A phylogeny of extant coleoid cephalopods with brain data

Kiran Basava[1], Theiss Bendixen[2], Alexander Leonhard Birk Sørensen[2], Nicole George[3], Zoé Vanhersecke[3], Joshua Omotosho[3], Jennifer Mather[4], Michael Muthukrishna[3]

[1] University of Arizona, Tucson, United States
[2] Aarhus University, Aarhus, Denmark
[3] London School of Economics and Political Science, London, United Kingdom
[4] University of Lethbridge, Lethbridge, Canada

Abstract

Extant coleoid cephalopods include over 800 species of octopuses, squid, and cuttlefish, which have drawn scientific and public interest for their complex behavior and cognition. Of these, approximately 10% (79) species have adult specimens with recorded measures of central nervous system size distributed across various sources. Here, we use a combination of topological placements from previous phylogenetic studies, along with mitochondrial and nuclear gene sequences obtained from GenBank, to build a composite phylogenetic tree with estimated branch lengths of all species with available brain measurements. This phylogeny is used for analyses in a forthcoming paper on cephalopod brain evolution, and ideally will be of use to other researchers interested in conducting comparative studies of coleoid cephalopod brains.

Introduction

The origin of the subclass Coleoidea, containing today's octopuses, squid, and cuttlefish, is dated to roughly 300 million years ago in the Devonian period, with the two major orders of Decapodiformes (containing squid and cuttlefish) and Octopodiformes (comprised of octopuses and the vampire squid) diverging in the mid-Triassic (Schweigert and Fuchs 2012; Tanner et al. 2017). Although these high-level splits have been established by fossil and genetic evidence, relationships within them including the placement of cuttlefish and bobtail squid, and the relationships between octopus species, have been more difficult to clarify (Lindgren & Anderson 2018; Sanchez et al. 2018).

In our paper 'Coleoid Cephalopods Demonstrate Asocial Path to the Evolution of Big Brains' (Basava & Bendixen et al. forthcoming), we test for correlates of brain evolution across coleoid cephalopods using a newly constructed trait database and the updated phylogeny described here. Often, phylogenetic analyses of trait evolution use a phylogeny broadly agreed upon by the field to conduct analyses, discarding taxa in the dataset which do not match those species on the phylogeny. For instance, a recent study analysing brain evolution in cephalopods by Ponte et al. (2021) includes 38 species in their phylogenetic analysis out of 78 in their total dataset, as these were the species that matched the cephalopod phylogeny from Lindgren et al. (2012). As we did not wish to exclude any taxa for which brain data were available from our analyses, we sought to build a phylogeny that would include all of these species.

Methods

Rather than inferring a phylogeny from genetic data, we manually compiled a topology of species for which brain measurements were available based on the results of studies selected by Jan Strugnell, who lent her expertise as a cephalopod geneticist. We then added branch lengths to this topology using a phylogenetic inference of sequence data. This approach (manually compiling an informal topology from previously published results and adding branch lengths based on a constrained phylogenetic inference from genetic data) was taken on the advice of Dr. Strugnell based on the availability and accuracy of phylogenetic results and genetic data for coleoid cephalopods.

The starting point for this tree was the phylogeny from Lindgren et al. (2012), which is a relatively complete and widely-referenced phylogeny of cephalopods (López-Córdova et al. 2022; Sanchez et al. 2018, 2021; Strugnell et al. 2013). Relationships within clades were specified further using more recent phylogenetic analyses. Taxa were added in Newick format, either by hand or by using the 'tree.merger' function in the RRphylo package (Castiglione et al. 2018) depending on the number of new species added at a time. The Newick topology, as well as the placement for each species on the topology with the citation for the referenced study, are in https://github.com/kcbasava/ceph-brain-evolution/tree/main/phylogeny.

To add branch lengths, we conducted a topology-constrained phylogenetic inference from molecular data using BEAST 2.7.4 via the BEAUti GUI (Bouckaert et al. 2019). We constructed a dataset by downloading sequences for three nuclear genes: 18S rRNA (18S), 28S rRNA (28S), and rhodopsin (RHO), and three mitochondrial genes: 12S rRNA (12S), 16S rRNA (16S), and cytochrome oxidase I (COI) for all available species. Genetic data were not available for six species in our brain dataset; in those cases, sequences for closely related species (based on other studies using morphological characters) in the same genus were used[1]. The table containing GenBank accession numbers for all sequences used for each species is in https://github.com/kcbasava/ceph-brain-evolution/tree/main/phylogeny with notes indicating substituted species.

The genes were aligned separately using Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004) with default settings as implemented on the NGPhylogeny.fr website (Lemoine et al. 2019). They were trimmed to obtain conserved blocks with Gblocks (Castresana 2000) likewise implemented on the NGPhylogeny.fr website using less restrictive settings allowing gaps in up to half the blocks (b1=50%, b2 = 85%, b3 = 10, b4 = 5, b5 =50%). Alignments were processed further to correct sequence names and add empty sequences for missing taxa for each gene in R using the packages seqinr (Charif et al. 2023) and ape (Paradis et al. 2019). Using BEAUti, the dataset was partitioned into noncoding and coding nuclear and mitochondrial genes, with four partitions total (COI; RHO; 18S and 28S; and 16S and 12S). Substitution models for the partitions were selected by bModelTest (Bouckaert & Drummond 2017). We set a Yule model of speciation as the tree prior and an optimized relaxed clock model (Douglas et al. 2021). The topology we compiled from previous phylogenetic studies described above was added as a multiple monophyletic constraint with substituted species names replaced. It contained multiple polytomies (non-binary speciation events where order of speciation is unclear) which were also resolved based on the BEAST analysis.

The divergence times were calibrated following the priors used by López-Córdova et al. (2022): for the root of the tree at the split between Decapodiformes and Octopodiformes using the Triassic coleoid fossil *Germanoteuthis* (Schweigert & Fuchs 2012) with a gamma distribution (offset=236, shape=2.0, scale=3.0, the split between Vampyromorpha and Octopoda using the Jurassic fossil *Loligosepia* (Fuchs & Weis 2008) with a gamma distribution (offset=195, shape=2.0, scale=3.0), the split between incirrate and cirrate octopods using the fossil *Styletoctopus annae* (Fuchs et al. 2009) with a gamma distribution (offset=93, shape=2.0, scale=3.0), and the origin of Loliginidae with the fossil *Loligo applegatei* (Clarke & Fitch 1979) with a gamma distribution (offset=48, shape=2.0, scale=3.0).

The MCMC was run for 30 million generations sampled every 3000 generations with a 10% burn-in. Tracer v1.7.2 (Rambaut et al. 2018) was used to ascertain effective sample sizes (>200)

[1] While some genera, e.g. Octopus, are likely paraphyletic and genus title is not a reliable indicator of sister relationships, this was not the case for any of the substituted species.

and chain convergence. The posterior distributions of trees from two independent BEAST runs were combined using LogCombiner (Bouckaert et al. 2019). A consensus tree (maximum clade credibility tree, highest product of posterior node probabilities) was also obtained using the software TreeAnnotator v2.7.4 (Fig. 2).

Ancestral state reconstruction

The ancestral states of brain size (CNS) and relative brain size (CNS/ML) was estimated for the root and fossilized nodes using the anc.Bayes function from the phytools package in R (Revell 2024). This was estimated on the consensus phylogeny with the MCMC run for 100000 generations with a 20% burn-in.

Results

Estimates for divergence times of major clades are shown in Table 1 with corresponding fossil calibrations. The origin for Decapodiformes (which was not fossil constrained) was estimated at 91.90 Mya (120.91 - 68.25) with *Idiosepius paradoxus* assigned as the first diverging group in the topological constraint.

Taxa	**Fossil calibration**	**MRCA mean age**	**HPD 95% (Mya)**	**Estimated EQ (log)**	**Estimated CNS (log)**
Coleoidea (root)	*Germanoteuthis* ~236 Mya	240.88	236.08 - 247.63	0.973	3.564
Decapodiformes	-	91.90	120.91 - 68.25	1.077	4.181
Octopodiformes	*Loligosepia* ~195 Mya	200.66	195.10 - 208.33	0.931	3.487
Octopoda	*Styletoctopus annae* ~93 Mya	103.93	93.45 - 117.26	0.859	3.633
Loliginids	*Loligo applegatei* ~48 Mya	54.02	48.21 - 61.40	1.098	4.644

Table 1. Estimated divergence times (95% highest posterior density interval) and means for fossil-calibrated nodes.

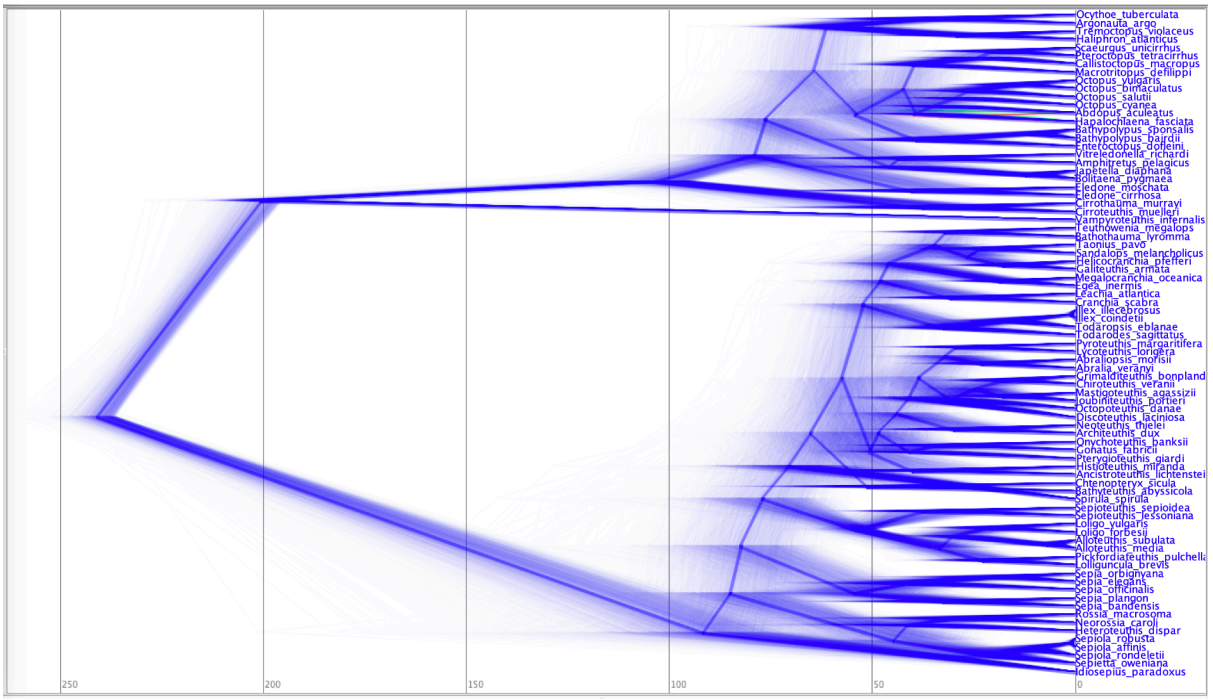

Figure 1. The posterior distribution of trees from the topology-constrained analysis in BEAST visualized with DensiTree v2.7.4 (Bouckaert & Helped 2014).

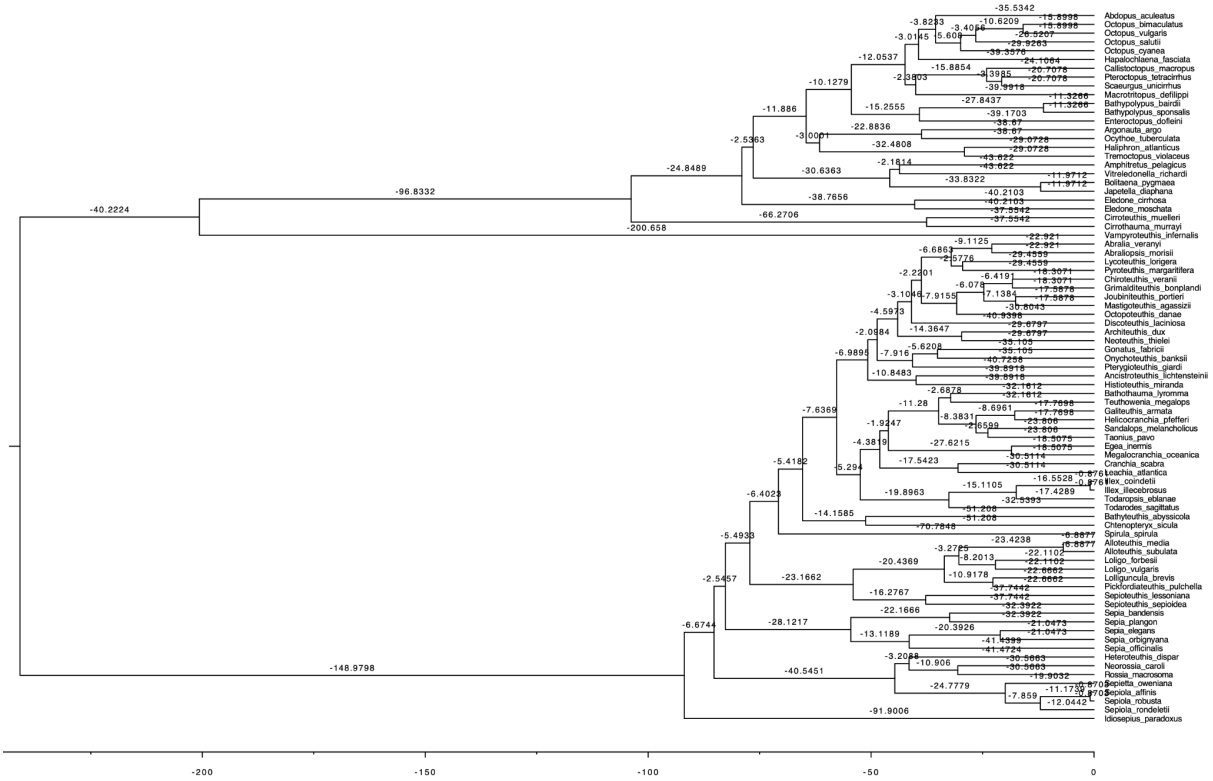

Figure 2. Estimated divergence times on consensus phylogeny using FigTree.

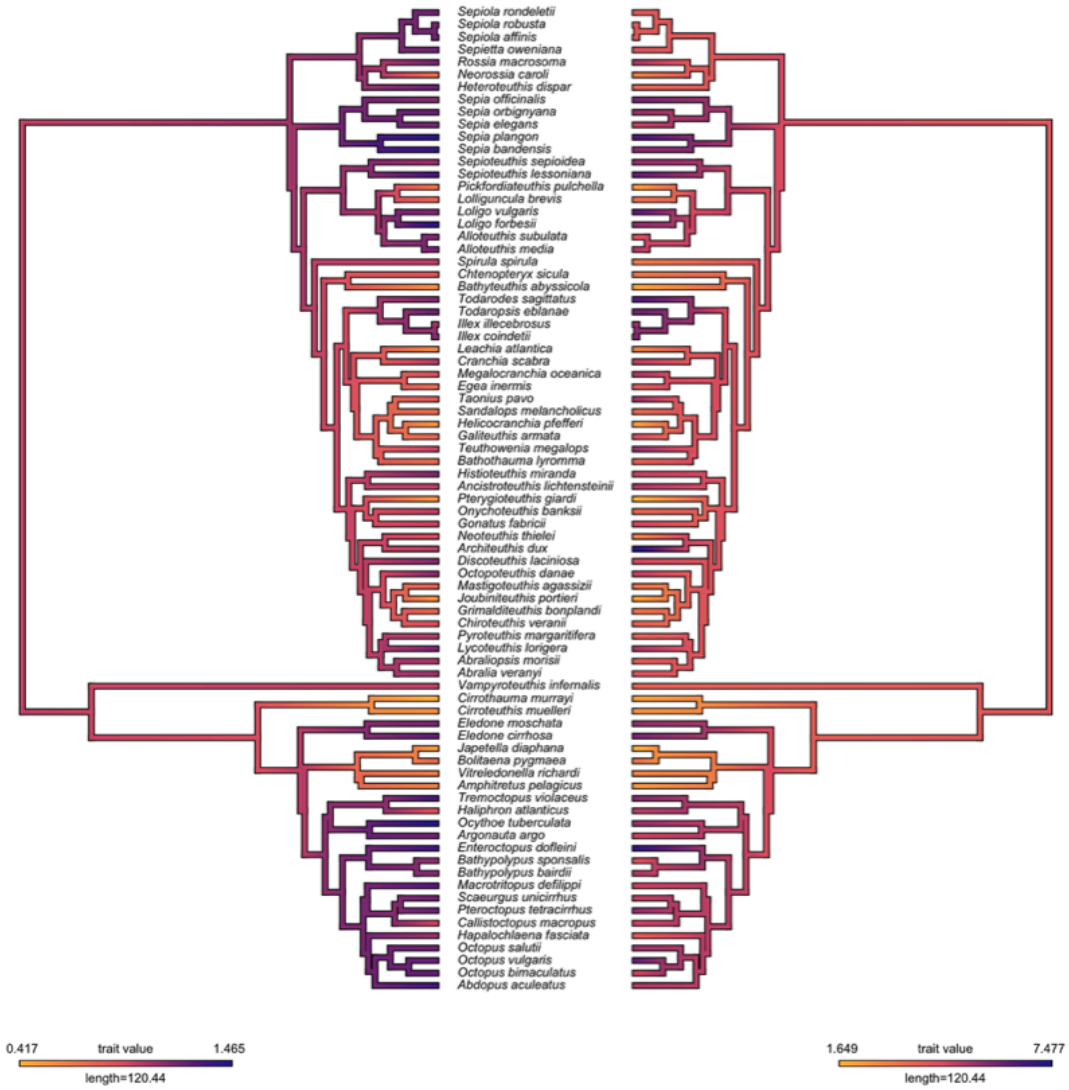

Figure 3. Ancestral state estimates for relative brain size (CNS/ML, left) and total brain size (CNS, right) on a log scale.

Data

The XML files for the BEAST analyses, file containing GenBank accession numbers for each species, brain and body data, output posterior trees, and consensus tree are available at https://github.com/kcbasava/ceph-brain-evolution/.

Acknowledgements

We thank the Templeton World Charity Foundation for supporting this work (Grant TWCF0464). We would also like to thank Jan Strugnell for assistance and advice in compiling and constructing the phylogeny.

References

Bouckaert, R., & Heled, J. (2014). *DensiTree 2: Seeing Trees Through the Forest*. https://doi.org/10.1101/012401

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., Maio, N. D., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Plessis, L. du, Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Braid, H. E., & Bolstad, K. S. R. (2019). Cephalopod biodiversity of the Kermadec Islands: implications for conservation and some future taxonomic priorities. *Invertebrate Systematics*, *33*(2), 402–425. https://doi.org/10.1071/IS18041

Castiglione, S., Tesone, G., Piccolo, M., Melchionna, M., Mondanaro, A., Serio, C., Di Febbraro, M., & Raia, P. (2018). A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology and Evolution*, *9*(4), 974–983. https://doi.org/10.1111/2041-210X.12954

Charif, D., Clerc, O., Frank, C., Lobry, J. R., Necşulea, A., Palmeira, L., Penel, S., & Perrière, G. (2023). *seqinr: Biological Sequences Retrieval and Analysis* (4.2-30). https://cran.r-project.org/web/packages/seqinr/index.html

Clarke, M. R., & Fitch, J. E. (1979). Statoliths of Cenozoic teuthoid cephalopods from North America. *Palaeontology*, *22*(2), 479–511. https://oceanrep.geomar.de/id/eprint/34692/

Douglas, J., Zhang, R., & Bouckaert, R. (2021). Adaptive dating and fast proposals: Revisiting the phylogenetic relaxed clock model. *PLOS Computational Biology*, *17*(2), e1008322. https://doi.org/10.1371/journal.pcbi.1008322

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Evans, A. (2018). *A Systematic Review of the Squid Family Cranchiidae (Cephalopoda: Oegopsida) in the Pacific Ocean* [Auckland University of Technology]. https://hdl.handle.net/10292/11934

Fernández-Álvarez, F. Á., Taite, M., Vecchione, M., Villanueva, R., & Allcock, A. L. (2022). A phylogenomic look into the systematics of oceanic squids (order Oegopsida). *Zoological Journal of the Linnean Society*, *194*(4), 1212–1235. https://doi.org/10.1093/zoolinnean/zlab069

Fuchs, D., Bracchi, G., & Weis, R. (2009). NEW OCTOPODS (CEPHALOPODA: COLEOIDEA) FROM THE LATE CRETACEOUS (UPPER CENOMANIAN) OF HÂKEL AND HÂDJOULA, LEBANON. *Palaeontology*, *52*(1), 65–81. https://doi.org/10.1111/j.1475-4983.2008.00828.x

Fuchs, D., & Weis, R. (2008). Taxonomy, morphology and phylogeny of Lower Jurassic loligosepiid coleoids (Cephalopoda). *Neues Jahrbuch Für Geologie Und Paläontologie - Abhandlungen*, *249*(1), 93–112. https://doi.org/10.1127/0077-7749/2008/0249-0093

Judkins, H., Rose-Mann, L., Lindgren, A., Taite, M., Bush, S., & Vecchione, M. (2022). A newly discovered *Helicocranchia* species (Cephalopoda: Cranchiidae: Taoniinae) in the northern Gulf of Mexico. *Bulletin of Marine Science*, *98*(3), 419–430. https://doi.org/10.5343/bms.2021.0048

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research*, *47*(W1), W260–W265. https://doi.org/10.1093/nar/gkz303

Lima, F. D., Strugnell, J. M., Leite, T. S., & Lima, S. M. Q. (2020). A biogeographic framework of octopod species diversification: the role of the Isthmus of Panama. *PeerJ*, *8*, e8691. https://doi.org/10.7717/peerj.8691

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 Evolutionary Biology*, *12*(1), 129. https://doi.org/10.1186/1471-2148-12-129

López-Córdova, D. A., Avaria-Llautureo, J., Ulloa, P. M., Braid, H. E., Revell, L. J., Fuchs, D., & Ibáñez, C. M. (2022). Mesozoic origin of coleoid cephalopods and their abrupt shifts of diversification patterns. *Molecular Phylogenetics and Evolution*, *166*, 107331. https://doi.org/10.1016/j.ympev.2021.107331

Lupše, N., Reid, A., Taite, M., Kubodera, T., & Allcock, A. L. (2023). Cuttlefishes (Cephalopoda, Sepiidae): the bare bones—an hypothesis of relationships. *Marine Biology*, *170*(8), 93. https://doi.org/10.1007/s00227-023-04195-3

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Ponte, G., Taite, M., Borrelli, L., Tarallo, A., Allcock, A. L., & Fiorito, G. (2021). Cerebrotypes in Cephalopods: Brain Diversity and Its Correlation With Species Habits, Life History, and Physiological Adaptations. *Frontiers in Neuroanatomy*, *14*. https://www.frontiersin.org/articles/10.3389/fnana.2020.565109

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Revell, Liam J. "Phytools 2.0: An Updated R Ecosystem for Phylogenetic Comparative Methods (and Other Things)." PeerJ 12 (January 5, 2024): e16505. https://doi.org/10.7717/peerj.16505.

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. *Molecular Phylogenetics and Evolution*, *68*(2), 293–299. https://doi.org/10.1016/j.ympev.2013.03.027

Sanchez, G., Fernández-Álvarez, F. Á., Taite, M., Sugimoto, C., Jolly, J., Simakov, O., Marlétaz, F., Allcock, L., & Rokhsar, D. S. (2021). Phylogenomics illuminates the evolution of bobtail and bottletail squid (order Sepiolida). *Communications Biology*, *4*(1), 1–9. https://doi.org/10.1038/s42003-021-02348-y

Sanchez, G., Setiamarga, D. H. E., Tuanapaya, S., Tongtherm, K., Winkelmann, I. E., Schmidbaur, H., Umino, T., Albertin, C., Allcock, L., Perales-Raya, C., Gleadall, I., Strugnell, J. M., Simakov, O., & Nabhitabhata, J. (2018). Genus-level phylogeny of cephalopods using molecular markers: current status and problematic areas. *PeerJ*, *6*, e4331. https://doi.org/10.7717/peerj.4331

Sánchez-Márquez, A., Navarro, J., Kaliontzopoulou, A., Farré, M., Taite, M., Escolar, O., Villanueva, R., Allcock, A. L., & Fernández-Álvarez, F. Á. (2023). Unravelling the phylogenetic and ecological drivers of beak shape variability in cephalopods. *Reviews in Fish Biology and Fisheries*, *33*(1), 221–239. https://doi.org/10.1007/s11160-022-09744-5

Schweigert, G., & Fuchs, D. (2012). First record of a true coleoid cephalopod from the Germanic Triassic (Ladinian). *Neues Jahrbuch Für Geologie Und Paläontologie - Abhandlungen*, *266*(1), 19–30. https://doi.org/10.1127/0077-7749/2012/0258

Strugnell, J. M., Norman, M. D., Vecchione, M., Guzik, M., & Allcock, A. L. (2014). The ink sac clouds octopod evolutionary history. *Hydrobiologia*, *725*(1), 215–235. https://doi.org/10.1007/s10750-013-1517-6

Taite, M., Fernández-Álvarez, F. Á., Braid, H. E., Bush, S. L., Bolstad, K., Drewery, J., Mills, S., Strugnell, J. M., Vecchione, M., Villanueva, R., Voight, J. R., & Allcock, A. L. (2023). Genome skimming elucidates the evolutionary history of Octopoda. *Molecular Phylogenetics and Evolution*, *182*, 107729. https://doi.org/10.1016/j.ympev.2023.107729

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. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), 20162818. https://doi.org/10.1098/rspb.2016.2818

Uribe, J. E., & Zardoya, R. (2017). Revisiting the phylogeny of Cephalopoda using complete mitochondrial genomes. *Journal of Molluscan Studies*, *83*(2), 133–144. https://doi.org/10.1093/mollus/eyw052