ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Molecular Phylogenetics and Evolution xxx (2005) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework

Jan Strugnell [a,*], Mark Norman [b], Jennifer Jackson [a], Alexei J. Drummond [a], Alan Cooper [a]

[a] *Molecular Evolution, Department of Zoology, South Parks Road, Oxford, OX1 3PS, UK*
[b] *Sciences, Museum Victoria GPO Box 666E Melbourne, Vic 3001, Australia*

Received 9 December 2004; revised 3 March 2005

Abstract

The resolution of higher level phylogeny of the coleoid cephalopods (octopuses, squids, and cuttlefishes) has been hindered by homoplasy among morphological characters in conjunction with a very poor fossil record. Initial molecular studies, based primarily on small fragments of single mitochondrial genes, have produced little resolution of the deep relationships amongst coleoid cephalopod families. The present study investigated this issue using 3415 base pairs (bp) from three nuclear genes (*octopine dehydrogenase*, *pax-6*, and *rhodopsin*) and three mitochondrial genes (12S rDNA, 16S rDNA, and *cytochrome oxidase I*) from a total of 35 species (including representatives of each of the higher level taxa). Bayesian analyses were conducted on mitochondrial and nuclear genes separately and also all six genes together. Separate analyses were conducted with the data partitioned by gene, codon/rDNA, gene + codon/rDNA or not partitioned at all. In the majority of analyses partitioning the data by gene + codon was the appropriate model with partitioning by codon the second most selected model. In some instances the topology varied according to the model used. Relatively high posterior probabilities and high levels of congruence were present between the topologies resulting from the analysis of all Octopodiform (octopuses and vampire "squid") taxa for all six genes, and independently for the datasets of mitochondrial and nuclear genes. In contrast, the highest levels of resolution within the Decapodiformes (squids and cuttlefishes) resulted from analysis of nuclear genes alone. Different higher level Decapodiform topologies were obtained through the analysis of only the 1st + 2nd codon positions of nuclear genes and of all three codon positions. It is notable that there is strong evidence of saturation among the 3rd codon positions within the Decapodiformes and this may contribute spurious signal. The results suggest that the Decapodiformes may have radiated earlier and/or had faster rates of evolution than the Octopodiformes. The following taxonomic conclusions are drawn from our analyses: (1) the order Octopoda and suborders Cirrata, Incirrata, and Oegopsida are monophyletic groups; (2) the family Spirulidae (Ram's horn squids) are the sister taxon to the family Sepiidae (cuttlefishes); (3) the family Octopodidae, as currently defined, is paraphyletic; (4) the superfamily Argonautoidea are basal within the suborder Incirrata; and (5) the benthic octopus genera *Benthoctopus* and *Enteroctopus* are sister taxa.

Keywords: Cephalopoda; Data partitioning; Molecular systematics; Bayesian analysis

1. Introduction

The class Cephalopoda, phylum Mollusca, is a morphologically diverse group that originated about 500 mya in the Upper Cambrian period from a mono-

* Corresponding author.
E-mail address: j.strugnell@qub.ac.uk (J. Strugnell).

1055-7903/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2005.03.020

placophoran ancestor (Salvini-Plawen, 1980; Young et al., 1996). Extant cephalopods can be divided into two subclasses; Nautiloidea, (*Nautilus* and *Allonautilus*) and the Coleoidea, containing the octopuses, squids, and cuttlefishes. Coleoid cephalopods differ from nautiloids most notably through the reduction and internalisation (or complete loss) of the shell (Teichert, 1998). As a result, coleoid cephalopods rarely fossilise well and the fossil record provides us with very little information about the origin and relationships of the extant orders.

At present the Coleoidea are divided into two superorders; Decapodiformes and Octopodiformes (Berthold and Engeser, 1987). Morphological studies have been useful in classifying species within subfamilies and/or genera (e.g., Berthold and Engeser, 1987; Clarke, 1988; Khromov, 1990; Voight, 1993; Young and Vecchione, 1996; Voss, 1988;), but have proved to be problematic when determining higher level relationships.

The Decapodiformes currently contains the orders Teuthoidea [suborders Myopsida (closed-eye squids, Fig. 1A) and Oegopsida (open-eye squids, Fig. 1B)] and Sepioidea [families Idiosepiidae (pygmy squid, Fig. 1C), Sepiidae (cuttlefishes, Fig. 1D), Spirulidae (Ram's horn squid, Fig. 1E), Sepiolidae (bobtail squids, Fig. 1F), and Sepiadariidae (bottletail squids, Fig. 1G)] (Supplementary Appendix 1) (Voss, 1977; Young et al., 1996). Although each of these families and suborders are understood to form monophyletic groups, there is much debate concerning the validity of the ordinal level of classification (Berthold and Engeser, 1987; Naef, 1921–1923; Voss, 1977; Young and Vecchione, 1996).

The Octopodiformes contain the orders Vampyromorpha (vampire "squid") and Octopoda (pelagic and benthic octopuses). A sister taxa relationship between these two orders is now well understood from both morphological, (Boletzky, 1992; Engeser, 1997; Pickford, 1939; Young and Vecchione, 1996; Young et al., 1998) and molecular studies (Carlini et al., 2000). The Octopoda is understood to contain the suborders Cirrata (deep-sea finned octopuses, Fig. 1H) and Incirrata [benthic octopuses, Fig. 1I, and pelagic octopuses including the argonautoids (i.e., argonauts, Fig. 1J, and blanket octopuses, Fig. 1K)]. A sister taxa relationship between these suborders is also widely accepted (Grimpe, 1921; Naef, 1921–1923; Voight, 1997; Young and Vecchione, 1996). However, phylogenetic relationships between the nine Incirrata families (Supplementary Appendix 1) are less clear and have been debated extensively in the literature (Naef, 1921–1923; Robson, 1929, 1931; Voight, 1997; Voss, 1977; Young and Vecchione, 1996).

Analysis of coleoid cephalopod relationships using molecular sequence data began in the mid 1990s. Bonnaud et al. (1994, 1996, 1997) and Boucher-Rodoni and Bonnaud (1996) conducted a number of molecular studies to resolve relationships within the Decapodiformes. These studies used a small number of taxa (9–18), mitochondrial genes (16S rDNA, COII, and COIII), and primarily used distance and maximum parsimony (MP) methods of analysis. The data were able to resolve relationships at the intrafamilial level, but branching patterns at higher levels were not well supported.

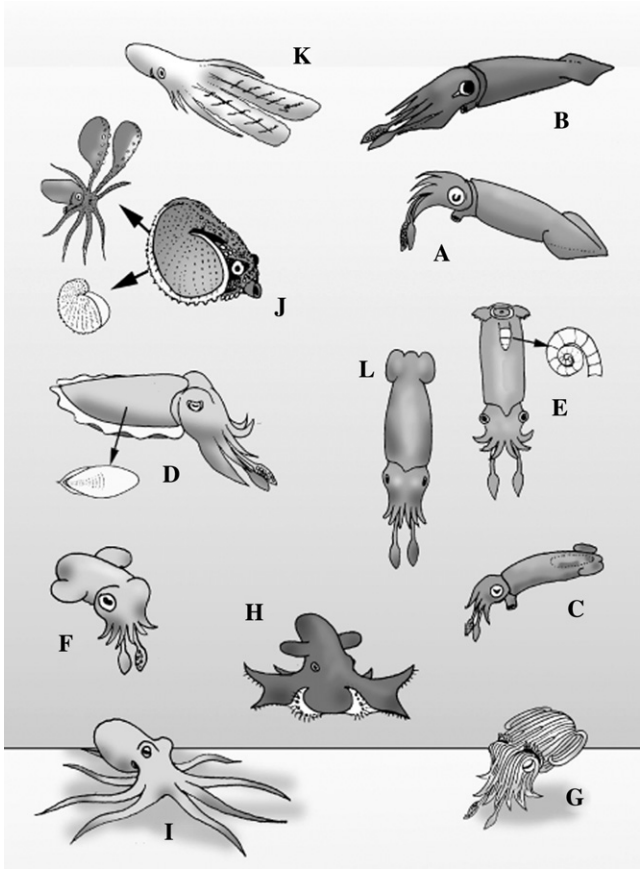

Fig. 1. Illustration depicting the various higher level cephalopod groups. The Decapodiformes is comprised of the order Teuthoidea (suborders Myopsida (A) and Oegopsida (B)) and the order Sepioidea (families Idiosepiidae (C), Sepiidae (D), Spirulidae (E), Sepiolidae (F), and Sepiadariidae (G)). The family Sepiidae possess an internal white shell known as a cuttlebone (D). The family Spirulidae also possess an internal chambered shell (E). The Octopodiformes contains the order Vampyromorpha and order Octopoda (suborders Cirrata (H) and Incirrata). The suborder Incirrata contains benthic octopuses (I) and pelagic octopuses; including families Argonautidae (J) and Tremoctopodidae (K).

Carlini and Graves (1999), Carlini et al. (2000, 2001) recognised that slower evolving genes and a larger number of species were required to resolve higher level relationships within the coleoid cephalopods. These studies used the mitochondrial gene COI (~650 bp) (Carlini and Graves, 1999; Carlini et al., 2001) and *actin* (~780 bp), in the first nuclear gene used to study higher level coleoid phylogenetics (Carlini et al., 2000). Over 40 cephalopod species were sequenced and phylogenetically analysed using MP and maximum likelihood (ML)

methods. Unfortunately both genes were found to be less informative than expected and higher level relationships could not be resolved.

The main aim of this study was to resolve the higher order relationships within Octopodiformes and Decapodiformes. Three nuclear genes; *pax-6*, *rhodopsin*, and *octopine deydrogenase* (ODH), and three mitochondrial genes; 12S rDNA, 16S rDNA, and *cytochrome oxidase I* (COI), were sequenced (Supplementary Appendix 1).

The most appropriate method for analysing multiple genes in phylogenetic analyses is an issue of some debate (DeBry, 2003; Huelsenbeck et al., 1996). It is well recognised that the three codon positions within protein-coding genes evolve at different rates, (Kafatos et al., 1977) while different genes can evolve both at different rates and exhibit markedly different evolutionary properties (Mouchiroud et al., 1995; Ticher and Graur, 1989). The practice of treating all genes as if they are undergoing the same substitution process is being replaced by the partitioning of datasets by gene and/or codon (DeBry, 2003) and even assigning separate partitions to stem and loop regions of rRNA. The recent development of Bayesian inference of phylogeny using Markov chain Monte Carlo (MCMC) methods has allowed very complex and parameter rich models of sequence evolution to be implemented due to its relative computational efficiency in comparison to ML methods. Suchard et al. (2003) presented a method detailing the use of multipartite sequence data in a single analysis, whilst allowing different phylogenetic parameters in individual partitions. Subsequently, Nylander et al. (2004) presented a Bayesian approach to the analysis of a combined dataset of morphological data and four genes and investigated a number of complex substitution models including independent parameters for each gene. The second aim of this study was to use a Bayesian approach to investigate the various methods of partitioning sequence data in a multigene analysis; by gene, by codon or by both gene and codon, and to examine the effects on the fit of the model to the data and the resulting topology.

2. Materials and methods

2.1. Molecular techniques

Thirty-five species were used in the present study, including representatives from each higher level taxa within the Subclass Coleoidea (Supplementary Appendix 1). Tissue samples were fixed in ethanol (70–90%) and stored at –20 °C until DNA extractions were carried out. The DNA extraction protocol followed that of Carlini and Graves (1999), except that centrifugal dialysis (Centricon, Millipore) was used to concentrate and desalt the DNA instead of ethanol precipitation.

2.2. Primers

PCR primers for the three mitochondrial genes were taken from the literature (Hillis et al., 1996). Primers for the nuclear genes were designed in the conserved regions of cephalopod and invertebrate sequences present on GenBank and are available on request.

2.3. Molecular techniques

Twenty-five microlitre PCRs were used to amplify the regions of the genes of interest. Each 25 μL reaction contained the following reagents: 0.25 μL of 25 mM dNTPs (AB gene), 13.5 μL of ddH_20, 2.5 μL of 10× buffer (AB gene), 0.25 μL *Taq* DNA polymerase (AB gene), 2.5 μL of 20 mM Mg^{2+} (AB gene), and 2.5 μL of 10 μM of both the forward and reverse primers (Invitrogen). PCR thermal cycling programs were typically 94 °C for 40 s, 50 °C for 40 s, and 72 °C for 90 s for 40 cycles, followed by a final extension step of 72 °C for 10 min. Annealing temperatures varied according to the primers used and are available on request. Amplified products were purified using the QiaGen PCR purification kit (Qiagen, UK).

Sequencing reactions were performed on both forward and reverse strands using a PRISM BigDye terminator v3 or v3.1 cycle sequencing ready reaction kit (Applied Biosystems, UK). The pellet was then frozen and transferred to the DNA sequencing facility (Zoology Department, Oxford University) for sequencing on an ABI 377, 310 or 3700.

2.4. Phylogenetic analysis

DNA sequences were compiled and aligned by eye in Se-Al v2.0a11 Carbon (Rambaut, 2002). PAUP*4.0b10 (Swofford, 1998) was used for χ^2 tests of composition homogeneity of the nucleotide data. Tests of base homogeneity were based on variable sites only.

Initially the sequences of each gene were analysed individually using Bayesian inference (MrBayes 3) (Ronquist and Huelsenbeck, 2003). The topological congruence between each of the individual genes was then evaluated using crossed SH tests (Shimodaira and Hasegawa, 1999) implemented in PAUP* (Swofford, 1998). This test was applied in place of the Kishino–Hasegawa (KH) test (Kishino and Hasegawa, 1989) because the latter is invalid when topologies being compared are specified a posteriori (Shimodaira and Hasegawa, 1999).

The sequence data from the individual genes were also concatenated into a single dataset. The dataset was partitioned (using MrBayes 3) into either; individual genes (termed 'gene'), into each of the three codon positions and rDNA (termed 'codon'), or by both individual genes and codon positions/rDNAs (termed 'g + c'). Each of these partitions were investigated separately for datasets of (1) all six genes concatenated (termed '6'), (2) three nu-

clear genes concatenated (termed 'nuc'), and (3) three mitochondrial genes concatenated (termed 'mt'). This terminology is used throughout the manuscript to enable clear description of the datasets and models used. For e.g., '6: gene' indicates a dataset using all six genes that were partitioned by gene. Each of these models was investigated for (1) all taxa, (2) Decapodiformes only, and (3) Octopodiformes only.

MrBayes 3 (Ronquist and Huelsenbeck, 2003) was used to calculate marginal posterior probabilities using the GTR + I + Γ model of nucleotide substitution for each partition. The substitution and branch length estimates were allowed to vary independently between each partition. Although each partition shared the same overall topology, the substitution parameters and branch lengths were therefore estimated independently allowing the evolution of each partition to be modelled separately. Model parameter values were treated as unknown and were estimated in each analysis. Random starting trees were used and analyses were run between 1 and 60 million generations, sampling the Markov chain every 1000 generations.

Three methods were used to ensure that analyses were not trapped in local optima: (1) analysis was performed twice, starting from a different random tree. Log-likelihood values at stationarity were compared for convergence (Huelsenbeck and Bolback, 2001); (2) the topologies and clade posterior probabilities from each of the two analyses were compared for congruence (Huelsenbeck and Imennov, 2002); (3) Metropolis-coupled Markov chain Monte Carlo (MCMCMC) was used, with one cold and three incrementally heated Markov chains run simultaneously (default MrBayes heating values) to allow a more extensive exploration of parameter space (Huelsenbeck and Ronquist, 2001). Tracer v1.0.1 (Rambaut and Drummond, 2003) was used to ensure that the Markov chains had reached stationarity and to also determine the correct 'burn-in' for the analysis (i.e., the number of initial generations that must be discarded before stationarity is reached).

The program 'reducetrees' (Drummond, 2004) was used to compare the posterior probabilities of the various higher order taxonomic classifications. 'Reducetrees' assigns higher level classifications (e.g., Myopsida, Sepiidae) to each species and then sum together all the topologies (from the .t file from MrBayes) that share the same higher level topology, thereby calculating posterior probabilities for each of the higher level relationships sampled.

2.5. Evaluation of the saturation of nucleotide substitutions

The inferred number of substitutions between each pair of sequences was estimated from the Bayesian topology (GTR + I + Γ) as the sum of the lengths of the branches joining these two sequences. Saturation was estimated by plotting the uncorrected *p* value (mean distance) of each pair of sequences as a function of the number of substitutions inferred for all comparisons for 35 sequences (Mindell and Honeycutt, 1990). Saturation is present when the data points plateau.

2.6. Model testing

The Akaike Information Criterion (AIC) (Akaike, 1973) was used to test between different substitution models. This method of model selection was selected as it simultaneously allows the comparison of multiple nested models and accounts for model selection uncertainty (Posada and Buckley, 2004) and can also validly use the estimated marginal likelihood (EML) as a starting value. The EML value of an analysis is penalised by the number of parameters in the model. The smaller the AIC value, the better the fit of the model to the data.

3. Results

Sequences generated in this study were deposited in GenBank under Accession Nos. AY545077–AY545194, AY616863–AY616941. Alignment of ODH, *pax-6*, and COI sequences required no insertion/deletion events (indels). Indels were introduced into aligned sequences of 12S rDNA, 16S rDNA, and *rhodopsin*. Highly variable loop regions within 12S rDNA (168 bp in total) and 16S rDNA (240 bp in total) that were unalignable were removed prior to analysis.

Gene duplications of the ODH were identified within two species of Oegopsida (*Pterygioteuthis microlampas*, *Joubiniteuthis* sp.) neither of which contained stop codons, however only a fragment of ODH was sequenced and therefore it is not known whether start and stop codons were present in sequences. Three sequences of *P. microlampas* were recovered which are 12% different (maximally) from one another, and two sequences of *Joubiniteuthis* sp. which are 3% different from one another. For both species, 64% of the sequence differences are transitions and 36% are transversions. Additional primers were developed to attempt to identify further duplications within other species, but none were found. Bayesian analysis of the multiple ODH sequences shows that the gene duplications clustered together within their species groups (Supplementary Appendix 1) and are closer to one another than any other species. Consensus sequences were formulated for the species for use in phylogenetic analyses.

3.1. Evolutionary properties of the six genes

Striking differences in base composition are evident between mitochondrial and nuclear genes. Each of the three nuclear genes has a similar proportion of G + C

to A + T (Table 1) whereas the three mitochondrial genes are all A + T rich. The four protein coding genes all show the expected pattern of sequence variability between codon positions, with third positions containing the highest number of variable sites (ranging from 85% in *rhodopsin* to 96% in ODH and *pax-6*) (Table 1). Second positions are least variable (ranging from 3% in *pax-6* to 31% in *rhodopsin*).

Plots of Bayesian estimated GTR + I + Γ parameters provide a clear means of comparing the evolutionary dynamics of the six genes (Supplementary Appendix 5). The three mitochondrial genes each show a much greater mean number of transitions than transversions (Supplementary Appendix 5). Each of the mitochondrial genes also demonstrate relatively strong rate heterogeneity between nucleotide sites, which generates low values of the Γ-shape parameter. For 12S rDNA and 16S rDNA this may be due to differing functional constraints on stem and loop regions (Springer and Douzery, 1996). Each of the mitochondrial genes also possess a notably larger mean tree length than each of the nuclear genes (Supplementary Appendix 5), indicating greater evolutionary change than in nuclear genes.

The parameters for each of the three nuclear genes are very similar (Supplementary Appendix 5). *Rhodopsin* has the lowest average substitution rate, although *pax-6* and ODH are only marginally greater (Supplementary Appendix 5).

χ^2 homogeneity tests of each of the genes shows that ODH and COI have significant base frequency heterogeneity even after Bonferroni correction for multiple tests (Table 1). RY-coding the third positions of these genes was used to resolve base composition heterogeneity (Table 1). RY coding pools purines (adenine and guanine: R) and pyrimidines (cytosine and thymine: Y) into two-state categories (R,Y) and helps resolve bias resulting from differences in the relative frequency of either the two purines or pyrimidines (Phillips et al., 2001).

3.2. Nucleotide substitution saturation analyses

Saturation plots were generated for the third codon positions of each protein-coding gene and for the whole rDNA genes (Supplementary Appendix 2) Some substitution saturation is evident within each of the six genes with third codon positions of *rhodopsin* appearing the least saturated.

3.3. Phylogenetic results

3.3.1. Results from individual genes

No two topologies obtained from the Bayesian analyses of each of the six individual genes are identical, with major differences present in the branching order of the higher level Decapodiform and Octopodiform taxa (Supplementary Appendix 1). The topologies resulting from analysis of the three mitochondrial genes contain numerous polytomies. Topological incongruence between the individual genes was evaluated by crossed SH tests (Shimodaira and Hasegawa, 1999) whereby the consensus topologies obtained from individual datasets were compared against one another (Supplementary

Table 1
Nucleotide composition and character information for each gene fragment

Gene	Codon position	A	C	G	T	$\chi^2(p)$	Total	Variable	Parsimony informative
ODH	1	0.30	0.20	0.29	0.20	1.000	261	107 (41%)	84
	2	0.30	0.23	0.16	0.31	0.999	261	71 (27%)	47
	3	0.18	0.30	0.20	0.32	**0.000**	260	249 (96%)	244
	All	0.26	0.24	0.22	0.27	**0.000**	783	428 (55%)	375
	All 3rds RY	0.27	0.24	0.21	0.28	0.833	783	333 (43%)	259
pax-6	1	0.27	0.24	0.33	0.17	1.000	98	18 (18%)	12
	2	0.22	0.22	0.29	0.27	0.999	98	3 (3%)	2
	3	0.27	0.27	0.25	0.20	0.879	99	95 (96%)	91
	All	0.25	0.24	0.29	0.21	0.994	296	116 (39%)	105
Rhodopsin	1	0.26	0.21	0.31	0.23	0.999	318	132 (42%)	92
	2	0.28	0.29	0.14	0.29	1.000	318	98 (31%)	62
	3	0.25	0.30	0.21	0.24	**0.032**	317	271 (85%)	251
	All	0.26	0.27	0.23	0.24	0.534	954	502 (53%)	405
COI	1	0.27	0.19	0.28	0.26	0.336	202	71 (35%)	57
	2	0.14	0.26	0.17	0.42	1.000	202	28 (14%)	8
	3	0.44	0.13	0.04	0.39	**0.000**	202	198 (98%)	188
	All	0.29	0.19	0.16	0.36	**0.000**	606	299 (49%)	253
	All 3rds RY	0.22	0.22	0.24	0.33	0.999	606	238 (39%)	199
12S rDNA		0.41	0.18	0.09	0.33	0.959	347	221 (64%)	170
16S rDNA		0.32	0.11	0.21	0.36	1.000	430	173 (40%)	113

Tests of base homogeneity (χ^2) across taxa were based on variable sites only. $\chi^2(p) < 0.05$ are in bold.

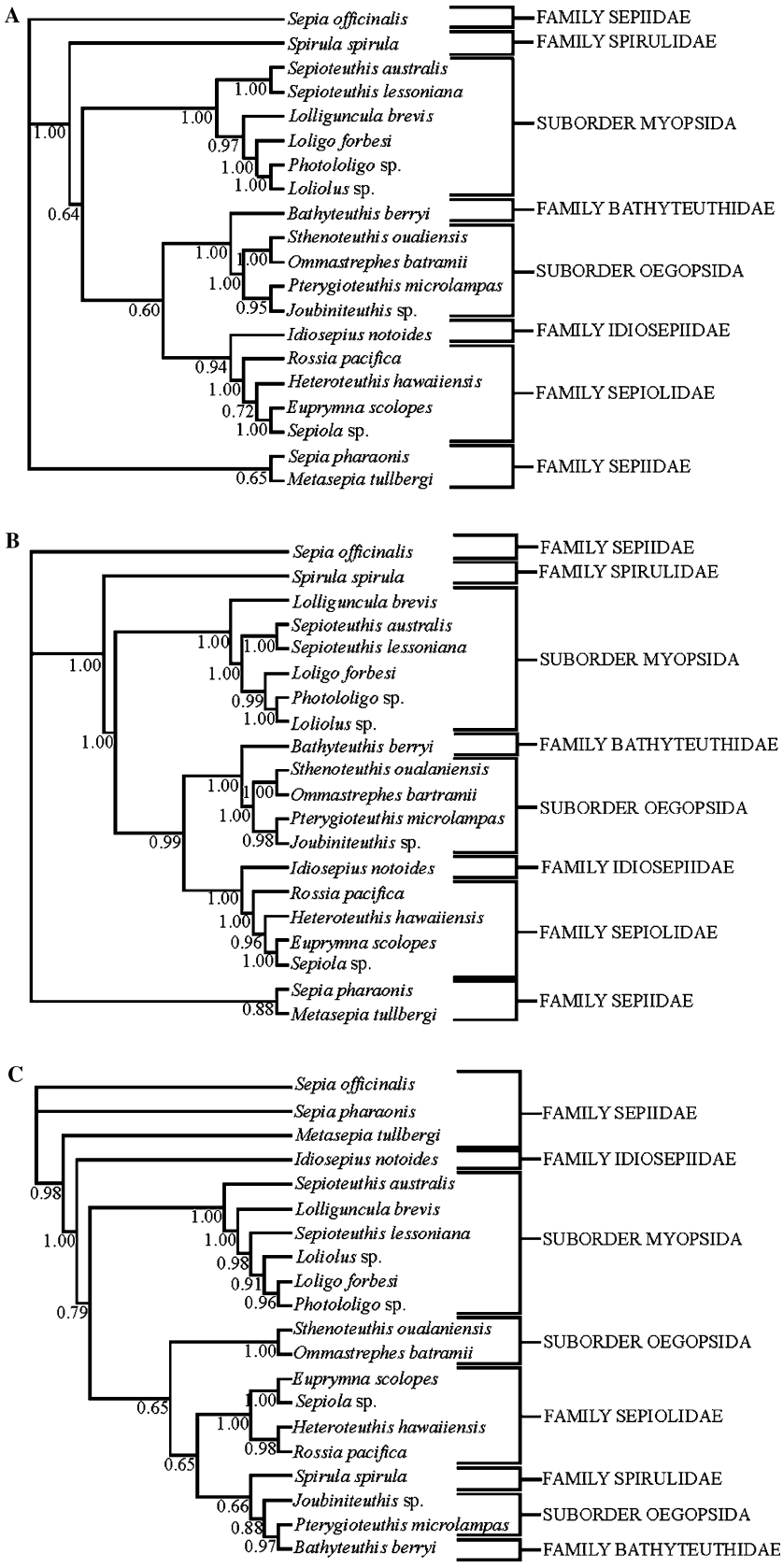

Fig. 2. Bayesian topologies of Decapodiform relationships obtained using the best fitting models for (A) all six genes (partitioned by codon) (6: codon) (B) three nuclear genes (partitioned by gene and codon) (nuc: g + c) and (C) three mitochondrial genes (partitioned by codon) (mt: codon) using the GTR + I + Γ model. Bayesian posterior probabilities are indicated beneath each node.

Appendix 4). In the majority of cases each individual dataset significantly rejects the Bayesian consensus topology of the other genes.

3.3.2. Decapodiformes—all genes

The optimal model for the dataset of all six genes for the 20 Decapodiform species is that partitioned by codon (6: codon) (Fig. 2A, Table 2). The suborders Myopsida and Oegopsida and families Sepiidae and Sepiolidae are each monophyletic clades and each is supported by a posterior probability (PP) of 1.00 regardless of the model used. The family Idiosepiidae is the sister taxa to the Sepiolidae (PP = 0.94) and together this clade forms the sister taxa to the Oegopsida (PP = 0.60). The family Spirulidae is basal to the Sepiidae (PP = 1.00) and together this clade is adjacent to the Myopsida (PP = 0.64). This topology has a posterior probability of 0.36 (Table 2). A slightly different topology is recovered when models are partitioned by gene (6: gene) only or by gene + codon (6: g + c) (tree not shown). In this topology Spirulidae is basal to the Oegopsida.

3.3.3. Decapodiformes—nuclear genes

Each of the four dataset partition models used to analyse the nuclear genes of the 20 Decapodiform spe-

Table 2
Estimated marginal likelihood values (arithmetic mean) and calculated Akaike Information Criterion (AIC) values for each model

	Analysis	EML	Penalty	AIC	PP of consensus topology
Decapodiformes	6: nopart	−21001.38	27	42056.76	0.72
	6: gene	−18071.47	162	36466.94	0.5
	6: codon	**−18112.49**	**108**	**36440.98**	**0.36**
	6: g + c	−17916.5	378	36589.00	0.30
	Nuc: nopart	−10571.88	27	21197.76	0.48
	Nuc: gene	−10424.02	81	21010.04	0.54
	Nuc: codon	−10203.67	81	20569.34	0.90
	Nuc: g + c	**−10032.63**	**243**	**20551.26**	**0.81**
	Mt: nopart	−9249.04	27	18552.08	0.26
	Mt: gene	−7613.24	81	15388.48	0.28
	Mt: codon	**−7485.06**	**81**	**15186.12**	**0.28**
	Mt: g + c	−7393.46	243	15272.92	0.22
Octopodiformes	6: nopart	−16826.45	22	33696.90	1.00
	6: gene	−15025.12	132	30314.24	0.98
	6: codon	−14952.24	88	30080.48	0.55
	6: g + c	**−14430.50**	**308**	**29477.00**	**0.89**
	Nuc: nopart	−8372.07	22	16788.14	0.86
	Nuc: gene	−8214.99	66	16561.98	0.81
	Nuc: codon	−7981.35	66	16094.70	0.63
	Nuc: g + c	**−7826.04**	**198**	**16048.08**	**0.29**
	Mt: nopart	−7027.90	22	14099.80	0.38
	Mt: gene	−6794.92	66	13721.84	0.82
	Mt: codon	**−6709.96**	**66**	**13551.92**	**0.54**
	Mt: g + c	**−6580.71**	**198**	**13557.42**	**0.60**
All taxa	6: nopart	−31233.23	77	62620.46	0.21 (D) 0.99 (O)
	6: gene	−29787.39	462	60498.78	0.55 (D) 1.00 (O)
	6: codon	−30075.21	308	60766.42	0.75 (D) 1.00 (O)
	6: g + c	**−28942.81**	**1078**	**60041.62**	**0.99 (D) 1.00 (O)**
	Nuc: nopart	−17270.97	77	34695.94	0.42 (D) 0.65 (O)
	Nuc: gene	−16909.88	231	34281.76	0.82 (D) 0.77 (O)
	Nuc: codon	**−16673.43**	**231**	**33808.86**	**0.81 (D) 0.49 (O)**
	Nuc: g + c	−16293.57	693	33973.14	0.82 (D) 0.54 (O)
	Mt: nopart	−13218.41	77	26590.82	0.21 (D) 0.35 (O)
	Mt: gene	−12809.58	231	26081.16	0.26 (D) 0.90 (O)
	Mt: codon	−12745.52	308	26107.04	0.25 (D) 0.40 (O)
	Mt: g + c	**−12557.01**	**385**	**25884.02**	**0.15 (D) 0.49 (O)**
	Nuc12: nopart	**−7950.11**	**77**	**16054.22**	**0.51 (D) 0.14 (O)**
	Nuc12: gene	−7828.78	231	16119.56	0.60 (D) 0.84 (O)
	Nuc12: codon	−7893.80	154	16095.60	0.51 (D) 0.13 (O)
	Nuc3: gene	**−11984.43**	**231**	**24430.86**	**0.21 (D) 0.29 (O)**
	Nuc3: nopart	−12317.26	77	24788.92	0.30 (D) 0.34 (O)

Posterior probabilities (PP) are shown for the higher level topologies. The best fitting model for each dataset is shown in bold. The codes for each of the datasets are as follows: 6, all six genes; nuc, nuclear genes (all three codon positions); nuc12, 1st + 2nd codon positions of nuclear genes; nuc3, 3rd codon positions of nuclear genes. The partitioning models are indicated as follows: no part, no partitioning; gene, partitioned by gene; codon, partitioned by codon/rDNA; and g + c, partitioned by gene + codon/rDNA.

cies resulted in identical topologies with the nuc: g + c model providing the best fit to the data (Fig. 2B, Table 2). This topology is the same as the 6: codon model (described directly above, apart from the arrangement of *Loligunculla* and *Sepioteuthis* within the Myopsida), although the former topology has a notably larger posterior probability (0.81) comparing the posterior probabilities of the first and second highest topologies (Table 2). All of the posterior probabilities on nuc: g + c topology are 0.88 or higher (Fig. 2B), which are notably larger than those on the same topology resulting from the 6: codon analysis.

3.3.4. Decapodiformes—mitochondrial genes

The best fitting model for the three mitochondrial genes for the 20 Decapodiform species is mt: codon (Fig. 2C, Table 2). The topology resulting from this analysis differed from that obtained from the 6: codon and nuc: g + c models (described above). The Oegopsida are polyphyletic in the topology resulting from mt: codon analysis. This configuration of higher level relationships has a posterior probability of 0.28 (Table 2) which is less than half of those obtained by nuc: g + c analysis of the Decapodiform taxa.

As the ancestral Decapodiform is unknown, trees were not able to be rooted in these analyses. Midpoint rooting is not particularly useful since each of the above topologies generated different rooting positions depending upon the method of partitioning used.

3.3.5. Octopodiformes—all genes

The Octopodiform topologies are rooted with *V. infernalis*, as morphological and molecular studies have confirmed the order Vampyromorpha as the sister taxa to the Octopoda (Carlini et al., 2000; Young and Vecchione, 1996). The best fitting model for the dataset of all six genes of the 15 Octopodiform species is 6: g + c (Fig. 3A, Table 2). The suborders Cirrata and Incirrata are sister taxa (PP = 1.00) and the superfamily Argonautoidea is basal within the Incirrata (PP = 1.00). The rest of the Incirrata is divided into two clades (PP = 0.88), one containing Octopodinae species (apart from *Enteroctopus*) (PP = 1.00) and the other containing *Benthoctopus* and *Enteroctopus* which together form the sister taxa to a clade containing *Pareledone*, *Vitreledonella*, and *Japetella* (P = 0.88) (Fig. 3A). The topology resulting from the 6: g + c model has a posterior probability of 0.887 (Table 2).

3.3.6. Octopodiformes—nuclear genes

The best fitting model for the nuclear genes of the 15 Octopodiform species is nuc: g + c (Fig. 3B, Table 2). The resultant topology is very similar to that described above for the 6: g + c topology for Octopodiformes, except that less resolution is observed within the Incirrata. The nuc: g + c topology has a notably lower posterior probability than that for 6: g + c (0.29).

3.3.7. Octopodiformes—mitochondrial genes

Partitioning the Octopodiform mitochondrial genes by codon (mt: codon) and by gene + codon (mt: g + c) produced similar fits (Table 2, Fig. 3C). The topologies resulting from these models are identical and have posterior probabilities of 0.54 and 0.60, respectively (Table 2). This topology is the same as that generated with all genes for Octopodiformes except that the superfamily Argonautoidea do not fall as a monophyletic group; *Tremoctopus* is basal within the Octopodiformes (PP = 1.00) and *Argonauta* groups within the Octopodinae (PP = 0.76).

3.3.8. All genes—all taxa

The best fitting model for partitioning the dataset containing all genes and all taxa is 6: g + c (Fig. 4A, Table 2). Six Decapodiform clades are joined by a polytomy indicating very little signal at higher level relationships for this dataset. However, a sister taxa relationship between the families Sepiidae and Spirulidae is resolved in this topology. The Octopodiform part of the topology is identical for each of the models, and is very highly supported (Fig. 4A, Table 2). This topology is identical to that resulting from analysis of all genes for only the Octopodiform taxa (Fig. 3A). Each bipartition within the Octopodiformes has a posterior probability of 1.00. Furthermore the arrangement of higher level relationships has an estimated posterior probability of 1.00 (Table 2).

3.3.9. All taxa—nuclear genes

Nuc: codon produced the best fit for the nuclear gene dataset for all taxa (Fig. 4B, Table 2). The Decapodiform part of this topology is identical to the 6: codon (Fig. 2A) and nuc: g + c (Fig. 2B) models for only Decapodiformes sequences. The nuc: codon analysis of Octopodiform and Decapodiform taxa together place the Oegopsida basal within the Decapodiformes (PP = 1.00) (Fig. 4B). The nuc: codon arrangement of higher level relationships has a posterior probability of 0.81 (Table 2).

In the Octopodiform part of the topology resulting from the nuc: codon for all taxa, a clade containing *Vitreledonella* and *Japetella* is the next most basal incirrate group after the Argonautoidea (PP = 1.00) (Fig. 4B). The higher level relationships of this topology have a posterior probability of 0.49 (Table 2). This placement of *Vitreledonella*/*Japetella* differs notably from each of the previous topologies where *Vitreledonella*/*Japetella* group as the sister taxa to *Pareledone* (Figs. 3A–C, 4A). It is notable that this latter configuration was also recovered when the nuclear gene dataset for all taxa is partitioned by gene + codon.

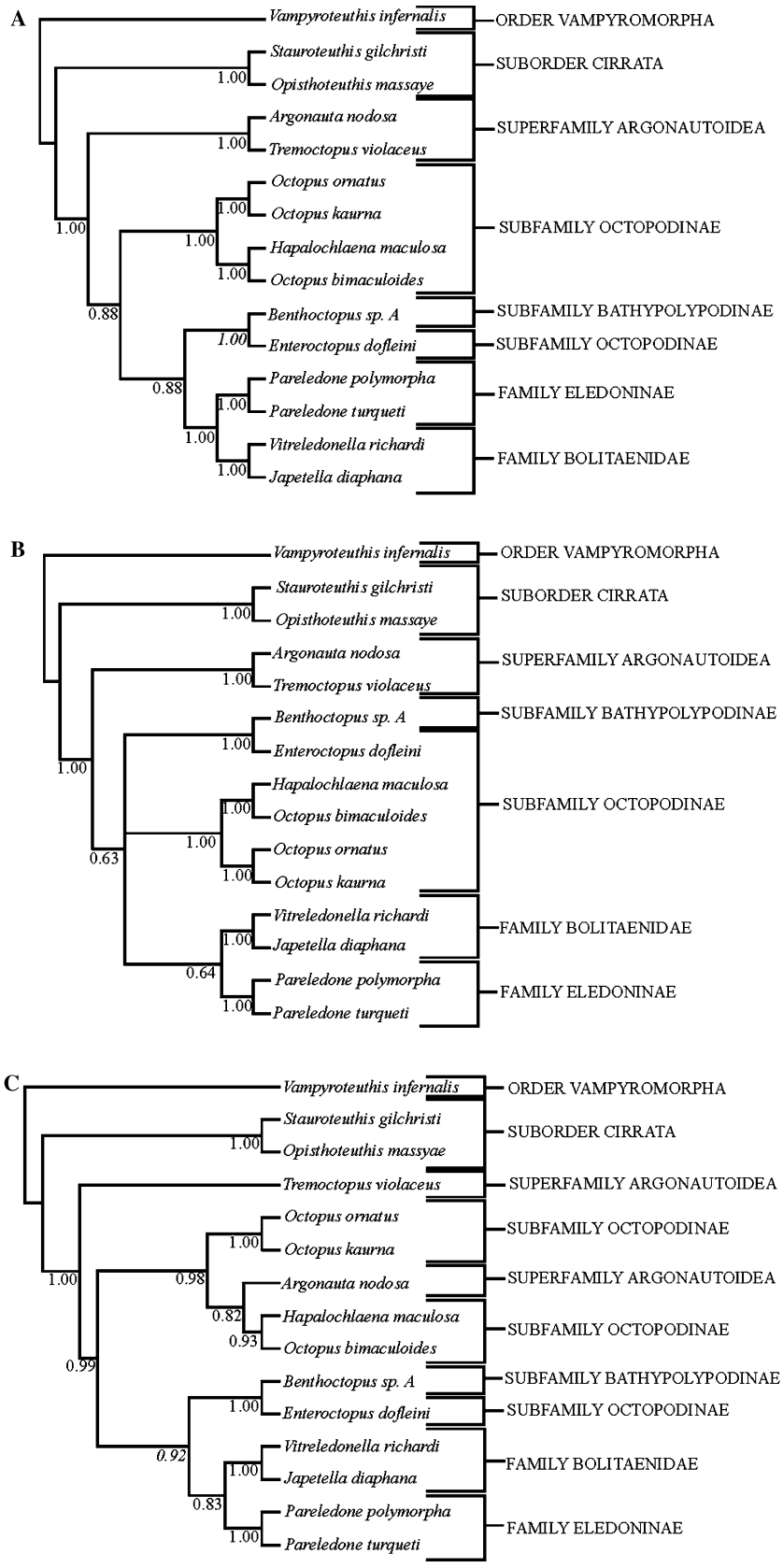

Fig. 3. Bayesian topologies of Octopodiform relationships obtained using the best fitting models for (A) all six genes (partitioned by gene + codon) (6: g + c) (B) three nuclear genes (partitioned by gene + codon) (nuc: g + c), and (C) three mitochondrial genes (partitioned by codon) (mt: codon) using the GTR + I + Γ model. Bayesian posterior probabilities are indicated beneath each node.

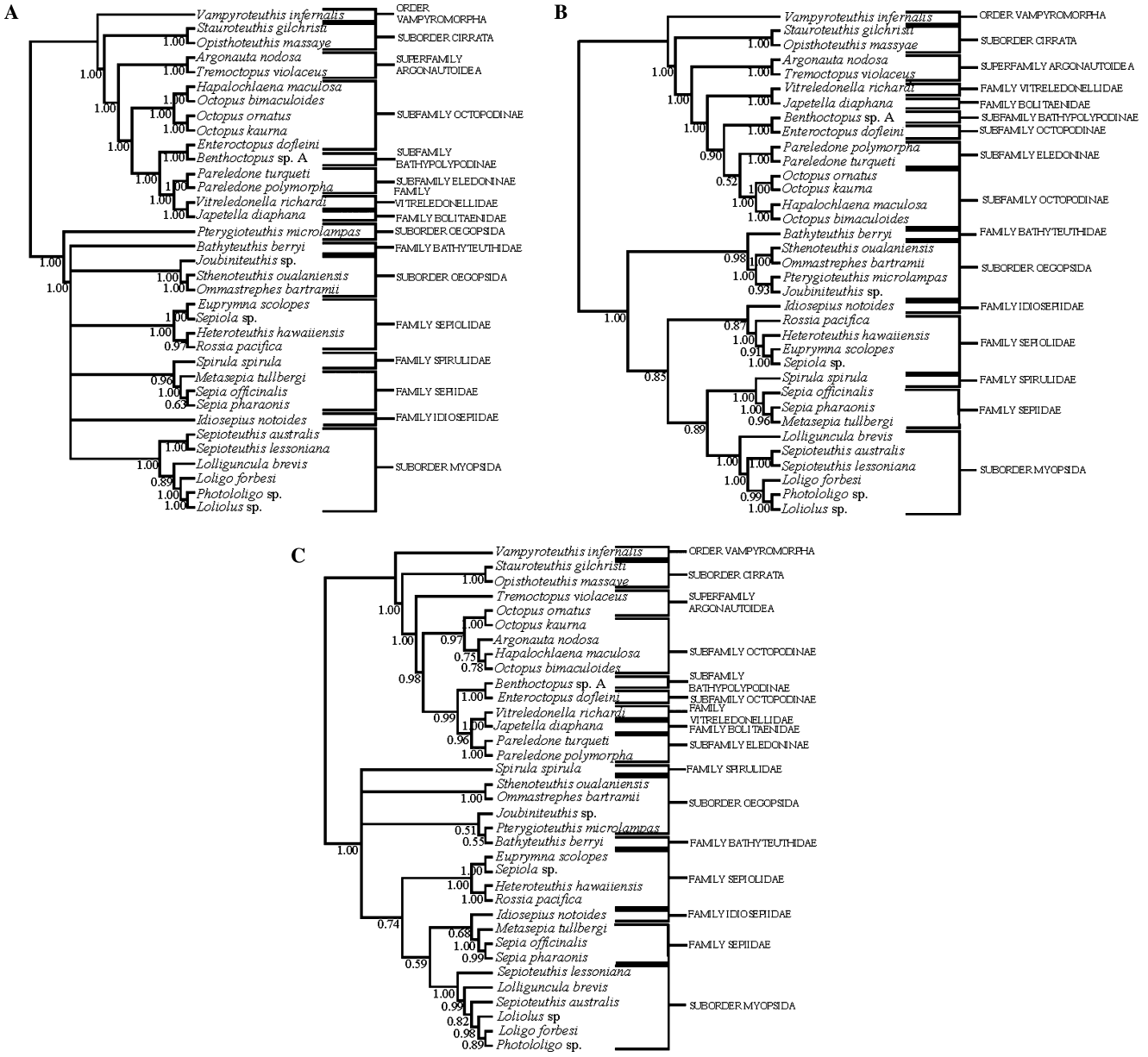

Fig. 4. Bayesian topologies of coleoid relationships obtained using the best fitting models for (A) all six genes (partitioned by gene and codon) (6: g + c) (B) three nuclear genes (partitioned by codon) (nuc: codon), and (C) three mitochondrial genes (partitioned by gene and codon) (mt: g + c) using the GTR + I + Γ model. Bayesian posterior probabilities are indicated at each node.

3.3.10. All taxa—mitochondrial genes

The best fitting model for partitioning the mitochondrial gene dataset for all taxa is mt: g + c (Table 2, Fig. 4C). The Decapodiform part of this topology is characterised by very short internal branches and higher level relationships with a posterior probability of 0.15 (Table 2, Fig. 4C). The Octopodiform part of this topology shows much greater resolution, and the same topology resulted from each of the models. This topology is identical to that resulting from the analysis of mitochondrial genes for the Octopodiformes only (Fig. 3C). The higher level relationships of this topology has a posterior probability of 0.49.

3.3.11. Nuclear 1st + 2nd codon positions

It is relatively common practice to either downweight or remove third positions in studies investigating phylogenies with deep divergences (Huchon et al., 2002). Since the divergence of the Octopodiformes and Decapodiformes occurred at least in the Carboniferous (290–354 mya) (Kluessendorf and Doyle, 2000), it was decided to analyse the slower evolving nuclear dataset using only 1st + 2nd codon positions to investigate the effect of this on the topology, particularly with respect to the position of the root to the Decapodiformes.

The rooting position and topology of the Decapodiformes is identical for all of the models for the dataset

of 1st + 2nd codon positions (Table 2, Fig. 5A) with the model without partitioning providing the best fit to the dataset (nuc12: nopart) (Table 2). Each of the higher level taxa are monophyletic for the nuc: nopart analysis, Oegopsida (PP = 1.00), Myopsida (PP = 1.00), Sepiidae (BP = 1.00), and Sepiolidae (PP = 1.00). The higher level relationships of these taxa differed from each of the topologies described above. In this topology the Idiosepiidae is the most basal taxon within the Decapodiformes (PP = 1.00) followed by the Oegopsida (PP = 0.68).

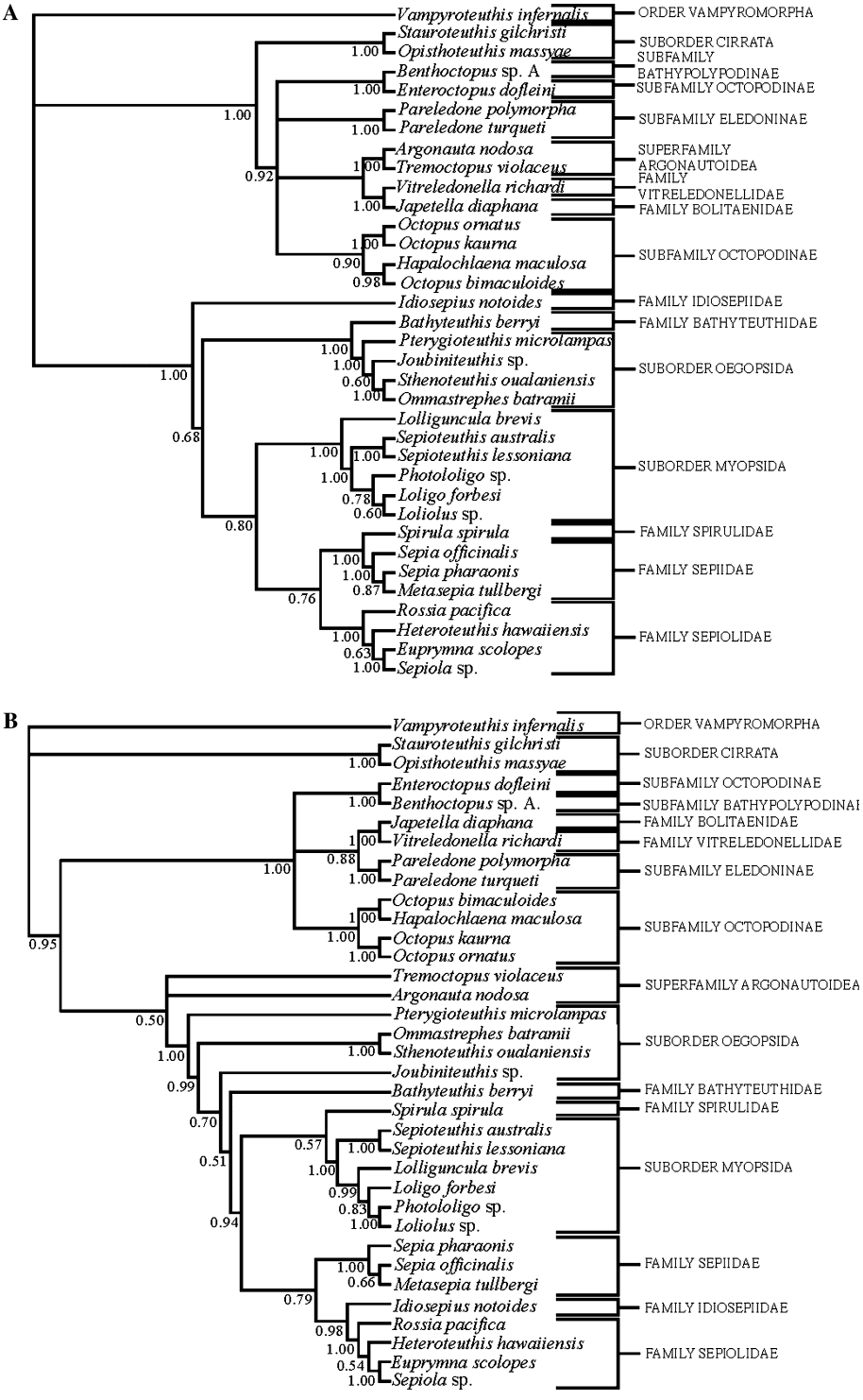

Fig. 5. Bayesian topologies of coleoid relationships obtained using the best fitting models for the three nuclear genes only using the GTR + I + Γ model. (A) 1st + 2nd codon positions only (no partition) (nuc12: nopart) and (B) 3rd positions only (by gene) (nuc3: gene). Bayesian posterior probabilities are indicated at each node.

The Spirulidae and the Sepiidae are sister taxa in the most distal part of the tree (PP = 1.00) and together this clade is the sister taxa to the Sepiolidae (PP = 0.76). This higher level topology has a posterior probability of 0.51 in the analysis. Interestingly, this topology is the second most sampled topology (PP = 0.06) in the nuc: codon analysis (all three codon positions) for all taxa (Table 2). Little resolution was evident within the Octopodiform part of the topology when using 1st and 2nd codon positions only. A large polytomy joined 5 clades within the Incirrata (Fig. 5A).

3.3.12. Nuclear genes 3rd positions only

To further investigate the role of the 3rd positions, particularly with respect to the rooting position of the Decapodiformes, Bayesian analyses were performed on the 3rd positions of nuclear genes only (Fig. 5B). Nuc3: gene is the best fitting model (Table 2). Groupings of higher level taxa are again supported, although the Oegopsida are a notable exception. The most obvious difference between this and other topologies is the rooting position of the Decapodiformes. The Decapodiformes fall within the Octopodiformes and are most closely related to the Argonautoidea. This result is unlikely to be correct as a number of previous morphological (Young and Vecchione, 1996) and molecular (Carlini et al., 2000) studies suggest a sister taxon relationship between the Decapodiformes and Octopodiformes.

3.3.13. Investigation of the effect of base composition of 3rd positions on nuclear genes

The potential effect of base composition bias at 3rd positions of nuclear genes was investigated to determine whether this was driving the conflicting signals of the rooting of the Decapodiformes and the placement of the Idiosepiidae. Non-stationary base composition can cause species to group by compositional similarity rather than evolutionary history (Lockhart et al., 1994). To test for non-stationary base composition two approaches were taken. First, a tree was constructed using only base composition of the 3rd positions of nuclear genes. Second, analysis was also performed based on the LogDet (LD) distance correction (Lockhart et al., 1994) which allows for variation in rate and composition between branches.

The tree constructed using only base composition of the 3rd positions of nuclear genes shows *Idiosepius* to fall within the four Sepiolidae taxa (Supplementary Appendix 6A) suggesting that base composition heterogeneity is playing a role in the placement of this group. However, LogDet analysis of the 3rd position sites of the nuclear genes shows *Idiosepius* to fall between the Oegopsida and the Sepiolidae (Supplementary Appendix 6B). This positioning of *Idiosepius* next to the Sepiolidae suggests this grouping is not based on base composition alone.

4. Discussion

4.1. Data partitioning

For the majority of the analyses in the present study, the selected model is partitioning the data by gene together with codon position/rDNA. This suggests that the differences in the substitution process between each codon position/rDNA and gene are large enough that providing a separate partition improves the fit of the model to the data. However, caution should be taken when increasing the number of partitions in a model, as this decreases the amount of sequence data in each partition causing an increase in the variance of the substitution parameters. When the amount of data in a partition becomes too small, the prior distribution of a substitution model in a Bayesian approach can begin to affect the posterior distribution, as insufficient data are present in the partition for the parameters to be effectively estimated. This point was investigated more thoroughly by Nylander et al. (2004) who reported that more complex substitution models are associated with more topological uncertainty than are simple models.

In some cases, partitioning by codon/rDNA provides the best fit to the dataset (e.g., for Decapodiform taxa, all: codon, mt: codon, and also for all taxa, nuc: codon). This reflects the similarity in the substitution process between each codon position across different genes. This finding is also supported by Nylander et al. (2004) who partitioned nucleotide data by gene only (not by codon) and found that allowing within partition variation was by far the most important model component. They found that allowing rate variation across molecular partitions but not within them was far less successful than allowing rate variation within a single partition.

It is well understood that 3rd codon positions contain the highest number of variable sites, whilst second positions contain the lowest number. Partitioning sequence data by codon position groups together nucleotide positions with similar rates, effectively performing some of the function of the gamma distributed model of rate heterogeneity among sites. Therefore in a codon partitioned analysis, each partition (already containing data with similar rates) will be modelled by four rate categories. This is reflected in the significantly better fit of codon partitioned to gene partitioned models in 80% of comparisons in this study. In datasets that are partitioned by gene, the rate categories are largely modelling the differences between each of the codon positions.

The method by which sequence data are partitioned affects the overall topology for some datasets. Examples include the placement of *Japetella* and *Vitreledonella* and the rooting of the Decapodiformes, (discussed below). This shows the degree to which different codon

positions affect the resulting topology is dependent upon the method by which the data are partitioned.

4.2. Mitochondrial vs nuclear data

Comparison of the saturation plots, posterior probabilites, and 'reducetrees' percentages of nuclear and mitochondrial genes suggests that overall, the mitochondrial genes possess greater levels of saturation than nuclear genes. However, the levels of saturation appear to differ between the Octopodiformes and Decapodiformes. High posterior probabilities, 'reducetrees' values and high levels of congruence are present between the topologies for the Octopodiform taxa only using all genes, mitochondrial genes and nuclear genes (with the exception of the position of *Vitreledonella*/*Japetella* and the placement of *Argonauta*). Low levels of resolution within the Octopodiformes are obtained from analysis of 1st + 2nd nuclear protein coding genes only, with a more highly resolved topology obtained from analysis of 3rd codon positions. In contrast, the highest 'reducetrees' values and highest resolution within the Decapodiform taxa only results from analysis of nuclear genes alone. Lower levels of resolution are evident in the Decapodiform topology through the analysis of mitochondrial and nuclear genes together. Support for the 3rd positions within Decapodiformes being saturated is evident in the conflict in the Decapodiform topology between the 1st + 2nd codon positions and 3rd codon positions.

This difference in the ability of mitochondrial and nuclear data and also codon position to resolve Octopodiform and Decapodiform topologies suggests that the Decapodiformes either radiated earlier or have faster rates of evolution than the Octopodiformes.

4.3. Octopine dehydrogenase gene duplications

The apparent 'gene duplications' present within *P. microlampas* and *Joubinteuthis* sp. are difficult to explain considering that the sequences from each species always fell together as a monophyletic group. It is possible that some of the 'duplications' are pseudogenes, although in the short segment sequenced each of the genes appear to be coding. It is implausible that each of the different forms are alleles because of the large genetic distances. A gene duplication event(s) may have occurred prior to the divergence of these two taxa and that the genes evolved in a concerted manner, however there is no evidence at this stage to support this proposal.

4.4. Cephalopod phylogeny

4.4.1. Family spirulidae

The family Spirulidae is represented by just one living species, the Ram's Horn Squid, *Spirula spirula*. This small pelagic squid is the only cephalopod that possesses an internal coiled and chambered shell. These features have obscured the phylogenetic relationships of this taxon. Morphological studies have proposed different placements of this squid (see Donovan, 1977; Engeser and Bandel, 1988; Naef, 1912). In a molecular study using NJ analysis and a relatively small number of taxa Bonnaud et al. (1997) found the Spirulidae to fall with the Sepiidae and/or the Oegopsida although the topology varied depending upon the type of data analysed (i.e., all substitutions, only transversions or amino acids). Using a portion of the COI gene, Carlini et al. (2000) reported the Spirulidae to fall with Sepiidae or oegopsid species but without any bootstrap support. Warnke et al. (2002) found little resolution using 18S rDNA, and the placement of Spirulidae differed according to the methods of analysis.

The present study shows strong support for a sister taxa relationship between the Spirulidae and the Sepiidae. This relationship is supported by a number of morphological characteristics, including the possession of a phragmocone, the structure of the tentacular clubs, sperm storage (spermatheca) in females below the mouth, statolith shape (Clarke and Maddock, 1988; see Young et al., 1998 for a more complete list).

4.4.2. Suborder Myopsida

On the basis of a similar gladii and tentacular clubs Naef (1916, 1921–1923) placed the suborders Myopsida and Oegopsida together in the order Teuthoidea. However, the Myopsida have also been suggested to be derived from the 'Sepioidea' line (Berthold and Engeser, 1987; d'Orbigny, 1845; Engeser, 1997; Haas, 1997). Previous molecular work has found little support for the placement of the Myopsida (Bonnaud et al., 1996, 1997; Boucher-Rodoni and Bonnaud, 1996; Carlini et al., 2000).

Two possible alternative placements of the Myopsida are suggested from this study. Analysis of the nuc: codon model for all taxa place the Myopsida as the sister taxa to a clade containing the Spirulidae and Sepiidae whilst nuc12: nopart analysis results in the Myopsida falling as the sister taxa to a clade containing Sepiolidae, Spirulidae, and Sepiidae. This grouping of Sepiolidae, Spirulidae, and Sepiidae is reminiscent of the traditional taxonomic grouping 'Sepioidea' (Naef, 1916; Voss, 1977) which contain each of these taxa but also the Idiosepiidae and the Bottletail squid (Sepiadariidae) (not included in this study). If the nuc12: nopart topology proves correct, then the placement of Myopsida between this clade and the Oegopsida might help explain the conundrum outlined by Young et al. (1998), that the Myopsida appear 'intermediate between Sepioida and the Oegopsida.'

4.4.3. The nature of the suborder Oegopsida and the position of family Bathyteuthidae

Previous molecular studies have suggested the squid suborder Oegopsida to be a polyphyletic group with uncertain phylogenetic affinities (Bonnaud et al., 1994, 1997; Carlini and Graves, 1999). In a study using the gene actin, Carlini et al. (2000) found oegopsid taxa to fall basally within the Decapodiformes in parsimony and ML analyses. Similarly, when all three codon positions of nuclear genes were analysed in this study (except when partitioned by gene only), the Oegopsida (and Bathyteuthidae) fell basal within the Decapodiformes with strong support. The Oegopsida are strongly supported as being monophyletic in these analyses.

The phylogenetic position of the family Bathyteuthidae (Fig. 1L) has varied in past studies as they possess characteristics of both the Myopsida (Young et al., 1998; Carlini and Graves, 1999) and the Oegopsida (Naef, 1921–1923). Carlini and Graves (1999) found *Bathyteuthis* to fall within the Oegopsida although without support. Subsequently, Carlini et al. (2000) found *Bathyteuthis* to be the sister taxon to the Oegopsida in MP (but not ML) analyses of an actin I dataset, although with bootstrap support of only 55%. This study strongly supports a sister taxa relationship between the Bathyteuthidae and Oegopsida, supporting Naef's (1921–1923) supposition that the Bathyteuthidae possess 'primitive characters for all Oegopsida.'

4.4.4. Family Idiosepiidae

In contrast to the basal position of the Oegopsida in the Decapodiformes found above, nuc12: nopart analysis of all taxa shows the pygmy squids (family Idiosepiidae) to fall basally within the Decapodiformes, followed by the Oegopsida. Given the evidence for saturation of third positions within the Decapodiform nuclear data, a basal Idiosepiidae might be expected to be effected especially since it would have a long unbranched edge in the phylogeny. The basal position of Idiosepiidae may suggest that the absence of accessory nidamental glands and tentacles at hatching may be due to the relatively 'primitive' position of this taxon rather than 'secondary simplification' due to their small size, as has been previously suggested (Young et al., 1998). Furthermore, the discovery of a fully developed but rather 'primitive' gladius led Hylleberg and Nateewathana (1991a,b) to suggest that the Idiosepiidae may be more closely related to the order Teuthoidea than to the order Sepioidea as supported by the nuc12: nopart analysis.

Analyses of nuc: codon and nuc3: gene for all taxa show high levels of support for the Idiosepiidae as the sister taxa to the Sepiolidae. This relationship is supported by morphological studies by Fioroni (1981), Berthold and Engeser (1987), Clarke (1988), and Guerra (1992) which suggested the Idiosepiidae as a specialised group derived from a sepiolid ancestor. Previous molecular studies have also found the Idiosepiidae to fall with either the family Sepiolidae (Carlini and Graves, 1999; Carlini et al., 2000) or the order Oegopsida (Bonnaud et al., 1996, 1997; Carlini et al., 2000). Both alternatives are considered viable options, but unfortunately no members of the family Sepiadariidae were able to be included in this analysis. The Sepiadariidae have been suggested to be closely related to the Idiosepiidae and also the Sepiolidae (Berthold and Engeser, 1987; Fioroni, 1981) and future inclusion of this group in analyses should help clarify the position of the Idiosepiidae.

4.4.5. Superfamily Argonautoidea

The pelagic octopuses of the superfamily Argonautoidea have been a difficult group to place on the basis of their morphology. Naef (1921–1923) believed the Argonautoidea were derived from the bottom-living octopuses of the family Octopodidae. In contrast Voight (1997) placed the Argonautoidea with two other pelagic families; Amphitretidae and Vitreledonellidae. Carlini et al. (2001) reported *Argonauta* to fall basally within the Incirrata in an MP analysis of a COI data set, yet *Haliphron* and *Tremoctopus* grouped together in a much more derived position within the tree. However, none of these placements were well supported.

A number of strong morphological characteristics unite the members of the superfamily Argonautoidea (Naef, 1921–1923; Young et al., 1998). As a consequence, separate origins for these taxa are unlikely. In support of this, analyses of all genes and nuclear genes in the present study found strong support for a basal placement of the superfamily Argonautoidea within the Incirrata. Although analysis of mitochondrial genes alone found the Argonautoidea to be polyphyletic there is strong evidence of saturation within the mitochondrial genes at this level.

4.4.6. Families Vitreledonellidae and Bolitaenidae

The placement of Vitreledonellae/Bolitaenidae within the Octopodiformes is dependent upon the model and/or datasets used. The fact that the sister taxa relationship between Vitreledonellidae/Bolitaenidae and *Pareledone* is recovered in most topologies and also in analyses of mitochondrial and nuclear genes separately lends support to this relationship. Furthermore, a study by Strugnell et al. (2004) included a further three species of Octopodiform in their analyses and also recovered strong support for this sister taxa relationship. The alternative placement of Vitreledonellidae/Bolitaenidae as the next most basal group after the Argonautoidea in the Incirrata is recovered only in nuc: codon analysis for all taxa. However, when this dataset is analysed by 3rd positions alone (nuc3: gene), the sister taxa relationships of Vitreledonellidae/Bolitaenidae and *Pareledone* is again supported whilst nuc12: nopart analysis resulted in very little resolution within the Incirrata, with five

clades linked by a polytomy. Therefore, the majority of analyses support a sister taxa relationship between Vitreledonellidae/Bolitaenidae and *Pareledone*.

Note added in press

Lindgren et al. (2004) published a study investigating coleoid cephalopod relationships, using morphological and molecular data, with emphasis on Decapodiformes, after the present study was accepted. In contrast to our findings of a sister taxa relationship between Sepiidae and Spirulidae, they found Spirulidae fell within the Oegopsida in most analyses. Furthermore, a relationship between the Idiosepiidae, Sepiidae, Sepiolidae, and Loliginidae was supported in some of their analyses.

Acknowledgments

We thank the following people for their generosity in donating tissue samples: David Carlini, Martin Collins, Stephen Craig, Eileen Dillane, C.C. Lu, Richard Stride, Kerstin Warnke, and Richard Young. This research was supported by BBSRC (43/G16942) and by the Rhodes trust.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2005.03.020.

References

Akaike, H., 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest.

Berthold, T., Engeser, T., 1987. Phylogenetic analysis and systematization of the Cephalopoda (Mollusca). Ver. Naturwissenschaftliche Vereins Hamburg 29, 187–220.

Boletzky, S.V., 1992. Evolutionary aspects of development, life style, and reproductive mode in incirrate octopods (Mollusca, Cephalopoda). Rev. Zool. 99, 755–770.

Bonnaud, L., Boucher-Rodoni, R., Monnerot, M., 1994. Phylogeny of decapod Cephalopods based on partial 16S rDNA nucleotide sequences. C. R. Acad. Sci. Paris 317, 575–580.

Bonnaud, L., Boucher-Rodoni, R., Monnerot, M., 1996. Relationship of some coleoid cephalopods established by 3′ end of the 16S rRNA and cytochrome oxidase III gene sequence comparison. Am. Malac. Bull. 12, 87–90.

Bonnaud, L., Boucher-Rodoni, R., Monnerot, M., 1997. Phylogeny of cephalopods inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol., 44–54.

Boucher-Rodoni, R., Bonnaud, L., 1996. Biochemical and molecular approach to cephalopod phylogeny. Am. Malac. Bull. 12, 79–85.

Carlini, D.B., Graves, J.E., 1999. Phylogenetic analysis of cytochrome *c* oxidase I sequences to determine higher-level relationships within the coleoid cephalopods. Bull. Mar. Sci. 64, 57–76.

Carlini, D.B., Reece, K.S., Graves, J.E., 2000. Actin gene family evolution and the phylogeny of coleoid cephalopods (Mollusca: Cephalopoda). Mol. Biol. Evol. 17, 1353–1370.

Carlini, D.B., Young, R.E., Vecchione, M., 2001. A molecular phylogeny of the octopoda (Mollusca: Cephalopoda) Evaluated in light of morphological evidence. Mol. Phylogenet. Evol. 21, 388–397.

Clarke, M.R., 1988. Evolution of recent cephalopods—A brief review. In: Clarke, M.R., Trueman, E.R. (Eds.), The Mollusca. Paleontology and Neontology of Cephalopods, vol. 12. Academic Press, San Diego, p. 355.

Clarke, M.R., Maddock, L., 1988, Statoliths from living species of cephalopods and evolution. In: Clarke, M.R., Trueman, E.R., (Eds.), The Mollusca. Paleontology and Neontology of Cephalopods, vol. 12. Academic Press. San Diego.

d'Orbigny, A., 1845. Mollusques Vivant et Fossiles, Vol 2nd part (Paris).

DeBry, R.W., 2003. Identifying conflicting signal in a multigene analysis reveals a highly resolved tree: the phylogeny of rodentia (Mammalia). Syst. Biol. 52, 604–617.

Donovan, D.T., 1977. Evolution of the Dibranchiate Cephalopoda. In: Nixon, M., Messenger, J.B. (Eds.), The Biology of Cephalopods. Academic Press, London, pp. 15–46.

Drummond, A.J., 2004. Reducetrees. Oxford University. (unpublished).

Engeser, T., 1997. The fossil Coleoidea page. <http://userpagefu-berlinde/~palaeont/fosscolhtml>

Engeser, T., Bandel, K., 1988. Phylogenetic classification of coleoid cephalopods. In: Wiedman, J., Kullmann, J. (Eds.), Cephalopods—Present and Past. Schweizerbart'sche-Verlagsbuchhandlung, Stuttgart, pp. 105–116.

Fioroni, P., 1981. Die sonderstellung der sepioliden, ein vergleich der ordnungen der rezenten Cephalopoden. Zool. J. Biol. Syst. 108, 178–228.

Grimpe, G., 1921. Teuthologische Mitteilungen. VII: systematiche Ubersicht der Nordsee-cephalopoden. Zool. Anz. 52, 297–305.

Guerra, A., 1992. Mollusca, cephalopoda. In: Ramos, M.A. (Ed.), Fauna Iberica. Museo Nacional de Ciencias Naturales, CSIC, Madrid.

Haas, W., 1997. Der Ablauf der Entwicklungsgeschichte der Decabrachia (Cephalopoda, Coleoidea). Palaeontographica 245A, 63–81.

Hillis, D.M., Moitz, C., Mable, B. (Eds.), 1996. Molecular Systematics, second ed. Sinauer Associates, Sunderland, MA.

Huchon, D., Madsen, O., Sibbald, M.J.J.B., Ament, K., Stanhope, M.J., Catzeflis, F., DeJong, W.W., Douzery, J.P., 2002. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. Mol. Biol. Evol. 19, 1053–1065.

Huelsenbeck, J.P., Bolback, J.P., 2001. Empirical and hierchical Bayesian estimation of ancestral states. Syst. Biol. 50, 351–366.

Huelsenbeck, J.P., Bull, J.J., Cunningham, C.W., 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11, 152–158.

Huelsenbeck, J.P., Imennov, N.S., 2002. Geographic origin of mitochondrial DNA: accommodating phylogenetic uncertainty and model comparison. Syst. Biol., 51.

Huelsenbeck, J.P., Ronquist, F., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Hylleberg, J., Nateewathana, A., 1991a. Morphology, internal anatomy, and biometrics of the cephalopod Idiosepius pygmaeus Voss, 1962. A new record for the Andam sea. Phuket Marine Biological Center, Res. Bull. 56, 1–9.

Hylleberg, J., Nateewathana, A., 1991b. Redescription of *Idiosepius pygmaeus* Steenstrup, 1881 (Cephalopoda:Idiosepiidae), with men-

tion of additional morphological characters. Phuket Marine Biological Center, Res. Bull. 55, 33–42.

Kafatos, F.C., Efstratiadis, B.G.F., Weissman, S.M., 1977. Molecular Evolution of human and rabbit B-globin mRNAs. Proc. Natl. Acad. Sci. USA 74, 5618–5622.

Khromov, D.N., 1990. Cuttlefishes in the systematics and phylogeny of Cephalopoda. Zool. Zhur. 69, 12–20.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate from the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.

Kluessendorf, J., Doyle, P., 2000. *Pohlsepia mazonensis*, an early 'octopus' from the Carboniferous of Illinois, USA. Palaeontology 43, 919–926.

Lindgren, A.R., Giribet, G., Nishiguchi, M.K., 2004. A combined approach to the phylogeny of Cephalopoda (Mollusca). Cladistics 20, 454–486.

Lockhart, P.J., Steel, M.A., Hendy, M.D., D., P., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol., 11.

Mindell, D.P., Honeycutt, R.L., 1990. Ribosomal RNA in vertebrates; evolution and phylogenetic applications. Annu. Rev. Ecol. Syst. 21, 541–566.

Mouchiroud, D., Gautier, C., Bernardi, G., 1995. Frequencies of synonymous substitution in mammals are gene-specific and correlated with frequencies of nonsynonymous substitutions. J. Mol. Evol. 40, 107–113.

Naef, A., 1912. Teuthologische Notizen. 2. Die Gattungen der Sepioliden. Zool. Anz. 39, 244–248.

Naef, A., 1916. Uber neue Sepioliden aus dem Golf von Neapel. Pubblicazioni della Stazione zooloica di Napoli 1, 1–10.

Naef, A., 1921–1923. Cephalopoda. Fauna e flora del Golfo di Napoli, Monograph. In (translated from German by the Israel program for Scientific translations, Jerusalem 1972), p. 917.

Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Systematic Bilogy 53 (1), 47–67.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.

Phillips, M.J., Lin, Y.-H., Harrison, G.L., Penny, D., 2001. Mitochondrial genomes of a bandicoot and a bruchtail possum confirm the monophyly of Australiadelphian marsupials. Proc. R. Soc. Lond. B 48, 76–93.

Pickford, G.E., 1939. The Vampyromorpha. A new order of dibranchiate Cephalopoda. Vestn. Csl. Zool. Spolecnosti 6–7, 346–358.

Rambaut, A., 2002. Se-Al v2.0a11 Carbon Oxford University.

Rambaut, A., Drummond, A.J., 2003. Tracer. 1.0.1. Oxford University.

Robson, G.C., 1929. A monograph of Recent Cephalopoda (London, Br. Mus. Nat. History).

Robson, G.C., 1931. A monograph of the Recent Cephalopoda, British Museum of Natural History.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Salvini-Plawen, L.V., 1980. A reconsideration of the systematics in the mollusca phylogeny and higher classification. Malacologia 19, 249–278.

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

Springer, M.S., Douzery, E.J.P., 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12s rRNA molecules. J. Mol. Evol. 43, 357–373.

Strugnell, J.M., Norman, M.D., Drummond, A.J., Cooper, A., 2004. The octopuses that never came back to earth: neotenous origins for pelagic octopuses. Curr. Biol. 18, R300–R301.

Suchard, M.A., Kitchen, C.M.R., Sinsheimer, J.S., Weiss, R.E., 2003. Hierarchical phylogenetic models for analyzing multipartite sequence data 52 (5), 649–664.

Swofford, D.L., 1998. PAUP*4.0- Phylogenetic Analysis Using Parsimony (* and Other Methods) (Sunderland, MA, Sinauer).

Teichert, C., 1988. Main features of cephalopod evolution. In: Clarke, M.R., Trueman, E.R. (Eds.), The Mollusca. Paleontology and Neontology of Cephalopods, vol. 12. Academic Press, San Diego, pp. 11–79.

Ticher, A., Graur, D., 1989. Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J. Mol. Evol. 28, 286–298.

Voight, J.R., 1993. A cladistic reassessment of Octopodid classification. Malacologia 35, 343–349.

Voight, J.R., 1997. Cladistic analysis of the Octopods based on anatomical characters. J. Molluscan Stud. 63, 311–325.

Voss, G.L., 1977. Present status and new trends in cephalopod systematics. In: Nixon, M., Messenger, J.B. (Eds.), The Biology of Cephalopods. Symp. Zool. Soc. Lond. Academic Press, London, pp. 49–60.

Voss, G.L., 1988. Evolution and phylogenetic relationships of deep-sea octopods (Cirrata and Incirrata). In: Clarke, M.R., Trueman, E.R., (Eds.), The Mollusca. Paleontology and Neontology of Cephalopods, vol. 12, pp. 253–276.

Warnke, K., Plötner, J., Santana, J.I., Rueda, M.J., Llinàs, O., 2002. Reflections on the phylogenetic position of *Spirula* (Cephalopoda): preliminary evidence from the 18S ribosomal RNA gene. In: Paper presented at: Coleoid cephalopods through time (Berlin, berliner palaobiologische abhandlungen), vol. 3, pp. 253–260.

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

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

Young, R.E., Vecchione, M., Mangold, K.M., 1996. Cephalopoda Cuvier 1797 (Tree of Life Web Project).