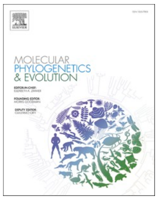
Molecular Phylogenetics and Evolution 156 (2021) 107038

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



Phylogenomic analyses recover a clade of large-bodied decapodiform cephalopods

Frank E. Anderson[a,*], Annie R. Lindgren[b]

[a] School of Biological Sciences, Southern Illinois University, Carbondale, IL 62901, USA
[b] The Center for Life in Extreme Environments, Department of Biology, Portland State University, 1719 SW 10th Ave, SRTC Rm 246, Portland, OR 97201, USA

ARTICLE INFO

Keywords:
Decapodiformes
Phylogenomics
Squid
Cuttlefish
Molecular phylogenetics

ABSTRACT

Phylogenetic relationships among the squids and cuttlefishes (Cephalopoda:Decapodiformes) have resisted clarification for decades, despite multiple analyses of morphological, molecular and combined data sets. More recently, analyses of complete mitochondrial genomes and hundreds of nuclear loci have yielded similarly ambiguous results. In this study, we re-evaluate hypotheses of decapodiform relationships by increasing taxonomic breadth and utilizing higher-quality genome and transcriptome data for several taxa. We also employ analytical approaches to (1) identify contamination in transcriptome data, (2) better assess model adequacy, and (3) account for potential biases. Using this larger data set, we consistently recover a clade comprising Myopsida (closed-eye squid), Sepiida (cuttlefishes), and Oegopsida (open-eye squid) that is sister to a Sepiolida (bobtail and bottletail squid) clade. Idiosepiida (pygmy squid) is consistently recovered as the sister group to all sampled decapodiform lineages. Further, a weighted Shimodaira-Hasegawa test applied to one of our larger data matrices rejects all alternatives to these ordinal-level relationships. At present, available nuclear genome-scale data support nested clades of relatively large-bodied decapodiform cephalopods to the exclusion of pygmy squids, but improved taxon sampling and additional genomic data will be needed to test these novel hypotheses rigorously.

1. Introduction

The cephalopod clade Decapodiformes (squids and cuttlefishes) comprises some of the most charismatic, economically and scientifically important invertebrate species. The largest non-colonial invertebrates on the planet are members of Decapodiformes—the giant squid *Architeuthis dux* and the colossal squid *Mesonychoteuthis hamiltoni* (both ~3 m mantle length and weighing ~500 kg) (Jereb and Roper, 2010). Several decapodiform species, particularly loliginids, sepiids and ommastrephids, are targets of major fisheries; the global squid catch totaled ~3 million metric tons in 2010 alone (Arkhipkin et al., 2015), with cephalopod populations appearing to be expanding (Doubleday et al., 2016), potentially fueling even larger catches in the future.

Decapodiformes includes seven major extant clades (typically ranked as orders)—Bathyteuthoidea (*Bathyteuthis* and the comb-finned squid *Chtenopteryx*), Idiosepiida (pygmy squids), Myopsida (the neritic “closed-eye” squids, primarily loliginids), Oegopsida (the mostly oceanic “open-eye” squids), Sepiida (cuttlefishes), Sepiolida (bobtail and bottletail squids) and Spirulida (the ram’s horn squid). Though monophyly of these groups is fairly well established, relationships among them have been extraordinarily difficult to clarify using morphological and PCR-based molecular phylogenetic analyses (Allcock et al., 2014, 2011; Lindgren et al., 2012; Lindgren, 2010; Strugnell and Nishiguchi, 2007; Young and Vecchione, 1996).

High-throughput sequencing technologies have more recently allowed researchers to assemble data from complete mitochondrial genomes (Strugnell et al., 2017; Uribe and Zardoya, 2017) and nuclear data sets (primarily transcriptomes) comprising hundreds to thousands of loci (Lindgren and Anderson, 2018; Tanner et al., 2017). Despite the significant increase in data set size, these studies followed a pattern similar to that seen in morphological and PCR-based studies, namely the recovery of conflicting topologies of relationships within Decapodiformes, often with seemingly strong (e.g., bootstrap values >90%) support. Taxon sampling likely plays an important role in uncovering higher-level decapodiform relationships. For example, Lindgren and Anderson (2018) did not include data for Bathyteuthoidea or Spirulida, and recovered topologies with Idiosepiida sister to a clade comprising the rest of Decapodiformes (Fig. 1). By contrast, three other studies, using either RNAseq or mitochondrial genomes, that included either limited data from Spirulida (Tanner et al., 2017) or data from both

---

[*] Corresponding author.
E-mail address: feander@siu.edu (F.E. Anderson).

https://doi.org/10.1016/j.ympev.2020.107038
Received 15 May 2020; Received in revised form 30 October 2020; Accepted 1 December 2020

Bathyteuthoidea and Spirulida (Strugnell et al., 2017; Uribe and Zardoya, 2017) recovered Sepiida as sister to a clade comprising all extant decapodiforms (Fig. 1).

Thanks in large part to the plummeting costs of high-throughput sequencing, new data are regularly becoming publicly available. In addition, new analytical approaches designed to ameliorate challenges in phylogenomic analysis (e.g., among-lineage compositional and substitution rate heterogeneity, model inadequacy/misspecification and heterotachy) are continually being developed (Crotty et al., 2020; Duchêne et al., 2018; Naser-Khdour et al., 2019; Schrempf et al., 2020). In light of continued conflicting hypotheses of higher-level relationships within Decapodiformes, we here build on previous studies (e.g., Lindgren and Anderson, 2018; Tanner et al., 2017) by utilizing more complete transcriptomes for several previously sampled taxa and by incorporating data for new taxa (including two bottletail squids and representatives of two additional oegopsid families). Further, we employ novel, recently developed analytical approaches to evaluate and improve substitution model fit and reduce the potential impact of heterotachy. Combined, these approaches allow us to better evaluate the ability of currently available transcriptome data to distinguish between two alternative hypotheses of relationships among the decapodiform subclades—Sepiida as sister to the rest of Decapodiformes *versus* Idiosepiida as sister to the rest of Decapodiformes.

2. Methods

2.1. Taxon sampling

To better evaluate relationships among the decapodiform orders, we focused our sampling on Decapodiformes, using members of Octopodiformes as outgroups. Overall taxon sampling was similar to Lindgren and Anderson (2018), with some exceptions: 1) data from *Nautilus* were not included, 2) only four members of Octopodiformes were used (*Vampyroteuthis infernalis*, *Cirroctopus glacialis*, *Octopus bimaculoides* and *Pareledone albimaculata*), and 3) nine additional decapodiform taxa were sampled from five lineages—Oegopsida: *Pterygioteuthis hoylei* (Pyroteuthidae) and *Octopoteuthis deletron* (Octopoteuthidae), Myopsida: *Loligo vulgaris* (Loliginidae), Sepiida: *Metasepia pfefferi*, *Sepia pharaonis* and *Sepiella maindroni*, Sepiolida: *Euprymna tasmanica* and Sepiadariidae (a sepiolid lineage that was unsampled in Lindgren and Anderson [2018]): *Sepioloidea lineolata* and *Sepiadarium austrinum*. Spirulida was excluded because very little nuclear sequence data is available for this taxon, as discussed in Lindgren and Anderson (2018). Most of the data used in this study were downloaded from the Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra/). In total, 27 decapodiforms were sampled and, for several taxa, data used in Lindgren and Anderson (2018) were replaced with new publicly available data to increase gene coverage (Table 1).

2.2. Transcriptome assembly

Transcriptome assemblies were generated from downloaded fastq/fasta files or raw 454 reads in Trinity v. 2.8.3 (22 August 2018 release) (Grabherr et al., 2011) using default parameters and quality trimming ("–trimmomatic" option), with *in silico* normalization for particularly large transcriptomes ("–normalize reads" option). Assemblies were produced on the Southern Illinois University BigDawg high-performance computing cluster or via the National Center for Genome Assembly Support (NCGAS) (Ganote et al., 2017) using the Carbonate cluster at Indiana University. TransDecoder 5.3.0 (http://transdecoder.github.io) was used to find open reading frames and translate nucleotide sequences into amino acid sequences at least 100 amino acids in length.

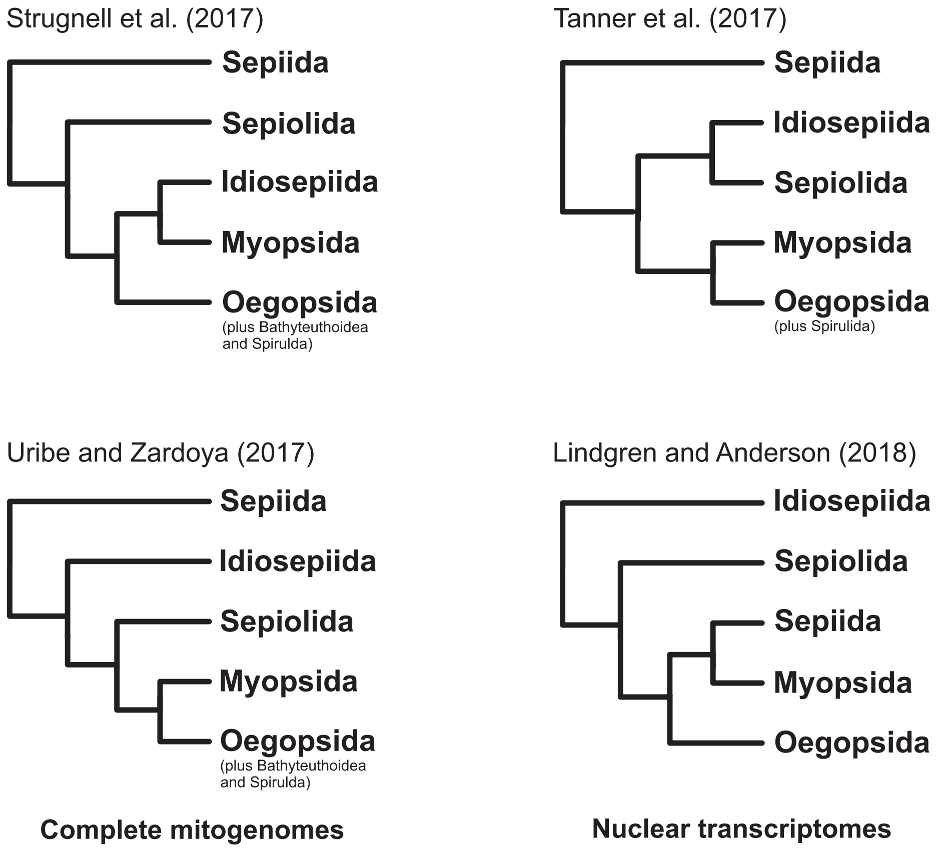


Fig. 1. Summary of recently published hypotheses of relationships among the major decapodiform lineages based on complete mitochondrial genome (left) and nuclear transcriptome sequence data (right).

Table 1
Taxon name, SRA run number and number of HaMStR orthologs recovered using the cephalopod or new lophotrochozoan core ortholog set; "–" = data provided by R. de Fonseca as described in Lindgren and Anderson (2018), * = prefix is DRR instead of SRR.

Species	Higher Taxon	NCBI SRA (SRR) number	# loci (Ceph)	# loci (Loph)
Decapodiformes				
Architeuthis dux	Oegopsida: Architeuthidae	–	2062	2138
Chiroteuthis calyx	Oegopsida: Chiroteuthidae	5527417	1813	1881
Doryteuthis pealei	Myopsida	1522987	1971	2024
Dosidicus gigas	Oegopsida: Ommastrephidae	1955488	1693	1790
Euprymna scolopes	Sepiolida: Sepiolidae	3495043–3495048	2108	2155
Euprymna tasmanica	Sepiolida: Sepiolidae	2985349	1819	1818
Galiteuthis armata	Oegopsida: Cranchiidae	2102359	569	656
Heteroloigo bleekeri	Myopsida	018274*, 018275*	1832	2011
Idiosepius notoides	Idiosepiida	2984342	1947	1991
Loligo vulgaris	Myopsida	3472303	1392	1473
Lolliguncula brevis	Myopsida	–	1625	1714
Metasepia pfefferi	Sepiida	5253657	1465	1539
Octopoteuthis deletron	Oegopsida: Octopoteuthidae	5527415	2098	2141
Onychoteuthis banksii	Oegopsida: Onychoteuthidae	–	1954	2110
Pterygioteuthis hoylei	Oegopsida: Pyroteuthidae	5527418	2030	2044
Sepiadarium austrinum	Sepiolida: Sepiadariidae	2962674, 2966910, 2966928, 2973270, 2973271	2098	2176
Sepia esculenta	Sepiida	1386223	1694	1728
Sepia officinalis	Sepiida	5204445	1999	2056
Sepia pharaonis	Sepiida	3011300	2077	2106
Sepiella japonica	Sepiida	2889753	2094	2150
Sepiella maindroni	Sepiida	5358819	2097	2133
Sepioloidea lineolata	Sepiolida: Sepiadariidae	5396788–5396792	2075	2093
Sepioteuthis lessoniana	Myopsida	1386192	1521	1528
Sthenoteuthis oualaniensis	Oegopsida: Ommastrephidae	7165537	2088	2139
Todarodes pacificus	Oegopsida: Ommastrephidae	3472305	1556	1597
Uroteuthis edulis	Myopsida	3537536–3537540, 3497883	2003	2125
Watasenia scintillans	Oegopsida	2960126–2960131	2129	2157
Octopodiformes				
Cirroctopus glacialis	Cirrata	–	1929	2040
Octopus bimaculoides	Incirrata: Octopodidae	predicted proteome	2204	2174
Pareledone albimaculata	Incirrata: Megaleledonidae	–	1934	2048
Vampyroteuthis infernalis	Vampyromorpha	5527416	2052	2083

2.3. Data matrix construction and preliminary data filtering

Initial transcriptome processing closely followed Lindgren and Anderson (2018). Amino acid sequences from TransDecoder and the *O. bimaculoides* and *A. dux* predicted proteomes were screened with HaMStR v.13.2.6 (Ebersberger et al., 2009), retaining all sequences that met the reciprocity requirement. We used two core ortholog sets—a cephalopod set (Lindgren and Anderson, 2018) and a revised lophotrochozoan set (Kocot et al., 2017). We converted the HaMStR output to a FASTA-formatted file for each orthogroup (locus) and aligned each locus using MAFFT (L-INS-i algorithm) (Katoh et al., 2005). We trimmed the ends of each sequence if any of the leading or trailing 20 positions were missing data. After alignment, we removed all gap-only columns and columns with four or fewer non-gap characters. To remove possible paralogs, we used FastTreeMP (Price et al., 2010) ("-slow" and "-gamma" settings) to infer a tree for each alignment, and then employed PhyloTreePruner (Kocot et al., 2013) to screen the resulting trees, as described in Lindgren and Anderson (2018) (though here, we collapsed nodes on each tree with SH-like local support values <0.7 into polytomies). All loci were then realigned with MAFFT (L-INS-i) and concatenated with FASconCAT v1.0.pl (Kück and Meusemann, 2010). This pipeline was run multiple times to generate concatenated matrices based in different core ortholog sets (cephalopod vs. lophotrochozoan) and different amounts of missing data (loci found in at least 25%, 50% and 75% of transcriptomes).

Investigations of single-locus trees revealed that the sepiadariid transcriptomes (especially *Sepioloidea lineolata*) contain sequences that appear to be of octopod origin. For example, of the 1065 loci in the 25% set (lophotrochozoan core orthologs) that included data for *Sepioloidea lineolata*, over a third (359) yielded ML gene trees in which *S. lineolata* was nested within Octopodiformes, usually as sister to *Octopus bimaculoides*. Preliminary phylogenetic analyses of concatenated data matrices strongly suggested that this conflicting signal was having important topological effects for some data matrices (most notably, by producing a paraphyletic Sepiolida). We used two approaches to ameliorate these effects. First, we used topological constraints in PAUP* version 4.0a (build 166) (Swofford, 2003) to find all loci whose ML trees contained a non-monophyletic Octopodiformes (i.e., a non-monophyletic Decapodiformes, as all taxa were either members of Decapodiformes or Octopodiformes) and deleted those loci. Second, for each locus that yielded an ML tree with a sepiadariid sequence nested within Octopodiformes, we simply deleted the misplaced sepiadariid sequence. Both of these "cleaned" sets of loci were then processed through the pipeline described above.

2.4. Additional filtering steps

Our pipeline included steps intended to minimize the impact of paralogy and missing data on our inferences and yielded eight sets of loci and concatenated matrices. To explore the impact that common challenges to systematics such as substitution-rate and compositional heterogeneity (Felsenstein, 1978; Foster and Hickey, 1999; Hendy and Penny, 1989; Saccone et al., 1990), substitution model violations (Ababneh et al., 2006; Shepherd and Klaere, 2018) and inadequate models (Brown, 2014; Doyle et al., 2015; Goldman, 1993) have on phylogenomic inference, we performed several additional filtering steps. We inferred best-fitting amino acid substitution models for each locus using the ProteinModelSelection.sh script (https://github.com/stamatak/standard-RAxML/blob/master/usefulScripts/ProteinModelSelection.pl), then inferred ML trees for each locus under the best-fitting model using RAxML 8.2.12 (Stamatakis, 2014) with 100 rapid bootstrap replicates. We then calculated multiple measures of phylogenetic signal and substitution rate and amino acid compositional heterogeneity for each locus: 1) the standard deviation of the tip-to-root distance, 2) the average patristic distance, 3) the average pairwise patristic distance of a taxon to all other taxa relative to the average pairwise patristic distance across all taxa, 4) the median RCFV (relative compositional frequency variability) value (Zhong et al., 2011), 5) a chi-square test of homogeneity, and 6) the relative tree certainty score, including all conflicting bipartitions (TCA).

The first three are indices of substitution rate (i.e., branch length) heterogeneity and were calculated using TreSpEx.v1.1 (Struck, 2014), while the fourth and fifth are measures of compositional heterogeneity and were calculated with BaCoCa v. 1.104r (Kück and Struck, 2014). For one set of analyses, we deleted any loci that were 1.5 times the interquartile range above the mean/median for the first four values, as well as any loci that returned a *p* value of less than 0.05 in the chi-square test of homogeneity. Combined, these filtering steps removed loci that showed either high levels of substitutional rate or compositional heterogeneity among lineages (or both).

The final measure (TCA) is an index of phylogenetic signal. Salichos and Rokas (2013) proposed that focusing on genes with strong phylogenetic signal could be helpful for reconstructing ancient divergences, and developed two measures—internode certainty (a measure of the amount of conflict between a given internode on a gene tree and the most prevalent conflicting bipartition in the set of gene trees) and tree certainty (the sum of internode certainty values across a gene tree) (Salichos et al., 2014; Shen et al., 2017)—to assess this signal. We used RAxML v. 8.2.12 to calculate tree certainty values for all loci using the options -L MR -x <name of file containing all single-locus ML trees> (Kobert et al., 2016). Loci with relative tree certainty including all conflicting bipartitions (i.e., TCA) scores 1.5 times the interquartile range *below* the median TCA score (i.e., loci that had especially *low* phylogenetic signal) were removed.

Standard substitution models such as those used in RAxML and IQ-TREE assume that sequence evolution is stationary, reversible and homogeneous (SRH; Ababneh et al., 2006), but empirical data often do not fit these assumptions (Naser-Khdour et al., 2019). To determine if our data meet SRH assumptions, we used the "–symtest" option in IQ-TREE 1.7 beta version 17 (Nguyen et al., 2015) to calculate p-values for three matched-pairs test of homogeneity (symmetry, marginal symmetry and internal symmetry) for all of our data sets.

2.5. Phylogenetic analyses

The filtering steps above yielded a final set of twenty-four concatenated data matrices (Table 2). All matrices were subjected to partitioned maximum likelihood phylogenetic analysis in RAxML 8.2.10 and 8.2.12 (Stamatakis, 2014) on CIPRES (Miller et al., 2010) with 200 rapid bootstrap pseudoreplicates, using these options: -f a -x <rapid bootstrapping random number seed; unique for each analysis> -p <parsimony random number seed; unique for each analysis> -# 200 -m PROTGAMMA<amino acid model> -s <inputfile> -n <outputfile> -q <partitionfile>.

Debate about the use of partitioned vs. site-heterogeneous models (Halanych et al., 2016; Pisani et al., 2016, 2015; Whelan et al., 2015; Whelan and Halanych, 2016) and recent work suggesting that partitioned analyses are vulnerable to long-branch attraction artifacts (Wang et al., 2019) led us to explore alternative analytical approaches. Unfortunately, analyses under the CAT-GTR model in PhyloBayes are computationally challenging. However, recent developments using distribution mixture models estimated via maximum likelihood (Le et al., 2008; Schrempf et al., 2020; Wang et al., 2018) have made analyses using site-heterogeneous models more tractable. Attempts to infer empirical site distributions for our data repeatedly failed due to computer memory limitations, so we used prepackaged universal distribution mixture (UDM) models produced by Schrempf et al. (2020), based on log center log ratio transformations of site distributions from the HOGENOM database (Dufayard et al., 2005), in IQ-TREE 1.6.12. Preliminary analyses of the C75-B1 matrix suggested that mixture models with 16, 32 and 64 components were all overfit. To minimize model overfitting while retaining a number of components comparable to a C20 model (Le et al., 2008), we used a model with 16 components (hogenom_lclr_016) for analyses of all 24 matrices described above, running each analysis twice to check for consistency.

2.6. Evaluating conflict among gene trees

Traditional measures of support (e.g., bootstrapping and Bayesian posterior probabilities) may mask substantial conflict among genes in phylogenomic data sets (Smith et al., 2015). To characterize the amount and distribution of conflict among genes, we used phyparts (Smith et al., 2015) to map single-locus ML trees to concatenated ML trees for four data sets (C25-A1, C75-A1, L25-A1 and L75-A1; Table 2) that span a range of included loci and amounts of missing data. These analyses require rooted trees, so we only mapped locus trees that included at least one member of Octopodiformes, and used Mesquite version 3.6 (Maddison and Maddison, 2018) to root the locus trees, designating octopodiforms as outgroups. Analyses were conducted by mapping all single-locus tree bipartitions and only single-locus bipartitions with 50% or greater bootstrap support. Pie charts reflecting conflict at each node of the concatenated ML tree were generated with the phypartspiecharts.py script (available at https://github.com/mossmatters/MJPythonNotebooks/blob/master/PhyParts_PieCharts.ipynb).

Table 2
Summary of all data matrices used in this study. "Ortholog set": the ortholog set used in HaMStR; "Coverage": loci included were represented in at least this percentage of transcriptomes (of 31); "Filtering 1": indices used to remove outlier loci (TB = heterogeneity indices calculated using TreSpEx and BaCoCa; TC = relative tree certainty score, including all conflicting bipartitions [TCA]); "Filtering 2": approach used to remove likely contaminant sequences ("Sepiadariidae" = deletion of individual sepiadariid sequences that grouped within Octopodiformes in single-locus ML trees; "Decapodiform monophyly" = deletion of loci for which Decapodiformes was non-monophyletic in single-locus ML trees). See text for details.

Matrix	Ortholog Set	Coverage	Filtering 1	Filtering 2
C25-A1	cephalopod	25%	TB	Sepiadariidae
C25-A2	cephalopod	25%	TB	Decapodiform monophyly
C25-B1	cephalopod	25%	TB + TC	Sepiadariidae
C25-B2	cephalopod	25%	TB + TC	Decapodiform monophyly
C50-A1	cephalopod	50%	TB	Sepiadariidae
C50-A2	cephalopod	50%	TB	Decapodiform monophyly
C50-B1	cephalopod	50%	TB + TC	Sepiadariidae
C50-B2	cephalopod	50%	TB + TC	Decapodiform monophyly
C75-A1	cephalopod	75%	TB	Sepiadariidae
C75-A2	cephalopod	75%	TB	Decapodiform monophyly
C75-B1	cephalopod	75%	TB + TC	Sepiadariidae
C75-B2	cephalopod	75%	TB + TC	Decapodiform monophyly
L25-A1	lophotrochozoan	25%	TB	Sepiadariidae
L25-A2	lophotrochozoan	25%	TB	Decapodiform monophyly
L25-B1	lophotrochozoan	25%	TB + TC	Sepiadariidae
L25-B2	lophotrochozoan	25%	TB + TC	Decapodiform monophyly
L50-A1	lophotrochozoan	50%	TB	Sepiadariidae
L50-A2	lophotrochozoan	50%	TB	Decapodiform monophyly
L50-B1	lophotrochozoan	50%	TB + TC	Sepiadariidae
L50-B2	lophotrochozoan	50%	TB + TC	Decapodiform monophyly
L75-A1	lophotrochozoan	75%	TB	Sepiadariidae
L75-A2	lophotrochozoan	75%	TB	Decapodiform monophyly
L75-B1	lophotrochozoan	75%	TB + TC	Sepiadariidae
L75-B2	lophotrochozoan	75%	TB + TC	Decapodiform monophyly

We also employed IQ-TREE 1.7 beta 17 to infer partitioned ML trees with 1000 ultrafast bootstrap replicates (Hoang et al., 2018) for each concatenated data matrix and each single-locus data matrix, using the substitution models selected for each locus with the ProteinModelSelection script, with gamma-distributed rates (four rate categories). We then used the concatenated matrix trees and single-locus trees to calculate gene concordance factors (gCFs) and site concordance factors (sCFs) (Ané et al., 2006; Gadagkar et al., 2005; Minh et al., 2020).

2.7. *Further assessments of potential sources of bias*

In some cases, the analyses described above returned differing topologies. To explore some possible causes for these differences across different data matrices, we performed additional analyses on a subset of the matrices. First, to further reduce the impact of composition bias for two of the smallest matrices that yielded different topologies (C75-B1 and L75-B2), we recoded amino acids into groups that minimize compositional heterogeneity using minmax-chisq (http://www.mathstat.dal.ca/tsusko/software.cgi; Susko and Roger, 2007) as described in Lindgren and Anderson (2018). Second, to assess the potential impact of heterotachy (i.e., variation in evolutionary rate at a given position over time; Lopez et al., 2002; Philippe and Lopez, 2001), we analyzed a version of the L75-B2 matrix under the GHOST mixture model (Crotty et al., 2020) using an MPI version of IQ-TREE v. 1.6.12. Because we cannot simultaneously use multiple substitution models and variable site classes, we only analyzed loci for which JTT was the best-fitting substitution model (108 of 154 loci had JTT or JTTF as best-fitting model). We inferred trees under the JTT model using two to twelve classes of variable sites, allowing amino acid frequencies, branch lengths and substitution rates to vary among classes (i.e., -m JTT + FO*H2 to -m JTT + FO*H12) and used the number of classes with the lowest Bayesian information criterion in a subsequent analysis (1000 ultrafast bootstrap replicates, employing the -bnni option). Finally, to assess whether different protein structural regions contain conflicting phylogenetic signal, we evaluated all single-locus alignments for the C75-B1 and L75-B2 data sets with the structure_assignment.pl script (Pandey and Braun, 2020), which uses a protein structure prediction package (SCRATCH-1D, release 1.2; Cheng et al., 2005; Magnan and Baldi, 2014) to infer secondary structure and solvent accessibility. The script parsed sites in each alignment into five structural classes—coil, helix, sheet, buried or exposed. Then, for each matrix, we concatenated sites by class and analyzed the concatenated coil, helix, etc., matrices in IQ-TREE v. 1.6.12 (inferring an unpartitioned best-fitting model and conducting 1000 ultrafast bootstrap replicates for each matrix). The resulting trees were visually compared to determine whether different classes of sites supported conflicting topologies.

2.8. *Hypothesis testing*

To further evaluate the conflicting phylogenomic hypotheses of decapodiform relationships (Fig. 1), we performed weighted Shimodaira-Hasegawa tests (WSH; Buckley et al., 2001; Shimodaira, 1998; Shimodaira and Hasegawa, 1999) in IQ-TREE using 3 datasets: one of our larger matrices (C50-B1), a matrix of comparable size (C75-B1) to Tanner et al. (2017) and the data matrix from Tanner et al. (2017). The WSH test assumes that all credible trees are included in the set of trees to be evaluated, which can constitute a major computational burden if the number of credible trees is large. To facilitate comparisons, we modified each data matrix. We modified Tanner et al.'s matrix by deleting *Spirula*, all but four members of Octopodiformes and all non-coleoid outgroups to more closely match our sampling. To generate a computationally manageable set of credible trees, we deleted *Galiteuthis armata* (the transcriptome with the largest amount of missing data; Table 1) from our C75-B1 and C50-B1 matrices. For all three data matrices, we used a constraint tree in which all orders and families were monophyletic and relationships within Myopsida (Anderson, 2000a, 2000b; Lindgren and Anderson, 2018; Sales et al., 2013; Tanner et al., 2017), Oegopsida (Lindgren, 2010), Ommastrephidae (Pardo-Gandarillas et al., 2018), Sepiida (Strugnell et al., 2017; Uribe and Zardoya, 2017) and Octopodiformes matched previous studies. This yielded a credible set of 11,025 trees for the Tanner et al. data matrix and 23,625 trees for our two matrices. Since we did not know the gene boundaries in the Tanner et al. matrix, we inferred the best-fitting unpartitioned substitution model with ModelFinder (Kalyaanamoorthy et al., 2017) for that matrix; for our matrices, we used the partitioned model inferred with RAxML, with gamma-distributed rates (four rate categories). For all tests, model parameters were estimated on an initial parsimony tree, and either 1000 (for C50-B1) or 10,000 (Tanner et al. and C75-B1) RELL replicates were used to find a WSH confidence set of trees.

3. Results

3.1. *Data matrix characteristics*

Data matrices ranged from 154 loci and ~86,000 characters to 1404 loci and ~720,000 characters, with amounts of missing data ranging from ~24% to ~58% (Table 3). Maximum test of symmetry (SymPval; Naser-Khdour et al., 2019) p-values for all loci and matrices were greater than 0.1; i.e., we cannot reject the null hypothesis that our data comply with SRH assumptions.

3.2. *Phylogenies*

We conducted an initial set of 48 analyses (two ortholog sets, three levels of missing data, two methods for removing contaminant sequences from the sepiadariids, two filtering methods and two types of ML analysis). Across all of the resulting topologies, several clades were recovered:

1) Octopodiformes and Decapodiformes were reciprocally monophyletic;
2) *Idiosepius* was sister to a clade comprising all other decapodiforms;
3) Sepiolida was monophyletic and sister to all other decapodiforms except *Idiosepius*;
4) Myopsida, Oegopsida and Sepiida were each monophyletic.

Two patterns of relationships among the major decapodiform subclades were recovered (Fig. 2; Supplementary Materials). In the majority topology (recovered in 44 analyses), Myopsida and Sepiida formed a clade sister to Oegopsida (Fig. 2). In the minority topology, recovered in four analyses (standard partitioned model and UDM model analyses of two of the smallest matrices, the L75-A2 and L75-B2 matrices), Myopsida was recovered as sister to Oegopsida instead of Sepiida. Relationships within the four decapodiform subclades represented by more

Table 3
Number of orthogroups (loci), characters and taxa and percentage of missing data for the concatenated data matrices. Values are for the loci recovered using the cephalopod core ortholog set.

Matrix	# Loci	# Characters	% Missing
C25-A1	1404	719,229	47.72
C25-A2	1102	586,555	50.18
C25-B1	1375	708,188	47.45
C25-B2	1079	577,641	49.99
C50-A1	909	472,928	35.94
C50-A2	659	359,367	37.30
C50-B1	893	468,482	35.95
C50-B2	645	355,264	37.35
C75-A1	512	245,142	24.57
C75-A2	357	178,394	25.37
C75-B1	496	241,453	24.69
C75-B2	346	175,587	25.50
L25-A1	1242	661,283	53.46
L25-A2	930	511,712	57.76
L25-B1	1225	652,623	53.18
L25-B2	916	504,440	57.55
L50-A1	630	343,883	37.94
L50-A2	410	231,954	41.17
L50-B1	622	340,530	37.87
L50-B2	406	229,896	41.06
L75-A1	293	147,476	24.48
L75-A2	159	88,134	27.51
L75-B1	289	146,254	24.52
L75-B2	154	86,268	27.57

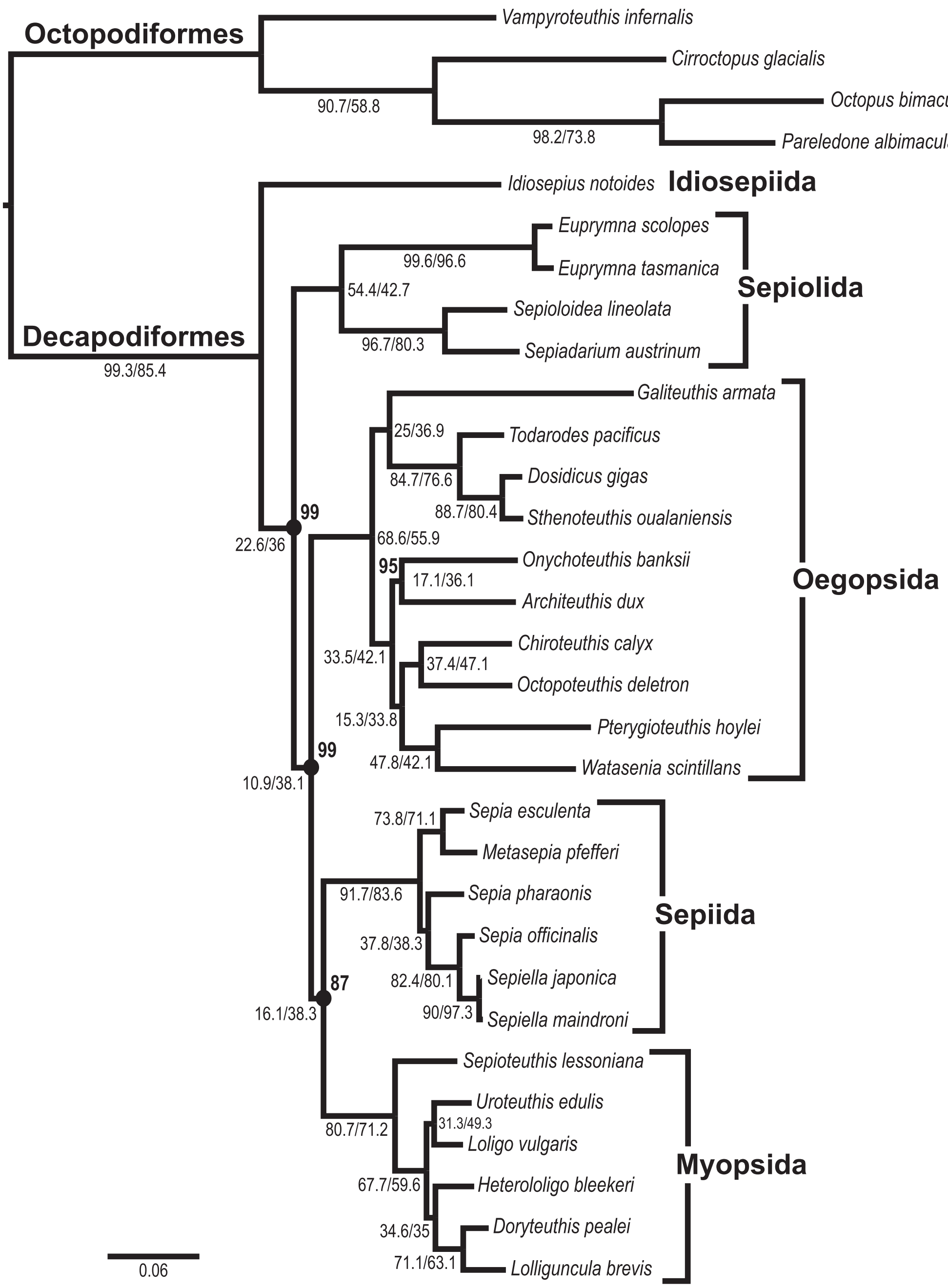


Fig. 2. RAxML maximum likelihood phylogram for the C75-A1 data matrix. Numbers on the branches are gene concordance factors/site concordance factors. ML bootstrap values are all 100% except where shown (in bold text). Nodes of interest along the decapodiform backbone are marked with black circles. See text for analytical details.

than one species (Myospida, Oegopsida, Sepiida and Sepiolida) were consistent across all trees, with one exception. In 46 of 48 trees, *Galiteuthis*, our single representative from Cranchiidae, was recovered as sister to Ommastrephidae, which was represented by three genera. In two trees, *Galiteuthis* was sister to a clade comprising all oegopsids except Ommastrephidae, though with low (~70%) bootstrap support (Supplementary Materials). It is noteworthy that while *Galiteuthis* is the smallest transcriptome included in this study (Table 1), previous molecular studies have recovered Cranchiidae as sister to Ommastrephidae and/or sister to other families within Oegopsida (Lindgren, 2010; Lindgren et al., 2012).

Bootstrap support for most clades was high, but examination of conflict (Figs. 2 and 3) showed substantial variation in the nature and amount of conflict among loci across nodes and data sets. All nodes in Octopodiformes and most nodes within the five sampled decapodiform subclades showed higher proportions of loci congruent with the concatenated majority ML topology than loci that support conflicting bipartitions. In relationships among the five decapodiform subclades (Fig. 3), the number of loci in conflict with the node present in the concatenated majority ML topology dwarfed the number of supporting loci. However, for both within and among subclade relationships, the number of loci supporting a given node was higher than the number supporting the most common alternative bipartition. Similarly, gene and site concordance factors were relatively high (>50) for many nodes, with the primary exceptions being along the decapodiform backbone (Fig. 2), where all gCF values were very low.

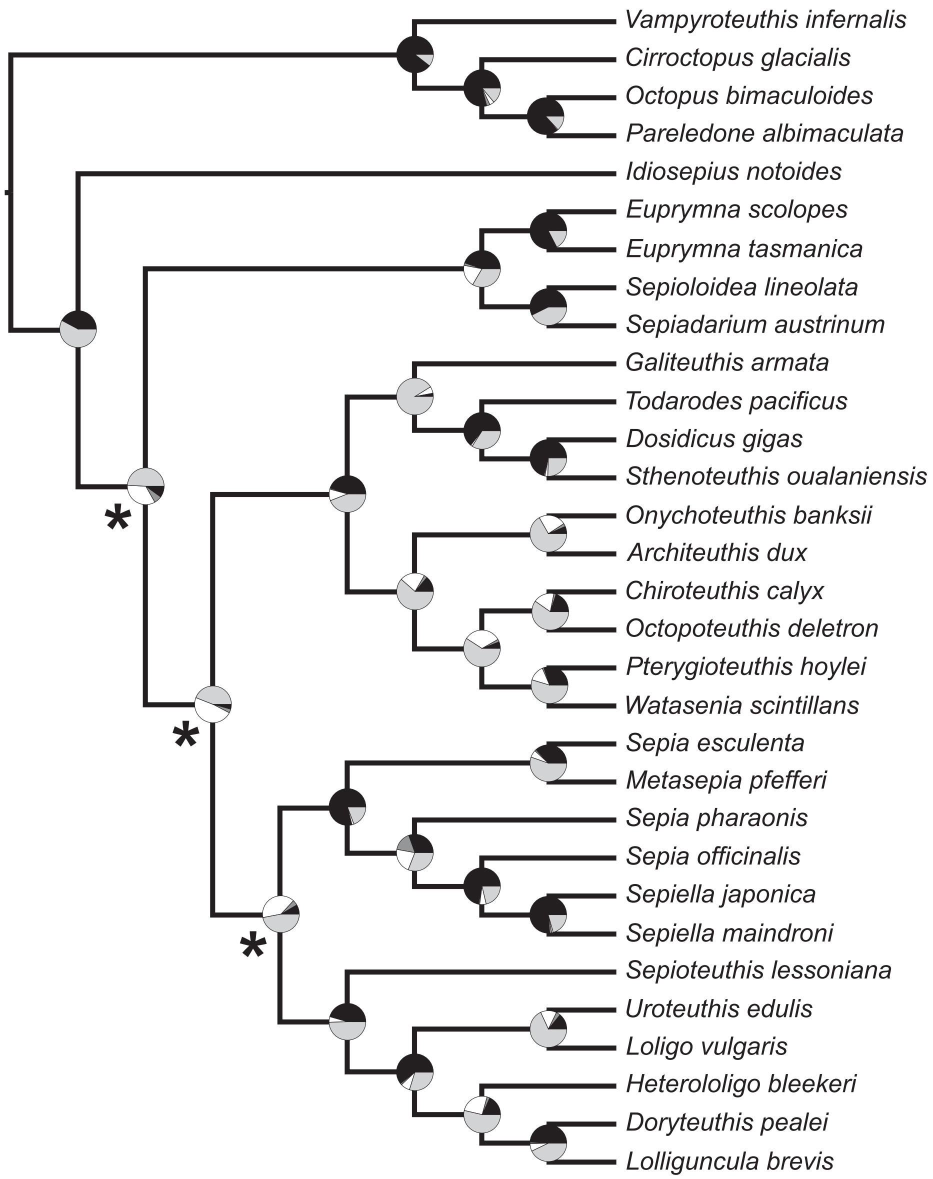


Fig. 3. PhyParts output for the C75-A1 data matrix, showing concordance and conflict among genes for each node on the concatenated ML topology. Pie chart colors are as follows: black = percentage of genes that are concordant with the node; dark gray = percentage of genes supporting the top alternative bipartition; white = percentage of genes supporting all other alternative bipartitions; light gray = percentage of genes that are uninformative for that node. Asterisks denote high-conflict nodes for relationships among the five major decapodiform lineages.

3.3. Sources of conflict

To assess potential sources of conflict, we conducted analyses focused on one data set (L75-B2) for which initial ML analysis returned the minority topology (i.e., Myopsida + Oegopsida). Recoding amino acids to further reduce compositional heterogeneity again returned the minority topology. GHOST analysis of only JTT loci extracted from this data set also returned the minority topology, with high (99%) bootstrap support for a myopsid + oegopsid clade. Analyses of sites sorted into inferred structural classes for the L75-B2 data set showed no obvious cases where different classes strongly supported alternative topologies. All structural classes supported a myopsid + oegopsid clade except sheet sites (Oegopsida + Sepiida, 46% bootstrap support) and helix sites (Myopsida + Sepiida, 38%). All structural classes had similar numbers of sites except helix (which comprised about three-quarters as many sites as buried, coil and exposed) and sheet sites (about one-third as many sites as buried, coil and exposed). Buried, sheet and helix sites had notably higher percentages of invariant sites than did coil and exposed sites. In short, analyzing a data set that initially returned a myopsid + oegopsid topology using methods designed to minimize compositional heterogeneity and heterotachy did not return the majority (Myopsida + Sepiida) topology, and exploring phylogenetic signal in different protein structural classes yielded no clear patterns, though analysis of helix sites alone did return the majority Myopsida + Sepiida topology with low bootstrap support.

3.4. Hypothesis tests

WSH tests of the similarly sized Tanner et al. data matrix and our C75-B1 data matrix yielded qualitatively similar results, though the number of trees in the credible sets and confidence sets differed substantially. 11,025 trees were evaluated for the Tanner et al. data set and 23,625 trees were evaluated for our C75-B1 data matrix, and the WSH test returned 90 trees as the confidence set for the Tanner et al. matrix and four trees as the confidence set for C75-B1. The much larger number of trees in the Tanner et al. matrix confidence set is suggestive of the comparatively low power of that matrix to discriminate topologies. Strict consensus trees of each confidence set were identical with respect to relationships among subclades—both showed a polytomy among all five decapodiform subclades (Fig. 4 a, b). However, the confidence set of trees for the WSH test of our (larger) C50-B1 data set comprised only two trees. In both trees, Idiosepiida was recovered as the sister to the rest of

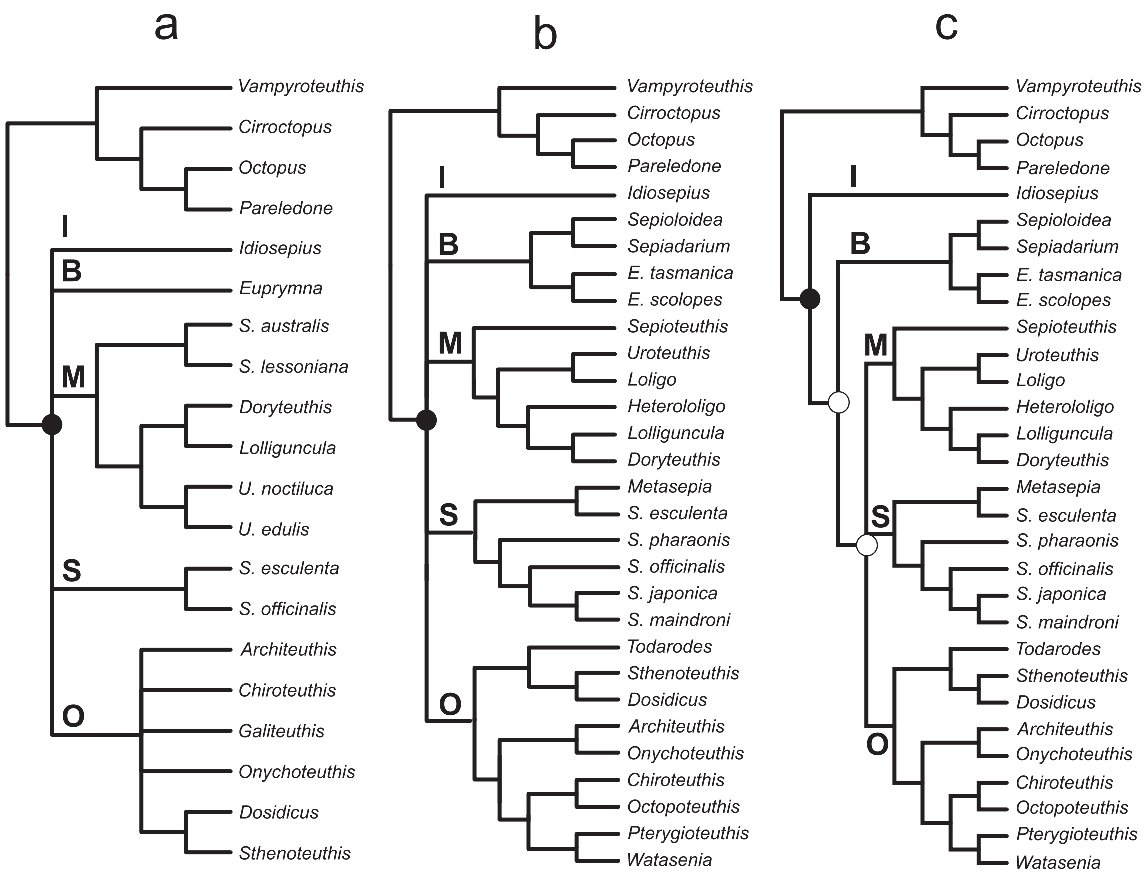


Fig. 4. Strict consensus trees summarizing the weighted Shimodaira-Hasegawa credible sets of trees. a) Credible set consensus for the Tanner et al. matrix; b) credible set consensus for the C75-B1 matrix with *Galiteuthis* deleted; c) credible set consensus for the C50-B1 matrix with *Galiteuthis* deleted. I = Idiosepiida, B = Sepiolida (bobtail and bottletail squids), M = Myopsida, S = Sepiida, O = Oegopsida.

Decapodiformes and Sepiolida was recovered as sister to a clade comprising Myopsida, Oegopsida and Sepiida (Fig. 4C). The two trees differed only in the relationships within this latter clade—in one, Myopsida and Oegopsida were sister taxa, and in the other, Myopsida and Sepiida were sister taxa, echoing our minority and majority topologies and resulting in a polytomy within this clade in the strict consensus tree.

4. Discussion

4.1. Relationships among the major decapodiform subclades

Our analyses consistently recovered a clade comprising all relatively large-bodied (>20 mm adult mantle length) decapodiform taxa—i.e., Myopsida, Oegopsida, Sepiida and Sepiolida—with Idiosepiida as sister to this clade, as well as a subclade comprising Myopsida, Oegopsida and Sepiida. Though this hypothesis of relationships had not been recovered prior to Lindgren and Anderson (2018), some overall patterns are coalescing across phylogenomic topologies within Decapodiformes. Most notably, Myopsida and Oegopsida appear to be derived decapodiform groups with respect to Idiosepiida, Sepiida and Sepiolida (Figs. 1 and 2). Though the current study did not include ancestral state reconstructions, this general finding has important implications for decapodiform evolution. First, the one-part corneal eye covering seen in Myopsida, Idiosepiida, Sepiida and Sepiolida appears to be ancestral for Decapodiformes (though distinct from the two-part cornea seen in Octopodiformes) and lost in the ancestor of Oegopsida, as suggested by Lindgren et al. (2012). Furthermore, most species in Idiosepiida, Sepiida and Sepiolida are demersal, and all of these taxa (except perhaps heteroteuthin sepiolids, whose egg-laying behavior is unknown), as well as all myopsid species, lay benthic eggs or egg masses (Boletzky, 1998). By contrast, many oegopsids are found in open-ocean habitats and produce pelagic egg masses or single eggs (Bjørke et al., 1997; Boletzky, 1998; O'Shea et al., 2004; Roberts et al., 2011; Young et al., 1985) or brood their eggs in their arms (Okutani et al., 1995). The status of Spirulida and Bathyteuthoidea in this respect is unclear—*Bathyteuthis* itself is a pelagic brooder, but the egg-laying behavior of *Chtenopteryx* (the other member of Bathyteuthoidea) remains unknown, and *Spirula* is assumed (but not known) to lay benthic eggs. All of the molecular phylogenies produced to date for Decapodiformes suggest a decapodiform ancestor that was closely associated with the benthos, particularly for reproduction (Young et al., 1998). By contrast, oegopsids have mostly (and perhaps completely) lost this connection with the benthos, in an evolutionary transition analogous to the evolution of the amniotic egg, which freed a lineage of tetrapod vertebrates from returning to water to reproduce.

4.2. Conflict and uncertainty

As noted by Lindgren and Anderson (2018), extant cephalopod diversity is likely the product of a series of ancient and rapid radiations, which can make phylogenetic resolution difficult even with immense amounts of genetic data. Previous attempts to clarify coleoid cephalopod phylogeny using transcriptome data (Tanner et al., 2017; Lindgren and Anderson, 2018) were somewhat limited, in part due to the paucity of data for some lineages. Tanner et al. (2017) used 129 loci that had been proposed as particularly suitable for metazoan phylogenetics (Philippe et al., 2011). Lindgren and Anderson (2018) used an orthology assessment approach that recovered far more loci, but some transcriptomes (most notably from *Nautilus*, *Vampyroteuthis* and *Idiosepius*) yielded far fewer orthologs than others (Lindgren and Anderson, 2018). The problem of limited/missing data caused by small transcriptomes is rapidly fading; the smallest transcriptome used here (other than *Galiteuthis*) yielded nearly 1400 orthologs using the cephalopod ortholog set in HaMStR (Table 1). Fewer missing data and more loci may have contributed to our consistent recovery in this study of the majority topology (Fig. 2), but some uncertainty remains. The WSH test of the largest data set we tested (C50-B1) rejected all topologies except two, but these differed with respect to internal relationships within the clade comprising Myopsida, Oegopsida and Sepiida. Our analyses suggest a Myopsida + Sepiida sister pair, but Myopsida was sometimes recovered as sister to Oegopsida, and the C50-B1 WSH 95% confidence set included trees containing Myopsida + Oegopsida and Myopsida + Sepiida. Furthermore, gene and site concordance factors showed substantial conflict along the decapodiform backbone. Thus, though we believe that the phylogeny shown in Fig. 2 is the best currently available estimate of relationships within Decapodiformes, at minimum, relationships among Myopsida, Oegopsida and Sepiida remain unclear.

4.3. Future prospects

If morphology, fossils, mitogenomes and transcriptomes have been unable to resolve relationships within Decapodiformes fully and robustly, what next? High-quality transcriptomes from *Spirula spirula* and at least one representative of Bathyteuthoidea, as well as for additional lineages within Oegopsida, are needed. However, the data collected to date (mitogenomes, plus some shotgun sequencing data for *Spirula*) (Strugnell et al., 2017; Tanner et al., 2017) strongly suggest that Spirulida and Bathyteuthoidea are closely related to Oegopsida. We question whether inclusion of transcriptomes from either or both of these taxa will greatly clarify decapodiform phylogeny, but acknowledge that their inclusion in future studies will be crucial for more rigorous hypothesis testing.

If mitochondrial genomes and transcriptome data continue to conflict, the obvious next step is complete genomes, sampled from at least one representative of the five decapodiform lineages represented here and, ideally, also from Spirulida and Bathyteuthoidea. Complete genome sequences would yield an immense amount of protein-coding gene data, but would also permit higher-order phylogenetic analyses of genome architecture such as synteny (Drillon et al., 2020; Sevillya and Snir, 2019), rare genomic changes (Rokas and Holland, 2000) and sequence string frequency (Edwards et al., 2002; Yu et al., 2005).

The necessary data will be available soon. Complete genome sequences are available for *Octopus bimaculoides* (Albertin et al., 2015), *Euprymna scolopes* (Sepiolida) (Belcaid et al., 2019) and *Architeuthis dux* (Oegopsida) (da Fonseca et al., 2020), and a genome for *Doryteuthis pealei* is forthcoming (Myopsida) (C. Albertin, pers. comm.). Genome sequences for representatives of Idiosepiida and Sepiida are sure to follow soon. Analyses of coding sequences and higher-order genomic features from these genomes may be necessary to clarify decapodiform phylogeny fully, if such clarification is possible.

5. Availability of data and material

The datasets generated during and analyzed in this study, custom scripts used in the phylogenomics pipeline and all trees are available in the Mendeley Data repository (https://doi.org/10.17632/v65ptjwd7j.1).

6. Authors' contributions

FEA designed the study and implemented the bioinformatics pipelines; FEA and ARL performed the phylogenetic analyses; FEA and ARL wrote and edited the manuscript.

CRediT authorship contribution statement

Frank E. Anderson: Conceptualization, Investigation, Methodology, Data curation, Formal analysis, Visualization, Project administration, Funding acquisition, Writing - original draft, Writing - review & editing. **Annie R. Lindgren:** Conceptualization, Investigation, Formal analysis, Funding acquisition, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

We have no competing interests.

Acknowledgements

We thank Mark Miller for his generous and repeated assistance with use of Cyberinfrastructure for Phylogenetic Research (CIPRES) resources, Alistair Tanner and Rute da Fonseca who allowed us to access their data and assemblies, Jan Strugnell and colleagues for making their *Sepioloidea* and *Sepiadarium* transcriptome data publicly available and the National Center for Genome Analysis Support at Indiana University. This research is based in part upon work supported by the National Science Foundation under Grant No. ABI-1759906 to Indiana University. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the National Center for Genome Analysis Support, or Indiana University. This work was supported by two U.S. National Science Foundation grants—DEB-1036516 to FEA and DEB-090633 to ARL.

References

Ababneh, F., Jermiin, L.S., Ma, C., Robinson, J., 2006. Matched-pairs tests of homogeneity with applications to homologous nucleotide sequences. Bioinformatics 22, 1225–1231.

Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., Rokhsar, D.S., 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. Nature 524, 220–224. https://doi.org/10.1038/nature14668.

Allcock, A.L., Cooke, I.R., Strugnell, J.M., 2011. What can the mitochondrial genome reveal about higher-level phylogeny of the molluscan class Cephalopoda? Zool. J. Linn. Soc. 161, 573–586. https://doi.org/10.1111/j.1096-3642.2010.00656.x.

Allcock, A.L., Lindgren, A., Strugnell, J.M., 2014. The contribution of molecular data to our understanding of cephalopod evolution and systematics: a review. J. Nat. Hist. 49, 1373–1421.

Anderson, F.E., 2000a. Phylogenetic relationships among loliginid squids (Cephalopoda: Myopsida) based on analyses of multiple data sets. Zool. J. Linn. Soc. 130, 603–633. https://doi.org/10.1111/j.1096-3642.2000.tb02203.x.

Anderson, F.E., 2000b. Phylogeny and historical biogeography of the loliginid squids (Mollusca: cephalopoda) based on mitochondrial DNA sequence data. Mol. Phylogenet. Evol. 15, 191–214. https://doi.org/10.1006/mpev.1999.0753.

Ané, C., Larget, B., Baum, D.A., Smith, S.D., Rokas, A., 2006. Bayesian estimation of concordance among gene trees. Mol. Biol. Evol. 24, 412–426. https://doi.org/10.1093/molbev/msl170.

Arkhipkin, A.I., Rodhouse, P.G.K., Pierce, G.J., Sauer, W., Sakai, M., Allcock, L., Arguelles, J., Bower, J.R., Castillo, G., Ceriola, L., Chen, C.-S., Chen, X., Diaz-Santana, M., Downey, N., González, A.F., Granados Amores, J., Green, C.P., Guerra, A., Hendrickson, L.C., Ibáñez, C., Ito, K., Jereb, P., Kato, Y., Katugin, O.N., Kawano, M., Kidokoro, H., Kulik, V.V., Laptikhovsky, V.V., Lipinski, M.R., Liu, B., Mariátegui, L., Marin, W., Medina, A., Miki, K., Miyahara, K., Moltschaniwskyj, N., Moustahfid, H., Nabhitabhata, J., Nanjo, N., Nigmatullin, C.M., Ohtani, T., Pecl, G., Perez, J.A.A., Piatkowski, U., Saikliang, P., Salinas-Zavala, C.A., Steer, M., Tian, Y., Ueta, Y., Vijai, D., Wakabayashi, T., Yamaguchi, T., Yamashiro, C., Yamashita, N., Zeidberg, L.D., 2015. World squid fisheries. Rev. Fish. Sci. Aquac. 23, 92–252. https://doi.org/10.1080/23308249.2015.1026226.

Belcaid, M., Casaburi, G., McAnulty, S.J., Schmidbaur, H., Suria, A.M., Moriano-Gutierrez, S., Pankey, M.S., Oakley, T.H., Kremer, N., Koch, E.J., 2019. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proc. Natl. Acad. Sci. 116, 3030–3035.

Bjørke, H., Hansen, K., Sundt, R.C., 1997. Egg masses of the squid *Gonatus fabricii* (Cephalopoda, Gonatidae) caught with pelagic trawl off Northern Norway. Sarsia 82, 149–152. https://doi.org/10.1080/00364827.1997.10413648.

Boletzky, S.V., 1998. Cephalopod eggs and egg masses. Oceanogr. Mar. Biol. 36, 341–372.

Brown, J.M., 2014. Detection of implausible phylogenetic inferences using posterior predictive assessment of model fit. Syst. Biol. 63, 334–348.

Buckley, T.R., Simon, C., Shimodaira, H., Chambers, G.K., 2001. Evaluating hypotheses on the origin and evolution of the New Zealand alpine cicadas (Maoricicada) using multiple-comparison tests of tree topology. Mol. Biol. Evol. 18, 223–234. https://doi.org/10.1093/oxfordjournals.molbev.a003796.

Cheng, J., Randall, A.Z., Sweredoski, M.J., Baldi, P., 2005. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res. 33, W72–W76. https://doi.org/10.1093/nar/gki396.

Crotty, S.M., Minh, B.Q., Bean, N.G., Holland, B.R., Tuke, J., Jermiin, L.S., Haeseler, A.V., 2020. GHOST: recovering historical signal from heterotachously evolved sequence alignments. Syst. Biol. 69, 249–264. https://doi.org/10.1093/sysbio/syz051.

da Fonseca, R.R., Couto, A., Machado, A.M., Brejova, B., Albertin, C.B., Silva, F., Gardner, P., Baril, T., Hayward, A., Campos, A., Ribeiro, Â.M., Barrio-Hernandez, I., Hoving, H.-J., Tafur-Jimenez, R., Chu, C., Frazão, B., Petersen, B., Peñaloza, F., Musacchia, F., Alexander, G.C., Osório, H., Winkelmann, I., Simakov, O., Rasmussen, S., Rahman, M.Z., Pisani, D., Vinther, J., Jarvis, E., Zhang, G., Strugnell, J.M., Castro, L.F.C., Fedrigo, O., Patricio, M., Li, Q., Rocha, S., Antunes, A., Wu, Y., Ma, B., Sanges, R., Vinar, T., Blagoev, B., Sicheritz-Ponten, T., Nielsen, R., Gilbert, M.T.P., 2020. A draft genome sequence of the elusive giant squid, *Architeuthis dux*. Gigascience 9. https://doi.org/10.1093/gigascience/giz152.

Doubleday, Z.A., Prowse, T.A.A., Arkhipkin, A., Pierce, G.J., Semmens, J., Steer, M., Leporati, S.C., Lourenço, S., Quetglas, A., Sauer, W., Gillanders, B.M., 2016. Global proliferation of cephalopods. Curr. Biol. 26, R406–R407. https://doi.org/10.1016/J.CUB.2016.04.002.

Doyle, V.P., Young, R.E., Naylor, G.J.P., Brown, J.M., 2015. Can we identify genes with increased phylogenetic reliability? Syst. Biol. 64, 824–837. https://doi.org/10.1093/sysbio/syv041.

Drillon, G., Champeimont, R., Oteri, F., Fischer, G., Carbone, A., 2020. Phylogenetic reconstruction based on synteny block and gene adjacencies. Mol. Biol. Evol. 37, 2747–2762. https://doi.org/10.1093/molbev/msaa114.

Duchêne, D.A., Duchêne, S., Ho, S.Y.W., 2018. PhyloMAd: efficient assessment of phylogenomic model adequacy. Bioinformatics 34, 2300–2301.

Dufayard, J.-F., Duret, L., Penel, S., Gouy, M., Rechenmann, F., Perriere, G., 2005. Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. Bioinformatics 21, 2596–2603. https://doi.org/10.1093/bioinformatics/bti325.

Ebersberger, I., Strauss, S., von Haeseler, A., 2009. HaMStR: Profile hidden markov model based search for orthologs in ESTs. BMC Evol. Biol. 9, 157. https://doi.org/10.1186/1471-2148-9-157.

Edwards, S.V., Fertil, B., Giron, A., Deschavanne, P.J., 2002. A genomic schism in birds revealed by phylogenetic analysis of DNA strings. Syst. Biol. 51, 599–613.

Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27, 401–410. https://doi.org/10.1093/sysbio/27.4.401.

Foster, P.G., Hickey, D.A., 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48, 284–290. https://doi.org/10.1007/PL00006471.

Gadagkar, S.R., Rosenberg, M.S., Kumar, S., 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. Part B Mol. Dev. Evol. 304, 64–74.

Ganote, C.L., Sanders, S.A., Papudeshi, B.N., Blood, P.D., Doak, T.G., 2017. A voice for bioinformatics. In: Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact. ACM, p. 36.

Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36, 182–198.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.

Halanych, K.M., Whelan, N.V., Kocot, K.M., Kohn, A.B., Moroz, L.L., 2016. Miscues misplace sponges. Proc. Natl. Acad. Sci. 113, E946–E947.

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309. https://doi.org/10.2307/2992396.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

Jereb, P., Roper, C.F.E., 2010. Cephalopods of the world. An annotated and illustrated catalogue of cephalopod species known to date. Volume 2. Myopsid and oegopsid squids. Food and Agricultural Organization of the United Nations, Rome.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518. https://doi.org/10.1093/nar/gki198.

Kobert, K., Salichos, L., Rokas, A., Stamatakis, A., 2016. Computing the internode certainty and related measures from partial gene trees. Mol. Biol. Evol. 33, 1606–1617.

Kocot, K.M., Citarella, M.R., Moroz, L.L., Halanych, K.M., 2013. PhyloTreePruner: A phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. Evol. Bioinform. Online 9, 429–435. https://doi.org/10.4137/EBO.S12813.

Kocot, K.M., Struck, T.H., Merkel, J., Waits, D.S., Todt, C., Brannock, P.M., Weese, D.A., Cannon, J.T., Moroz, L.L., Lieb, B., 2017. Phylogenomics of Lophotrochozoa with consideration of systematic error. Syst. Biol. 66, 256–282.

Kück, P., Meusemann, K., 2010. FASconCAT: convenient handling of data matrices. Mol. Phylogenet. Evol. 56, 1115–1118. https://doi.org/10.1016/j.ympev.2010.04.024.

Kück, P., Struck, T.H., 2014. BaCoCa–a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. Mol. Phylogenet. Evol. 70, 94–98. https://doi.org/10.1016/j.ympev.2013.09.011.

Le, S.Q., Gascuel, O., Lartillot, N., 2008. Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics 24, 2317–2323.

Lindgren, A.R., 2010. Molecular inference of phylogenetic relationships among Decapodiformes (Mollusca: Cephalopoda) with special focus on the squid order Oegopsida. Mol. Phylogenet. Evol. 56, 77–90. https://doi.org/10.1016/j.ympev.2010.03.025.

Lindgren, A.R., Anderson, F.E., 2018. Assessing the utility of transcriptome data for inferring phylogenetic relationships among coleoid cephalopods. Mol. Phylogenet. Evol. 118, 330–342. https://doi.org/10.1016/j.ympev.2017.10.004.

Lindgren, A.R., Pankey, M.S., Hochberg, F.G., Oakley, T.H., 2012. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. BMC Evol. Biol. 12, 129. https://doi.org/10.1186/1471-2148-12-129.

Lopez, P., Casane, D., Philippe, H., 2002. Heterotachy, an important process of protein evolution. Mol. Biol. Evol. 19, 1–7. https://doi.org/10.1093/oxfordjournals.molbev.a003973.

Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Version 3, 40.

Magnan, C.N., Baldi, P., 2014. SSpro/ACCpro 5: almost perfect prediction of protein secondary structure and relative solvent accessibility using profiles, machine learning and structural similarity. Bioinformatics 30, 2592–2597. https://doi.org/10.1093/bioinformatics/btu352.

Miller, M., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, in: Gateway Computing Environments Workshop (GCE), 2010. IEEE, pp. 1–8.

Minh, B.Q., Hahn, M.W., Lanfear, R., 2020. New methods to calculate concordance factors for phylogenomic datasets. Mol. Biol. Evol. 37, 2727–2733. https://doi.org/10.1093/molbev/msaa106.

Naser-Khdour, S., Minh, B.Q., Zhang, W., Stone, E.A., Lanfear, R., 2019. The prevalence and impact of model violations in phylogenetic analysis. Genome Biol. Evol. 11, 3341–3352. https://doi.org/10.1093/gbe/evz193.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

O`Shea, S., Bolstad, K.S., Ritchie, P.A., 2004. First records of egg masses of *Nototodarus gouldi* McCoy, 1888 (Mollusca: Cephalopoda: Ommastrephidae), with comments on egg-mass susceptibility to damage by fisheries trawl. New Zeal. J. Zool. 31, 161–166. https://doi.org/10.1080/03014223.2004.9518369.

Okutani, T., Nakamura, I., Seki, K., 1995. An unusual egg-brooding behavior of an oceanic squid in the Okhotsk Sea. Venus 54, 237–239.

Pandey, A., Braun, E.L., 2020. Phylogenetic analyses of sites in different protein structural environments result in distinct placements of the metazoan root. Biology (Basel). 9, 64. https://doi.org/10.3390/biology9040064.

Pardo-Gandarillas, M.C., Torres, F.I., Fuchs, D., Ibáñez, C.M., 2018. Updated molecular phylogeny of the squid family Ommastrephidae: Insights into the evolution of spawning strategies. Mol. Phylogenet. Evol. 120, 212–217. https://doi.org/10.1016/J.YMPEV.2017.12.014.

Philippe, H., Brinkmann, H., Copley, R.R., Moroz, L.L., Nakano, H., Poustka, A.J., Wallberg, A., Peterson, K.J., Telford, M.J., 2011. Acoelomorph flatworms are deuterostomes related to Xenoturbella. Nature 470, 255–258.

Philippe, H., Lopez, P., 2001. On the conservation of protein sequences in evolution. Trends Biochem. Sci. 26, 414–416. https://doi.org/10.1016/s0968-0004(01)01877-1.

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., Wörheide, G., 2016. Reply to Halanych et al.: Ctenophore misplacement is corroborated by independent datasets. Proc. Natl. Acad. Sci. 113, E948–E949.

Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., Wörheide, G., 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. 112, 15402–15407. https://doi.org/10.1073/pnas.1518127112.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490. https://doi.org/10.1371/journal.pone.0009490.

Roberts, M.J., Zemlak, T., Connell, A., 2011. Cyclonic eddies reveal Oegopsida squid egg balloon masses in the Agulhas Current, South Africa. African J. Mar. Sci. 33, 239–246. https://doi.org/10.2989/1814232X.2011.600294.

Rokas, A., Holland, P.W.H., 2000. Rare genomic changes as a tool for phylogenetics. Trends Ecol. Evol. 15, 454–459. https://doi.org/10.1016/S0169-5347(00)01967-4.

Saccone, C., Lanave, C., Pesole, G., Preparata, G., 1990. Influence of base composition on quantitative estimates of gene evolution. Methods Enzymol. 183, 570–583. https://doi.org/10.1016/0076-6879(90)83037-A.

Sales, J.B.de L., Shaw, P.W., Haimovici, M., Markaida, U., Cunha, D.B., Ready, J., Figueiredo-Ready, W.M.B., Schneider, H., Sampaio, I., 2013. New molecular phylogeny of the squids of the family Loliginidae with emphasis on the genus *Doryteuthis* Naef, 1912: Mitochondrial and nuclear sequences indicate the presence of cryptic species in the southern Atlantic Ocean. Mol. Phylogenet. Evol. 68, 293–299. https://doi.org/10.1016/j.ympev.2013.03.027.

Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327–331. https://doi.org/10.1038/nature12130.

Salichos, L., Stamatakis, A., Rokas, A., 2014. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol. Biol. Evol. 31, 1261–1271.

Schrempf, D., Lartillot, N., Szöllősi, G., 2020. Scalable empirical mixture models that account for across-site compositional heterogeneity. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msaa145.

Sevillya, G., Snir, S., 2019. Synteny footprints provide clearer phylogenetic signal than sequence data for prokaryotic classification. Mol. Phylogenet. Evol. 136, 128–137. https://doi.org/10.1016/J.YMPEV.2019.03.010.

Shen, X.-X., Hittinger, C.T., Rokas, A., 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1, 1–10.

Shepherd, D., Klaere, S., 2018. How well does your phylogenetic model fit your data? Syst. Biol. 68, 157–167.

Shimodaira, H., 1998. An application of multiple comparison techniques to model selection. Ann. Inst. Stat. Math. 50, 1–13. https://doi.org/10.1023/A:1003483128844.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Smith, S.A., Moore, M.J., Brown, J.W., Yang, Y., 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. BMC Evol. Biol. 15, 150. https://doi.org/10.1186/s12862-015-0423-0.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Struck, T.H., 2014. TreSpEx—Detection of misleading signal in phylogenetic reconstructions based on tree information. Evol. Bioinforma. 10, 51. https://doi.org/10.4137/EBO.S14239.

Strugnell, J., Nishiguchi, M.K., 2007. Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) inferred from three mitochondrial and six nuclear loci: a comparison of alignment, implied alignment and analysis methods. J. Molluscan Stud. 73, 399–410. https://doi.org/10.1093/mollus/eym038.

Strugnell, J.M., Hall, N.E., Vecchione, M., Fuchs, D., Allcock, A.L., 2017. Whole mitochondrial genome of the Ram's Horn Squid shines light on the phylogenetic position of the monotypic order Spirulida (Haeckel, 1896). Mol. Phylogenet Evol. 109, 296–301. https://doi.org/10.1016/j.ympev.2017.01.011.

Susko, E., Roger, A.J., 2007. On reduced amino acid alphabets for phylogenetic inference. Mol. Biol. Evol. 24, 2139–2150. https://doi.org/10.1093/molbev/msm144.

Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).

Tanner, A.R., Fuchs, D., Winkelmann, I.E., Gilbert, M.T.P., Pankey, M.S., Ribeiro, Â.M., Kocot, K.M., Halanych, K.M., Oakley, T.H., da Fonseca, R.R., Pisani, D., Vinther, J., 2017. Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. Proc. R. Soc. London B Biol. Sci. 284.

Uribe, J.E., Zardoya, R., 2017. Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. J. Molluscan Stud. 83, 133–144. https://doi.org/10.1093/mollus/eyw052.

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. Syst. Biol. 67, 216–235. https://doi.org/10.1093/sysbio/syx068.

Wang, H.-C., Susko, E., Roger, A.J., 2019. The relative importance of modeling site pattern heterogeneity versus partition-wise heterotachy in phylogenomic inference. Syst. Biol. 68, 1003–1019.

Whelan, N.V., Halanych, K.M., 2016. Who let the CAT out of the bag? Accurately dealing with substitutional heterogeneity in phylogenomic analyses. Syst. Biol. 66, 232–255.

Whelan, N.V., Kocot, K.M., Moroz, L.L., Halanych, K.M., 2015. Error, signal, and the placement of Ctenophora sister to all other animals. Proc. Natl. Acad. Sci. USA 112, 5773–5778. https://doi.org/10.1073/pnas.1503453112.

Young, R.E., Harman, R.F., Mangold, K.M., 1985. The common occurrence of oegopsid squid eggs in near-surface oceanic waters. Pacific Sci. 39 (4), 359–366.

Young, R.E., Vecchione, M., 1996. Analysis of morphology to determine primary sister-taxon relationships within coleoid cephalopods. Am. Malacol. Bull. 12, 91–112.

Young, R.E., Vecchione, M., Donovan, D.T., 1998. The evolution of coleoid cephalopods and their present biodiversity and ecology. South African. J. Mar. Sci. Tydskr. Vir Seewetenskap 20, 393–420.

Yu, Z.G., Zhou, L.Q., Anh, V.V., Chu, K.H., Long, S.C., Deng, J.Q., 2005. Phylogeny of prokaryotes and chloroplasts revealed by a simple composition approach on all protein sequences from complete genomes without sequence alignment. J. Mol. Evol. 60, 538–545. https://doi.org/10.1007/s00239-004-0255-9.

Zhong, M., Hansen, B., Nesnidal, M., Golombek, A., Halanych, K.M., Struck, T.H., 2011. Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. BMC Evol. Biol. 11, 369. https://doi.org/10.1186/1471-2148-11-369.