

Confusion will be my epitaph: Genome-scale discordance stifles phylogenetic resolution of Holothuroidea

Nicolás Mongiardino Koch¹, Ekin Tilic^{1,2}, Allison K. Miller³, Josefin Stiller⁴, Greg W. Rouse¹

¹ Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA.

² Department of Marine Zoology, Senckenberg Research Institute and Museum, Frankfurt, Germany.

³ Anatomy Department, University of Otago, Dunedin, Otago, New Zealand.

⁴ Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark.

Abstract

Sea cucumbers (Holothuroidea) are a diverse clade of echinoderms found from intertidal waters to the bottom of the deepest trenches. Their reduced skeletons and limited number of phylogenetically-informative traits have long obfuscated morphological classifications. Sanger-sequenced molecular datasets have also failed to constrain the position of major lineages. Noteworthy, topological uncertainty has hindered a resolution for Neoholothuriida, a highly diverse clade of Permo-Triassic age. We perform the first phylogenomic analysis of Holothuroidea, combining existing datasets with twelve novel transcriptomes. Using a highly-curated dataset of 1,100 orthologues, our efforts recapitulate previous results, struggling to resolve interrelationships among neoholothuriid clades. Three approaches to phylogenetic reconstruction (concatenation under both site-homogeneous and site-heterogeneous models, and coalescent-aware inference) result in alternative resolutions, all of which are recovered with strong support, and across a range of datasets filtered for phylogenetic usefulness. We explore this intriguing result using gene-wise log-likelihood scores, and attempt to correlate these with a large set of gene properties. While presenting novel ways of exploring and visualizing support for alternative trees, we are unable to discover significant predictors of topological preference, and our efforts fail to favor one topology. Neoholothuriid genomes seem to retain an amalgam of signals derived from multiple phylogenetic histories.

Keywords: Sea cucumbers, systematics, phylogenomics, phylogenetic signal, phylogenetic conflict

Introduction

Holothuroidea, commonly known as sea cucumbers, is arguably the most morphologically diverse major clade of extant Echinodermata (Fig. 1). The smallest adults can be less than 1 cm in length, as seen in the meiofaunal *Leptosynapta minuta* [1] and epibenthic *Incubocnus* [2]. The largest can be thin and elongate, reaching several meters length, as in the snake sea cucumber *Synapta maculata* [3], or they may be less than a meter but robust and weighing over 5 kg, as in the case of *Holothuria fuscopunctata* [4]. While predominantly benthic as adults, some taxa are capable of swimming and there are even forms that spend their entire lives in the water column [5]. While all holothuroids have a ring of tentacles and are deposit or filter feeders, some clades lack tube feet and have a substantially reduced water vascular system, traits otherwise developed across all echinoderms. They can also entirely lack calcareous elements (ossicles) in the body wall, or these can be expanded to form overlapping plates that build a rigid test [6]. There are currently 1,775 accepted extant species of holothuroids [7] found in ocean waters that range from the intertidal to the bottom of the deepest trenches [8, 9]. Especially in benthic deep-sea habitats, they can constitute the vast majority of total biomass and have a strong impact on ecosystem functioning, bioturbation, and nutrient cycling [10-12].

While multiple morphological attempts have been made to delineate major subdivisions within Holothuroidea, these have been limited by the extreme simplification of their skeleton (relative to other echinoderms), the delicate and fragile nature of their bodies, and the small number of traits that provide useful information at high taxonomic levels [13-15]. The most recent revision of the group's classification was based on a six-gene dataset including terminals from 25 of the 29 accepted family-ranked taxa [16]. This study recovered a basal split within sea cucumbers between Apodida, a clade characterized by a complete loss of tube feet, and Actinopoda (among which secondary reductions or loss of tube feet occur only within Molpadida). Actinopoda were further subdivided into Elasipodida and Pneumonophora, the latter of which includes all species with respiratory trees, a unique cloacal

invagination that plays an important (although not exclusive) role in respiration [17, 18]. Furthermore, the names Holothuriida and Neoholothuriida were applied to the main subdivisions within Pneumonophora, with four well-supported major clades inside Neoholothuriida: Dendrochirotida, Molpadida, Persiculida, and Synallactida. However, the relationships among these four lineages remained uncertain. A phylogenetic resolution for the major neoholothuriid lineages is necessary to explore the origins of the high morphological and ecological disparity harbored by this clade, as well as to establish a natural classification framework for a substantial fraction of sea cucumber diversity (62% of species-level diversity is contained within Neoholothuriida [7]). Miller et al. [16] concluded that meeting these objectives would likely require sequencing efforts of a different magnitude.

Here we present the first phylogenomic study of sea cucumbers, the last major eleutherozoan clade (which further includes echinoids [19], asteroids [20], and ophiuroids [21]) to have its phylogeny tackled using genome-scale datasets. Through the generation of novel transcriptomic resources for holothuroids we built a molecular dataset encompassing over a thousand orthologs. The goal was to resolve some of the lingering uncertainties in the holothuroid tree of life, yet a continuing lack of resolution encouraged novel ways to explore phylogenomic datasets.

Materials and methods

Taxon sampling, extraction and sequencing

Sea cucumber specimens were collected by SCUBA diving, snorkeling, dredging, and remotely operated vehicles (ROV), or purchased from aquarium suppliers, and vouchers were deposited at the Benthic Invertebrate Collection, Scripps Institution of Oceanography (SIO-BIC, see Table 1). Species identification was based on multiple lines of evidence, including anatomical (gross and ossicle morphology), biogeographical and molecular (mitochondrial cytochrome c oxidase subunit I, COI)

information. DNA extractions and COI amplifications followed protocols described in Miller et al. [16], and sequences are deposited in NCBI (accession numbers available in Table 1). Previous identifications of transcriptomic vouchers at SIO were also revised, including those sequenced and released as part of EchinoDB [22, 23].

For large specimens, tissue was dissected from the body wall or tube feet, while whole body sections were sampled for the remainder. Sampled tissues were finely chopped, placed in RNA*later* (Invitrogen) buffer solution, and stored at -80 °C. RNA extractions were performed from Trizol (ThermoFisher), using Direct-zol RNA Miniprep Kit with in-column DNase treatment (Zymo Research). mRNA was isolated with Dynabeads mRNA Direct Micro Kit (Invitrogen). mRNA concentration was estimated using Qubit RNA broad range assay kit (ThermoFisher), and quality was assessed using RNA ScreenTape with an Agilent 4200 TapeStation on an Agilent Bioanalyzer 2100. Most libraries were prepared using a KAPA-Stranded RNA-Seq kit targeting a 200-300 bp insert size, and results were assessed using DNA ScreenTape (Bioanalyzer 2100). Libraries were then sequenced in multiplexed (8 libraries per lane) pair-end runs using 150 bp paired end Illumina HiSeq 4000 at the UC San Diego IGM Genomics Center. To minimize read crossover, we employed 10 bp sequence tags designed to be robust to indel and substitution errors [24]. For four samples (*Benthodytes cf. sanguinolenta*, *Benthogone* sp., *Colochirus robustus*, and *Peniagone cf. vitrea*), library preparation and multiplexed pair-end sequencing on an Illumina NovaSeq 6000 PE150 was performed by Novogene.

Twelve novel transcriptomes were generated for this study and combined with publicly available genomic and transcriptomic datasets downloaded from NCBI and EchinoBase [25]. Final taxonomic sampling included 35 holothuroids as well as three echinoids and one asteroid outgroups (Table 1). Raw files for all novel datasets, as well as those so far available only on EchinoDB [23], are deposited in the NCBI sequence read archive (SRA) under Bioproject XXX. All assemblies are available at the Dryad data repository YYY.

Assembly, sanitation and matrix construction

Reads were trimmed or excluded based on quality scores using Trimmomatic v 0.3.6 under default settings [26]. Additional sanitation steps were implemented by the Agalma 2.0 pipeline [27], resulting in the removal of reads based on compositional and quality filtering criteria, as well as those mapping to rRNA sequences or retaining adapter sequences. Remaining reads were assembled *de novo* with Trinity v. 2.5.1 [28]. Assemblies were then screened for contaminants using alien_index v 3.0 [29]. Transcripts with substantially better BLAST+ [30] hits to a dataset of well-curated archaeal, bacterial and fungal genomes than to a metazoan database (both available from http://ryanlab.whitney.ufl.edu/downloads/alien_index/), defined as those exhibiting an alien index > 45 (see [31]), were excluded. Sanitized transcriptomes were imported back into Agalma for orthology inference [27, 32], alignment using MAFFT v. 7.305 [33], and quality-based trimming with GBLOCKS v. 0.91b [34]. The resulting supermatrix was reduced using a 70% occupancy threshold, resulting in a dataset composed of 1,159 orthologs coded as amino acids (from a total of 13,767).

Gene trees were inferred from each amino acid alignment with ParGenes v. 1.0.1 [35], using optimal models and 100 bootstrap replicates. These were analyzed with TreeShrink v. 1.3.1 [36] (using parameters `-q 0.01 -k 3 -b 25`), which employs taxon-specific distributions of root-to-tip distances to identify outlier sequences potentially suffering from errors in alignment or orthology inference. Identified outliers were removed from both gene trees and individual alignments, and a new supermatrix was concatenated. As a final sanitation step, the data was run using *genesortR* [37, 38], which ordered all loci based on decreasing estimates of phylogenetic usefulness (Fig. S1). The worst-ranked 59 loci (5.1% of supermatrix) were further discarded, resulting in a final dataset of 1,100 loci and

264,991 amino acid positions. Two smaller datasets, composed of the top-scoring (i.e., most phylogenetically useful) half and quarter of loci (550 and 225, respectively) were also output for analysis.

Phylogenetic inference and signal dissection

Phylogenies were inferred from all three datasets using a variety of approaches. First, gene trees were provided to the coalescent-aware summary method ASTRAL-III [39], which employs local posterior probabilities to estimate node support [40]. Second, tree inference was performed under a best-fit partitioned model with IQ-TREE 2 v. 2.1.3 [41-43], using the fast-relaxed clustering algorithm to merge individual loci (using parameters `-m MFP + MERGE -rclusterf 10 -rcluster-max 3000`). Finally, the site-heterogeneous model GTR+CAT-PMSF [44] was used as an efficient alternative to the computationally-demanding CAT model family [45]. For each dataset, short runs of 1,100 generations were done in PhyloBayes-MPI v. 1.8.1 [46] under a fixed topology (that obtained with ASTRAL-III) to approximate site-specific stationary distributions and amino acid exchangeabilities under the GTR+CAT model. Model parameters were summarized after discarding the initial 100 generations as burn-in, and reformatted using scripts available at <https://github.com/drenal/cat-pmsf-paper>. Tree inference was then performed in IQ-TREE 2 under maximum likelihood using the PMSF method [47], setting exchangeabilities and site-specific frequencies to the posterior mean estimates previously obtained with PhyloBayes. For both concatenation approaches, support was estimated using 1000 replicates of ultrafast bootstrap [48].

Given persistent discordance among methods of inference regarding the resolution of major neoholothuriid clades, phylogenetic signal for the alternative topologies was explored using site-wise log-likelihood scores. Scores were computed with IQ-TREE 2 under the same best-fit partitioned model mentioned above, and using three constrained topologies differing only in the position of Molpadida,

with the remaining incongruency fixed to the preferred resolution (see Results). Gene-wise log-likelihood scores were obtained through addition across alignment positions, and their differences for all pairs of topologies—known as Δ GLS values—were computed. Given linear dependency between the three Δ GLS values ($\Delta GLS_x = \Delta GLS_y - \Delta GLS_z$), these were visualized in a two-dimensional space which was rotated using principal components analysis (PCA). A nominal Δ GLS threshold of ± 2 log-likelihood units was used to categorize loci as either informative or uninformative with regards to a given topological comparison.

To explore the drivers of differences in Δ GLS across loci, fifteen gene properties were estimated and treated as potential determinants. These were all calculated by *genesortR* [38], and included commonly used metrics of phylogenetic signal, potential sources of systematic bias, and estimates of the overall information content and evolutionary rate of each individual loci. Further details on these metrics can be found in Table S1.

Potential links between gene properties and the phylogenetic support for alternative neoholothuriid relationships were explored using two different statistical approaches. Associations between the PCA axes derived from Δ GLS values (representing major aspects of phylogenetic signal for competing hypotheses) and explanatory variables were first tested using the 'envfit' function in R package *vegan* [49]. This approach overlaid vectors onto the ordination plot depicting the directions and magnitudes of maximum correlation between individual gene properties and PCA scores. Each predictor was analyzed separately, and the significance of the correlations tested using 10,000 random permutations. Given the possibility of non-linear relationships between predictor and response variables, a second approach was explored in which Δ GLS values were transformed into a single categorical factor with three levels. For this, loci were categorized into: A) uninformative, including those for which all Δ GLS were within ± 2 log-likelihood units; as well as those either B) supporting, or C) rejecting the resolution obtained using ASTRAL-III, defined as exhibiting at least one comparison

favoring or disfavoring such topology, respectively, by an absolute Δ GLS value > 2 . This categorization is supported by analyses showing that the main aspect of differences in phylogenetic signal across loci relates to their support for/against the topology obtained with ASTRAL-III, with very little ability to discriminate between the other alternatives (see Results). A conditional inference classification tree was fit to the data using function 'ctree' in R package *partykit* [50], assessing whether partitioning the data by values of any of the gene properties was able to generate subsets of loci that show similar topological preferences. A Bonferroni correction for multiple comparisons was applied, and significant predictors were visualized on the ordination plot using smooth surfaces fit using penalized regression splines [51].

All statistical analyses were performed in the R environment [52] using code reliant on the packages mentioned above, as well as *adephylo* [53], *ape* [54], *phangorn* [55], *phytools* [56], and those included in the *tidyverse* [57].

Results

Phylogenetic inference under all methods explored and for the three datasets of different sizes recovered highly congruent and well-supported topologies (Figs. S2-4), which were also in broad agreement with the most recent large-scale study based on Sanger-sequenced loci [16]. As summarized in figure 2A, Apodida, Elasipodida and Holothuriida formed successive and monophyletic sister groups to the remainder of sea cucumber diversity included within Neoholothuriida. The latter was further subdivided into four major lineages: Dendrochirotida, Molpadida, Persiculida and Synallactida. Nodes defining all aforementioned clades had maximum support across analyses.

Only two regions of the tree topology showed incongruent resolutions among the analyses performed (Figs. 2B-C). First, ASTRAL-III rejected a close relationship among two of the cucumariid species sampled, *Colochirus robustus* and *Cucumaria georgiana*, which otherwise formed a clade under

concatenation approaches (Fig. 2B). Given the otherwise unambiguous support for a close relationship between *Colochirus* and *Cucumaria*, as well as the poor node support for the ASTRAL-III topology when using the complete dataset, we tentatively favor here the results obtained under concatenation methods. We note, however, that a monophyletic Cucumariidae was not recovered by our analyses regardless of how these terminals are resolved, as they were only distantly related to the remaining cucumariids (*Abysso cucumis*, *Pseudocolochirus*). In fact, discrepancies between our trees and the current family-level classification of sea cucumbers are pervasive, and also included the non-monophyly of elaspodid families Psychropotidae (*Psychropotes*, *Benthodytes*) and Laetmogonidae (*Benthogone*, *Pannychia*), the dendrochirotid family Sclerodactylidae (*Cladolabes*, *Eupentacta*, *Sclerodactyla*), and the synallactid families Synallactidae (*Synallactes*, *Bathyploetes*, *Paelopatides*) and Stichopodidae (*Stichopus*, *Isostichopus*, *Apostichopus*).

A second and more striking topological discordance involved the organization of the four major lineages within Neoholothuriida (Fig. 2C). Each one of the different methods of inference proposed an alternative resolution for the clade, which were recovered regardless of dataset size and strongly supported (values > 95) when employing the complete supermatrix. While all three inference methods agreed on a subtree in which Dendrochirotida and Synallactida share a closer relationship than either one does with Persiculida, the position of Molpadida within this scaffold was highly unstable and methodologically sensitive. Supported alternatives included a placement of Molpadida as sister to either Synallactida, Persiculida, or Synallactida + Dendrochirotida (henceforth referred to as 'ASTRAL', 'partitioned', and 'CAT-PMSF' topologies, respectively). Despite this level of uncertainty, our analyses still reject the long-hypothesized close relationships between Molpadida and Dendrochirotida [13, 15], as well as the topology of Miller et al. [16] in which Dendrochirotida placed as sister group to all other neoholothuriids.

To further explore the phylogenetic signal for competing neoholothuriid topologies, we estimated gene-wise log-likelihood scores for the three alternative resolutions of this clade. A PCA of differences in the scores obtained for pairs of trees (Δ GLS) revealed that the topological preferences of loci could be summarized using a single major underlying axis which accounted for 85.3% of total variance (Fig. 3A). The scores of loci along this first PC axis represented the relative support either for or against the ASTRAL topology (Fig. 3B). On the other hand, the ability of loci to discern between the partitioned and CAT-PMSF trees was much weaker and mainly captured by the second PC axis, which explained only 14.7% of variance. The absolute values of Δ GLS were generally small, with most loci (615 loci, 55.9% of the complete dataset) being relatively uninformative regarding relationships among neoholothuriid clades (Figs. 4A and S5). Nonetheless, the remainder of the dataset was once again roughly evenly split into a fraction that supported the ASTRAL configuration (207 loci, 18.8%), and one that rejected it in favor of either one, or both, of the topological alternatives (278 loci, 25.3%).

None of the fifteen gene properties explored was recovered as a significant predictor of Δ GLS (Fig. 4B). Furthermore, these metrics correlated mostly with PC 2 (Table S1), leaving the major aspect of topological preference entirely unexplained. An alternative approach based on classification trees recovered one significant predictor: uninformative loci were significantly more likely to have a short alignment length, but this property also fails to explain which resolution was preferred by longer and more informative loci (Fig. S6).

Discussion

Phylogenetic incongruence is a hallmark of genome-scale datasets [58-60]. A wide range of biological processes and methodological artifacts can lead phylogenomic datasets to harbor a mixture of phylogenetic signals, which can be differentially amplified by methods of reconstruction to produce

conflicting, yet well-supported, topologies [61-63]. Different avenues have been proposed to ameliorate phylogenetic incongruence and favor a specific resolution for recalcitrant nodes. One strategy is to focus on data filtering, exploring the effects of removing sites and/or loci with unexpectedly high topological preferences [64, 65], or those showing evidence of contributing mostly phylogenetic noise or biases [38, 66]. Alternatively, methods have been developed to dissect alternative signals [67-69] in the hopes that one emerges as a better-justified option. Finally, exploring a range of inference methods, which vary in their realism, complexity, susceptibility to errors, and (potentially) relative fit, can also be used to justify favoring one among several alternative hypotheses [70-72]. However, even after exhaustive testing of these options, a robust resolution for some nodes on the tree of life remains elusive [73-76], awaiting the discovery of novel phylogenetic markers, improved taxon sampling, or methodological developments.

We propose here that the early diversification of neoholothuriid sea cucumbers, an ancient, diverse, and morphologically heterogeneous clade, constitutes another example of a group that defies phylogenetic resolution. Previous studies had acknowledged that a robust topology for Neoholothuriida was likely unattainable with the use of small molecular datasets [16], yet phylogenetic resolution remains out of reach even when employing more than a thousand loci. The reason underlying this uncertainty is not a lack of statistical power, but the presence of multiple signals supporting alternative trees. While a node uniting Dendrochirotida and Synallactida to the exclusion of Persiculida emerges from all our analyses, the position of Molpadida within this topology remains uncertain. A coalescent-aware method of reconstruction places Molpadida inside the clade containing Dendrochirotida and Synallactida, while concatenation-based methods place it outside, with further disagreement emerging depending on whether site-homogeneous or site-heterogeneous models are used. All three of these topological alternatives are well-supported and robust to gene subsampling, and thus represent an example of remarkable methodological sensitivity. Further exploration reveals that our dataset is

unlikely to contain enough information to disambiguate between the topologies supported by alternative concatenation methods. On the other hand, the placements of Molpadida either inside or outside of the node containing Dendrochirotida + Synallactida are each strongly supported by substantial fractions of the data (19 and 25% of loci, respectively).

Complex site-heterogeneous models, such as the CAT family, are likely to fit genome-scale datasets better [47, 77, 78], but the use of model fit statistics when comparing mixture models against other alternatives (such as partitioned models) has been criticized [79]. Furthermore, issues relating to convergence, missing data, and over-parameterization [80-82] have still led many to question the results obtained under CAT models. Similarly, coalescent-aware methods have outperformed concatenation in a number of simulation scenarios [83, 84], yet doubts remain regarding their usefulness to resolve ancient divergences, given that gene tree error is expected to surpass incomplete lineage sorting as the dominant source of incongruence for deep nodes [85]. The fit of summary methods (such as ASTRAL) is also impossible to evaluate relative to that of others, further complicating arriving at an objective way of preferring one method of inference from among those tested here.

In the absence of clear guidance as to which inference method can be preferred, we focused instead on evaluating the amount and quality of the signals supporting alternative placements of Molpadida. We used Δ GLS as proxies for the topological preference of loci (as in e.g., [19, 64, 86]), extending this type of analysis to simultaneously consider three alternative topologies. This allowed us to uncover a strong asymmetry in the ability of loci to distinguish between alternative trees and, as explained above, redirect our efforts to assessing two broad topological alternatives. Although many studies have succeeded in disentangling phylogenetic from non-phylogenetic signals by exploring loci quality [87-89], our attempts failed to find any determinants of topological preference: loci supporting alternative positions of Molpadida do not differ in their levels of phylogenetic signal, systematic biases, amounts of information or evolutionary rates. The only major pattern uncovered is that longer loci are

more likely to harbor some sort of signal, a predictable and relatively trivial result stemming from the increased statistical power of longer alignments. Given a lack of evidence showing that competing hypotheses can be ascribed to systematic biases, we tentatively conclude that biological processes possibly underlie phylogenetic incongruence in this case, and that neoholothuriid evolution might be better explained by ancient events of reticulation, as produced by processes such as ancient hybridization and incomplete lineage sorting.

Although the exact placement of Molpadida remains challenging to ascertain, phylogenomics reveals an otherwise robust higher-level topology for sea cucumbers. These efforts are major step towards a stable classification for the group, corroborating much of the most recent classification based on molecular data [16]. The results presented here constitute a necessary tool with which to elucidate the times of origin, morphological evolution and diversification dynamics of one of the major lineages of marine invertebrates. At the same time, they also show the extent to which the current family-level classification scheme of holothuroids is at odds with their evolutionary history, highlighting the need for phylogenomic investigations with much-expanded taxon sampling and consequent morphological reassessments.

Acknowledgements

Many thanks to Bob Vrijenhoek and David Clague of the Monterey Bay Aquarium Research Institute and the captain and crew of the *R/V Western Flyer* for sampling opportunities. Thanks also to the Schmidt Ocean Institute and chief scientists, Erik Cordes, Pete Girguis, and Lisa Levin and science parties of the *R/V Falkor* cruises FK181005, FK190106, and FK210726. We are also grateful to the captain and crew of the *R/V Robert Gordon Sproul* and the *R/V IB Nathaniel B. Palmer* and science participants during the cruise NBP-1303. Many thanks to the pilots of the remote operated vehicles *Doc*

Ricketts and *SuBastian* for crucial assistance in specimen collection. We also thank Avery Hiley and Marina McCowin for assistance in the laboratory and to Charlotte Seid (SIO-BIC) for specimen cataloguing.

Funding

This work was supported by the Schmidt Ocean Institute, The David and Lucille Packard Foundation, University of California Ship Funds, and the U.S. National Science Foundation (NSF) Division of Environmental Biology (DEB) grants DEB-1036368 and DEB-2036186, and the NSF Office of Polar Programs grant OPP-1043749.

Table 1: Terminals included in the phylogenomic dataset. Voucher numbers are provided for all novel transcriptomic datasets, as well as for those available on EchinoDB [23]. For the latter, some changes in species identifications are noted. NCBI accession numbers of COI sequences are noted when available. Higher-level classification follows Miller et al. [16]. Within major clades, terminals are ordered by their occupancy level. OG = outgroups.

Clade and classification	Species	Voucher number	Occupancy	SRA	Source	COI
Apodida	<i>Synapta</i> sp.	SIO-BIC E5599	68.5		This study	
	<i>Leptosynapta tenuis</i>		61.8	SRR3217898	[21]	
	<i>Leptosynapta clarki</i>		55.2	SRR1695478	[90]	
	<i>Synapta maculata</i>	FLMNH 11293	15.5	SRR2846103	[22, 23]	
Actinopoda, Elasipodida	<i>Pannychia</i> sp.	SIO-BIC E6792	83.3		This study	
	<i>Benthodytes</i> cf. <i>sanguinolenta</i>	SIO-BIC E7146	82.3		This study	
	<i>Benthogone</i> sp.	SIO-BIC E7958	54.0		This study	
	<i>Psychropotes longicauda</i>	SIO-BIC E6788	49.1		This study	
	<i>Peniagone</i> cf. <i>vitrea</i>	SIO-BIC E7961	21.8		This study	
	Actinopoda, Pneumonophora,	<i>Holothuria forskali</i>		94.8	SRR5109955	[91]
	<i>Holothuria scabra</i>		90.6	SRR5713070	[92]	

Holothuriida			SRR5755244	
	<i>Holothuria</i>	88.7	SRR8893063	[93]
	<i>arguinensis</i>			
	<i>Holothuria</i>	82.0	DRR023763	[94]
	<i>leucospilota</i>			
Actinopoda,	<i>Sclerodactyla</i>	90.7	SRR1139189	[95]
Pneumonophora,	<i>briareus</i>			
Neoholothuriida,	<i>Cucumaria</i>	89.9	SRR3190098	[96]
Dendrochirotida	<i>georgiana</i>		SRR3190099	
			SRR3190100	
			SRR3190101	
			SRR3190102	
			SRR3190103	
			SRR3190104	
	<i>Eupentacta</i>	87.9	SRR8297985	[97]
	<i>fraudatrix</i>			
	<i>Colochirus</i>	SIO-BIC 86.8		This study
	<i>robustus</i>	E11213		
	<i>Pseudocolochirus</i>	SIO-BIC 81.0		This study
	<i>violaceus</i>	E7108		
	<i>Cladolabes</i>	69.4	SRR6023959	[98]
	<i>schmeltzii</i>			
	<i>Psolus fabricii</i>	SIO-BIC 65.1		[22] as <i>Psolus</i> sp.
		E6832		11
				[23] as <i>Psolus</i> sp.
	<i>Ypsilothuria</i> n. sp.	SIO-BIC 54.1		This study
		E6922		
	<i>Psolus</i> sp. 41	SIO-BIC 51.3		[22] as <i>Psolus</i> sp.
		E5792		41

					[23] as <i>Psolus</i> sp.
	<i>Abyssocucumis</i>	SIO-BIC	42.3	SRR2830762	[22] as
	<i>albatrossi</i>	E6816			<i>Abyssocucumis</i> sp. [23] as <i>Abyssocucumis</i> cf. <i>albatrossi</i>
Actinopoda,	<i>Molpadia</i>	SIO-BIC	91.9		This study
Pneumonophora,	<i>amorpha</i>	E6366			
Neoholothuriida,	<i>Molpadia</i>	SIO-BIC	12.9	SRR2845419	[22] as <i>Molpadia</i>
Molpadida	<i>granulata</i>	E5790			<i>intermedia</i> [23] as <i>Molpadia</i> <i>granulata</i>
Actinopoda,	<i>Pseudostichopus</i>	SIO-BIC	58.0		[22, 23] KX874388
Pneumonophora,	<i>sp.</i>	E6363			
Neoholothuriida,					
Persiculida					
Actinopoda,	<i>Stichopus</i>	FLMNH	93.3	SRR2846098	[22, 23]
Pneumonophora,	<i>chloronotus</i>	11289			
Neoholothuriida,	<i>Apostichopus</i>		91.1	SRR2484238	[25]
Synallactida	<i>parvimensis</i>				
	<i>Bathyploetes</i> cf.	SIO-BIC	90.9		This study
	<i>moseleyi</i>	E7308			
	<i>Isostichopus</i>		90.2	SRR12552908	[99]
	<i>badionotus</i>			SRR12552909	
				SRR12552910	
				SRR12552911	
	<i>Apostichopus</i>		86.0	SRR6075436	[100]
	<i>japonicus</i>				

	Synallactidae	SIO-BIC	74.7		This study
		E6955			
	<i>Apostichopus</i>		60.4	SRR1695477	[90]
	<i>californicus</i>				
	<i>Synallactes</i> cf.	SIO-BIC	57.9	SRR2895367	[22] as KX874365
	<i>chuni</i>	E5607			<i>Pannychia</i>
					<i>moseleyi</i>
					[23] as
					<i>Synallactes</i> sp.
	<i>Paelopatides</i>	SIO-BIC	48.6		[22] as KX874355
	<i>confundens</i>	E5609			<i>Peniagone</i> sp.
					[23] as
					<i>Paelopatides</i> sp.
Echinoidea (OG)	<i>Strongylocentrotus</i>		79.0	GCA_000002235.4	[25]
	<i>purpuratus</i>			(Spur_5.0)	
	<i>Stereocidaris</i>		75.2	SRR16134559	[101]
	<i>nascaensis</i>				
	<i>Fellaster zelandiae</i>		75.2	SRR16134559	[101]
Astroidea (OG)	<i>Patiria miniata</i>		78.6	GCA_000285935.1	[25]
				(Pmin_1.0)	



Figure 1: Representative holothuroid diversity included in this study. **A.** *Synapta* sp. **B.** *Peniagone* cf. *vitrea*. **C.** *Benthogone* sp. **D.** *Pseudocolochirus violaceus*. **E.** *Abyssocucumis albatrossi*. **F.** *Colochirus robustus*. **G.** *Ypsilothuria* n. sp. (SIO-BIC E6221). **H.** *Molpadia amorpha*. **I.** *Pseudostichopus* sp. **J.** Synallactidae. **K.** *Bathyploetes* cf. *moseleyi*. The classification of these terminals can be found in Table 1; colors for panel letters follow those used for major lineages in Figure 2. All photos except **G** are of the voucher specimens sequenced (catalogue numbers can be found in Table 1; further sampling information is accessible through the SIO-BIC online database, <https://sioapps.ucsd.edu/collections/bi/>).

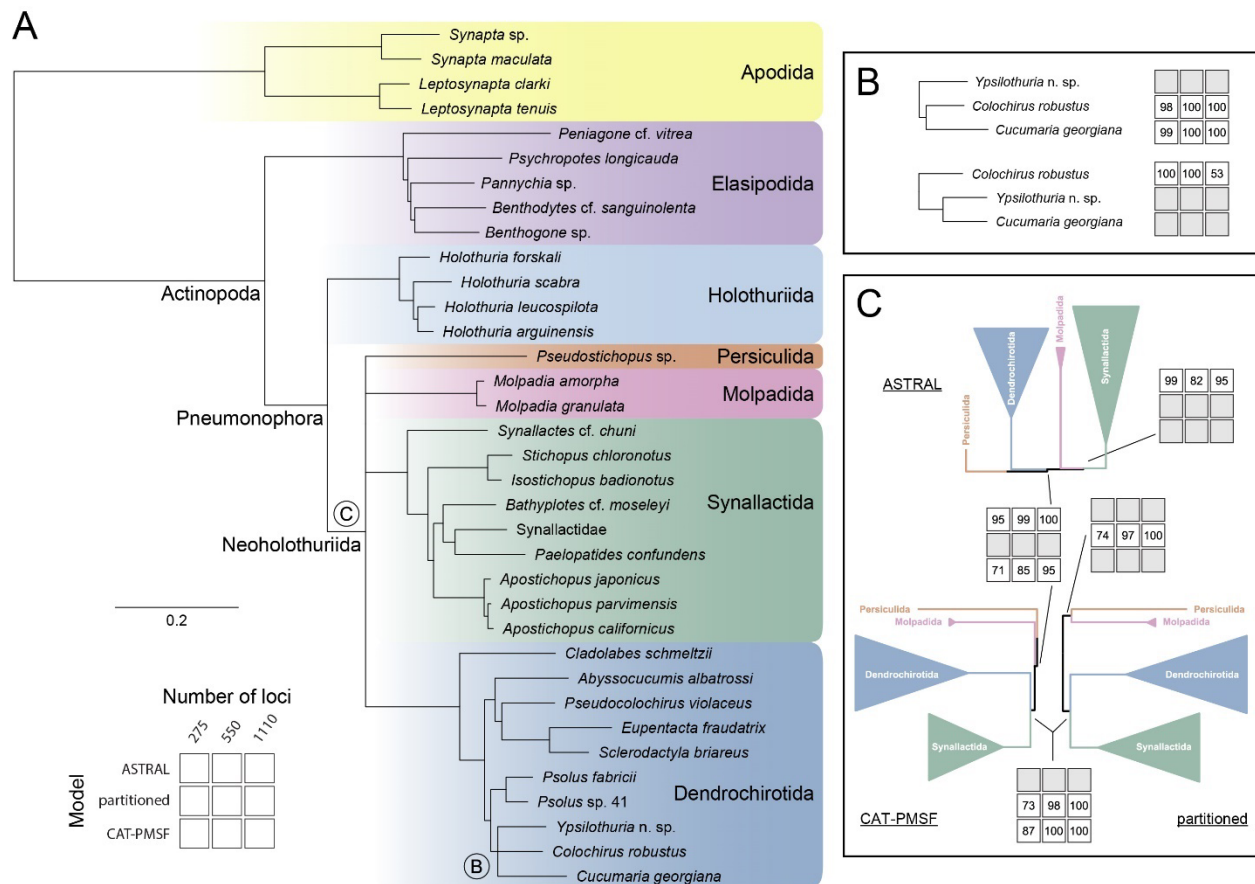


Figure 2: Summary of phylogenetic inference results. **A.** Strict consensus of the nine inference conditions explored, varying both the number of loci and the method of inference. Nodes disagreeing between analyses are collapsed and labeled (see panels **B-C** for further details); branch lengths are otherwise taken from the CAT-PMSF analysis of the full dataset. **B.** Monophyly of a clade composed of two cucumariid terminals, *Colochirus robustus* and *Cucumaria georgiana*, is rejected by ASTRAL-III, but upheld by the other methods (legend for support value grid is shown in **A**). **C.** Systematic disagreement between all methods of inference regarding relationships among major neoholothuriid clades. The resolution favored by each method is found across datasets of different sizes. Topologies, branch lengths and support values for each individual analysis are shown in Figs. S2-4.

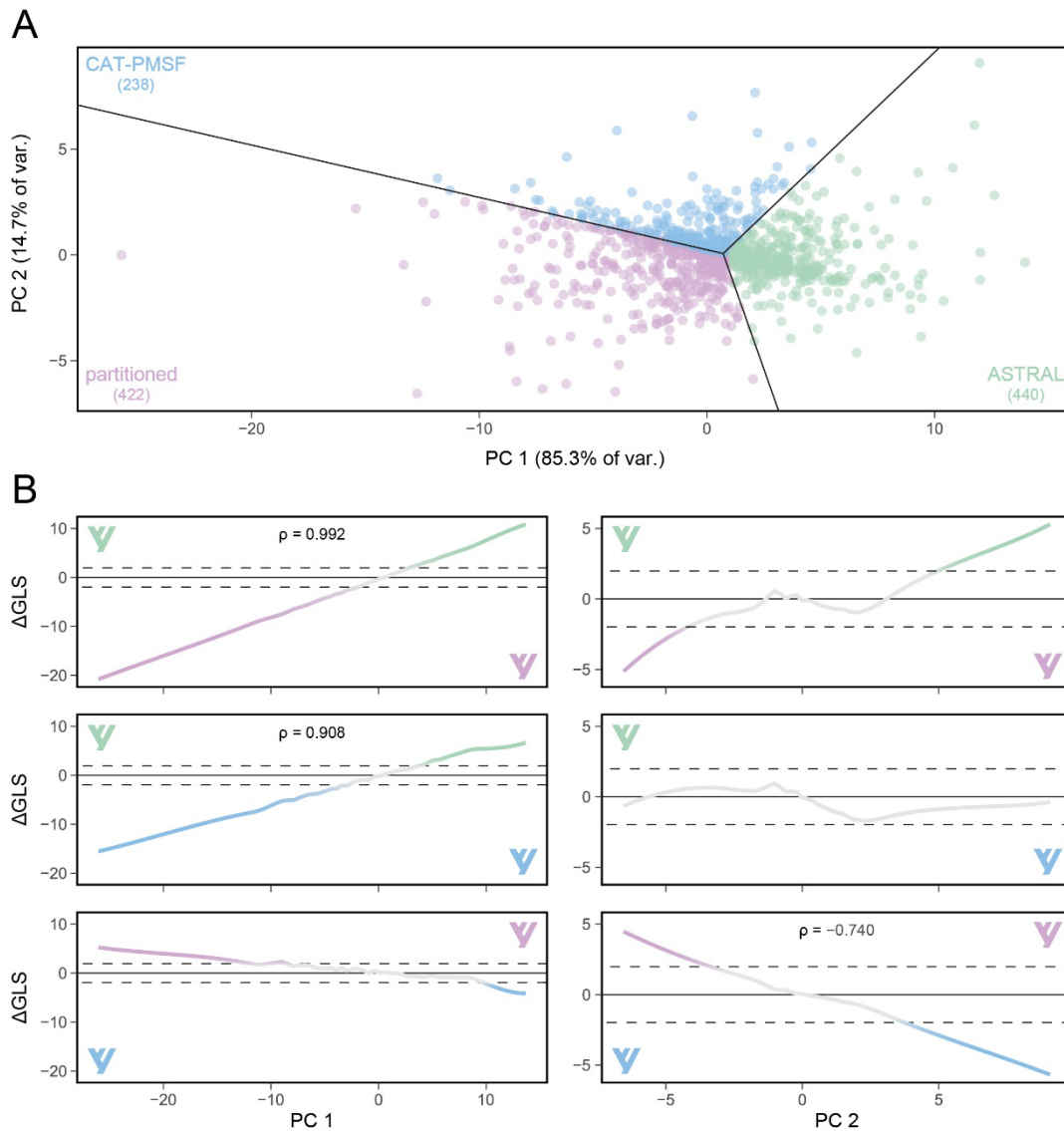


Figure 3: Exploration of support for alternative neoholothuriid topologies across loci. **A.** Principal components (PC) axes obtained from the three Δ GLS. Percentages of explained variance are shown on axis labels. Loci are color coded depending on their favored topology. **B.** Relationship between the PC axes and the scores of individual Δ GLS. Trendlines correspond to LOESS smoothing curves, and ρ values (Spearman's rank correlation coefficients) are shown when absolute values > 0.7 , taken to represent strong correlations. The area included within ± 2 log-likelihood units is highlighted, and considered an area of weak support. Note the markedly different scales of the y-axes for PCs 1 and 2. Topologies are color coded as in A.

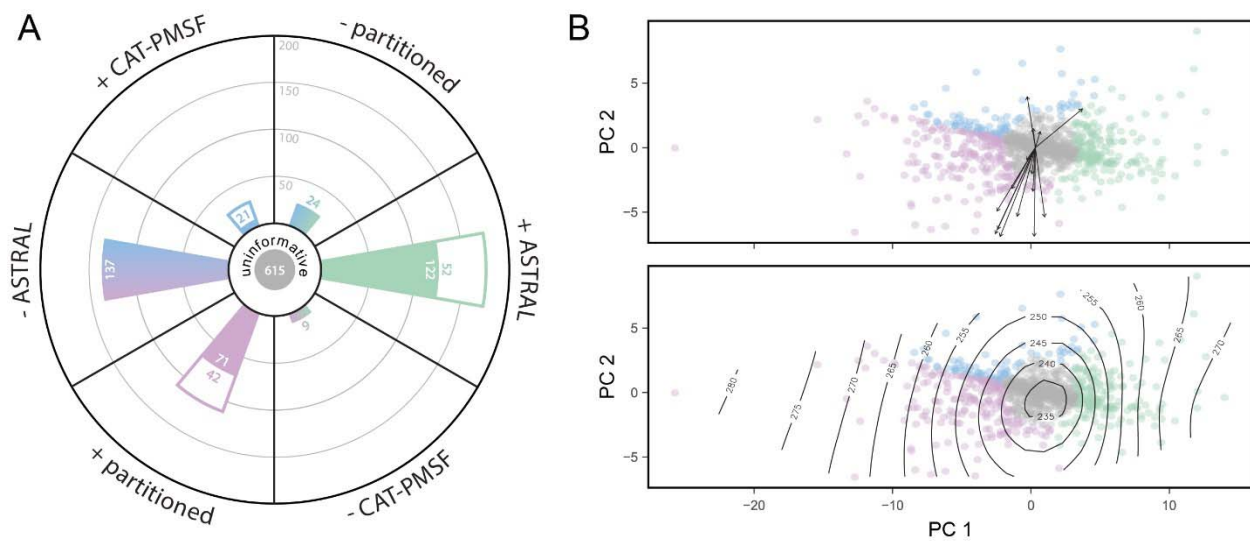


Figure 4: Categorization of loci depending on their favored topology, and exploration of potential determinants. Coloring scheme follows that of Figure 3. **A.** Most loci (615, 55.9% of the full dataset) can be considered uninformative regarding relationships among neoholothuriid clades. The remainder can be classified into those supporting a given topology (denoted using a +) if they favor a given resolution against both alternatives (colored section of bar chart) or only one (white section of bar chart) with a $\Delta\text{GLS} \geq 2$; or rejecting a given topology (denoted using a -). The number of loci either supporting (right side of wheel) or rejecting (left side of wheel) the ASTRAL topology are comparable in number: 207 (18.8%) vs. 278 (25.3%), respectively. Further details on loci categorization can be found in Figure S5. **B.** Top: Exploration of 15 potential determinants of ΔGLS . Arrows indicate directions of maximum correlation between scores and determinants; their length is scaled to the strength of the correlation. Predictors mostly load onto PC 2. R^2 and p -values are shown in Table S1, but no correlation is significant. Bottom: Smoothed surface of alignment length, the only significant determinant found using a classification tree. Longer loci are more likely to be informative, yet alignment length does not predict which topology is preferred (see Fig. S6).

References

- [1] Becher, S. 1906 Über *Synapta minuta* n. sp., eine brutpflegende Synaptide der Nordsee, und über die contractilen Rosetten der Holothurien. *Zoologischer Anzeiger* **30**, 505-509.
- [2] O'Loughlin, P.M. & O'Hara, T.D. 1992 New cucumariid holothurians (Echinodermata) from southern Australia, including two brooding and one fissiparous species. *Memoirs of the Museum of Victoria* **53**, 227-266.
- [3] Flammang, P. & Conand, C. 2004 Functional morphology of the tentacles in the apodid holothuroid *Synapta maculata*. In *Echinoderms: München (Proceedings of the 11th International Echinoderm Conference, Munich, Germany)*, pp. 327-333.
- [4] Purcell, S.W., Samyn, Y. & Conand, C. 2012 *Commercially important sea cucumbers of the world*, FAO.
- [5] Selig, G.M., Netburn, A.N. & Malik, M. 2019 Distributions of the Pelagic Holothurian Pelagothuria in the Central Pacific Ocean as Observed by Remotely-Operated Vehicle Surveys. *Frontiers in Marine Science* **6**, 684.
- [6] O'Hara, T. & Byrne, M. 2017 *Australian echinoderms: biology, ecology and evolution*, CSIRO publishing.
- [7] WoRMS Editorial Board. 2022 World Register of Marine Species. Available from <https://www.marinespecies.org> at VLIZ. Accessed 2022-11-16.
- [8] Gallo, N.D., Cameron, J., Hardy, K., Fryer, P., Bartlett, D.H. & Levin, L.A. 2015 Submersible-and lander-observed community patterns in the Mariana and New Britain trenches: influence of productivity and depth on epibenthic and scavenging communities. *Deep Sea Research Part I: Oceanographic Research Papers* **99**, 119-133.
- [9] Souza Junior, J., Ponte, I., Melo Coe, C., Lobo Farias, W.R., Vieira Feitosa, C., Hamel, J.-F. & Mercier, A. 2017 Sea cucumber fisheries in Northeast Brazil. *SPC Beche-de-mer Information Bulletin* **37**, 43-47.

- [10] Belyaev, G. 1972 *Hadal bottom fauna of the world ocean*, National Science Foundation and Smithsonian Institution.
- [11] Slater, M. & Chen, J. 2015 Sea cucumber biology and ecology. In: *Echinoderm Aquaculture*. New Jersey: Wiley-Blackwell, pp. 47-56.
- [12] Billett, D. 1988 The ecology of deep-sea holothurians, University of Southampton.
- [13] Smirnov, A. 2012 System of the class Holothuroidea. *Paleontological Journal* **46**, 793-832.
- [14] Smirnov, A. 2016 Parallelisms in the evolution of sea cucumbers (Echinodermata: Holothuroidea). *Paleontological Journal* **50**, 1610-1625.
- [15] Kerr, A.M. & Kim, J. 2001 Phylogeny of Holothuroidea (Echinodermata) inferred from morphology. *Zoological journal of the Linnean Society* **133**, 63-81.
- [16] Miller, A.K., Kerr, A.M., Paulay, G., Reich, M., Wilson, N.G., Carvajal, J.I. & Rouse, G.W. 2017 Molecular phylogeny of extant Holothuroidea (Echinodermata). *Molecular Phylogenetics and Evolution* **111**, 110-131.
- [17] Brown, W.I. & Shick, J.M. 1979 Bimodal gas exchange and the regulation of oxygen uptake in holothurians. *The Biological Bulletin* **156**, 272-288.
- [18] Jaekle, W.B. & Strathmann, R.R. 2013 The anus as a second mouth: anal suspension feeding by an oral deposit-feeding sea cucumber. *Invertebrate Biology* **132**, 62-68.
- [19] Mongiardino Koch, N., Coppard, S.E., Lessios, H.A., Briggs, D.E., Mooi, R. & Rouse, G.W. 2018 A phylogenomic resolution of the sea urchin tree of life. *BMC Evolutionary Biology* **18**, 189.
- [20] Linchangco Jr, G.V., Foltz, D.W., Reid, R., Williams, J., Nodzak, C., Kerr, A.M., Miller, A.K., Hunter, R., Wilson, N.G. & Nielsen, W.J. 2017 The phylogeny of extant starfish (Asteroidea: Echinodermata) including *Xyloplax*, based on comparative transcriptomics. *Molecular Phylogenetics and Evolution* **115**, 161-170.

- [21] O'Hara, T.D., Hugall, A.F., Thuy, B. & Moussalli, A. 2014 Phylogenomic resolution of the class Ophiuroidea unlocks a global microfossil record. *Current Biology* **24**, 1874-1879.
- [22] Clouse, R.M., Linchangco Jr, G.V., Kerr, A.M., Reid, R.W. & Janies, D.A. 2015 Phylotranscriptomic analysis uncovers a wealth of tissue inhibitor of metalloproteinases variants in echinoderms. *Royal Society Open Science* **2**, 150377.
- [23] Janies, D.A., Witter, Z., Linchangco, G.V., Foltz, D.W., Miller, A.K., Kerr, A.M., Jay, J., Reid, R.W. & Wray, G.A. 2016 EchinoDB, an application for comparative transcriptomics of deeply-sampled clades of echinoderms. *BMC Bioinformatics* **17**, 48.
- [24] Faircloth, B.C. & Glenn, T.C. 2012 Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLoS One* **7**, e42543.
- [25] Kudtarkar, P. & Cameron, R.A. 2017 Echinobase: an expanding resource for echinoderm genomic information. *Database* **2017**, bax074 <https://doi.org/10.1093/database/bax074>.
- [26] Bolger, A.M., Lohse, M. & Usadel, B. 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120.
- [27] Dunn, C.W., Howison, M. & Zapata, F. 2013 Agalma: an automated phylogenomics workflow. *BMC Bioinformatics* **14**, 330.
- [28] Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R. & Zeng, Q. 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652.
- [29] Ryan, J.F. 2014 Alien Index: identify potential non-animal transcripts or horizontally transferred genes in animal transcriptomes. <https://doi.org/10.5281/zenodo.21029>.
- [30] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. 2009 BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

- [31] Gladyshev, E.A., Meselson, M. & Arkhipova, I.R. 2008 Massive horizontal gene transfer in bdelloid rotifers. *Science* **320**, 1210-1213.
- [32] Guang, A., Howison, M., Zapata, F., Lawrence, C.E. & Dunn, C. 2017 Revising transcriptome assemblies with phylogenetic information. *PLoS One* **16**, e0244202.
- [33] Katoh, K. & Standley, D.M. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-780.
- [34] Talavera, G. & Castresana, J. 2007 Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**, 564-577.
- [35] Morel, B., Kozlov, A.M. & Stamatakis, A. 2018 ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. *Bioinformatics* **35**, 1771-1773.
- [36] Mai, U. & Mirarab, S. 2018 TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19**, 272.
- [37] Mongiardino Koch, N. & Thompson, J.R. 2021 A Total-Evidence Dated Phylogeny of Echinoidea Combining Phylogenomic and Paleontological Data. *Systematic Biology* **70**, 421-439.
- [38] Mongiardino Koch, N. 2021 Phylogenomic subsampling and the search for phylogenetically reliable loci. *Molecular Biology and Evolution* **38**, 4025-4038.
- [39] Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. 2018 ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153.
- [40] Sayyari, E. & Mirarab, S. 2016 Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* **33**, 1654-1668.
- [41] Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A. & Lanfear, R. 2020 IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**, 1530-1534.

- [42] Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K., von Haeseler, A. & Jermiin, L.S. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**, 587.
- [43] Chernomor, O., von Haeseler, A. & Minh, B.Q. 2016 Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65**, 997-1008.
- [44] Szánthó, L.L., Lartillot, N., Szöllősi, G.J. & Schrempf, D. 2022 Compositionally constrained sites drive long branch attraction. bioRxiv 2022.03.03.482715. <https://doi.org/10.1101/2022.03.03.482715>.
- [45] Lartillot, N. & Philippe, H. 2004 A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Molecular Biology and Evolution* **21**, 1095-1109.
- [46] Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. 2013 PhyloBayes MPI. Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biology* **62**, 611-615.
- [47] Wang, H.-C., Minh, B.Q., Susko, E. & Roger, A.J. 2018 Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. *Systematic Biology* **67**, 216-235.
- [48] Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q. & Vinh, L.S. 2018 UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518-522.
- [49] Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. 2019 vegan: community ecology package. R package version 2.5-4. <https://CRAN.R-project.org/package=vegan>.
- [50] Hothorn, T. & Zeileis, A. 2015 partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* **16**, 3905-3909.
- [51] Wood, S.N. 2003 Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65**, 95-114.
- [52] R Core Team. 2019 R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

- [53] Jombart, T. & Dray, S. 2010 adephylo: exploratory analyses for the phylogenetic comparative method. *Bioinformatics* **26**, 1-21.
- [54] Paradis, E. & Schliep, K. 2018 ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528.
- [55] Schliep, K.P. 2011 phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593.
- [56] Revell, L.J. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223.
- [57] Wickham, H. 2017 tidyverse: Easily install and load 'tidyverse' packages. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>.
- [58] Mongiardino Koch, N. 2019 The phylogenomic revolution and its conceptual innovations: a text mining approach. *Organisms Diversity & Evolution* **19**, 99-103.
- [59] Jeffroy, O., Brinkmann, H., Delsuc, F. & Philippe, H. 2006 Phylogenomics: the beginning of incongruence? *TRENDS in Genetics* **22**, 225-231.
- [60] Lozano-Fernandez, J. 2022 A practical guide to design and assess a phylogenomic study. *Genome Biology and Evolution* **14**, evac129.
- [61] Delsuc, F., Brinkmann, H. & Philippe, H. 2005 Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics* **6**, 361-375.
- [62] Kapli, P., Yang, Z. & Telford, M.J. 2020 Phylogenetic tree building in the genomic age. *Nature Reviews Genetics* **21**, 428-444.
- [63] Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Wörheide, G. & Baurain, D. 2011 Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* **9**, e1000602.
- [64] Shen, X.-X., Hittinger, C.T. & Rokas, A. 2017 Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution* **1**, 0126.

- [65] Francis, W.R. & Canfield, D.E. 2020 Very few sites can reshape the inferred phylogenetic tree. *PeerJ* **8**, e8865.
- [66] Dornburg, A., Su, Z. & Townsend, J.P. 2019 Optimal rates for phylogenetic inference and experimental design in the era of genome-scale data sets. *Systematic Biology* **68**, 145-156.
- [67] Arcila, D., Ortí, G., Vari, R., Armbruster, J.W., Stiassny, M.L., Ko, K.D., Sabaj, M.H., Lundberg, J., Revell, L.J. & Betancur-R, R. 2017 Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology & Evolution* **1**, 0020.
- [68] Smith, S.A., Walker-Hale, N., Walker, J.F. & Brown, J.W. 2020 Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology* **69**, 579-592.
- [69] Walker, J.F., Smith, S.A., Hodel, R.G. & Moyroud, E. 2022 Concordance-based approaches for the inference of relationships and molecular rates with phylogenomic data sets. *Systematic Biology* **71**, 943-958.
- [70] Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., Wörheide, G. & Pisani, D. 2017 Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Current Biology* **27**, 3864-3870.
- [71] Song, S., Liu, L., Edwards, S.V. & Wu, S. 2012 Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences* **109**, 14942-14947.
- [72] Redmond, A.K. & McLysaght, A. 2021 Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nature Communications* **12**, 1-14.
- [73] Cunha, T.J., Reimer, J.D. & Giribet, G. 2022 Investigating sources of conflict in deep phylogenomics of vetigastropod snails. *Systematic Biology* **71**, 1009-1022.
- [74] King, N. & Rokas, A. 2017 Embracing uncertainty in reconstructing early animal evolution. *Current Biology* **27**, R1081-R1088.

- [75] Simon, S., Blanke, A. & Meusemann, K. 2018 Reanalyzing the Palaeoptera problem—the origin of insect flight remains obscure. *Arthropod Structure & Development* **47**, 328-338.
- [76] Ontano, A.Z., Gainett, G., Aharon, S., Ballesteros, J.A., Benavides, L.R., Corbett, K.F., Gavish-Regev, E., Harvey, M.S., Monsma, S. & Santibáñez-López, C.E. 2021 Taxonomic sampling and rare genomic changes overcome long-branch attraction in the phylogenetic placement of pseudoscorpions. *Molecular Biology and Evolution* **38**, 2446-2467.
- [77] Si Quang, L., Gascuel, O. & Lartillot, N. 2008 Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317-2323.
- [78] Schrempf, D., Lartillot, N. & Szöllősi, G. 2020 Scalable empirical mixture models that account for across-site compositional heterogeneity. *Molecular Biology and Evolution* **37**, 3616-3631.
- [79] Crotty, S.M. & Holland, B.R. 2022 Comparing partitioned models to mixture models: Do information criteria apply? *Systematic Biology* **71**, 1541-1548.
- [80] Whelan, N.V. & Halanuch, K.M. 2016 Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. *Systematic Biology* **66**, 232-255.
- [81] Li, Y., Shen, X.-X., Evans, B., Dunn, C.W. & Rokas, A. 2021 Rooting the animal tree of life. *Molecular Biology and Evolution* **38**, 4322-4333.
- [82] Al Jewari, C. & Baldauf, S.L. 2022 Conflict over the eukaryote root resides in strong outliers, mosaics and missing data sensitivity of site-specific (CAT) mixture models. *Systematic Biology*, syac029. <https://doi.org/10.1093/sysbio/syac029>.
- [83] Kubatko, L.S. & Degnan, J.H. 2007 Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56**, 17-24.
- [84] Mirarab, S., Bayzid, M.S. & Warnow, T. 2016 Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology* **65**, 366-380.

- [85] Gatesy, J. & Springer, M.S. 2014 Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution* **80**, 231-266.
- [86] Ballesteros, J.A. & Sharma, P.P. 2019 A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. *Systematic Biology* **68**, 896-917.
- [87] Ballesteros, J.A., Santibáñez-López, C.E., Baker, C.M., Benavides, L.R., Cunha, T.J., Gainett, G., Ontano, A.Z., Setton, E.V., Arango, C.P. & Gavish-Regev, E. 2022 Comprehensive species sampling and sophisticated algorithmic approaches refute the monophyly of Arachnida. *Molecular Biology and Evolution* **39**, msac021.
- [88] Burbrink, F.T., Grazziotin, F.G., Pyron, R.A., Cundall, D., Donnellan, S., Irish, F., Keogh, J.S., Kraus, F., Murphy, R.W. & Noonan, B. 2020 Interrogating genomic-scale data for Squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. *Systematic Biology* **69**, 502-520.
- [89] Smith, B.T., Merwin, J., Provost, K.L., Thom, G., Brumfield, R.T., Ferreira, M., Mauck, W.M., Moyle, R.G., Wright, T.F. & Joseph, L. 2022 Phylogenomic analysis of the parrots of the world distinguishes artifactual from biological sources of gene tree discordance. *Systematic Biology*, syac055.
<https://doi.org/10.1093/sysbio/syac055>.
- [90] Cannon, J.T., Kocot, K.M., Waits, D.S., Weese, D.A., Swalla, B.J., Santos, S.R. & Halanych, K.M. 2014 Phylogenomic resolution of the hemichordate and echinoderm clade. *Current Biology* **24**, 2827-2832.
- [91] Demeuldre, M., Hennebert, E., Bonneel, M., Lengerer, B., Van Dyck, S., Wattiez, R., Ladurner, P. & Flammang, P. 2017 Mechanical adaptability of sea cucumber Cuvierian tubules involves a mutable collagenous tissue. *Journal of Experimental Biology* **220**, 2108-2119.

- [92] Mitu, S.A., Bose, U., Suwansa-Ard, S., Turner, L.H., Zhao, M., Elizur, A., Ogbourne, S.M., Shaw, P.N. & Cummins, S.F. 2017 Evidence for a saponin biosynthesis pathway in the body wall of the commercially significant sea cucumber *Holothuria scabra*. *Marine Drugs* **15**, 349.
- [93] Marquet, N., Cardoso, J.C., Louro, B., Fernandes, S.A., Silva, S.C. & Canário, A.V. 2020 Holothurians have a reduced GPCR and odorant receptor-like repertoire compared to other echinoderms. *Scientific Reports* **10**, 3348.
- [94] Chieu, H.D., Suwansa-ard, S., Abramov, T., Elizur, A. & Cummins, S.F. 2018 In vitro oocyte maturation by radial nerve extract and early development of the black sea cucumber (*Holothuria leucospilota*). *Aquaculture* **495**, 247-254.
- [95] Reich, A., Dunn, C., Akasaka, K. & Wessel, G. 2015 Phylogenomic analyses of Echinodermata support the sister groups of Asterozoa and Echinozoa. *PLoS One* **10**, e0119627.
- [96] Clark, M.S., Sommer, U., Sihra, J.K., Thorne, M.A., Morley, S.A., King, M., Viant, M.R. & Peck, L.S. 2017 Biodiversity in marine invertebrate responses to acute warming revealed by a comparative multi-omics approach. *Global Change Biology* **23**, 318-330.
- [97] Boyko, A.V., Girich, A.S., Tkacheva, E.S. & Dolmatov, I.Y. 2020 The *Eupentacta fraudatrix* transcriptome provides insights into regulation of cell transdifferentiation. *Scientific Reports* **10**, 1522.
- [98] Dolmatov, I.Y., Afanasyev, S.V. & Boyko, A.V. 2018 Molecular mechanisms of fission in echinoderms: Transcriptome analysis. *PLoS One* **13**, e0195836.
- [99] Martín-Hernández, R., Rodríguez-Canul, R., Kantún-Moreno, N., Olvera-Novoa, M.A., Medina-Contreras, O., Garikoitz-Legarda, C., Triviño, J.C., Zamora-Briseño, J.A., May-Solis, V. & Poot-Salazar, A. 2021 Comparative Transcriptomes of the Body Wall of Wild and Farmed Sea Cucumber *Isostichopus badiyonotus*. *International Journal of Molecular Sciences* **22**, 3882.

[100] Boyko, A.V., Girich, A.S., Eliseikina, M.G., Maslennikov, S.I. & Dolmatov, I.Y. 2019 Reference assembly and gene expression analysis of *Apostichopus japonicus* larval development. *Scientific Reports* **9**, 1131.

[101] Mongiardino Koch, N., Thompson, J.R., Hiley, A.S., McCowin, M.F., Armstrong, A.F., Coppard, S.E., Aguilera, F., Bronstein, O., Kroh, A. & Mooi, R. 2022 Phylogenomic analyses of echinoid diversification prompt a re-evaluation of their fossil record. *eLife* **11**, e72460.

Supplementary Information – Confusion will be my epitaph: Genome-scale discordance stifles phylogenetic resolution of Holothuroidea

Nicolás Mongiardino Koch, Ekin Tilic, Allison K. Miller, Josefin Stiller, Greg W. Rouse

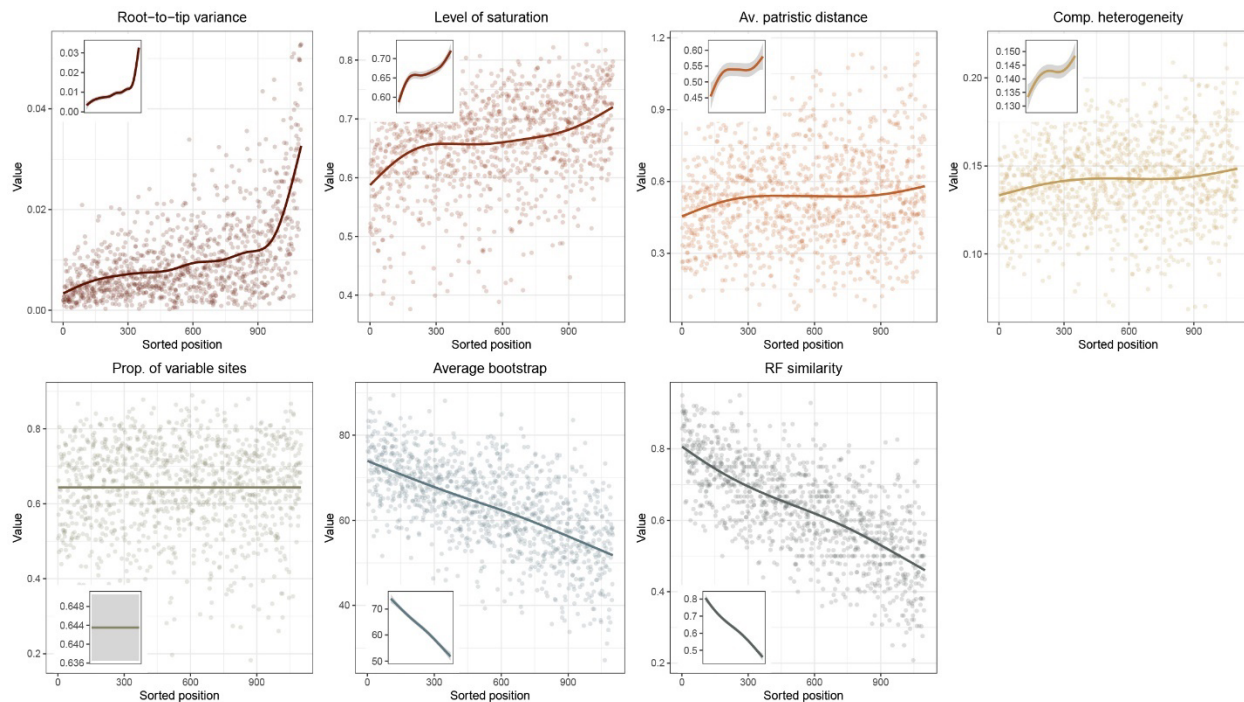


Figure S1: Distribution of gene properties after sorting the dataset for phylogenetic usefulness. Only ingroup taxa were considered for the estimation of these variables. Sorting was performed on PC 2, which explained 18.92% of total variance. Highly useful loci show limited evidence of suffering from systematic biases (top row), while having high phylogenetic signal (bottom row). The proportion of variable sites does not strongly load onto this axis.



Figure. S2: Topologies inferred using ASTRAL-III using the most phylogenetically useful quarter and half of loci (top left and top right, respectively), as well as the full dataset (bottom). Node values represent local posterior probabilities scaled to a maximum value of 100. Major holothuroid clades are colored as in Figure 2. Note that external branch lengths are meaningless in ASTRAL trees.

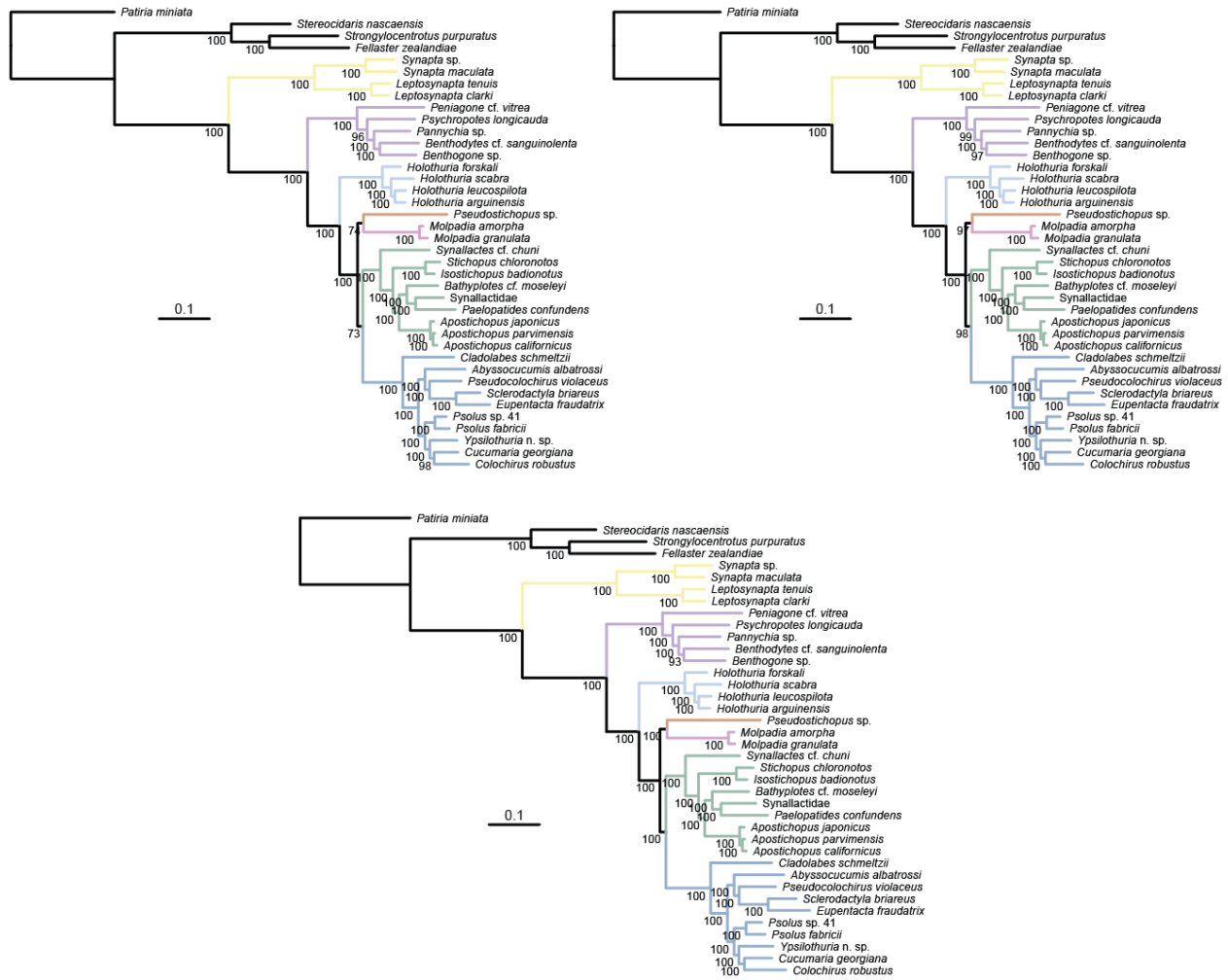


Figure S3: Topologies inferred using a best-fit partitioned model in IQ-TREE2 using the most phylogenetically useful quarter and half of loci (top left and top right, respectively), as well as the full dataset (bottom). Node values represent ultrafast bootstrap frequencies. Major holothuroid clades are colored as in Figure 2.

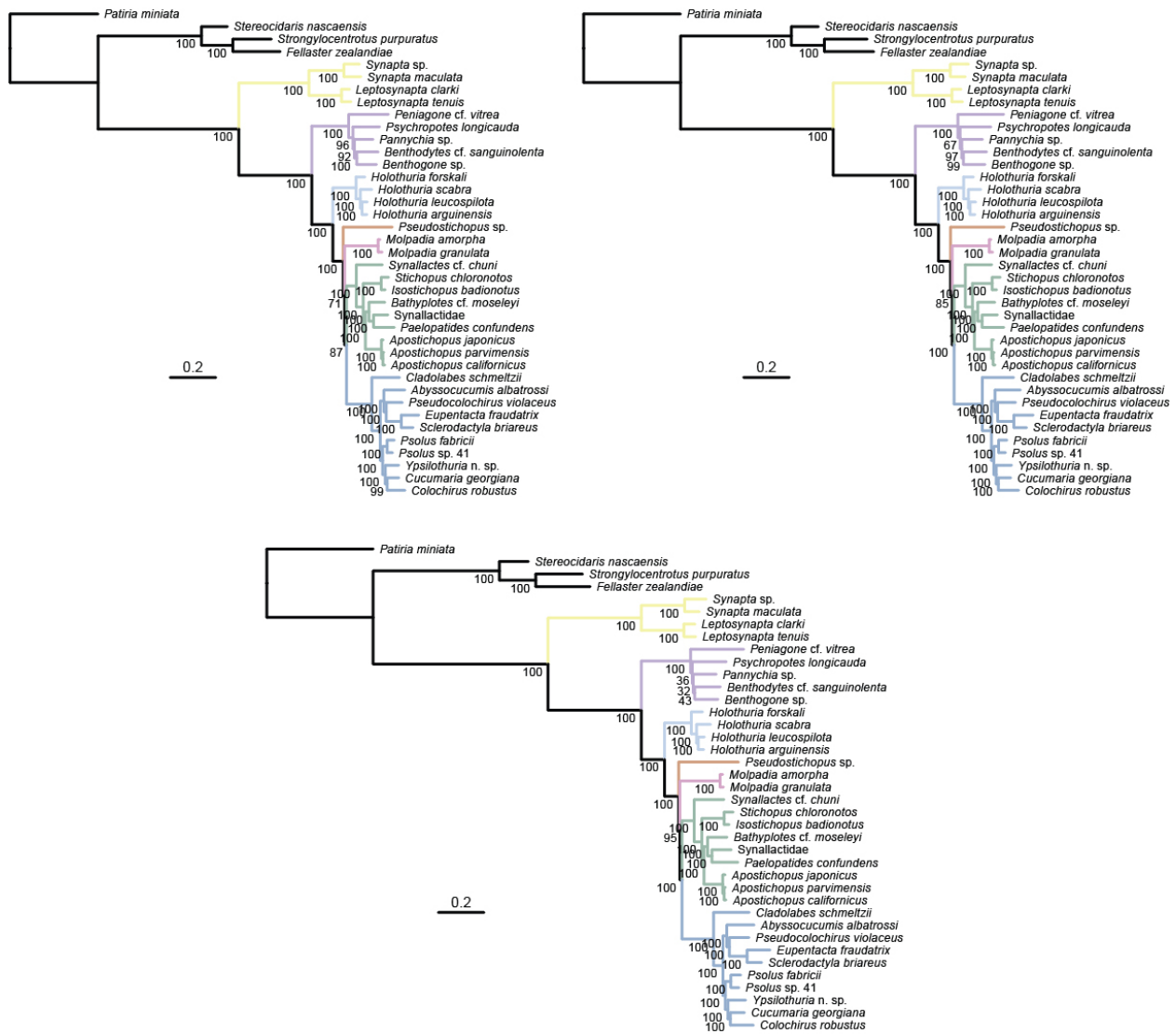


Figure S4: Topologies inferred using the site-heterogeneous model GTR+CAT-PMSF in IQ-TREE2 (with exchangeabilities and site-specific frequencies fixed to those obtained with PhyloBayes) using the most phylogenetically useful quarter and half of loci (top left and top right, respectively), as well as the full dataset (bottom). Node values represent ultrafast bootstrap frequencies. Major holothuroid clades are colored as in Figure 2.

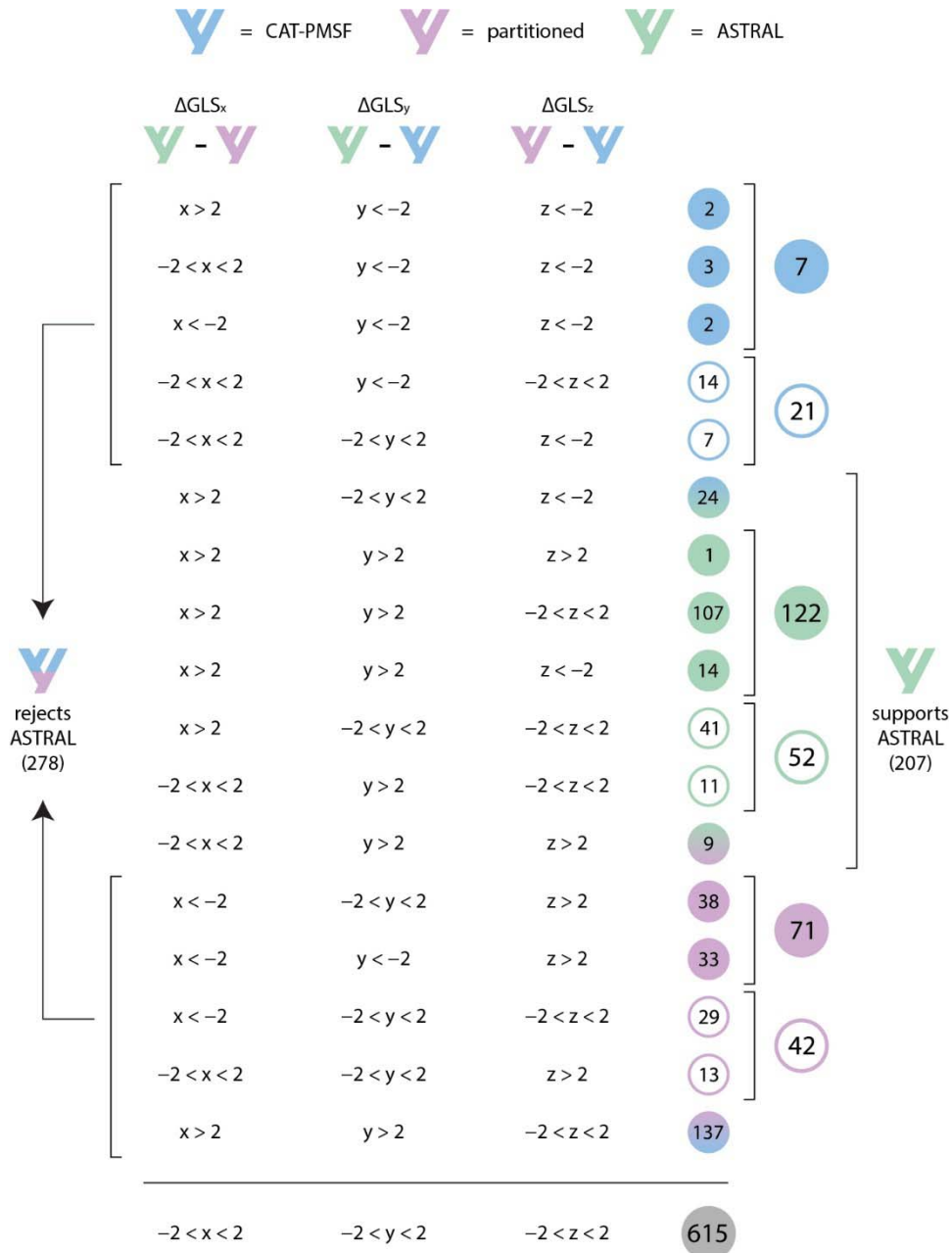


Figure S5: Categorization of loci following the application of ΔGLS thresholds of ± 2 log-likelihood units.

Although 27 possible combinations exist, only 18 are realized. Filled and empty circles represent total or partial support for a given topology, respectively. Circles with a color gradient denote loci that provide support for two topological alternatives (i.e., consistently reject the third option). This same data is further summarized in Figure 4A.

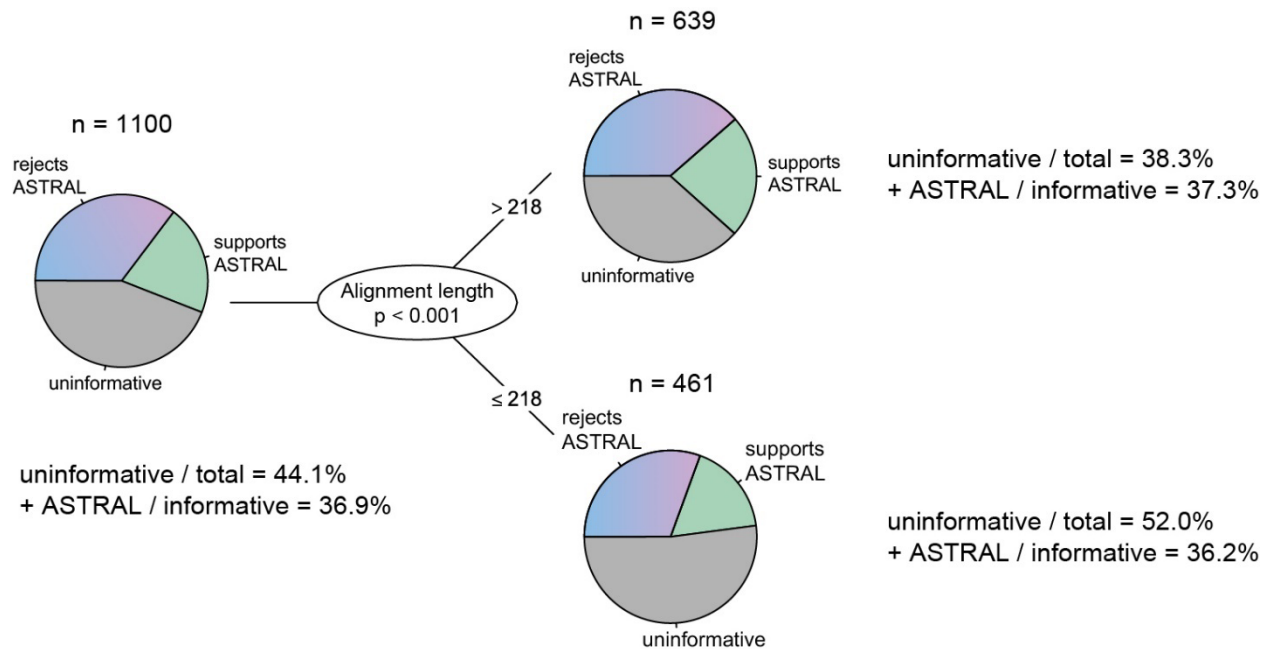


Figure S6: Results of the analysis of topological determinants using a classification tree. A single variable, alignment length, was found to be significantly associated with whether loci support or reject the ASTRAL resolution of Neoholothuriida, or remain topologically uninformative. Partitioning the data into small and large loci (with an automatically detected threshold of 218 base pairs) results in subsets that are enriched and depleted, respectively, in uninformative loci. The relative support for/against the topological alternatives, however, is not modified by this partitioning.

Table 1: Results of the exploration of topological determinants. Fifteen gene properties obtained with *genesortR* (see [38] for a definition of each) where correlated against the PC scores of Δ GLS. The magnitude of correlation against each is expressed as the endpoint of a vector of length = 1. R^2 values represent squared correlation coefficients, p -values were obtained using 10,000 random permutations.

Gene property	Type	Correlation against PC 1	Correlation against PC 2	R^2	p -value
Root-to-tip variance	Bias	-0.012	-1.000	0.002	0.311
Level of saturation	Bias	-0.469	-0.883	0.001	0.731
Average pairwise patristic distance	Bias	-0.393	-0.919	0.002	0.317
Compositional heterogeneity (RCFV)	Bias	0.123	-0.992	0.001	0.489
Average bootstrap support	Signal	-0.547	-0.837	0.000	0.969
Robinson-Foulds similarity	Signal	-0.122	-0.993	0.000	0.908
Phylogenetic usefulness (PC 2 of <i>genesortR</i>)	Signal	-0.052	-0.999	0.000	0.754
Proportion of missing data	Information	-0.111	0.994	0.000	0.951
Tree length	Information	-0.344	-0.939	0.002	0.270
Treeness	Information	0.747	0.664	0.001	0.613
Alignment length	Information	-0.147	0.989	0.001	0.689
Level of occupancy	Information	0.272	0.962	0.000	0.962
Proportion of variable sites	Information	-0.490	-0.872	0.001	0.460
Tree length divided by number of terminals	Rate	-0.397	-0.918	0.002	0.278
PC 1 of <i>genesortR</i>	Rate	-0.238	-0.971	0.001	0.476