees. Our molecular phylogeny of *Thersites* based on analyses of nuclear loci revealed that two species were non-monophyletic, however. First, we showed that *T. novaehollandiae* encompassed several lineages, one of which is represented by *T. mitchellae* rendering *T. novaehollandiae* a paraphyletic assemblage. This finding corresponds well with the mitochondrial phylogeny of Hugall and Stanisic (2011), who also recovered *T. novaehollandiae* as paraphyletic with respect to *T. mitchellae*. Second, we also found that *T. darlingtoni* and *T. richmondiana* were mutually non-monophyletic (Fig. 13, Fig S5). Accordingly, neither *T. mitchellae*

nor *T. darlingtoni* as currently delimited should be accepted as distinct species based on the criterium of monophyly. However, before a final taxonomic conclusion can be drawn from the available tree, we need to test if taxa can be delineated in alternative ways, such as being monophyletic groups that are supported by apomorphies. To this end, we have embarked on a strategy of character optimization as outlined in the following.

Morphological character optimization

We use character optimization, i.e., the mapping of observed character states at the tree tips and the reconstruction of ancestral character states across the given tree, to identify morphological apomorphies across the tree. By doing that, we simultaneously test any possible alternative taxonomic hypothesis through the mapping of both discrete and continuous morphological characters.

Traditionally, discrete characters have been used to identify apomorphies. However, in the present case the number of discrete characters is insufficient even to support the revised taxonomic hypothesis that there are just two accepted species, *T. novaehollandiae* and *T. richmondiana*, and two undescribed species, *Thersites* sp1 and sp2. When ignoring the missing observations of this character, there is a single character with two states (umbilicus open/closed), which distinguishes the two undescribed species from the other members of the genus as a synapomorphy (= open umbilicus).

In search of additional apomorphies, we categorized the landmark datasets for shell and genital features. This has not been done before, although there are some methods to use landmarks in the framework of cladistic analyses (e.g. Goloboff and Catalano, 2015). Our Silhouette score analyses suggested that the most robust categorizations for these landmark datasets always yielded two states, with individual species often displaying both. Consequently, our attempt to derive apomorphies by discretizing the landmark covariance matrix failed because of low homology in the shapes described, echoing earlier efforts by Zelditch et al. (1995).

Other methods, such as Contrasting Composite and Reductive Coding (Torres-Montúfar et al., 2018), have been proposed to resolve homoplasy issues that represent interpretation errors of characters (Nixon and Carpenter, 2002). As noted by Torres-Montúfar et al. (2018), complex characters should be iteratively decomposed into simpler, more homologous units. However, when the traits (such as shell shape) are heavily influenced by environmental factors (Goodfriend, 1986), this process becomes particularly challenging.

Statistical tests did not recognize natural groups when applied globally to all samples (Figs 5, 9). However, Krustal-Wallis tests applied at specific phylogenetic nodes supported specific lineages with distinct morphological character. In our phylogeny, three nodes supported the four morphologically distinct lineages according to the Krustal-Wallis tests: First, the species pair *T.* sp1 + *T.* sp2 and the

species pair *T. novaehollandiae* + *T. richmondiana* differed in the sizes of all six measured organs (i.e., shell width and height, lengths of phallus, epiphallus1, epiphallus2, and flagellum; $p < 0.05$). Second, the species in each of these two sister pairs could be distinguished from one another by means of shell width and height as well as the lengths of epihphallus2 and flagellum ($p < 0.05$).

However, it is important to recognize that statistical clustering primarily reflects similarity rather than true homology (Sneath and Sokal, 1973). We try to avoid this caveat here by applying statistical tests only to natural groups that have been recognized in a phylogenetic framework. While across the entire group, most morphological characters overlap in their ranges, we found that they can differ with statistical significance at specific nodes in the phylogeny. We argue that these morphological differences can serve as apomorphies.

Finally, our findings challenge the gen–morph species concept (Hong, 2020), which requires species to possess at least two statistically discontinuous apomorphies. This criterion is unlikely to be met in groups that are characterized by few, overlapping characters. In a saturated morphospace (Budd, 2021), morphological variation tends to be continuous, reflecting gradual shifts rather than clear-cut, discrete differences (Foote, 1994; Paup, 1966; Erwin, 2006). Such gradual shifts are also encountered in the present case. A strict application of the gen-morph species concept to the case of *Thersites* would lead to the recognition of just a single species as an artifact of lacking characters rather than a realistic reflection of the true diversity of the group. This illustrates that by enforcing a prescribed set of diagnostic criteria, one may only detect a subset of the actual species-level diversity (see also Wiley and Lieberman, 2011). Our results underscore that attempts to integrate all aspects of species delimitation into a single framework are unlikely to be universally applicable, such as the gen-morph species concept. Therefore, we agree in principle with the critique of Mayden's (1997), that species concepts should not be operational.

How to use morphological characters

We used *Thersites* as an example to explore how best to incorporate analyses of continuous characters to recognize apomorphies for taxa with a limited number of discrete characters. To this end, we aim to objectively define character status. As Pimentcl & Riggins (1987) and Stevens (1991) pointed out, that without delimiting character status objectively there can be no meaningful morphological character analyses ("garbage in, garbage out"). Our study makes steps toward the mathematization of systematics, aligning with the ideal of the '*mathesis universalis*', which Hennig (1949) described as 'all goals of all true science'. Hence, systematics is not a 'meeting ground of art and science' as Davis & Heywood (1963) have claimed.

Acknowledgements

Thanks go to Alison Miller (AM), Jennifer Caiza (AM), John Stanisic (QM), Darryl Potter (QM), Jennifer Trimble (MCZ), Gonzalo Giribet (MCZ), Jon Ablett (NHM) for making material under their care available to this study and/or for providing photographs of specimens. We thank Craig Stehn for helping with field work. We thank the Willi Hennig Society for providing access to the software TNT. The molecular component of this study was funded by the NSW Department of Climate Change, Energy, the Environment and Water, which is thankfully acknowledged.

References

Bishop, M.J. 1978. A revision of the genus *Thersites* Pfeiffer (Pulmonata: Camaenidae). Journal of the Malacological Society of Australia, 4, 9–21. https://doi.org/10.1080/00852988.1978.10673904

Blackwelder, R.E. 1967. Taxonomy: A text and reference book. J. Wiley, New York.

Budd, G.E. 2021. Morphospace. Current Biology, 31, R1181–R1185. https://doi.org/10.1016/j.cub.2021.08.040

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T. 2019. ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. Molecular Biology and Evolution, 37, 291–294. https://doi.org/10.1093/molbev/msz189

Davis, P.H., Heywood, V.H. 1963. Principles of angiosperm taxonomy. Oliver & Boyd.

Erwin, D.H. 2006. Evolutionary contingency. Current Biology, 16, R825–R826. https://doi.org/10.1016/j.cub.2006.08.076

Farris, J. 1988. Hennig86, version 1.5.

Foote, M. 1994. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. Paleobiology, 20, 320–344. https://doi.org/10.1017/s009483730001280x

Goloboff, P.A., Catalano, S.A. 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. Cladistics, 32, 221–238. https://doi.org/10.1111/cla.12160

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

Goloboff, P.A., Morales, M.E. 2023. TNT version 1.6, with a graphical interface for macOS and Linux, including new routines in parallel. Cladistics, 39, 144–153. https://doi.org/10.1111/cla.12524

Goloboff, P.A., Torres, A., Arias, J.S. 2017. Weighted parsimony outperforms other methods of phylogenetic inference under models appropriate for morphology. Cladistics, 34, 407–437.

https://doi.org/10.1111/cla.12205

Goodfriend, G.A. 1986. Variation in land-snail shell form and size and its causes: A review. Systematic Zoology, 35, 204–223. https://doi.org/10.1093/sysbio/35.2.204

Gray, J.E. 1834. Exhibition of an extensive series of land and freshwater shells regarded as hitherto undescribed. Proceedings of the Zoological Society of London, 2, 63–68.

Hennig, W. 1950. Grundzüge einer Theorie der Phylogenetischen Systematik. Deutschen Entomologischen Institut Berlin, Berlin.

Hennig, W. 1949. Zur Klärung einiger Begriffe der Phylogenetischen Systematik. Forschungen und Fortschritte, 25, 136–138.

Hong, D. 2020. Gen-morph species concept—a new and integrative species concept for outbreeding organisms. Journal of Systematics and Evolution, 58, 725–742. https://doi.org/10.1111/jse.12660

Hugall, A.F., Stanisic, J. 2011. Beyond the prolegomenon: A molecular phylogeny of the australian camaenid land snail radiation: Australian camaenid phylogeny. Zoological Journal of the Linnean Society, 161, 531–572. https://doi.org/10.1111/j.1096-3642.2010.00644.x

Jaccoud, D., Peng, K., Feinstein, D., Kilian, A. 2001. Diversity arrays: A solid state technology for sequence information independent genotyping. Nucleic Acids Research, 29, 25e–225. https://doi.org/10.1093/nar/29.4.e25

Klingenberg, C.P. 2010. MorphoJ: An integrated software package for geometric morphometrics. Molecular Ecology Resources, 11, 353–357. https://doi.org/10.1111/j.1755-0998.2010.02924.x

Mayden, R.L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem, in: Claridge, M.F., Dawah, H.A., Wilson, M.R. (Eds.), Species: The units of Diversity, Chapman & Hall, pp. 381–423.

Melville, J., Haines, M.L., Boysen, K., Hodkinson, L., Kilian, A., Smith Date, K.L., Potvin, D.A., Parris, K.M. 2017. Identifying hybridization and admixture using SNPs: Application of the DArTseq platform in phylogeographic research on Vertebrates. Royal Society Open Science, 4, 161061. https://doi.org/10.1098/rsos.161061

Murphy, M.J., Shea, M. 2023. Survey and conservation assessment of the land snail fauna of coolah tops national park in the hunter valley area of New South Wales, Australia. Molluscan Research, 43, 61–73. https://doi.org/10.1080/13235818.2023.2183538

Murphy, M.J., Shea, M. 2015. Survey of the land snail fauna (gastropoda: Pulmonata) of mount kaputar national park in northern inland new south wales, australia, including a description of the listing of australia's first legally recognised endangered land snail community. Molluscan Research, 35, 51–

64. https://doi.org/10.1080/13235818.2014.948147

Nixon, K. 2021. WinClada, version 2.0. published by the author, Ithaca, New York,.

Nixon, K.C., Carpenter, J.M. 2002. On homology. Cladistics, 28, 160–169. https://doi.org/10.1111/j.1096-0031.2011.00371.x

Odenbaugh, J. 2022. Species problems and beyond: Contemporary issues in philosophy and practice, in: Wilkins, J.S., Zachos, F.E., Pavlinov, I.Ja. (Eds.), Species and Systematics. CRC Press, Boca Raton, pp. 91–103.

Pimentcl, R.A., Riggins, R. 1987. The nature of cladistic data. Cladistics, 3, 201–209. https://doi.org/10.1111/j.1096-0031.1987.tb00508.x

Queiroz, K. de. 1998. The General Lineage Concept of Species, Species Criteria, and the Process of Speciation: A Conceptual Unification and Terminological Recommendations, in: Howard, D.J., Berlocher, S.H. (Eds.). Oxford University Press, p. 0. https://doi.org/10.1093/oso/9780195109009.003.0005

R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rae, T.C. 1998. The logical basis for the use of continuous characters in phylogenetic systematics. Cladistics, 14, 221–228. https://doi.org/10.1111/j.1096-0031.1998.tb00335.x

Raup, D.M. 1966. Geometric Analysis of Shell Coiling: General Problems. Journal of Paleontology, 40, 1178–1190.

Revell, L.J. 2024. Phytools 2.0: An updated r ecosystem for phylogenetic comparative methods (and other things). PeerJ, 12, e16505. https://doi.org/10.7717/peerj.16505

Rohlf, F.J. 2021b. tpsRelw, version 1.75.

Rohlf, F.J. 2021a. tpsDig, version 2.32.

Ronquist, F., Teslenko, M., Mark, P. van der, Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology, 61, 539–542. https://doi.org/10.1093/sysbio/sys029

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A. 2012. Fiji: An open-source platform for biological-image analysis. Nature Methods, 9, 676–682. https://doi.org/10.1038/nmeth.2019

Simpson, G.G. 1951. The species concept. Evolution, 5, 285–298. https://doi.org/10.1111/j.1558-5646.1951.tb02788.x

Sneath, P.H.A., Sokal, R.R. 1973. Numerical taxonomy: The principles and practice of numerical classification, A series of books in biology. Freeman, San Francisco.

Stanisic, J., Shea, M., Potter, D., Griffiths, O. 2010. Australian land snails. Bioculture Press, Mauritius.

Stevens, P.F. 1991. Character states, morphological variation, and phylogenetic analysis: A review. Systematic Botany, 16, 553–583. https://doi.org/10.2307/2419343

Swofford, D. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.

Thiele, K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. Cladistics, 9, 275–304. https://doi.org/10.1111/j.1096-0031.1993.tb00226.x

Torres-Montúfar, A., Borsch, T., Ochoterena, H. 2017. When homoplasy is not homoplasy: Dissecting trait evolution by contrasting composite and reductive coding. Systematic Biology, 67, 543–551. https://doi.org/10.1093/sysbio/syx053

Wiley, E.O., Lieberman, B.S. (Eds.). 2011. Phylogenetics: Theory and practice of phylogenetic systematics, 2nd ed. ed. Wiley-Blackwell, Hoboken, N.J.

Wilkins, J.S. 2018. Species: The evolution of the idea, Second edition. ed, Species and systematics. CRC Press, Boca Raton.

Zapata, F., Jiménez, I. 2011. Species delimitation: Inferring gaps in morphology across geography. Systematic Biology, 61, 179. https://doi.org/10.1093/sysbio/syr084

Zelditch, M.L., Fink, W.L., Swiderski, D.L. 1995. Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. Systematic Biology, 44, 179–189. https://doi.org/10.1093/sysbio/44.2.179

Zhang, G. 2025. Guoyi.run: TNT script for maximum parsimony analysis. https://doi.org/10.5281/ZENODO.8428197

Zhang, G., Ristola, P., Su, H., Kumar, B., Zhang, B., Hu, Y., Elliot, M.G., Drobot, V., Zhu, J., Staal, J., Larralde, M., Wang, S., Yi, Y., Yu, H. 2025. BioArchLinux: Community-driven fresh reproducible software repository for life sciences. Bioinformatics, 41, btaf106. https://doi.org/10.1093/bioinformatics/btaf106

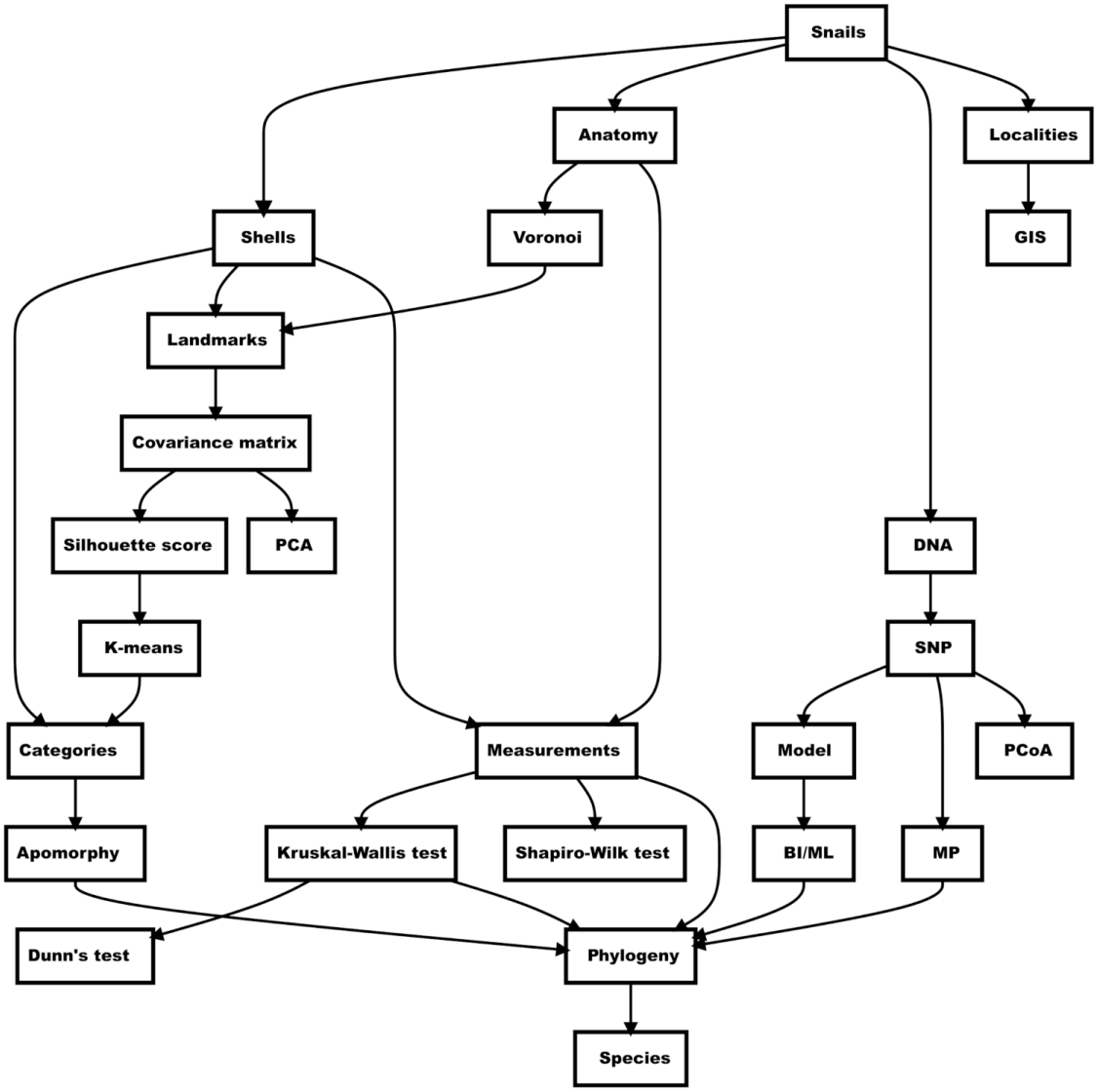


Fig. 1 Diagram of methodology used in this study.

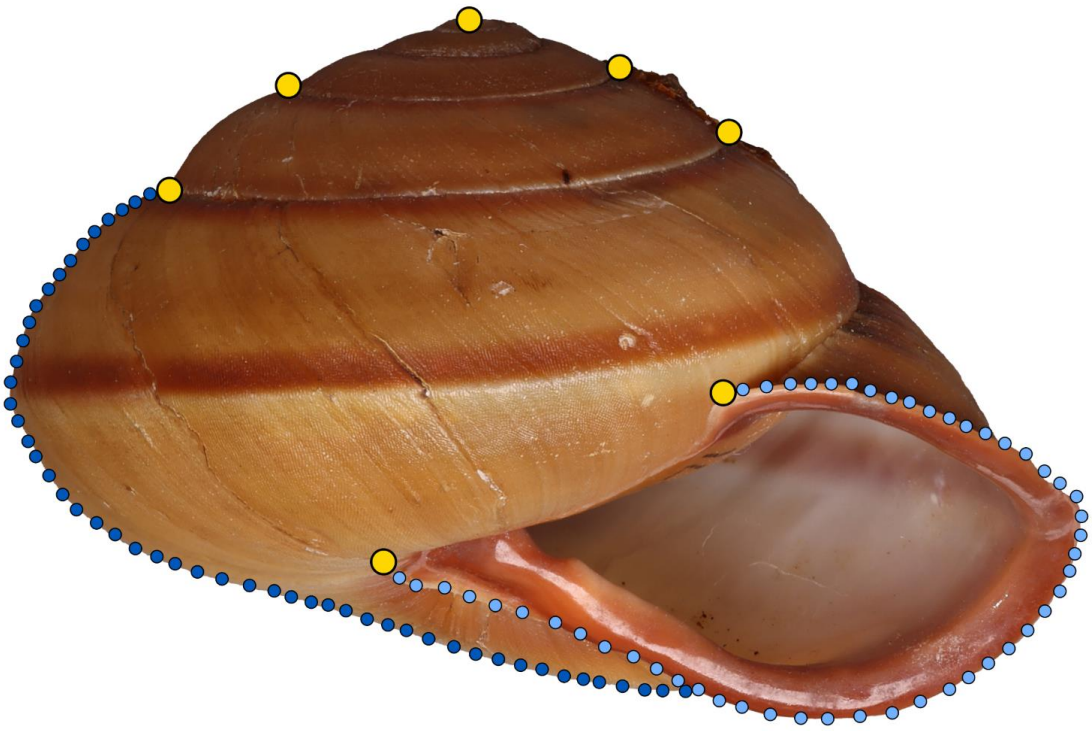


Fig. 2. Landmarks used for morphometric geometric analyses. Yellow dots: Landmarks; dark blue dots: Semi-landmark curve 1 along periphery of body whorl; blue dots: Semi-landmark curve 2 along apertural lip.

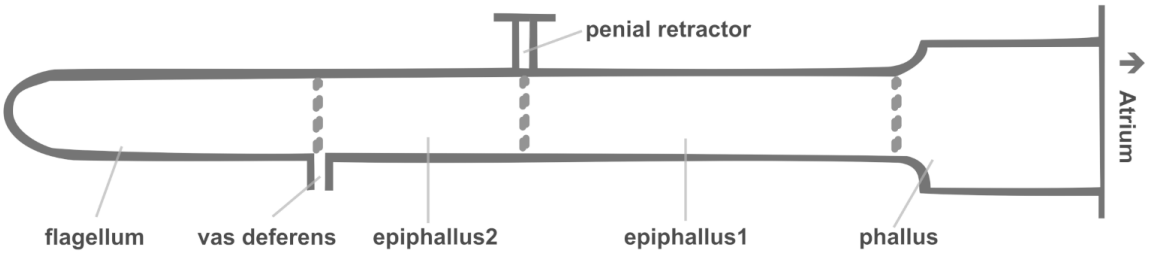


Fig. 3. Diagram of genital terminology.

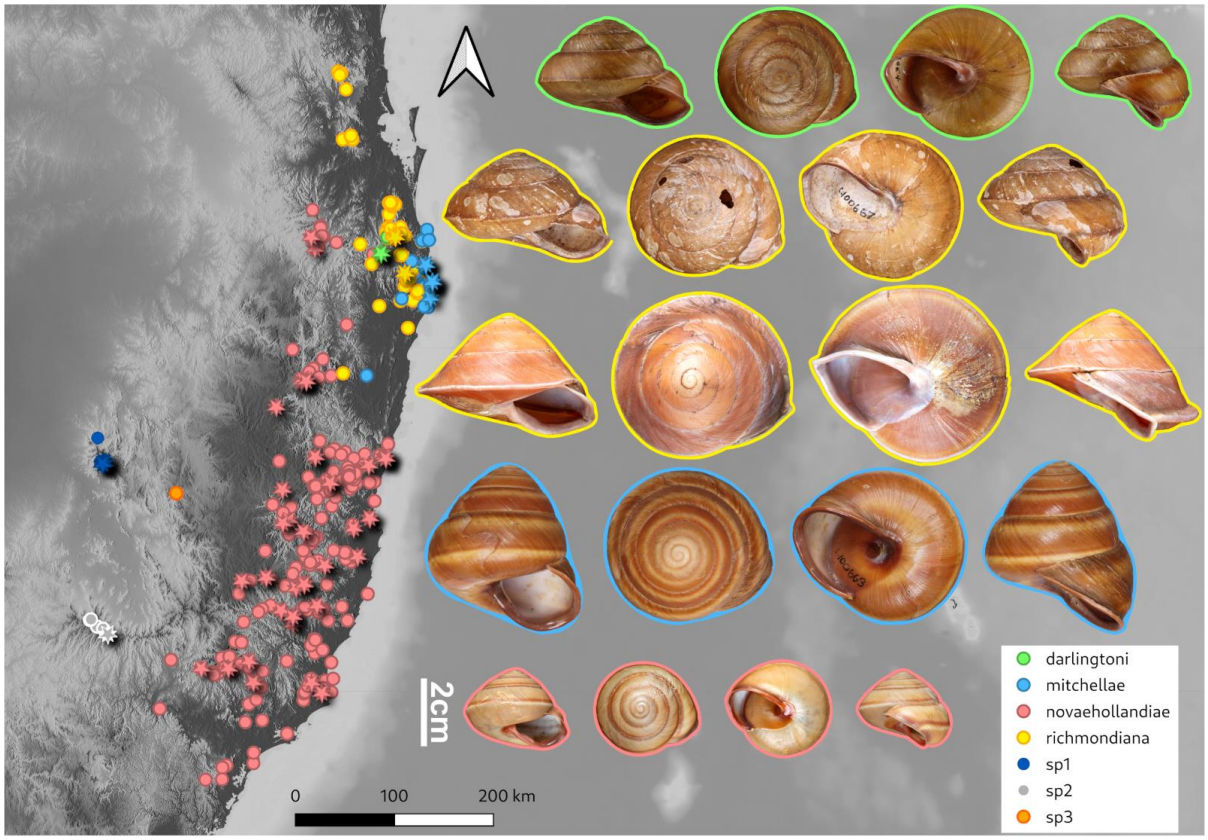


Fig. 4. Distribution of *Thersites* specimens based on collections from Australian Museum, Queensland Museum and Museum Victoria. Type specimens of *Thersites*, from top to bottom: *Thersites darlingtoni*, MCZ99054; *Annakelea tympanum*, AM C. 100667; *Helix richmondiana*, NHM 1983079; *Annakelea peragrans*, AM C. 100663; *Helix dupuyana*, NHM 1977032.

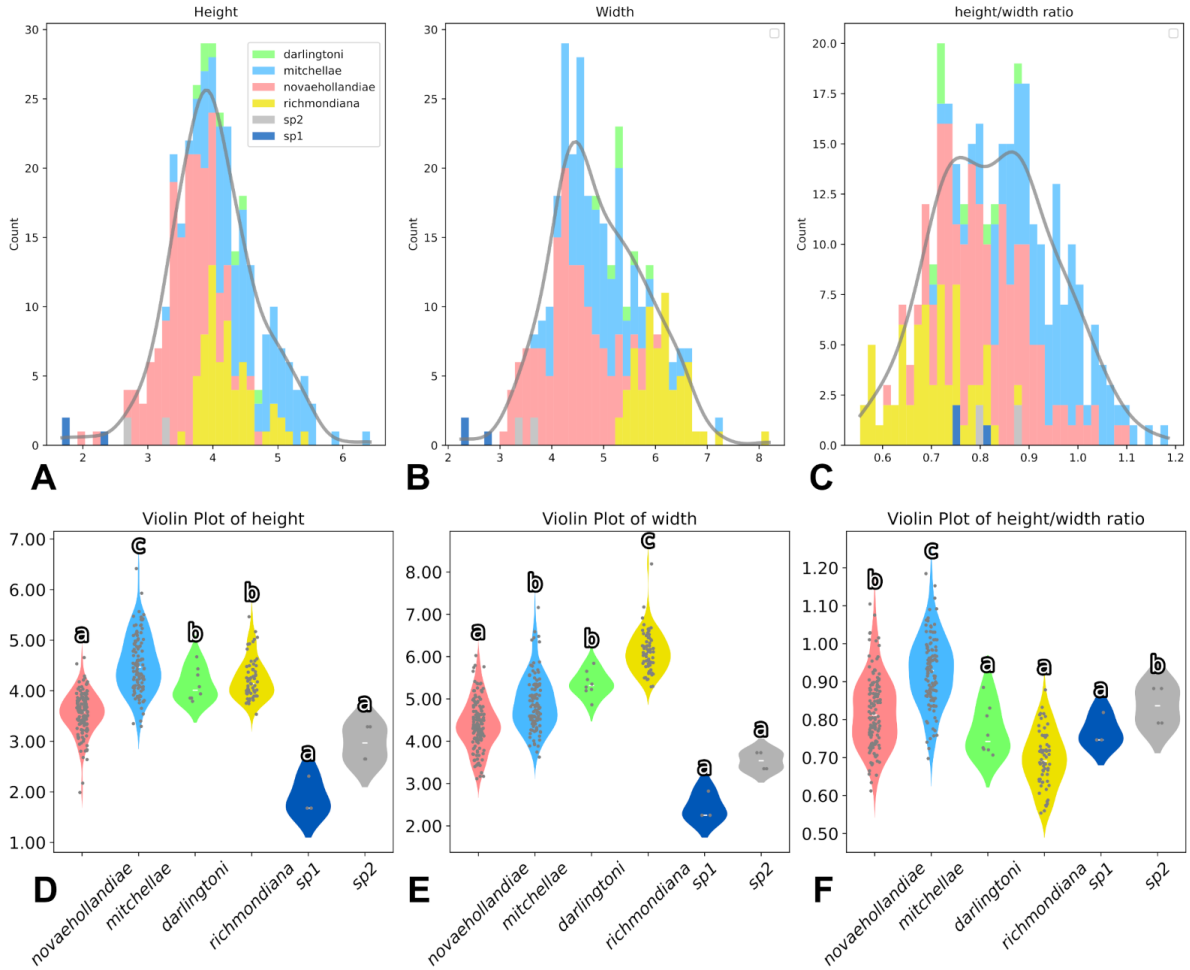
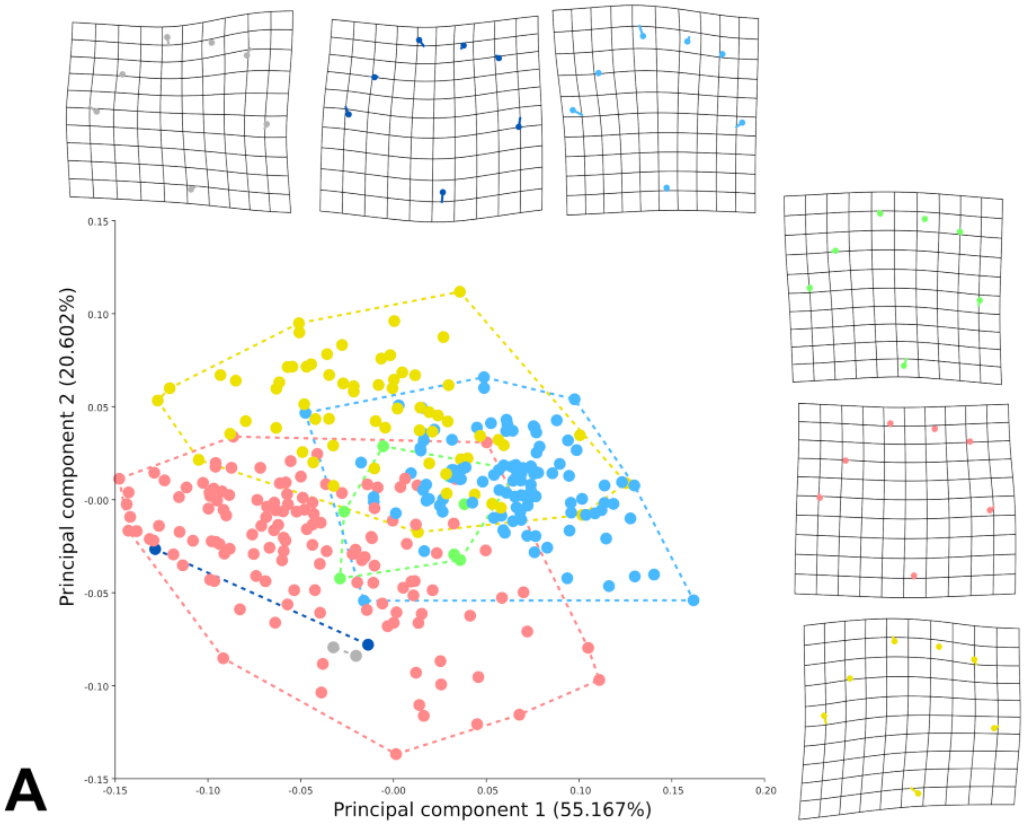


Fig. 5 A-C Histograms of shell height, width and height/width ratio. D-F Violin maps of shell height, width and height/width ratio. Dunn's test results are marked using letters on violin maps.


A
Principal component 2 (20.602%)
Principal component 1 (55.167%)
0.15
0.10
0.05
-0.00
-0.05
-0.10
-0.15
-0.15
-0.10
-0.05
-0.00
0.05
0.10
0.15
0.20

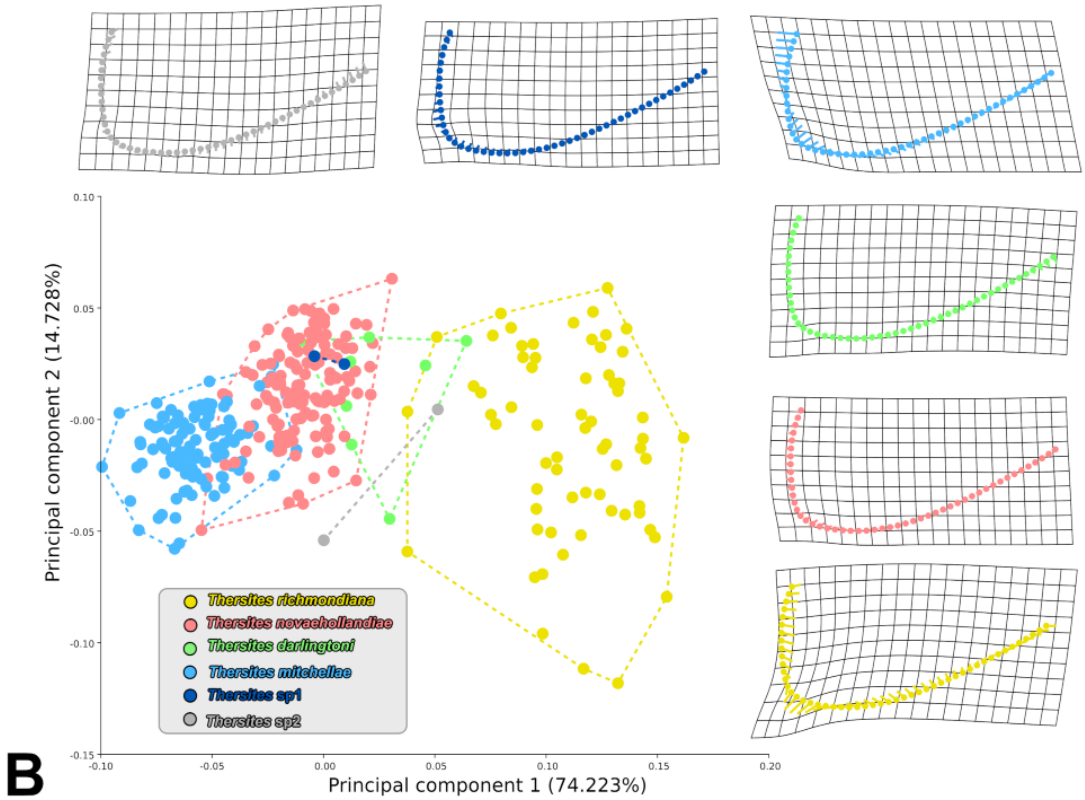
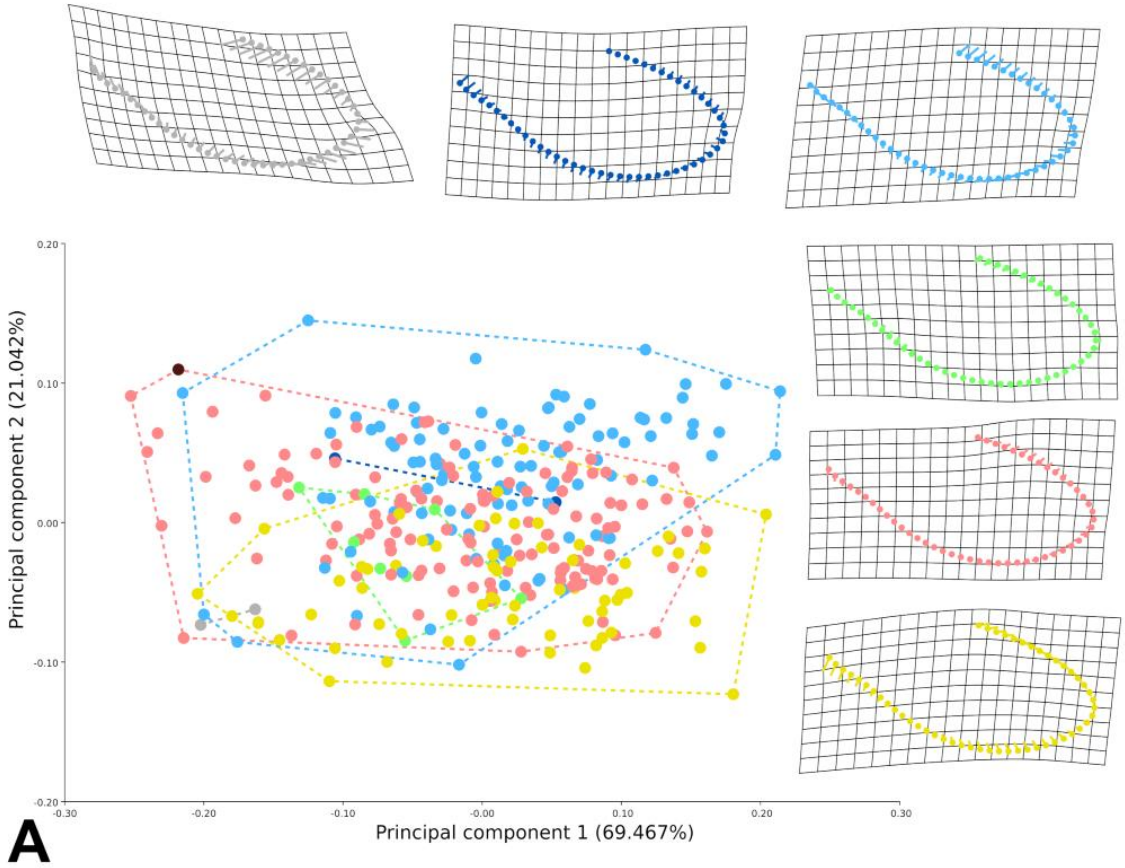

B
Principal component 2 (14.728%)
Principal component 1 (74.223%)
Thersites richmondiana
Thersites novaehollandiae
Thersites darlingtoni
Thersites mitchellae
Thersites sp1
Thersites sp2
0.10
0.05
-0.00
-0.05
-0.10
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20

Fig. 6. A. Scatter plots of principal component analysis for landmarks with thin plate spline grid map for every species. B. Scatter plots of principal component analysis for semi-landmark curve 1 (periphery of body whorl) with thin plate spline grid maps for every species.


Principal component 2 (21.042%)
Principal component 1 (69.467%)
A

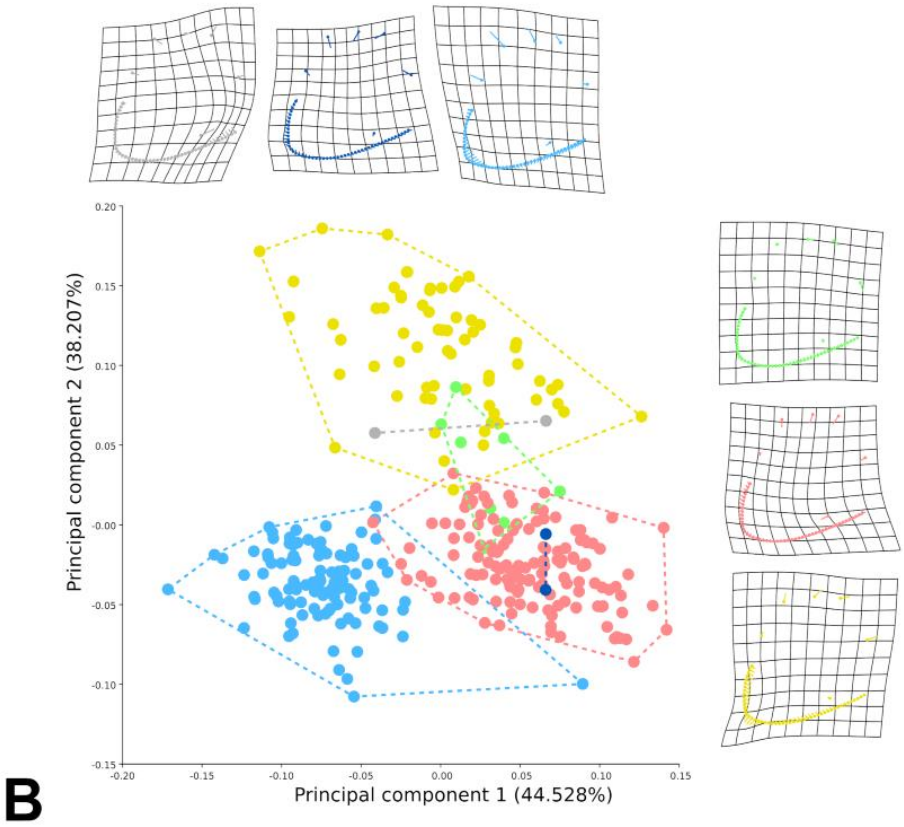

Principal component 2 (38.207%)
Principal component 1 (44.528%)
B

Fig. 7. A. Scatter plots of principal component analysis for semi-landmarks2 (aperture curve) with thin plate spline grid map for each species. B. Scatter plots of principal component analysis for combination of landmarks and semi-landmarks1 (body whorl curve) with thin plate spline grid maps for each species.

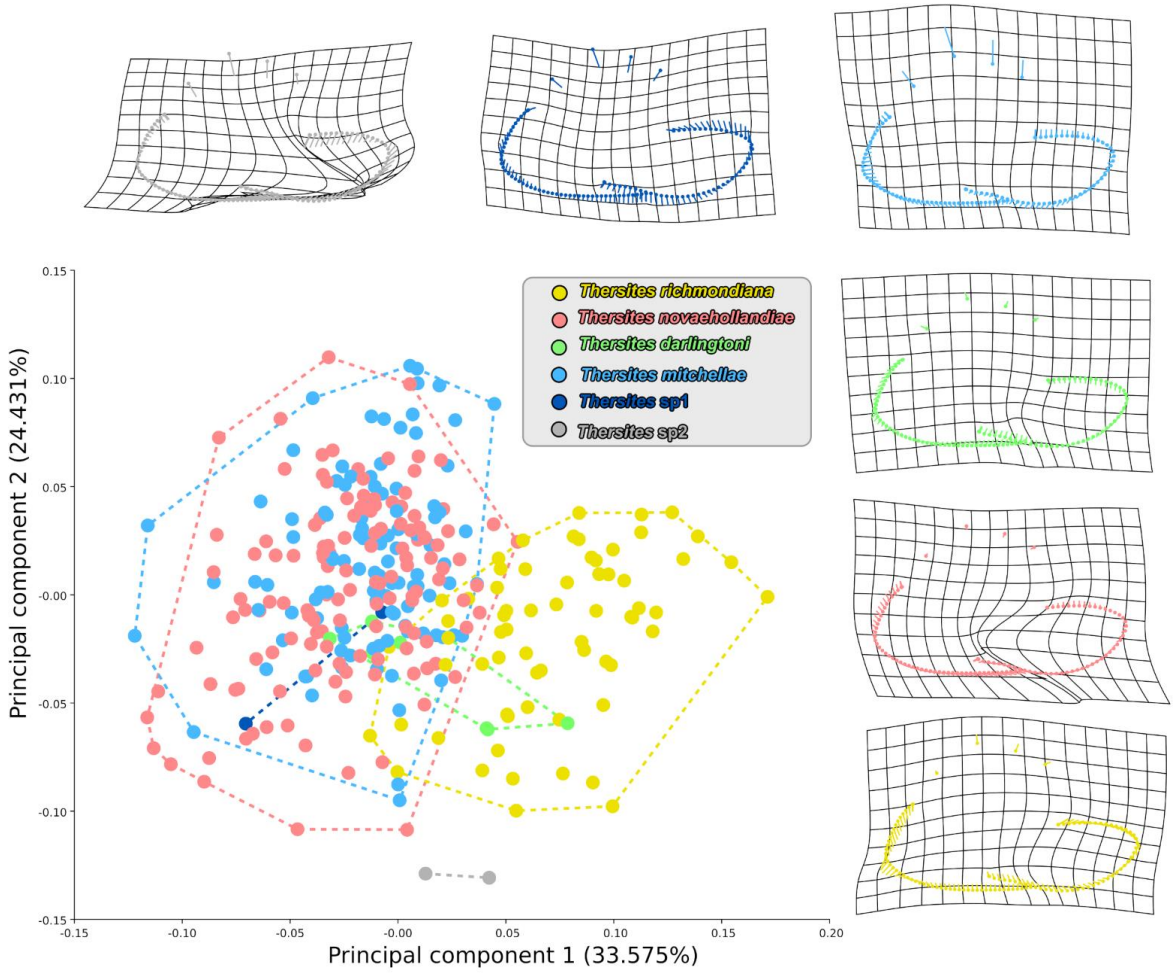


Fig. 8. Scatter plots of principal component analysis for all landmarks and semi-landmarks with thin plate spline grid map for each species.

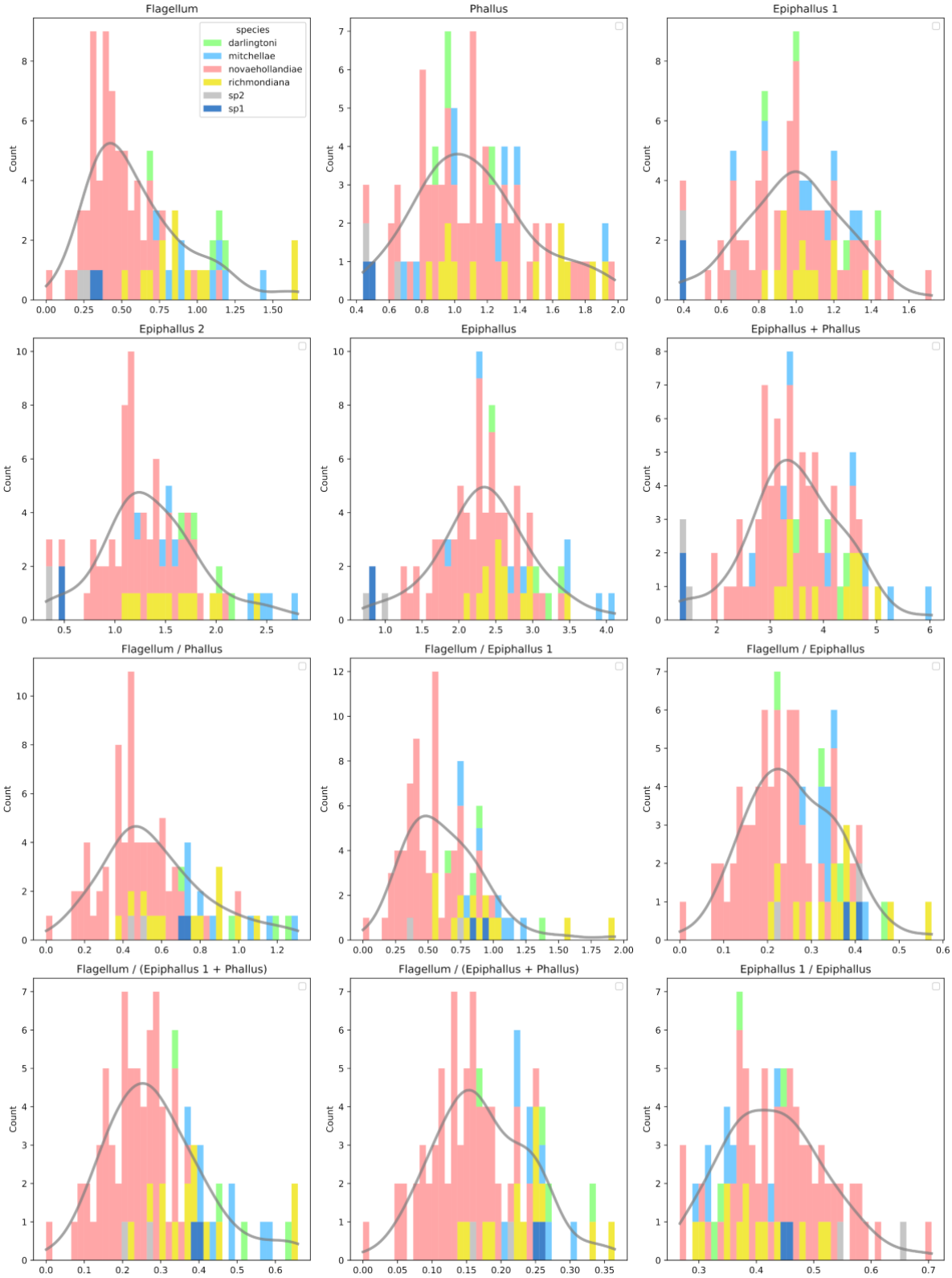


Fig. 9 Histograms of genital measurements and their ratios.

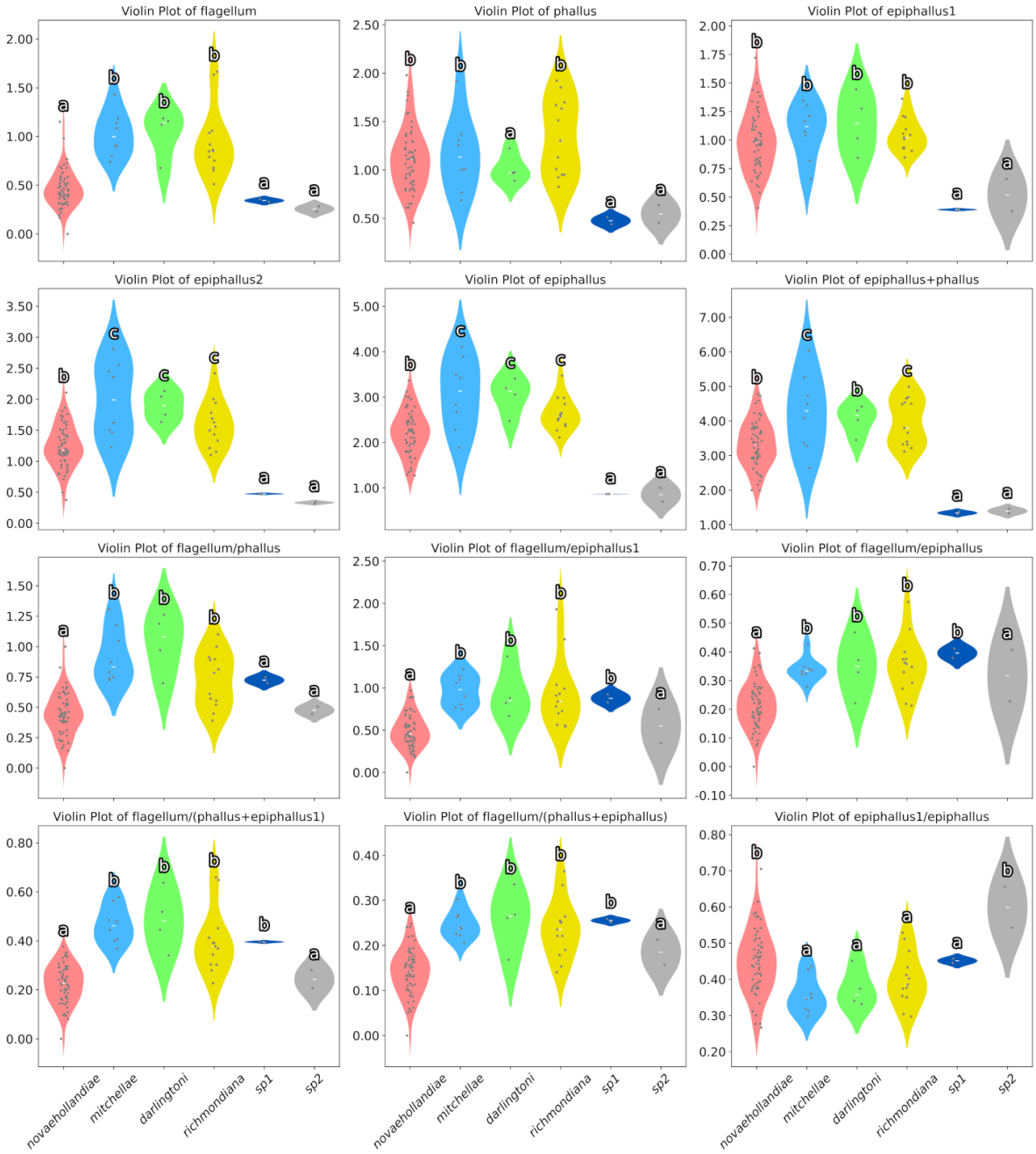


Fig. 10 Violin maps of genital measurements and their ratio. Dunn's test results are marked using letters on violin maps.

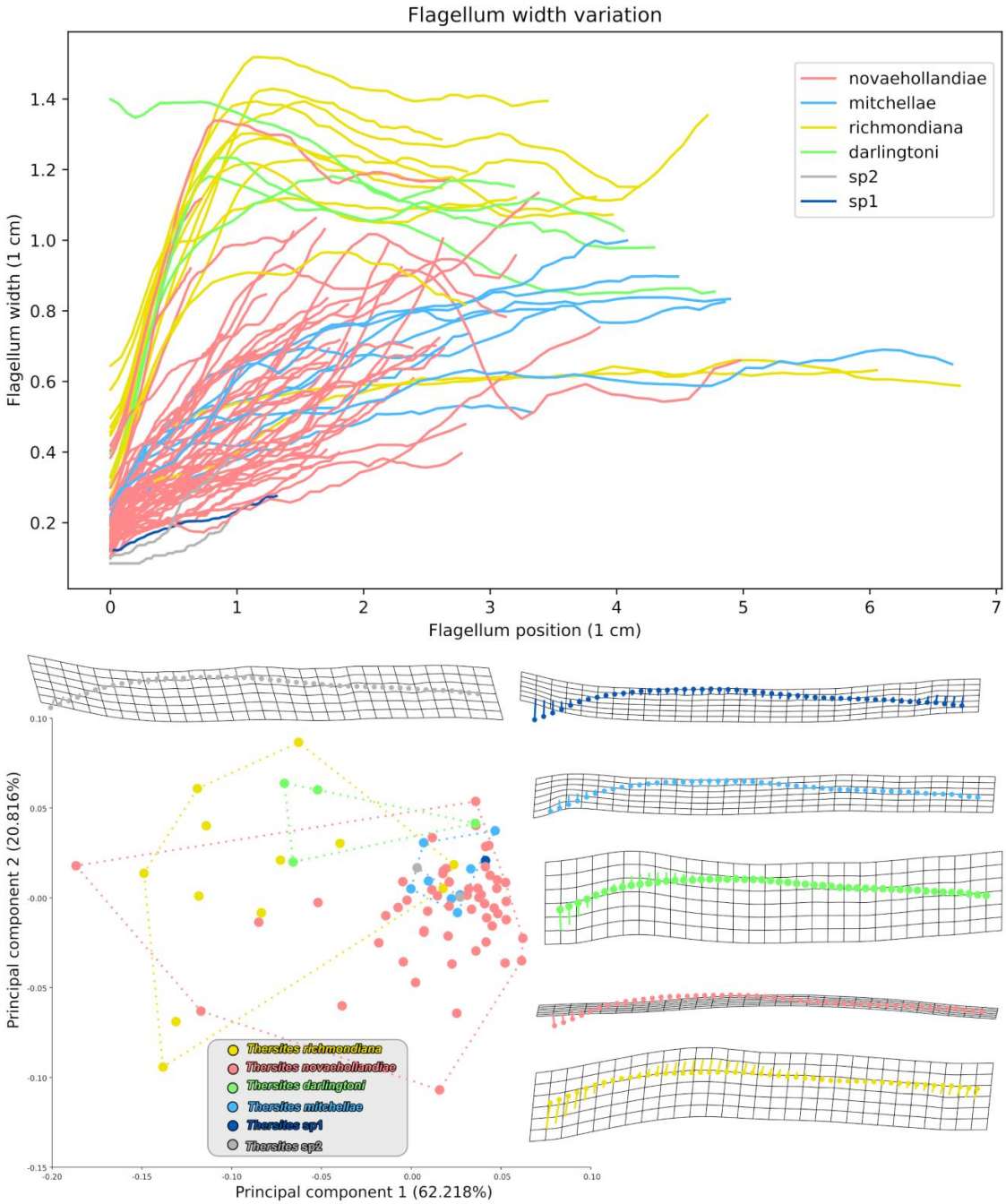


Fig. 11 Curves of 50% length flagellum shapes and scatter plots of principal component analysis for 50% length flagellum with thin plate spline grid maps for each species.

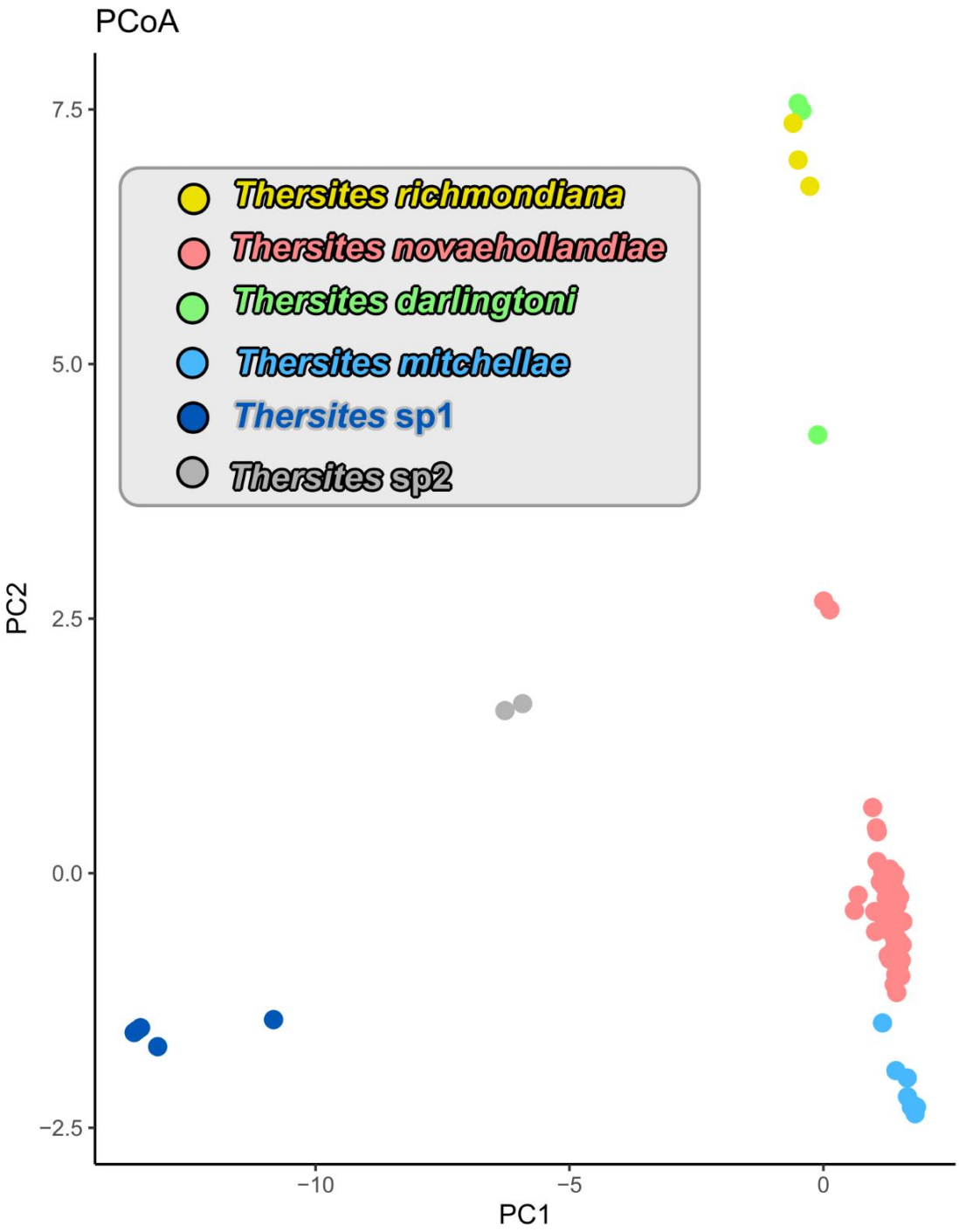


Fig. 12. Scatter plots of Principal Coordinate Analysis (PCoA) for DArT sequences

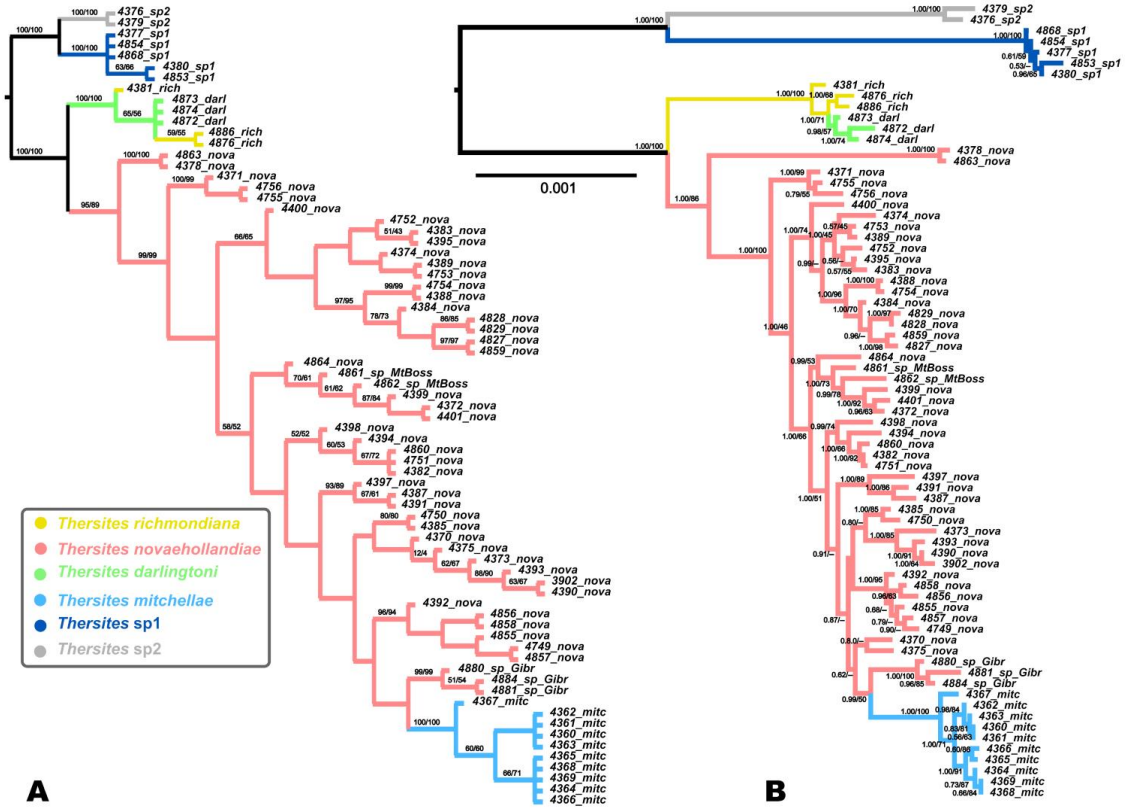


Fig. 13. A. Maximum Parsimony tree under Extended implied weighting or implied weighting with K value as 12. The Jackknifing and symmetric resampling were marked on the phylogenetics. B. Bayesian inference tree with Bayesian posterior probabilities and bootstrap from maximum likelihood marked on the branch.

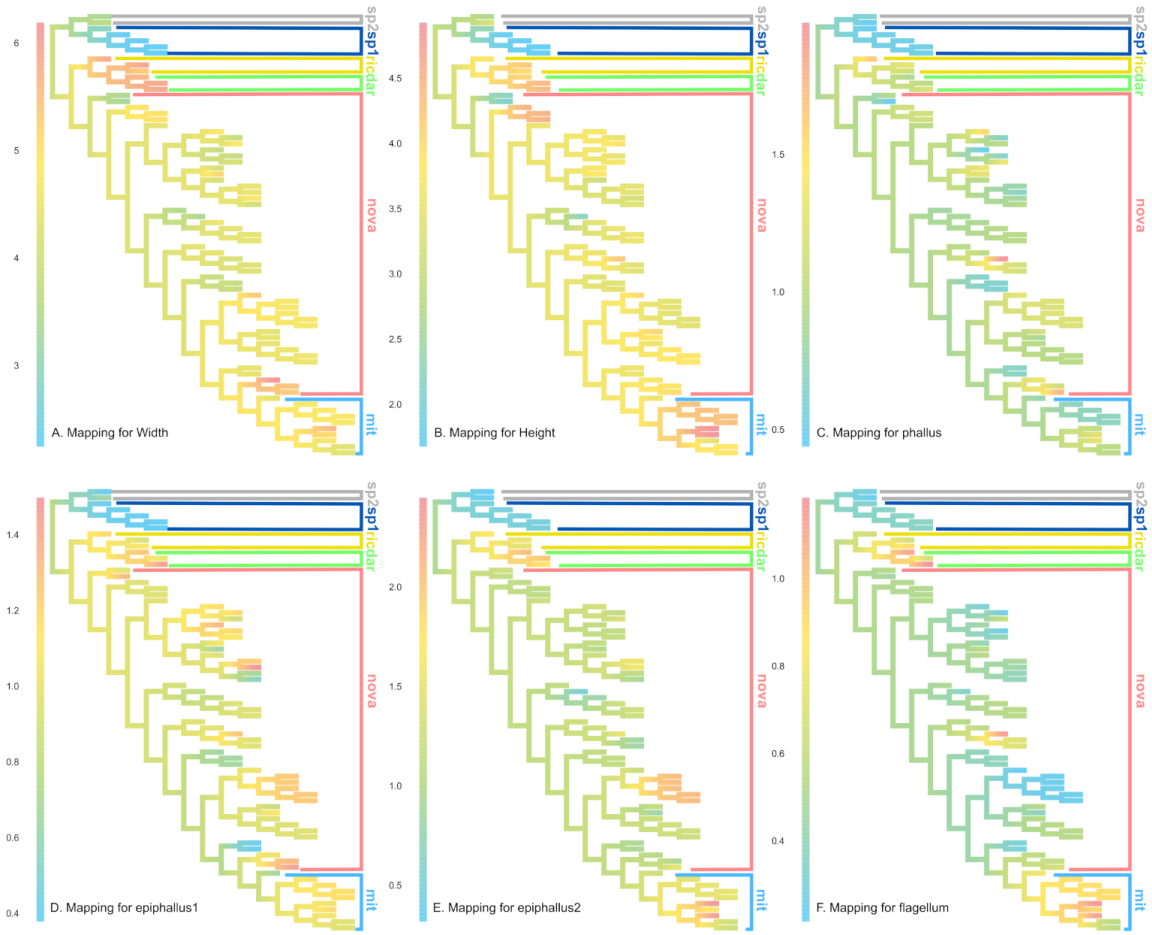


Fig. 14. Continuous characters changes mapping on the Bayesian inference topology.

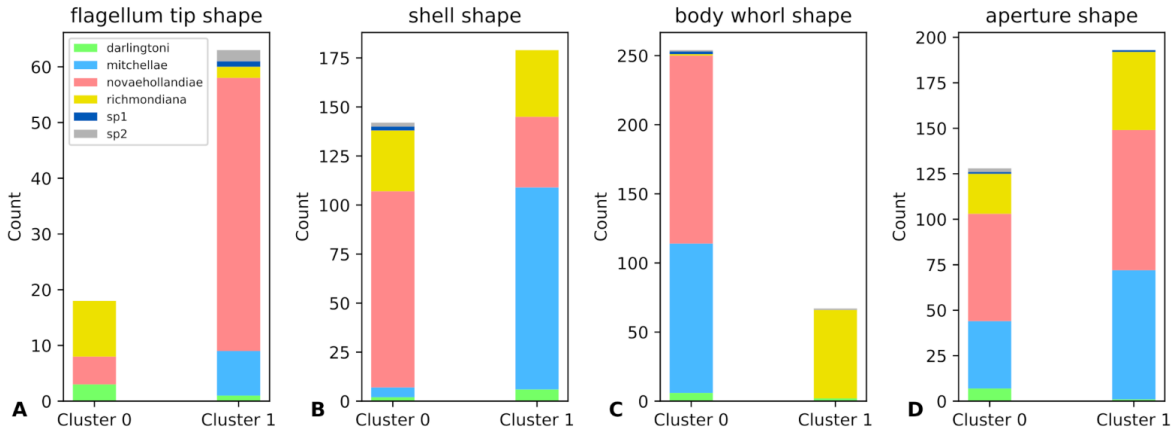


Fig. 15. Bart charts of K-means clustering for landmarks and semi-landmarks. The flagellum tip shape represents 50% flagellum length landmarks. Shell, body whorl and aperture shapes separately represent the landmarks, semi-landmark curve 1 and semi-landmark curve 2.

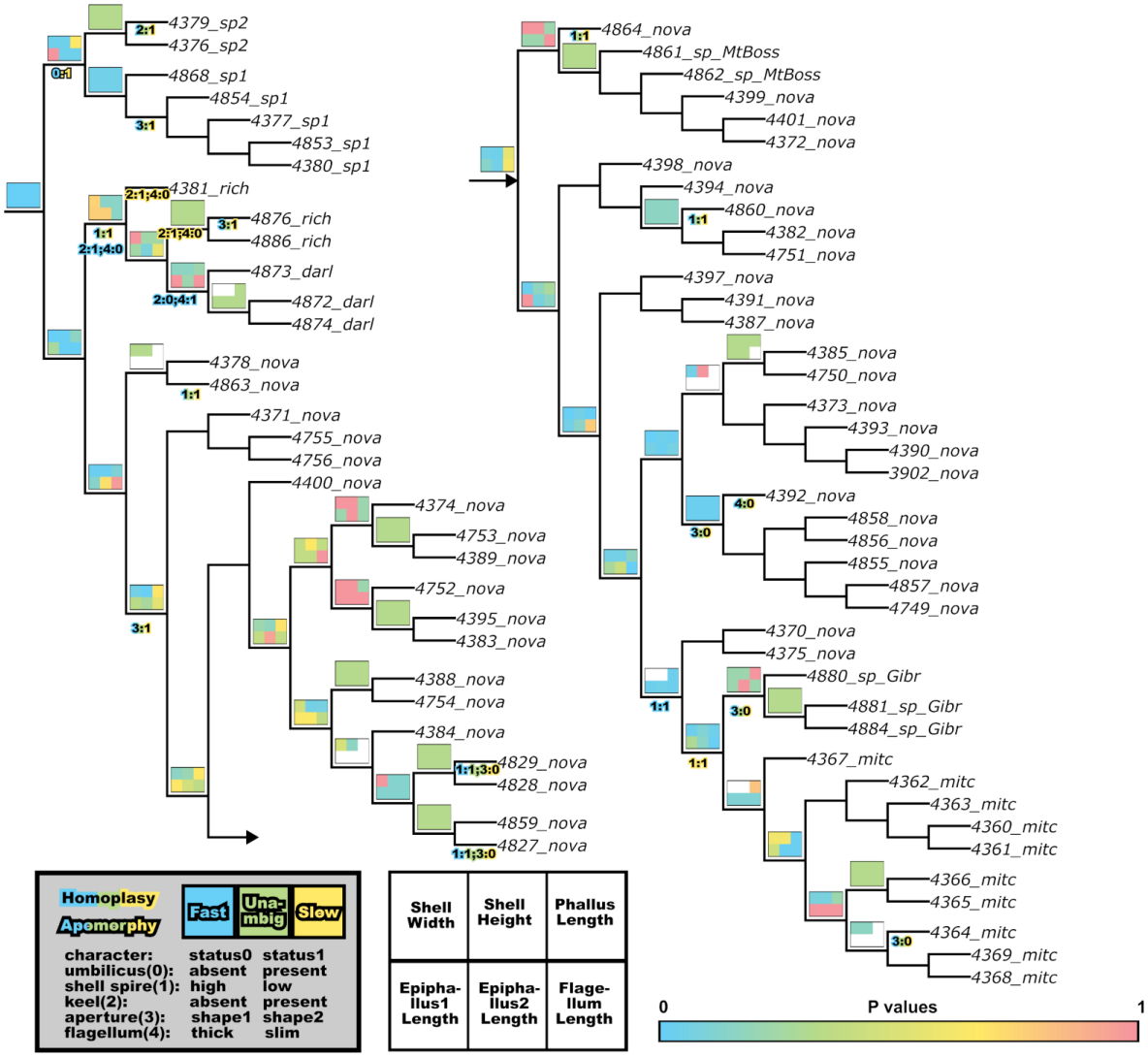
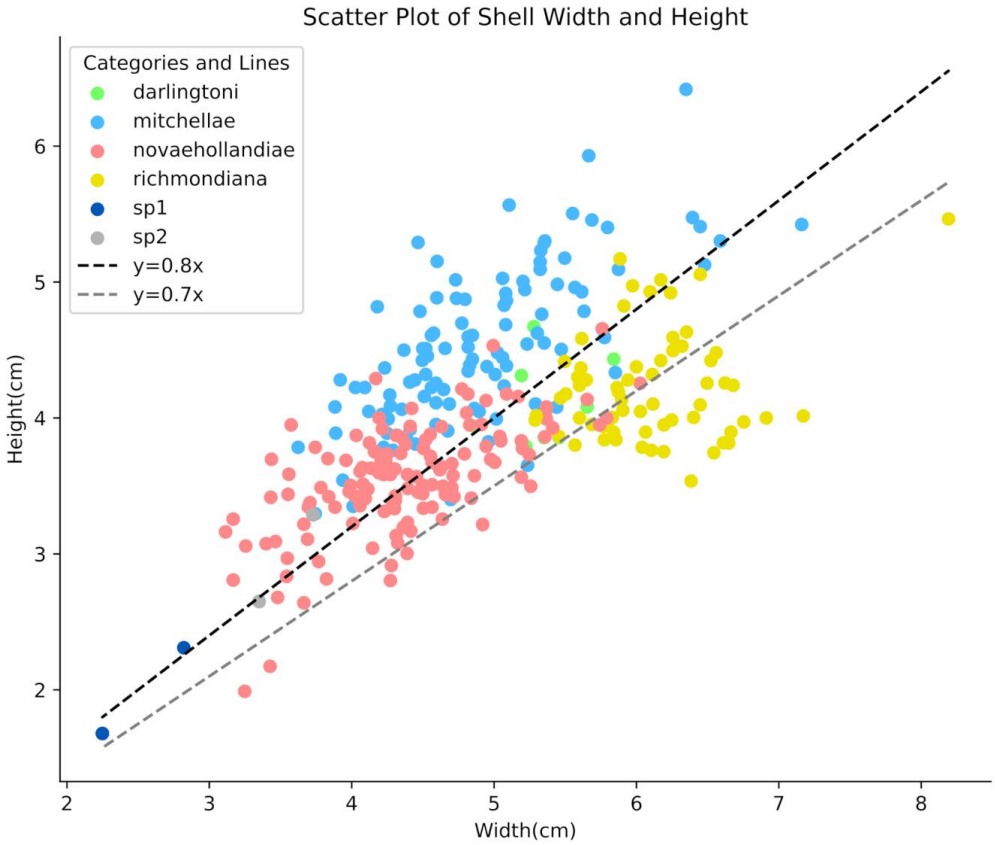
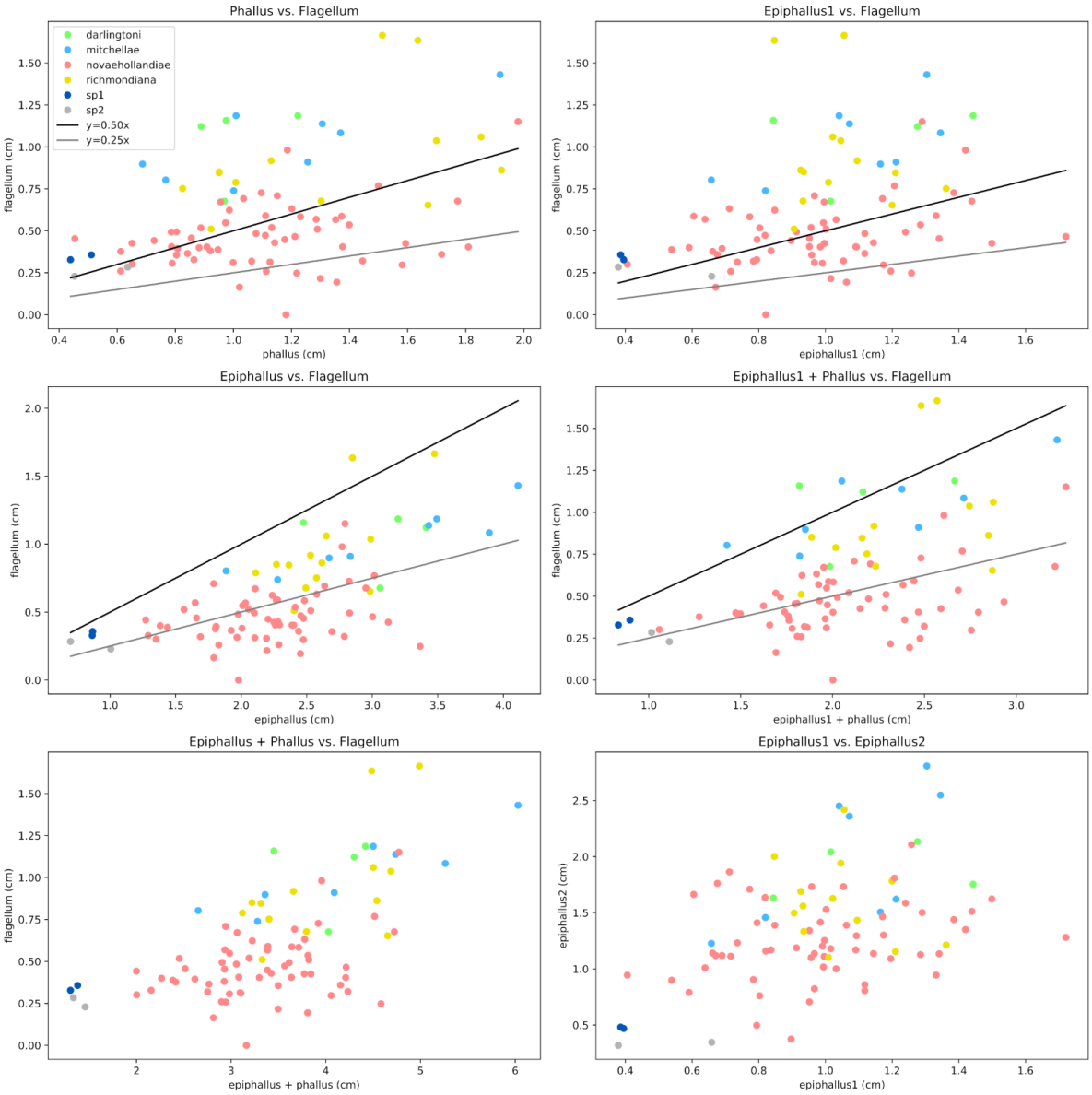


Fig. 16. Kruskal-Wallis test heatmaps of character measurements are marked beyond branches of the phylogeny. The homoplasic and apomorphic characters under fast, slow, and unambiguous character optimization are marked under the branches of phylogeny. Six-box heatmaps show the p-values calculated for measurements of shell width and height, measuremetns of phallus, epiphallus1, epiphallus2, and flagellum. These p-values are derived by comparing the branch's immediate child nodes to all of its descendant nodes. Homoplasic and apomorphic characters are annotated under branches: apomorphies are indicated by black text with colored outlines, while homoplasies are shown in colored text with black outlines. Fast optimization is represented in blue, slow in yellow, and unambiguous in green. Characters are labeled in the 'character number: status number' format. Discrete characters were derived from landmarks categorized using K-means clustering, along with one qualitative character (character 0, indicating the presence or absence of an umbilicus).

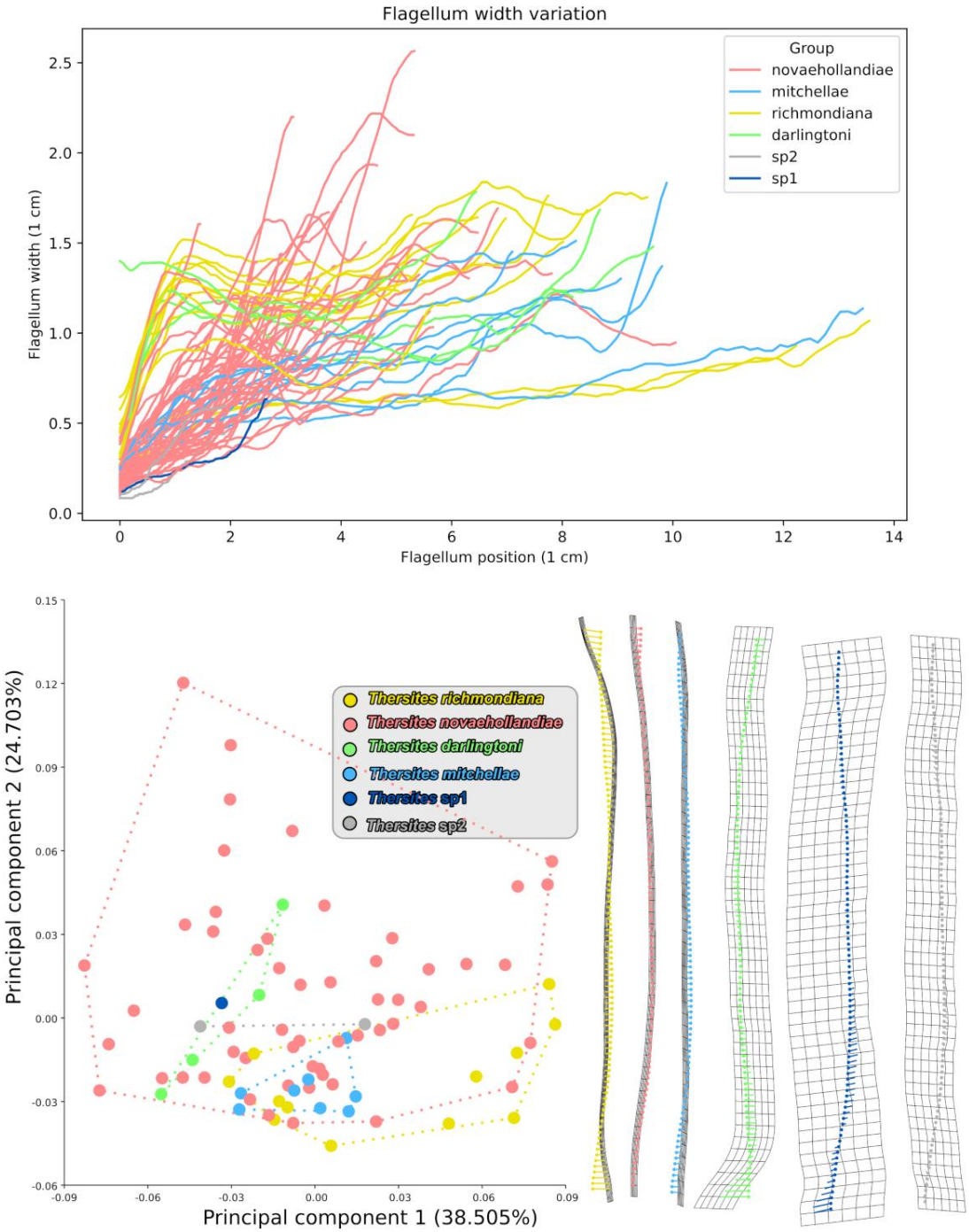
Supplementary File



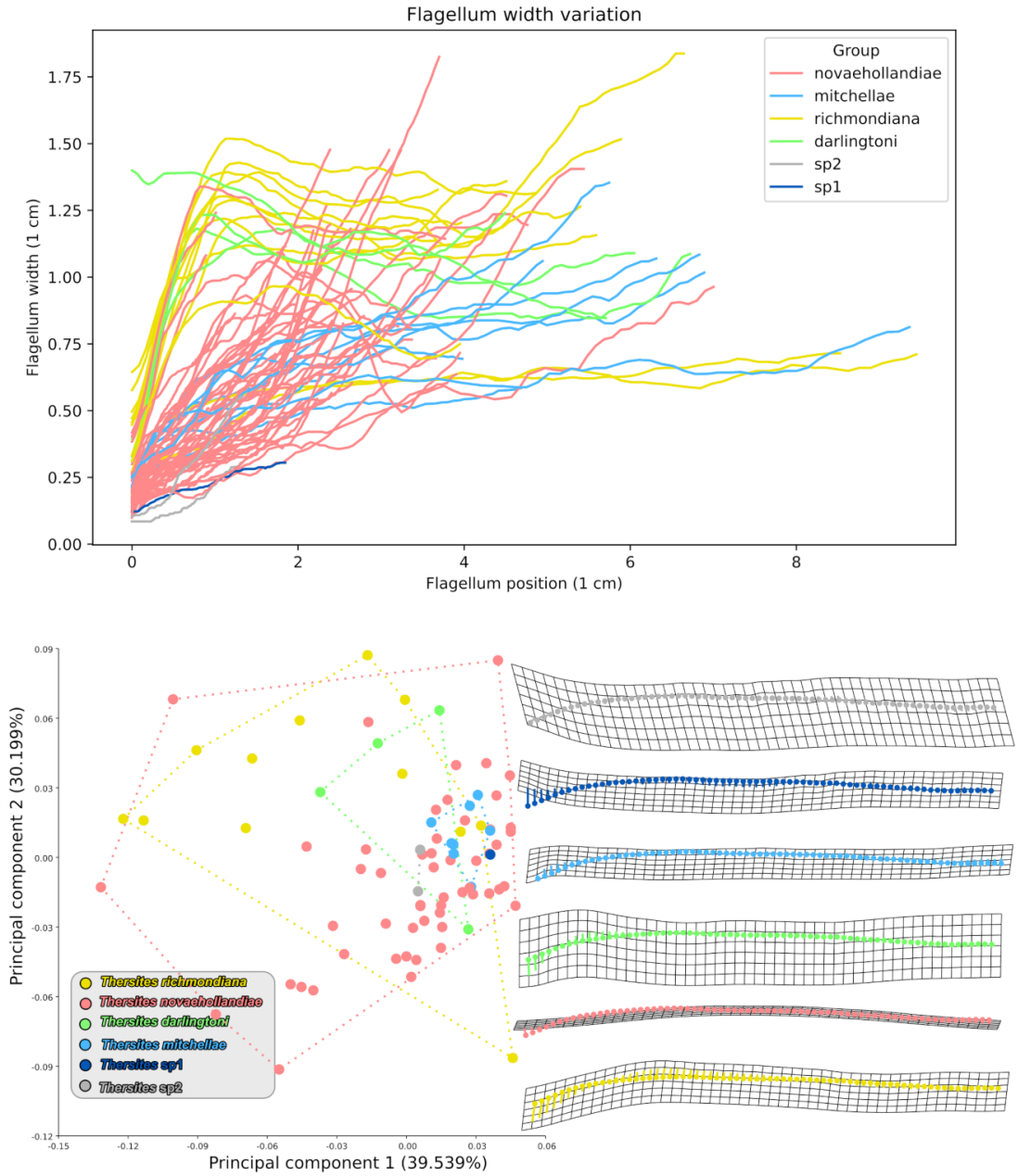
Supplementary Fig. 1. Scatter plots of shell width and height.



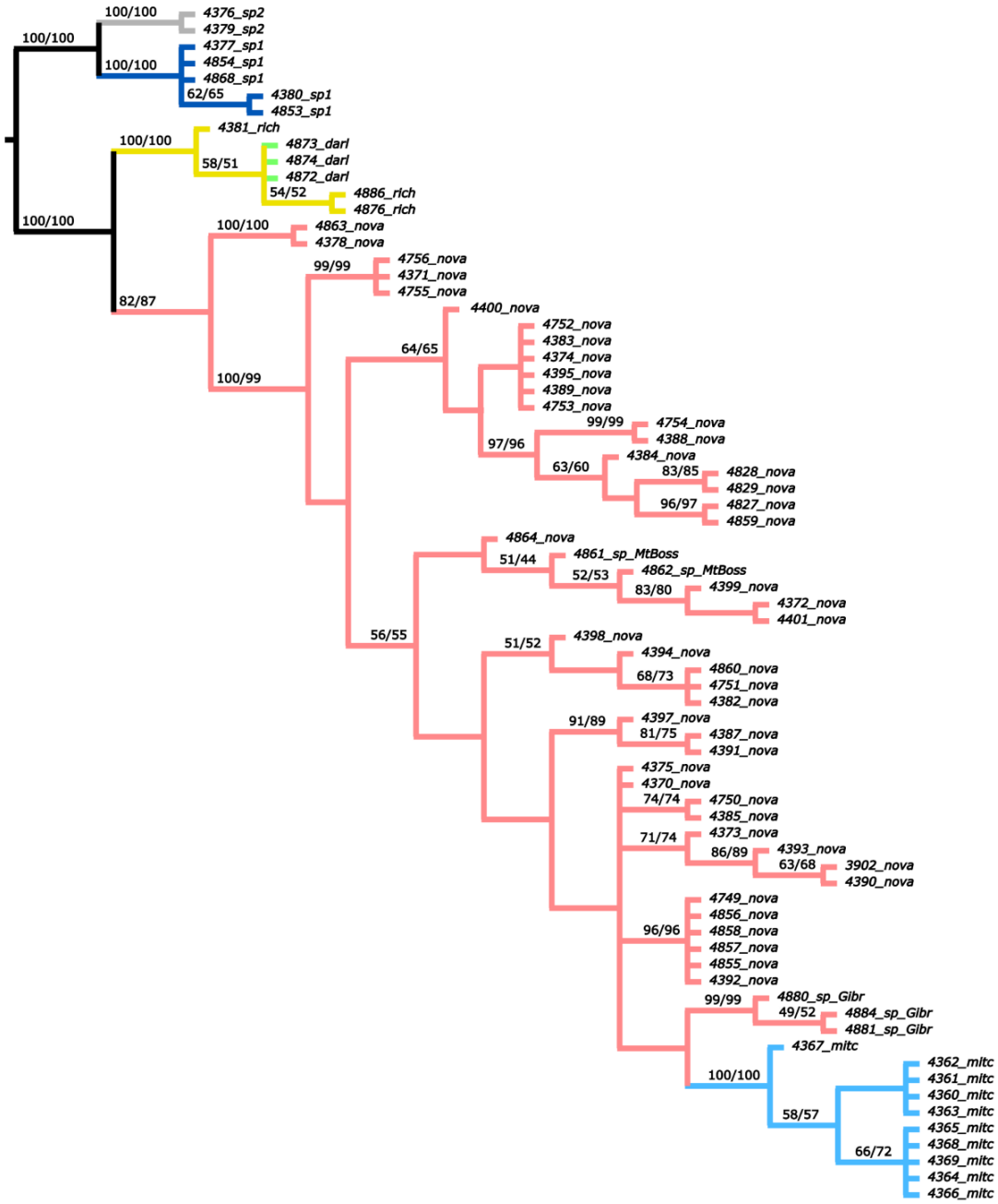
Supplementary Fig. 2. Scatter plots of genital measurements.



Supplementary Fig. 3. Curves of 100% length flagellum shapes and scatter plots of principal component analysis for 100% length flagellum with thin plate spline grid maps for each species.



Supplementary Fig. 4. Curves of 70% length flagellum shapes and scatter plots of principal component analysis for 70% length flagellum with thin plate spline grid maps for each species.



Supplementary Fig. 5. Maximum Parsimony tree under equal weighting. The Jackknifing and symmetric resampling were marked on the phylogenetics.