

RESEARCH ARTICLE

# Phylogenetic Signal Dissection Identifies the Root of Starfishes

Roberto Feuda<sup>1\*</sup>, Andrew B. Smith<sup>2</sup>

**1** Division of Biology and Biological Engineering, California Institute of Technology Pasadena, California, United States of America, **2** Department of Earth Sciences, The Natural History Museum, London, United Kingdom

\* [rfeuda@caltech.edu](mailto:rfeuda@caltech.edu)

## Abstract

Relationships within the class Asteroidea have remained controversial for almost 100 years and, despite many attempts to resolve this problem using molecular data, no consensus has yet emerged. Using two nuclear genes and a taxon sampling covering the major asteroid clades we show that non-phylogenetic signal created by three factors - Long Branch Attraction, compositional heterogeneity and the use of poorly fitting models of evolution – have confounded accurate estimation of phylogenetic relationships. To overcome the effect of this non-phylogenetic signal we analyse the data using non-homogeneous models, site stripping and the creation of subpartitions aimed to reduce or amplify the systematic error, and calculate Bayes Factor support for a selection of previously suggested topological arrangements of asteroid orders. We show that most of the previous alternative hypotheses are not supported in the most reliable data partitions, including the previously suggested placement of either Forcipulatida or Paxillosida as sister group to the other major branches. The best-supported solution places Velatida as the sister group to other asteroids, and the implications of this finding for the morphological evolution of asteroids are presented.



## OPEN ACCESS

**Citation:** Feuda R, Smith AB (2015) Phylogenetic Signal Dissection Identifies the Root of Starfishes. PLoS ONE 10(5): e0123331. doi:10.1371/journal.pone.0123331

**Academic Editor:** Axel Janke, BiK-F Biodiversity and Climate Research Center, GERMANY

**Received:** December 4, 2014

**Accepted:** February 20, 2015

**Published:** May 8, 2015

**Copyright:** © 2015 Feuda, Smith. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

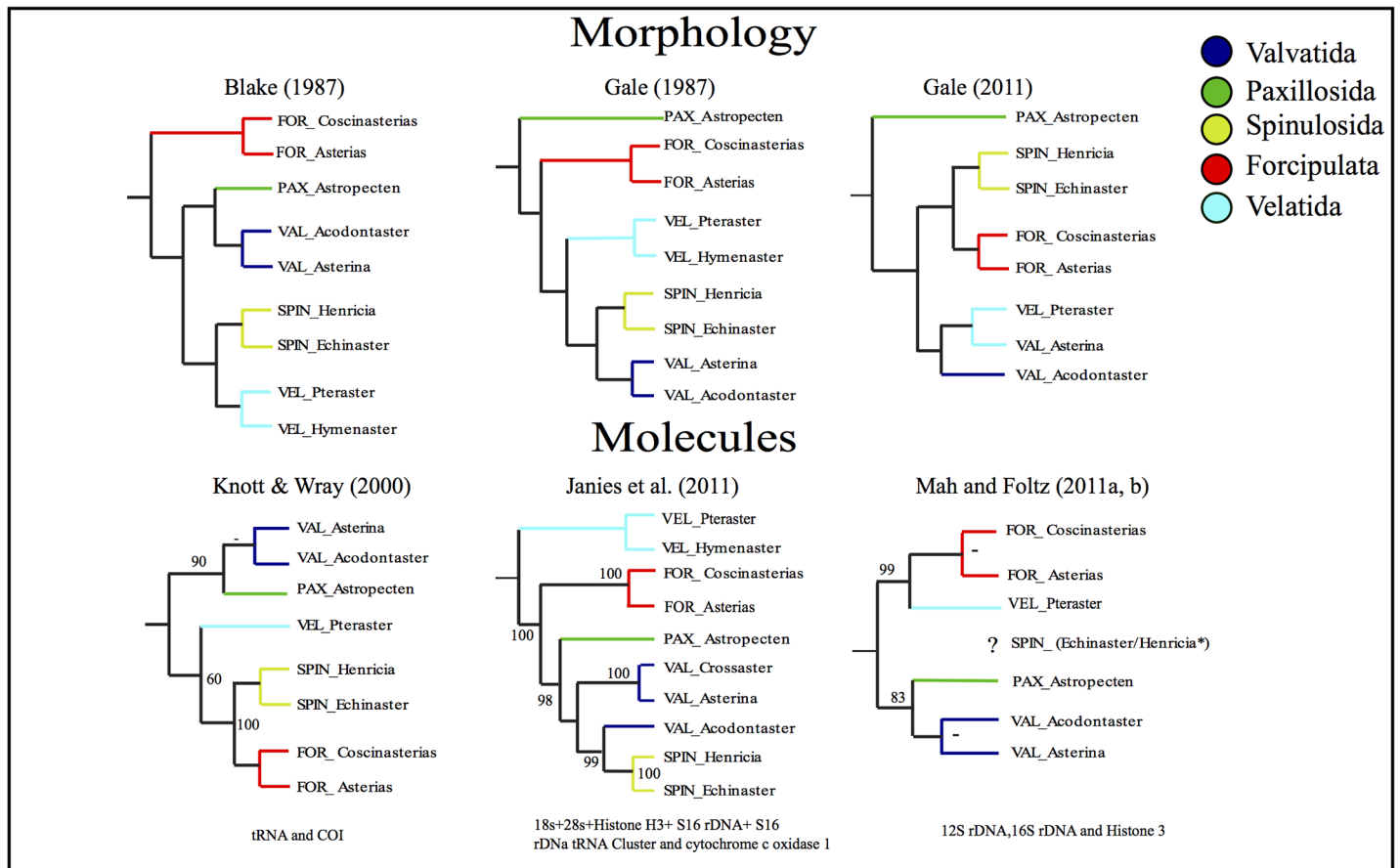
**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Starfishes (Asteroidea) are a morphologically well-defined clade, the most iconic of the five extant classes of echinoderm. This group includes around 1900 extant species classified into five major orders: Paxillosida, Spinulosida, Velatida, Valvatida and Forcipulatida [1]. While there has never been any doubt about the monophyly of the crown group from a morphological [2–5] or molecular [6–9] perspective, relationships among the orders are far from settled. Disagreement continues in particular about how the crown group should be rooted and, consequently, the relationships of the various orders. This argument started in 1921 with the debate between Mortensen and MacBride [10–11], and ignited again in 1987 when two morphology-based phylogenies of the Asteroidea were published that came to very different conclusions (Fig 1). Blake [4] identified the order Forcipulatida as sister group to other asteroids whereas Gale [2] followed traditional interpretations placing Paxillosida in that position. Since then Blake [5,



**Fig 1. Current competing hypotheses of relationships derived from morphological or molecular data for the major starfish clades considered in this paper.**

doi:10.1371/journal.pone.0123331.g001

[12], Gale [3] and others [13] have continued to debate the relative merits of each interpretation from a morphological perspective.

Given this striking disagreement amongst morphologists, various attempts have been made to resolve asteroid relationships using molecular data [7–8, 14–19]. Unfortunately these too have failed to arrive at a consistent answer and, depending upon taxa included, genes analysed and method of analysis employed, have identified a variety of possible taxa at the base of crown group asteroids (Fig 1). Initial analyses based on small data sets [14–15] identified Paxillosida as sister group to other asteroids. However, Knott and Wray [17] using two mitochondrial genes found neither paxillosids nor forcipulatids as basal, instead placing a paraphyletic valvatids basal with velatids nested within this grade. Janies’ [7] combined morphological and molecular investigation of echinoderm class relationships included 16 asteroid genera. Although not specifically examining asteroid rooting this presented a fourth topology again with valvatids as paraphyletic but with velatids as basal. A later analysis [8], again focusing on class relationships but including 35 asteroids and partial sequences from 7 genes, identified a clade comprising two velatids (*Pteraster* and *Hymenaster*) and the highly divergent *Xyloplax* as sister group to other asteroids with Forcipulata as the next major clade to branch after that. The recent extensive studies of asteroid relationships by Mah & Foltz [18–19] grouped Velatida with Forcipulata and Paxillosida with Valvatida but left the Spinulosida unplaced. Thus both the

branching pattern and the root position of the asteroid tree remains disputed and [Fig 1](#) summarises the major competing topologies that have been proposed.

There are many potential reasons why molecular data have generated different results, as each analysis has used different approaches applied to different suites of taxa and gene sequences of varying completeness. Significantly, none have seriously attempted to correct for potential systematic bias in their data. It is now widely recognized that non-phylogenetic signal is a common problem that can distort molecular phylogenies [20–21]. Non-phylogenetic signal has multiple and disparate sources. The use of suboptimal models of evolution, missing data [22], the presence of fast evolving taxa and compositional heterogeneity are all widely recognized as potential sources of false signal, with the latter two effects causing sequences to be erroneously grouped according to their relative branch lengths or analogous nucleotide or amino acid composition [23–24]. Here we apply a suite of newly developed statistical techniques to a data set that includes representatives of the major starfish clades in order to study the effect of non-phylogenetic signal on our perception of starfish relationships.

## Methods and Materials

### Choice of taxa and genes

Representative species from each of the five major asteroid clades were selected: *Astropecten* (Paxillosida); *Asterias* and *Coscinasterias* (Forcipulatida); *Henricia* and *Echinaster* (Spinulosida); *Asterina* and *Acodontaster* (Valvatida) and *Pteraster* and *Hymenaster* (Velatida)—classification follows [18–19]. As the statistical power of our approach improves with larger data sets and missing data could generate artifacts [22], we selected the two representatives with the most complete gene sequences in each clade. Two orders, Brisingida and Cocentricycloidea, each encompassing relatively few species, could not be included in our analyses as gene sequence data for these deep-sea taxa was very incomplete. A selection of hemichordates and representatives of the four other echinoderm classes were included as outgroups (Table A in [S1 File](#)). Sequence data were assembled for two nuclear ribosomal genes (18S and 28S rRNA) and two mitochondrial ribosomal genes (12S and 16S rRNA). However, applying posterior probability analysis, as implemented in Phylobayes [25], we found that the amount of homoplasy was significantly higher in the combined dataset than in the dataset of nuclear genes (Table D in [S1 File](#)). Consequently, mitochondrial ribosomal genes were excluded from the analysis because of the high levels of saturation they displayed. [S1 Dataset](#) lists taxa and the sequences used. After alignment any regions that could not be unambiguously aligned across both in-group and outgroup were deleted. The final data set comprises 31 species and 3017 positions and is available as Electronic Supplementary Data.

### Phylogenetic analysis

We first identified the best fitting model using a 12-fold Bayesian Cross-validation as implemented in Phylobayes 3.3e [25]. We compared the site-heterogeneous models CAT-GTR- $\Gamma$  and CAT- $\Gamma$  versus the site homogeneous model GTR- $\Gamma$ . Results of the cross validation (Table B in [S1 File](#)) suggest that CAT-GTR- $\Gamma$  generates the best fit. Accordingly our phylogenetic reconstruction for the full data set was performed using Bayesian Analysis under the optimal CAT-GTR- $\Gamma$  model and the worst performing GTR- $\Gamma$  model using Phylobayes3.3e [25]. Among site rate variation was modeled using a discrete Gamma distribution (4 rate categories). For all Phylobayes analyses two runs were performed and convergence was investigated using the bpcomp option (part of the Phylobayes package). Phylogenetic reconstruction was performed using maximum likelihood methods as implemented in PhyML under GTR- $\Gamma$ . Node supports were evaluated using aBayes [26].

### Phylogenetic signal dissection and compositional heterogeneity

Phylogenetic signal dissection [27] was performed to assess the effect of fast evolving sites on tree topology. Site-specific rates of evolution were estimated using the program TIGER [28], which assigns evolutionary rates to characters and places them in bins of approximately equal rate. This is tree independent, eliminating the need for *a priori* tree specification [28]. Two partitions were created: the first containing the fastest evolving sites plus invariant sites (our ‘heterogeneous rates partition’- 2284 positions); the second contained all the remaining variant sites (our ‘homogeneous rates partition’- 733 positions). Because the first data partition combines sites with extreme rate variation this partition poses an extreme problem for tree reconstructing methods and is expected to be misled more readily by systematic bias [29].

The presence of compositional heterogeneity was evaluated using posterior predictive analysis (PPA) as implemented in Phylobayes [25]. As composition heterogeneity was found to be an important issue for the Asteroidea (see Table C in S1 File) we applied two approaches to minimize the problem. First we performed the phylogenetic reconstruction of the full data set using the non-stationary CAT-break point model (CATBP) [30], as implemented in Nh-phylobayes [30]. This model, by implementing a multiple break point along the lineages, is able to account directly for compositional heterogeneity [30]. Among site variation was modeled using four discrete gamma categories. Convergence among chains was evaluated using comp-chain software, which is part of the Nh-phylobayes package. Additionally we performed site stripping as implemented in the program BMGE [31]. BMGE used a Stuart’s test of marginal homogeneity to remove compositionally heterogeneous sites. This removed 278 heterogeneous sites from our original data and generated a partition contain only compositionally homogeneous sites.

As heterogeneity of both rates and composition has been identified as potential problems, we generated a final partition removing the heterogeneous compositional sites from the homogenous rates partition (our ‘composition+rate homogeneous partition’- 556 positions). This data set comprises the best set of sites with a *bona-fide* phylogenetic signal. Since that performance of complex model such as CAT-GTR seems to be less efficient on small alignment (less than 1000 position—[32]) we decided to analyse all subpartitions under GTR- $\Gamma$  using Phylobayes3.3e. However, CAT-GTR- $\Gamma$  was used to confirm the topology obtained using the most reliable composition+rate homogeneous partition.

The overall quality of the partition was evaluated using statistical criteria (see Table C in S1 File) and biological criteria (Table 1). While the phylogenetic relationships amongst asteroid taxa remain uncertain, some relationships in other echinoderm classes are securely founded,

**Table 1. This table summarizes the ability of the various partitions to identify well-established relationship amongst echinoderms other than the Asteroidea.**

	Class relations	Echinoids	Ophiuroids	Holothurians	Crinoids
CAT-GTR-full	✓	x	x	✓	✓
CAT-BP	✓	✓	✓	✓	✓
GTR-full	✓	✓	✓	✓	✓
GTR hetero rates	x	x	x	x	✓
GTR homo comp	✓	✓	✓	✓	✓
GTR homo rates	✓	✓	✓	✓	✓
CAT-GTR, GTR rates+comp homo	✓	✓	✓	✓	✓

X = the partition failed to recover the expected topology within the clade indicated; ✓ = the partition identified the expected topology within the clade indicated.

doi:10.1371/journal.pone.0123331.t001

supported by both morphological and molecular data. The position of cidaroids (*Calocidaris*, *Stereocidaris*) as sister group to euechinoids (*Arbacia*, *Paracentrotus* and *Strongylocentrotus*) is unambiguously supported [33–34], as is the sister group relationship of Euryalida (*Asteronyx*, *Gorgonocephalus*) to Ophiurida (*Ophioderma*, *Ophiopsammus*, *Ophiocoma*, *Ophiothrix*, *Ophiophilis*) [35]. Similarly both morphological and molecular data place Elaspoda (*Psychropotes*) as sister group to the Aspidochirotida (*Cucumaria*, *Holothuria*) [36–37]. The relationships amongst some of the five classes of echinoderms are also largely uncontroversial, with crinoids as sister group to the other four classes and holothurians and echinoids as sister groups [29, 38]. We used these relationships as an additional criterion: data and methods that do not recover these well-founded relationships cannot be relied upon to have identified the correct root position for asteroids. Conversely, only when our data and methods identify these clades can we start to have faith in the root position identified for asteroids.

## Testing support for competing hypotheses

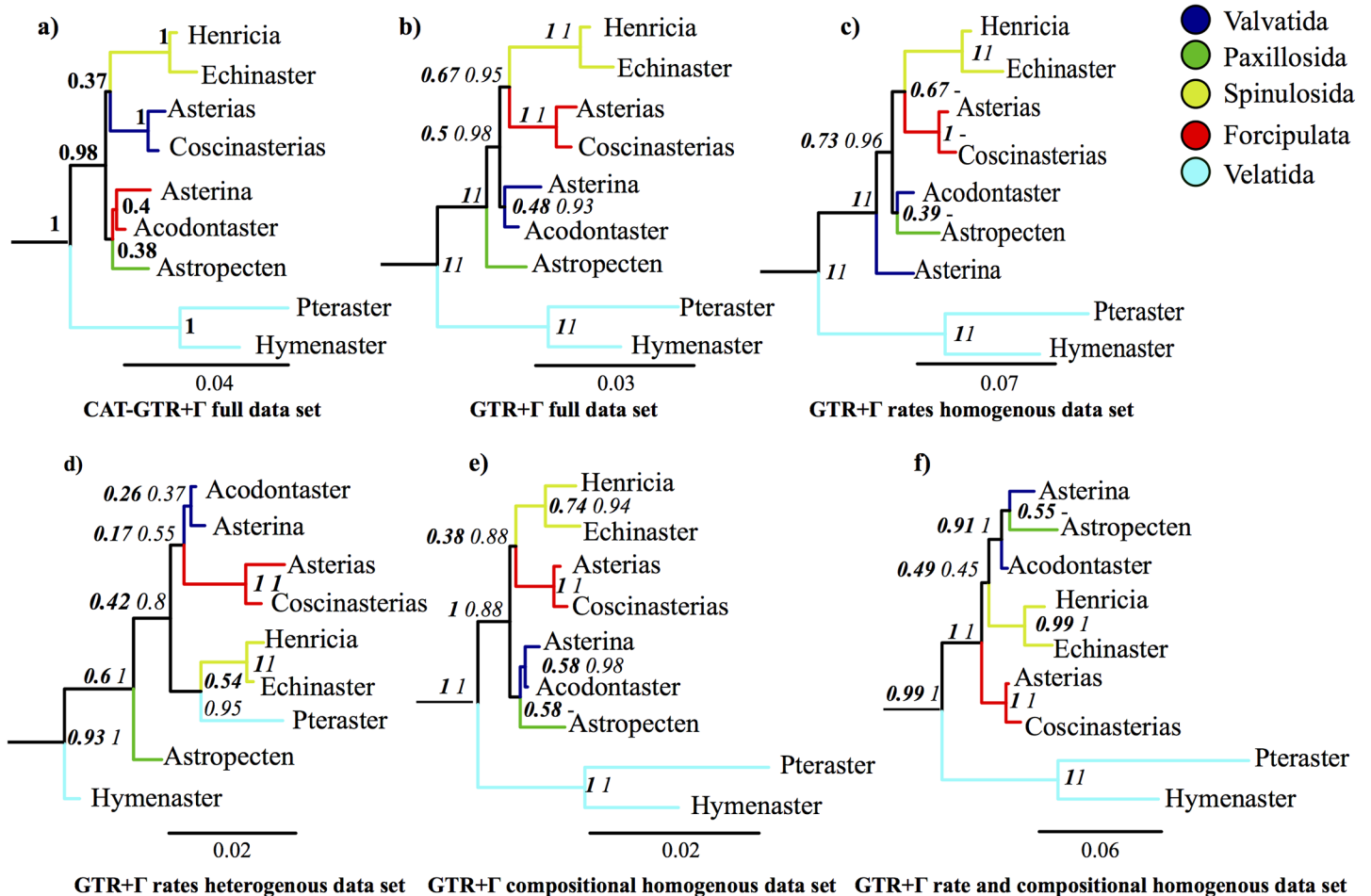
Support amongst competing hypotheses was assessed using a Bayes Factor (BF) approach [39]. The marginal likelihood for each constraint tree was estimated under GTR- $\Gamma$  using the stepping-stone procedure, as implemented in MrBayes3.2 (with 10 million generations and sampling every 2500 generations). The stepping-stone method has recently been developed and is known to outperform the traditional harmonic mean approach for estimating the marginal likelihood support [40]. We also compared the support for each of the topologies shown in Fig 1 (excluding [1], because of its incomplete coverage of asteroids orders) and Fig 2, in each data partition.

## Results

The data partitions we generated differ markedly in their composition homogeneity and saturation (Table C in S1 File) and in their ability to recover well-founded relationships within Echinoderms (Table 1). Furthermore, analysis of the same data partition under different evolutionary models (GTR vs. CAT-GTR) also generates different topologies (Fig 2). Two points emerge clearly from these analyses. First, there is strong support for Velatida at the root of Asterozoa that is rarely masked by non-phylogenetic signal. Second, the relationships among the remaining orders are partition dependent and any surviving phylogenetic signal is weak.

The support for Velatida as sister group to other Asterozoa has the highest posterior probability (PP = 1) in all the reliable partitions and under the best fitting model (Fig 2A, 2B, 2C, 2E and 2F and Figs S1–S4, S6 and S7). Only the least reliable rate heterogeneous partition does not support this topology, instead suggesting that Velatida are paraphyletic with one species monophyletic with Spinulosida (PP = 0.95) (Fig 2D and S5 Fig). The same pattern is observed in the maximum likelihood trees (Figs S8–S12).

Regarding the relationships among starfishes other than Velatida, our results are partition and model dependent (Fig 2 and Figs S1–S12). Analysis of the full data set under the GTR- $\Gamma$  model finds a poorly supported topology where Forcipulata and Spinulosida together form a monophyletic group (PP = 0.67), the Valvatida are monophyletic, but without support (PP = 0.48), and Paxillosida lie nested between the Velatida and all other starfishes (PP = 0.5). The same partition analysed under CATBP- $\Gamma$  (which takes account of compositional heterogeneity—see Table C in S1 File) suggests a polytomy for Spinulosida and Forcipulata (PP < 0.5) and the paraphyly of Valvatida, with Paxillosida nested within this group (PP = 0.7) (S3 Fig). Finally, when the best fitting CAT-GTR- $\Gamma$  model is used all nodes, with the exception of the root position of Velatida, have insignificant support (Fig 2A).



**Fig 2. Cladograms summarizing inferred asteroid relationships obtained from the different partitions and under different substitution models.** Numbers at nodes indicate Bayesian posterior probability support under CAT-GTR-  $\Gamma$  (bold), Bayesian posterior probability under GTR-  $\Gamma$  (italic bold) and maximum likelihood abayes bootstrap under GTR-  $\Gamma$  (italic). For all the trees except 2a, the branch lengths are estimated under Bayesian GTR-  $\Gamma$ .

doi:10.1371/journal.pone.0123331.g002

Focusing on the more reliable data partitions only, the main differences we observe concern the level of support for the monophyly of Forcipulata and Spinulosida (ranging from PP = 0.67 in the rate-homogeneous partition (Fig 2C and S4 Fig) to PP = 0.38 in the composition-homogeneous partition (Fig 2E and S6 Fig), and the placement of the Paxillosida, which either nests within the Valvatida (Fig 2E and S6 Fig; PP = 0.58) or lies between Velatida and the other starfishes, as in the full data set (PP = 0.5, Fig 2B and S2 Fig). However, our composition+rates homogeneous data set, under Bayesian CAT-GTR-  $\Gamma$ , GTR-  $\Gamma$  and maximum likelihood GTR-  $\Gamma$  (Fig 4, S7 Fig and S12 Fig) suggests a different topology where Valvatida, Forcipulata and Spinulosida are monophyletic, albeit without support (PP = 0.49), Paxillosida and Valvatida form a monophyletic group (PP = 0.91) and Forcipulata now nests between Velatida and all the other Asterozoa.

We also calculated the support in our seven data sets for the nine alternative phylogenetic hypotheses (Figs 1 and 2). The results (shown as a heat map in Fig 3) indicate that some of the previous published hypotheses i.e. Blake (1987), Gale (1987) Knott and Wray (2000) and Gale (2011) are not supported in any of the more reliable partitions (BF running from -21.7 in the rate homogeneous data set to -19.05 in the composition+rate homogeneous data set).

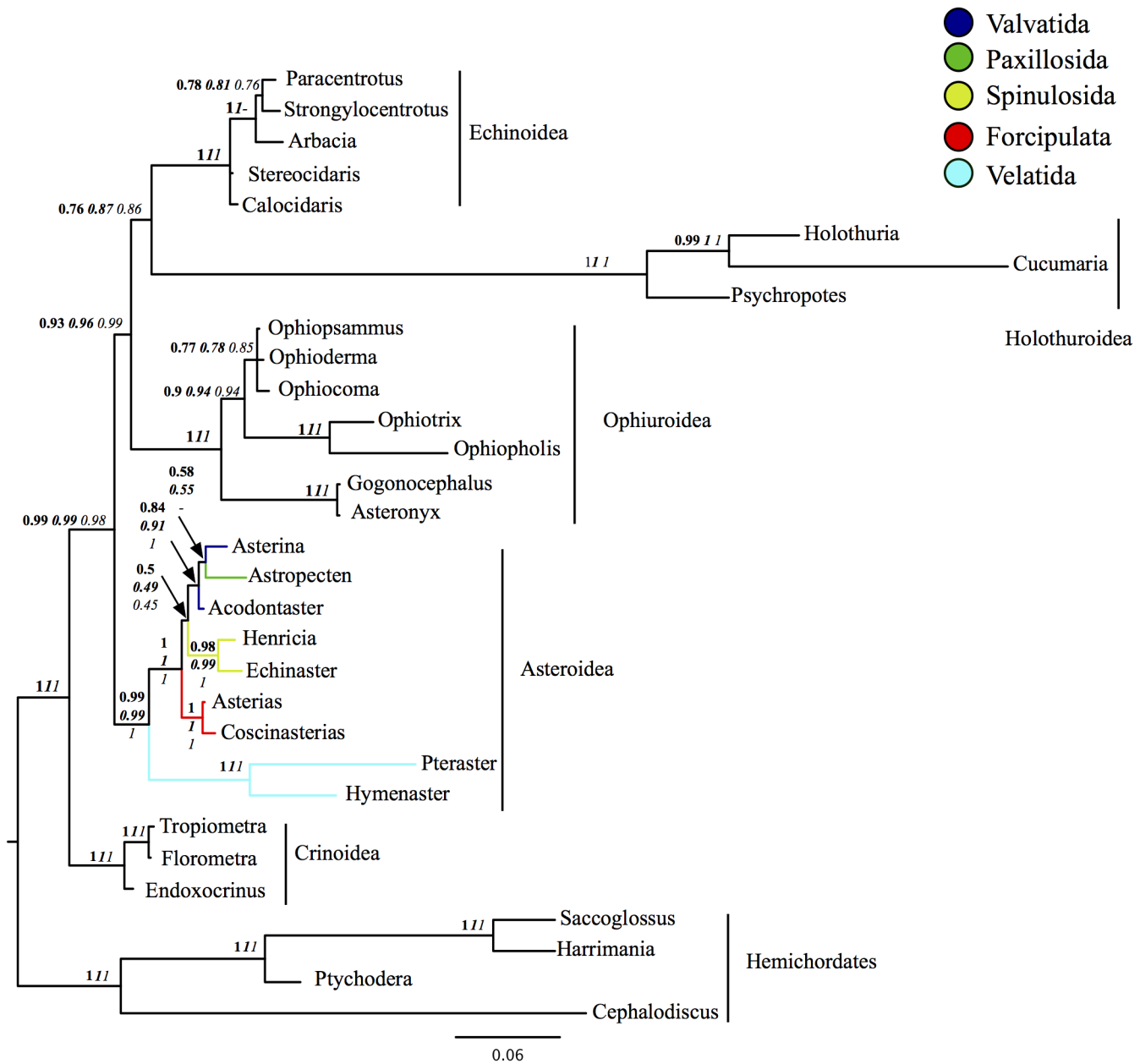
		Competing hypothesis									
		constrain tree	topology 2b	topology 2c	topology 2a and e	topology 2f	Gale 2011	Janies et al. 2012	Knott and Wray 2000	Gale 1987	Blake 1987
full data set	topology 2b		9.25	4.11	3.84	22.42	9.64	26.31	26.59	25.94	
	topology 2c	-9.25		-5.15	-5.41	13.17	0.39	17.06	17.34	16.69	
	topology 2a and 2e	-4.11	5.14		-0.27	18.31	5.53	22.2	22.48	21.83	
	topology 2f	-3.84	5.41	0.27		18.58	5.8	22.47	22.75	22.1	
	Gale 2011	-22.42	-13.17	-18.31	-18.58		-12.78	3.89	4.17	3.52	
	Janies et al. 2012	-9.64	-0.39	-5.53	-5.8	12.78		16.67	16.95	16.3	
	Knott & Wray 2000	-26.31	-17.06	-22.2	-22.47	-3.89	-16.67		0.28	-0.37	
	Blake 87	-25.94	-16.69	-21.83	-22.1	3.52	16.3	0.37	0.65		
rates hom.	topology 2b		3.81	-0.19	-0.18	25.5	4.55	27.47	29.51	26.43	
	topology 2c	-3.81		-4	-3.99	21.69	0.74	23.66	25.7	22.62	
	topology 2a and 2e	0.19	4		0.01	25.69	4.74	27.66	29.7	26.62	
	topology 2f	0.18	3.99	-0.01		25.68	4.73	27.65	29.69	26.61	
	Gale 2011	-25.5	-21.69	-25.69	-25.68		-20.95	1.97	4.01	0.93	
	Janies et al. 2012	-4.55	-0.74	-4.74	-4.73	20.95		22.92	24.96	21.88	
	Knott & Wray 2000	-27.47	-23.66	-27.66	-27.65	-1.97	-22.92		2.04	-1.04	
	Blake 87	-29.51	-25.7	-29.7	-29.69	-4.01	-24.96	-2.04		-3.08	
comp. hom.	topology 2b		4.18	2.6	7.47	3.67	-0.68	12.16	6.82	7.81	
	topology 2c	-4.18		-0.51	-1.58	-0.51	-4.86	7.98	2.64	3.63	
	topology 2a and 2e	-2.6	1.58		1.07	1.07	-3.28	9.56	4.22	5.21	
	topology 2f	-2.31	1.87	0.29		1.36	-2.99	9.85	4.51	5.5	
	Gale 2011	-3.67	0.51	-1.07	-4.35		-4.35	8.49	3.15	4.14	
	Janies et al. 2012	0.68	0.68	4.86	3.28	4.35		12.84	7.5	8.15	
	Knott & Wray 2000	-12.16	-7.98	-9.56	-12.16	-8.49	-12.84		-5.34	-4.69	
	Blake 87	-6.82	-2.64	-4.22	-4.51	-3.15	-7.5	5.34		0.99	
rate+comp. hom	topology 2b		9.01	-3.89	-7.64	7.9	-0.08	7.05	11.41	4.7	
	topology 2c	-9.01		-12.9	-16.65	-1.11	-9.09	-1.96	-4.31	-4.31	
	topology 2a and 2e	3.89	12.9		-3.75	11.79	3.81	10.94	15.3	8.59	
	topology 2f	7.64	16.65	3.75		15.54	7.56	14.69	12.34	12.34	
	Gale 2011	-7.9	1.11	-11.79	-15.54		-7.98	-0.85	3.51	-3.2	
	Janies et al. 2012	0.08	9.09	-3.81	-7.56	7.98		7.13	11.49	4.78	
	Knott & Wray 2000	-7.05	1.96	-10.94	-14.69	0.85	-7.13		4.36	-2.35	
	Blake 87	-11.41	-2.4	-15.3	-19.05	-3.51	-11.49	2.35		-6.71	
rates heter.	topology 2b		4.23	-4.57	-4.77	4.47	-3.31	0.52	-8.29	-8.05	
	topology 2c	-4.23		-8.8	-9	0.24	-7.54	-3.71	-12.52	-12.28	
	topology 2a and 2e	4.77	8.8		-0.2	9.04	1.26	5.09	-3.72	-3.48	
	topology 2f	4.77	9	0.2		9.24	1.46	5.29	3.52	-3.28	
	Gale 2011	-4.47	-0.24	-9.04	-9.24		-7.78	-3.95	12.76	-12.52	
	Janies et al. 2012	3.31	7.54	-1.26	-1.46	7.78		3.83	-4.98	-4.74	
	Knott & Wray 2000	-0.52	3.71	-5.09	-5.29	3.95	-3.83		-8.81	-8.57	
	Blake 87	8.29	12.52	3.72	3.52	12.76	4.98	8.81		0.24	

Interpretation of the Bayes factor

from 20 to 150	Data strongly favour constrained tree
from 3 to 20	Data weakly favour constrained tree
from -3 to 3	Barely worth mentioning
from -3 to -20	Data favour competing tree
from -20 to -150	Data strongly favour competing tree

**Fig 3. Bayes Factor support for the alternative topologies in the different partitions.** Supports are coded according to [45]. Rates hom. = homogeneous rates partition; comp. hom. = compositionally homogeneous partition; rates+comp. hom. = homogeneous rates and composition partition; rates heter. = heterogeneous rates partition. In this table the red color implies that competing tree is rejected and the darker the red the stronger the evidence for rejection.

doi:10.1371/journal.pone.0123331.g003



**Fig 4. Resultant tree from analysis of the homogenous rates and compositional data set under CAT-GTR- $\Gamma$ .** Support at the nodes indicates posterior probability under CAT-GTR- $\Gamma$  (bold), posterior probability GTR- $\Gamma$  (italic bold) and maximum likelihood abayes bootstrap under GTR- $\Gamma$  (italic).

doi:10.1371/journal.pone.0123331.g004

Significantly, the topologies of Gale (1987) and Blake (1987) increase their support in the less reliable rate heterogeneous partition and can thus be firmly rejected. Interestingly, the composition+rates homogenous data set is the only one consistently better supported compared to all other topologies, with BF values all positive and running from barely in favour (BF = 3.75) to values above 10.



## Discussion

### Phylogenetic signal and data partition reliability

The phylogenetic relationships of asteroids, especially the identity of the earliest branching clade, is a question that should be solvable using molecular data, but previous attempts have come to different solutions and no clear consensus has emerged. The main reason of this lack of consensus appears to be the small amount of phylogenetic signal that can be easily swamped by a combination of systematic and stochastic error.

The comparison of phylogenies obtained using differently fitting substitution models and different partitions has previously been used to explore the influence of non-phylogenetic signal on phylogenetic analyses [27, 29, 41–43]. The logic behind this approach is that unreliable partitions and poorly fitting models of evolution have a much greater chance of supporting artifactual topologies than reliable partitions and better fitting models. This is clearly demonstrated in our data by the fact that only the more reliable partitions find the expected relationships amongst the other echinoderm classes (Table 1). Following this logic, the increase of support we observe in the unreliable rate heterogeneous partition for the topologies of Blake 1987, Gale 1987, Knott and Wray 2000 and Gale 2011 shows that the strongest support for these comes from non-phylogenetic signal. By the same reasoning, all our results from the more reliable partitions analysed using the best fitting model, indicates that there is a strong phylogenetic signal supporting the Velatida as sister group to other asteroids, in keeping with Janies' [8] findings. Only the most unreliable rate heterogeneous partition fails to support this topology.

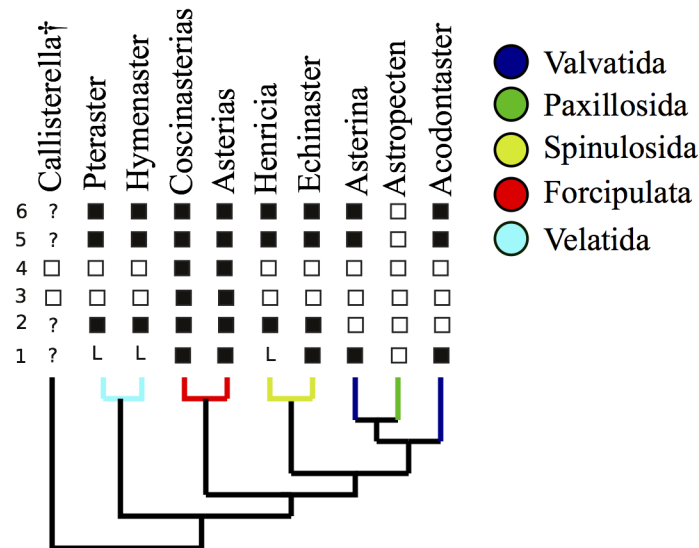
A good example of how the ratio between noise and good phylogenetic signal is an issue that needs careful consideration comes when considering the relationships of Spinulosida to Forcipulata. The tree obtained from the full dataset, applying the GTR- $\Gamma$  model, pairs Spinulosida and Forcipulata as a monophyletic group (Fig 2B). If the monophyly of this clade is genuine and not a phylogenetic artefact, we should see support for this clade increase in partitions with decreased compositional heterogeneity. Yet trees obtained under the best fitting model (CAT-GTR- $\Gamma$ ), non-homogeneous model (CATbp) and from compositional homogeneous data sets all fail to support the pairing of Spinulosida plus Forcipulata, indicating that the monophyly of this clade is most likely the result of non-phylogenetic signal.

The most reliable data sets are likely to be those minimizing rate and compositional heterogeneity and only these recover the well-established relationships in other echinoderm classes (Tables 1 and Table C in S1 File). The composition+rates homogeneous partition (Fig 2F and Fig 4) best meets all three criteria. Furthermore, our BF calculations suggest that overall the topology obtained from this data set is the best supported of all the topologies found (Fig 3). Based on these observations the tree in Figs 2F and 4 is taken as our best-supported topology.

While the reconstruction of secure and robust trees remains a major challenge for phylogeneticists where data are far from ideal, a more considered approach to rate and compositional heterogeneity in gene sequence data combined with the testing of competing hypotheses within a Bayesian framework, provide the best way forward, as demonstrated here.

### Morphological implications

Our analysis of molecular data allows us to confidently reject topologies that place either Paxillosida or Forcipulata as sister group to other asteroids, as favoured by morphological studies [2–4, 12]. This implies a radical rethink of the body plan evolution of asteroids. Traditionally the Paxillosida had been seen as the most primitive of extant starfish because their larvae do not develop to the brachiolaria stage, they lack suckered tube-feet, they have a simple digestive



**Fig 5. Key morphological characters relevant to asteroid rooting mapped onto our best-supported molecular phylogenetic tree.** 1, planktotrophic larval development; □ to bipinnaria stage; ■ to brachiolaria stage; L—lecithotrophic (no planktotrophic larval stage). 2, suckered tube feet; □ absent; ■ present. 3, pedicellariae; □ simple valves; ■ complex, crossed pedicellariae with basal element. 4, oral frame; □ ambulacral; ■ adambulacral. 5, eversible stomach; □ absent; ■ present. 6, anus; □ absent; ■ present.

doi:10.1371/journal.pone.0123331.g005

system often lacking an anus and with a stomach that cannot be everted, and they have a relatively simple organization of their ambulacral-adambulacral skeleton [2–3]. Blake [4–5, 12] however, has argued that the paxillosids are highly specialised starfishes that have secondarily lost a number of characters, and our analysis supports the latter view (Fig 5). The presence of brachiolarian larvae (larvae with anterior arms and a sucker for attachment) in velatids, valvatids and forcipulatids must be seen as a shared plesiomorphy. As the brachiolarian larva is preceded developmentally by a bipinnaria larval stage [14, 44], and a bipinnaria stage is found in paxillosids, our tree implies that paxillosids must have truncated their development.

Suckered tube feet are fully developed only in Spinulosida and Forcipulata so their absence in Velatida, Paxillosida and many Valvatida is a shared plesiomorphy. Pedicellariae are present in some paxillosids, valvatids and forcipulatids but are often wanting; they are absent in spinulosids and velatids. Forcipulata possess complex, derived pedicellariae that are pedunculate [3] which contrasts with the more simple pedicellariae in paxillosids and valvatids. The fact that simple pedicellariae also occur in stem group asteroids [3] indicates that velatids have secondarily lost pedicellariae.

Gale [3] lists a number of features of the ambulacral-adambulacral skeleton in paxillosids that he considers primitive. Blake [4] also noted that paxillosids and many velatids had a primitive pattern in which adambulacrals are strongly overlapping with large muscles linking successive ambulacrals and with ambulacrals with long heads. However, in Gale’s [3] cladogram there is only a single characters relating to this complex that places paxillosids as primitive and that is the symmetry of the processes on ambulacral ossicles for adambulacral attachment. This is best explained as a retained plesiomorphy that other asteroids have lost through evolution. Finally, the oral frame of paxillosids and velatids has a similar organization, both having prominent narrow elongate oral plates, which must be plesiomorphic for crown group asteroids.

Sadly, from a morphologist’s viewpoint, optimizing the characters in Gale [3] onto our best-supported molecular tree identifies no reliable morphological traits that can be considered as unambiguously supporting a Valvatida-Paxillosida-Forcipulata-Spinulosida clade. Of the 15

characters separating the two clades all are either homoplasious or autapomorphies of Pterasteridae. For the present our best evidence for higher asteroid relationships must come from molecular data.

## Supporting Information

### S1 Dataset.

(PHY)

### S1 Fig. CAT-GTR+ $\Gamma$ tree of the full data set.

(TIFF)

### S2 Fig. GTR+ $\Gamma$ tree of the full data set.

(TIFF)

### S3 Fig. CAT-BP+ $\Gamma$ tree of the full dataset.

(TIFF)

### S4 Fig. GTR+ $\Gamma$ tree of the homogenous rates data set.

(TIFF)

### S5 Fig. GTR+ $\Gamma$ tree of the heterogeneous rates data set.

(TIFF)

### S6 Fig. GTR+ $\Gamma$ tree of the homogenous compositional data set.

(TIFF)

### S7 Fig. GTR+ $\Gamma$ tree of the homogenous rate+compositional data set.

(TIFF)

### S8 Fig. Maximum likelihood GTR+ $\Gamma$ tree of the full data set.

(TIFF)

### S9 Fig. Maximum likelihood GTR+ $\Gamma$ tree of the homogenous rates data set.

(TIFF)

### S10 Fig. Maximum likelihood GTR+ $\Gamma$ tree of the heterogeneous rates data set.

(TIFF)

### S11 Fig. Maximum likelihood GTR+ $\Gamma$ tree of homogenous compositional data set.

(TIFF)

### S12 Fig. Maximum likelihood GTR+ $\Gamma$ tree homogenous rate+compositional data set.

(TIFF)

**S1 File. Supporting tables.** Table A, List of taxa and gene sequences used in this study. Consensus genus sequences were constructed from these. Table B, Result of the 12-fold Bayesian cross-validation. Positive scores indicate that the model compared is better than the reference model (CATGTR). Table C, Statistical comparison of the compositional heterogeneity in the various partitions. Z-scores quantify the amount of compositional heterogeneity (the greater the z-score the higher the compositional heterogeneity), p-values indicate the statistical significance of test with \* indicating significant values. Compositional heterogeneity clearly is not an issue in the compositional+rates homogenous data set based on these tests. Table D, Results of the posterior predictive estimation of the homoplasy in the nuclear and nuclear+mitochondrial genes data sets. Although the p-values are significant for both data sets, the observed homoplasy is higher in the nuclear+mitochondrial data set.

(DOCX)

## Acknowledgments

We thank Davide Pisani for discussion and suggestions throughout the progress of this research. All analyses were performed using the infrastructures provided by the University of Bristol.

## Author Contributions

Conceived and designed the experiments: RF ABS. Performed the experiments: RF. Analyzed the data: RF ABS. Contributed reagents/materials/analysis tools: RF ABS. Wrote the paper: RF ABS.

## References

1. Mah CL, Blake DB. Global diversity and phylogeny of the Asterozoa (Echinodermata). *PLoS one*. 2012; 7(4):e35644. doi: [10.1371/journal.pone.0035644](https://doi.org/10.1371/journal.pone.0035644) PMID: [22563389](https://pubmed.ncbi.nlm.nih.gov/22563389/)
2. Gale AS. Phylogeny and classification of the Asterozoa (Echinodermata). *Zoological Journal of the Linnean Society*. 1987; 89(2):107–132.
3. Gale AS. The phylogeny of post-Palaeozoic Asterozoa (Neosterozoa, Echinodermata). *Spec Pap Palaeontol*. 2011; 85.
4. Blake DB. A classification and phylogeny of post-Palaeozoic sea stars (Asterozoa: Echinodermata). *Journal of Natural History*. 1987; 21(2):481–528.
5. Blake DB, editor Paxillosidans are not primitive asteroids: a hypothesis based on functional considerations. *Echinoderm Biology: Proceedings of the Sixth International Echinoderms Conference*; 1988; Victoria, Australia. Rotterdam: A. A. Balkema.
6. Littlewood D, Smith A, Clough K, Emson R. The interrelationships of the echinoderm classes: morphological and molecular evidence. *Biological Journal of the Linnean Society*. 1997; 61(3):409–438.
7. Janies D. Phylogenetic relationships of extant echinoderm classes. *Canadian Journal of Zoology*. 2001; 79(7):1232–1250.
8. Janies DA, Voight JR, Daly M. Echinoderm phylogeny including *Xyloplax*, a progenetic asteroid. *Systematic biology*. 2011; 60(4):420–438. doi: [10.1093/sysbio/syr044](https://doi.org/10.1093/sysbio/syr044) PMID: [21525529](https://pubmed.ncbi.nlm.nih.gov/21525529/)
9. Perseke M, Bernhard D, Fritsch G, Brummer F, Stadler PF, Schlegel M. Mitochondrial genome evolution in Ophiurozoa, Echinozoa, and Holothurozoa: insights in phylogenetic relationships of Echinodermata. *Molecular phylogenetics and evolution*. 2010; 56(1):201–211. doi: [10.1016/j.ympev.2010.01.035](https://doi.org/10.1016/j.ympev.2010.01.035) PMID: [20152912](https://pubmed.ncbi.nlm.nih.gov/20152912/)
10. Mortensen T. Studies of the development and larval forms of echinoderms. Reitzel CA, editor. Copenhagen 1921.
11. MacBride EW. Echinoderm larvae and their bearing on evolution. *Nature*. 1921; 108:529–530.
12. Blake DB. Asterozoa: functional morphology, classification and phylogeny. Jangoux ML, J. M., editor. Rotterdam: A. A. Balkema; 1989.
13. Heddle D, Emson R, Smith A, Campbell A. The descent of the Asterozoa and the reaffirmation of paxillosidan primitiveness. *Echinoderm research*. 1995:179–183.
14. Wada H, Komatsu M, Satoh N. Mitochondrial rDNA phylogeny of the Asterozoa suggests the primitiveness of the Paxillosida. *Molecular phylogenetics and evolution*. 1996; 6(1):97–106. PMID: [8812310](https://pubmed.ncbi.nlm.nih.gov/8812310/)
15. Lafay B, Smith A, Christen R. A combined morphological and molecular approach to the phylogeny of asteroids (Asterozoa: Echinodermata). *Systematic Biology*. 1995; 44(2):190–208.
16. Matsubara M, Komatsu M, Araki T, Asakawa S, Yokobori S, Watanabe K, et al. The phylogenetic status of Paxillosida (Asterozoa) based on complete mitochondrial DNA sequences. *Molecular phylogenetics and evolution*. 2005; 36(3):598–605. PMID: [15878829](https://pubmed.ncbi.nlm.nih.gov/15878829/)
17. Knott KE, Wray GA. Controversy and consensus in asteroid systematics: new insights to ordinal and familial relationships. *American Zoologist*. 2000; 40(3):382–392.
18. Mah C, Foltz D. Molecular phylogeny of the Forcipulatacea (Asterozoa: Echinodermata): systematics and biogeography. *Zoological Journal of the Linnean Society*. 2011; 162(3):646–660.
19. Mah C, Foltz D. Molecular phylogeny of the Valvatatacea (Asterozoa: Echinodermata). *Zoological Journal of the Linnean Society*. 2011; 161(4):769–788.
20. Felsenstein J. Cases in Which Parsimony or Compatibility Methods Will Be Positively Misleading. *Syst Zool*. 1978; 27(4):401–410.

21. Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Worheide G, et al. Resolving difficult phylogenetic questions: why more sequences are not enough. *Plos Biology*. 2011; 9(3):e1000602. doi: [10.1371/journal.pbio.1000602](https://doi.org/10.1371/journal.pbio.1000602) PMID: [21423652](https://pubmed.ncbi.nlm.nih.gov/21423652/)
22. Roure B, Baurain D, Philippe H. Impact of missing data on phylogenies inferred from empirical phylogenomic datasets. *Molecular biology and evolution*. 2012.
23. Rota-Stabelli O, Lartillot N, Philippe H, Pisani D. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. *Systematic biology*. 2013; 62(1):121–133. doi: [10.1093/sysbio/sys077](https://doi.org/10.1093/sysbio/sys077) PMID: [22962005](https://pubmed.ncbi.nlm.nih.gov/22962005/)
24. Foster PG, Hickey DA. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of molecular evolution*. 1999; 48(3):284–290. PMID: [10093217](https://pubmed.ncbi.nlm.nih.gov/10093217/)
25. Lartillot N, Lepage T, Blanquart S. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. 2009; 25(17):2286–2288. doi: [10.1093/bioinformatics/btp368](https://doi.org/10.1093/bioinformatics/btp368) PMID: [19535536](https://pubmed.ncbi.nlm.nih.gov/19535536/)
26. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic biology*. 2011; 60(5):685–699. doi: [10.1093/sysbio/syr041](https://doi.org/10.1093/sysbio/syr041) PMID: [21540409](https://pubmed.ncbi.nlm.nih.gov/21540409/)
27. Pisani D. Identifying and Removing Fast-Evolving Sites Using Compatibility Analysis: An Example from the Arthropoda. *Systematic Biology*. 2004; 53(6):978–989. PMID: [15764565](https://pubmed.ncbi.nlm.nih.gov/15764565/)
28. Cummins CA, McInerney JO. A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic inference, resolve deep divergence and correct systematic biases. *Systematic biology*. 2011; 60(6):833–844. doi: [10.1093/sysbio/syr064](https://doi.org/10.1093/sysbio/syr064) PMID: [21804093](https://pubmed.ncbi.nlm.nih.gov/21804093/)
29. Pisani D, Feuda R, Peterson KJ, Smith AB. Resolving phylogenetic signal from noise when divergence is rapid: a new look at the old problem of echinoderm class relationships. *Molecular phylogenetics and evolution*. 2012; 62(1):27–34. doi: [10.1016/j.ympev.2011.08.028](https://doi.org/10.1016/j.ympev.2011.08.028) PMID: [21945533](https://pubmed.ncbi.nlm.nih.gov/21945533/)
30. Blanquart S, Lartillot N. A site- and time-heterogeneous model of amino acid replacement. *Molecular biology and evolution*. 2008; 25(5):842–858. doi: [10.1093/molbev/msn018](https://doi.org/10.1093/molbev/msn018) PMID: [18234708](https://pubmed.ncbi.nlm.nih.gov/18234708/)
31. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC evolutionary biology*. 2010; 10:210. doi: [10.1186/1471-2148-10-210](https://doi.org/10.1186/1471-2148-10-210) PMID: [20626897](https://pubmed.ncbi.nlm.nih.gov/20626897/)
32. Quang LS, Gascuel O, Lartillot N. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics*. 2008; 24(20):2317–2323. doi: [10.1093/bioinformatics/btn445](https://doi.org/10.1093/bioinformatics/btn445) PMID: [18718941](https://pubmed.ncbi.nlm.nih.gov/18718941/)
33. Smith AB, Pisani D, Mackenzie-Dodds JA, Stockley B, Webster BL, Littlewood DT. Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinozoa (Echinodermata). *Molecular biology and evolution*. 2006; 23(10):1832–1851. PMID: [16777927](https://pubmed.ncbi.nlm.nih.gov/16777927/)
34. Kroh A, Smith AB. The phylogeny and classification of post-Palaeozoic echinoids. *Journal of Systematic Palaeontology*. 2010; 8(2):147–212.
35. Okanishi M, O'Hara TD, Fujita T. Molecular phylogeny of the order Euryalida (Echinodermata: Ophiuroidea), based on mitochondrial and nuclear ribosomal genes. *Molecular phylogenetics and evolution*. 2011; 61(2):392–399. doi: [10.1016/j.ympev.2011.07.003](https://doi.org/10.1016/j.ympev.2011.07.003) PMID: [21798356](https://pubmed.ncbi.nlm.nih.gov/21798356/)
36. Kerr AM, Kim J. Phylogeny of Holothuroidea (Echinodermata) inferred from morphology. *Zoological Journal of the Linnean Society*. 2001; 133(1):63–81.
37. Smith AB, Reich M. Tracing the evolution of the holothurian body plan through stem-group fossils. *Biological Journal of the Linnean Society*. 2013; 109(3):670–681.
38. Telford MJ, Lowe CJ, Cameron CB, Ortega-Martinez O, Aronowicz J, Oliveri P, et al. Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proc Biol Sci*. 2014; 281(1786).
39. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Systematic biology*. 2011; 60(2):150–160. doi: [10.1093/sysbio/syq085](https://doi.org/10.1093/sysbio/syq085) PMID: [21187451](https://pubmed.ncbi.nlm.nih.gov/21187451/)
40. Bergsten J, Nilsson AN, Ronquist F. Bayesian tests of topology hypotheses with an example from diving beetles. *Systematic biology*. 2013; 62(5):660–673. doi: [10.1093/sysbio/syt029](https://doi.org/10.1093/sysbio/syt029) PMID: [23628960](https://pubmed.ncbi.nlm.nih.gov/23628960/)
41. Sperling EA, Peterson KJ, Pisani D. Phylogenetic-Signal Dissection of Nuclear Housekeeping Genes Supports the Paraphyly of Sponges and the Monophyly of Eumetazoa. *Molecular Biology and Evolution*. 2009; 26(10):2261–2274. doi: [10.1093/molbev/msp148](https://doi.org/10.1093/molbev/msp148) PMID: [19597161](https://pubmed.ncbi.nlm.nih.gov/19597161/)
42. Rota-Stabelli O, Campbell L, Brinkmann H, Edgecombe GD, Longhorn SJ, Peterson KJ, et al. A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. *Proc Biol Sci*. 2011; 278(1703):298–306. doi: [10.1098/rspb.2010.0590](https://doi.org/10.1098/rspb.2010.0590) PMID: [20702459](https://pubmed.ncbi.nlm.nih.gov/20702459/)

43. Sperling EA, Pisani D, Peterson KJ. Molecular paleobiological insights into the origin of the Brachiopoda. *Evolution & development*. 2011; 13(3):290–303.
44. McEDWARD LR, Janies DA. Life cycle evolution in asteroids: what is a larva? *The Biological bulletin*. 1993; 184(3):255–268.
45. Kass RE, Raftery AE. Bayes factors. *Journal of the american statistical association*. 1995; 90(430):773–795.