Chromosome-scale genome assembly of the European common cuttlefish *Sepia officinalis*

Simone Rencken[1,2] https://orcid.org/0009-0003-9341-2898, Georgi Tushev[1] https://orcid.org/0000-0002-3340-9422, David Hain[1,3] https://orcid.org/0000-0002-8979-7938, Elena Ciirdaeva[1] https://orcid.org/0009-0000-7360-7725, Oleg Simakov[4] https://orcid.org/0000-0002-3585-4511, Gilles Laurent[1] https://orcid.org/0000-0002-2296-114X

[1] Max Planck Institute for Brain Research, Frankfurt am Main, Germany
[2] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands
[3] Faculty of Biological Sciences, Goethe University, Frankfurt am Main, German
[4] Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria

Abstract

Coleoid cephalopods, a subclass of mollusks that includes octopuses, cuttlefish and squid, exhibit sophisticated biological features such as dynamic and neurally driven camouflage behavior, inter-individual communication, single-lens camera-like eyes, the largest brains among invertebrates and a distinctive embryonic development.

The common cuttlefish *Sepia officinalis* has served as a model organism in various research fields, spanning biophysics, neurobiology, behavior, evolution, ecology and biomechanics. More recently, it has become a model to investigate the neural mechanisms underlying cephalopod camouflage, using quantitative behavioral approaches alongside molecular techniques to characterize the identity, evolution and development of neuronal cell types.

Despite significant interest in this animal, a high-quality, annotated genome of this species is still lacking. To address this, we sequenced and assembled a chromosome-scale genome for *S. officinalis.* Our assembly spans 5.68 billion base pairs and comprises 1n=47 repeat-rich chromosome scaffolds. This was unexpected because the haploid karyotypes of other decapods indicate 46 chromosomes. Detailed comparisons of our data to those from published decapod genome assemblies and to another recent genome assembly of *S. officinalis* (itself suggesting 1n=49 chromosomes) in fact revealed clear homologies between 46 scaffolds across all the datasets. The discrepancies between datasets are explained by highly repetitive regions, impairing proper read alignments. We conclude that the true karyotype of *S. officinalis* is probably 1n=46 chromosomes, a likely ancestral and if true, conserved decapod karyotype.

Our results include a comprehensive gene annotation and full-length transcript prediction, which we used to characterize orthologous gene families across mollusks. We identified several large-scale expansions specific to cephalopods, with many genes specific to neural or non-neural tissues of adult *S. officinalis.* In summary, this genome should provide a valuable resource for future research on the evolution, brain organization, information processing, development and behavior in this important clade.

Introduction

Coleoid cephalopods (octopus, squid, cuttlefish) are a highly derived group of mollusks, characterized by the largest nervous systems among all invertebrates (ca. 500 million neurons in adult octopus of which 200 million are in the central brain[1,2], compared to ca. 140,000 in the fruit fly[3] or 70 million in the mouse[4]) and specializations with a great historical importance for neuroscience (e.g., “giant axons”[5] and “giant synapses”[6–8]). These animals exhibit very sophisticated behaviors such as dynamic camouflage[9–15], learning[16–18], social communication[19–21] and hunting[22–24] as well as two-stage sleep[25–27]. Because of these characteristics, cephalopods have been the focus of many fundamental studies by biologists, biophysicists and physiologists over the past century[28–37]. Thanks to advances in sequencing technologies, coleoid cephalopods are now also emerging as animal models in molecular neurobiology, evolution and genomics[38]. Recent studies examined cephalopod biology from the perspectives of single-cell gene expression[39–42], genome topology and gene regulation[43,44].

Despite recent technical progress in genomics, the genomes of cephalopods remain challenging to assemble because of their large sizes and high repeat fractions. While the genomes of Octopodiformes (Octopus, Eledone, Argonauta) are either smaller than (1.1 Gigabases or Gb[45]) or comparable in size to that of humans (around 3 Gb[46,47]), the typical genomes of Decapodiformes (squids and cuttlefish) often reach 6 Gb[48,49]. The biggest contributing factors to this genome expansion are transposable elements (TEs), with different TE classes differentiating squid and cuttlefish genomes from those of octopods[50]. Besides such differences in genome sizes, karyotypes in coleoid cephalopods are also unusual when compared to those of other invertebrates, including other mollusks: for example, haploid chromosome numbers are around 30 in octopuses and around 46 in squids and cuttlefish[49,51], while they are between 8 and 19 in non-cephalopod mollusks[52–59]. Karyotype reconstruction and validation techniques in animals with such large genomes are also not as developed as they are for more familiar invertebrate and vertebrate species.

The common cuttlefish *Sepia officinalis* has been used as a model organism for studies in neural control and development[60,61], behavior[9,17], evolution[62], environmental studies[63,64] and biomechanics[65,66]. Found in the Mediterranean, the English Channel and the European coastal Atlantic, it can produce large egg clutches containing hundreds of embryos, several times throughout its semelparous reproductive cycle[67,68]. More recently, this species has emerged as an important model to study the neural basis of cephalopod camouflage[69,70] building on extensive behavioral studies[9–15] and neuronal cell type characterization, evolution and development[71–73].

Despite a widespread interest in this animal, including the recent sequencing of several cephalopod species as part of the Aquatic Symbiosis Genomics project[74], a detailed, annotated genomic resource for this species is still lacking. We describe here the sequencing, assembly, and annotation of the *Sepia officinalis* genome.

Our data provided initial evidence for the existence of 1n=47 chromosomes. We compared them with available genome assemblies (i) from another *S. officinalis* individual by the Darwin Tree of Life project (DToL)[75], itself suggesting 49 chromosomes, and (ii) from other decapod species such as *Euprymna scolopes*, *Doryteuthis pealeii* and *Acanthosepion*

esculentum, each with estimates of 46 chromosomes[49]. When compared with the assemblies generated for the other species, both *S. officinalis* assemblies contained additional and independent chromosome splits. We thus investigated the repeat content and the alignment of raw data at the discrepant chromosome junctions to assess whether these differences were of a technical or biological nature.

Results

We sequenced the genome of a 6-month-old *Sepia officinalis* male individual, reared from eggs collected in the Portuguese Atlantic coast. The DNA was extracted from brain tissue and sequenced using long-read (PacBio HiFi) and chromatin conformation (Hi-C) methods (Figure 1A,B). The sex of the animal was confirmed by qPCR following a recently described protocol[76].

Genome size and heterozygosity

The haploid genome size of *S. officinalis* was estimated to be ~5.14 Gb based on k-mer estimation from the short-read data, with a high repeat content of 54 %[77,78]. The heterozygosity rate was estimated to be 1.03 %, which is higher than in octopus genomes, yet moderate among marine invertebrates[46,47,49].

The size of the scaffolded assembly was 5.68 Gb (Figure 1C), about 10% greater than our initial GenomeScope estimate. Indeed, the sizes of most published metazoan genome assemblies deviate by more than 10% from estimates, with proportionally greater deviations as assembly size increases[79].

Completeness of the genome was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO[80]) with *metazoa_odb12*. It was 94.3% [single copy: 93.2%, duplicated: 1.2%], with 4.5% fragmented and 1.2% missing BUSCOs. The BUSCO score with *mollusca_odb12* for the final assembly was 90.2% completeness [single copy: 89.1%, duplicated: 1.1%] with 4.9% fragmented and 4.9% missing BUSCOs.

Assembly-based karyotype

The assembly was generated from PacBio HiFi reads and chromosome conformation capture (Hi-C) reads with ca. 23-fold and 11-fold coverage, respectively. Scaffolding was performed using Hi-C reads, placing 95.8 % of bases in 47 scaffolds. The Hi-C heatmap in Figure 1D shows the chromatin contacts in clusters corresponding to these 47 scaffolds, suggesting that the 47 clusters each correspond to a single chromosome.

To further test the quality of our assembled *S. officinalis* karyotype, we used several scaffolding programs and manually curated the scaffolds using two independent approaches. The first rested on HapHiC[81], a tool based on Hi-C data that uses allele information from primary assembly programs, and scaffolds assemblies with the constraint of an expected number of chromosomes. The resulting Hi-C contact maps with an input of 46, 47, 48, 49 and 50 (Supplementary Figure 1) supported 47 scaffolds: an input of 46 scaffolds resulted in one clear chromosome merger (blue arrowhead); higher input numbers resulted in false chromosome splits (black arrowheads, 48, 49 and 50 scaffolds).

Second, we manually curated the scaffolds (obtained with YAHS[82]) using JBAT[83]. These two approaches, conducted independently and both based on Hi-C signal, again converged on 47 chromosomal scaffolds.

Karyotype comparisons with other Decapods

Several chromosome-scale genome assemblies of decapod cephalopods have been published recently. From those studies, a haploid karyotype of 46 chromosomes seems to be common and conserved among Decapodiformes[42,49,84] (Figure 3A). Thus, we sought to further investigate and confirm our estimated and different karyotype for *S. officinalis*.

Besides assembly-based karyotypes, several cephalopod species have been investigated using cytogenetic techniques. These studies, however, report widely varying estimates of chromosome numbers for decapods, reflecting the difficulty of resolving large (diploid) chromosome numbers *in situ*. For example, 1n=46 chromosomes has been reported for two species of cuttlefish (*Acanthosepion esculentum* and *Acanthosepion lycidas*) and three loliginid squids[85]; while 1n=34 chromosomes has been reported for *Aurosepina arabica*[86] and 1n=24 chromosomes in *Acanthosepion pharaonis*[87]. In *Sepia officinalis*, a karyotype of 1n=52 has even been described from testis samples[88].

Comparison with another chromosome-scale *Sepia officinalis* assembly

A chromosome-scale assembly for *Sepia officinalis* was released recently by the Wellcome Sanger Institute's Darwin Tree of Life project[75] (DToL, GCA_964300435.1). That genome was assembled from a male individual using high coverage PacBio Sequel II (~51x) and Arima2 Hi-C (~80x) data, with a final assembly size of 5.8 Gb. The haploid chromosome number was estimated to be 49. To compare both *S. officinalis* datasets directly, we downloaded the DToL data and created two new assemblies using the pipeline described above (hifiasm using PacBio HiFi and Hi-C data). The resulting assemblies were overall very similar, with the DToL assembly having a slightly higher contiguity (N50 length, see Table 1) and BUSCO completeness (Supplementary Figure 2A,B) due to their higher sequencing coverage.

Table 1: Statistics of *S. officinalis* assemblies from two independent datasets, assembled using a common pipeline.

	MPIBR reassembly			**DToL reassembly**		
version	**p_ctg**	**hap1**	**hap2**	**p_ctg**	**hap1**	**hap2**
number of contigs	8.289	10.651	10.425	8.783	11.026	11.089
raw length [bp]	6.049.669.443	5.675.386.986	5.662.586.038	6.053.996.452	5.721.157.269	5.950.565.264
N50 length [bp]	1.723.203	1.032.632	1.010.375	1.810.137	1.165.578	1.182.649
average contig length [bp]	729.843	532.850	543.173	689.285	518.878	536.618

After scaffolding with YAHS, both datasets reached the previously identified chromosome numbers (1n=47 for MPIBR and 1n=49 for DToL, Figure 2A,B). To further investigate this surprising discrepancy, we aligned both assemblies using Winnowmap[89,90] to locate the differences between them (Figure 2C). We observed four “breakpoints” (BP) of chromosome scaffolds: one in the MPIBR assembly compared to DToL (BP1: DToL_5 = MPIBR_40+44) and three in the DToL assembly compared to MPIBR (BP2: DToL_31+40 = MPIBR_2, BP3: DToL_41+46 = MPIBR_6, BP4: DToL_44+45 = MPIBR_7). We also aligned the assemblies to the chromosome-scale genome of another cuttlefish *Acanthosepion esculentum* (1n=46, GCA_964036315.1). In this alignment, all four breakpoints were collinear with single *A. esculentum* chromosomes (Figure 2D).

To better understand the potential cause of these divergent chromosome numbers, we analyzed the Hi-C and HiFi coverage in the breakpoint regions (Supplementary Figure 3A). First, we aligned the Hi-Fi reads to the scaffolds and extracted all alignments along the 200 kb terminal scaffold windows to find any notable drops in coverage, or reads spanning any of the scaffold junctions. We detected no spanning reads. This is not surprising given that no contigs were assembled at these sites, resulting in the observed scaffold junctions. More interestingly, we noted a ~5-fold decrease in HiFi coverage along the DToL scaffold_40 (part of BP2) relative to its flanking regions, indicating a highly repetitive, low-mappability region at this boundary.

Next, we realigned the Hi-C data to the scaffolded assemblies using bwa-mem2[91] and extracted all trans HiC pairs (between-scaffold contacts) using pairtools[92]. We normalized trans HiC contacts to the scaffold length and compared contact rates between breakpoint scaffolds to the baseline contact rate (computed from pairs of scaffolds with a clear 1-to-1 match between assemblies), and the contact rate within scaffolds (intra-scaffold pairs) (Supplementary Figure 3B,C). The contact rates within breakpoints were consistently lower than within scaffolds, likely falling below the threshold to be merged during assembly. However, the contact rates at three of four breakpoints (BP1, BP3, BP4) were significantly elevated above the genome-wide background distribution (empirical $p = 0.010$, 0.005, 0.005 respectively), suggesting that they may represent intra-chromosomal contacts disrupted by a misassembly. Notably, BP2 was not significant (empirical $p = 0.170$), likely due to the low coverage and mappability around the DToL scaffold_40 boundary. Considered jointly, the three DToL breakpoint scaffold pairs showed significantly higher trans contact rates than the background (Wilcoxon rank-sum, one-tailed, $U = 1771$, $p = 0.004$).

Lastly, we analyzed the repeat landscape around the 200 kb scaffold ends using RepeatMasker[93] and the custom repeat library that we had generated for *Sepia officinalis* (described further below). Compared to control scaffolds of the same assembly, we observed consistently elevated repeat content at the breakpoint junctions (mean 71.5% vs 67.6% masked bases), with an enrichment of unclassified repeats (32.1% vs 30.0%), which could explain a repeat-driven assembly fragmentation or scaffolding failure. The BP2 DToL scaffold_40 junction window was 99.99% masked (99.2% unclassified repeats), providing a likely mechanistic explanation for both the HiFi coverage drop and the absence of a significant trans Hi-C signal at this breakpoint. Taken together, these analyses suggest that the different chromosome numbers across the two *S. officinalis* assemblies are due to technical reasons, caused by repeat-rich scaffold boundaries that impair HiFi and Hi-C read alignment and in turn, correct assembly in these regions.

Comparisons with other chromosome-scale Decapodiformes genomes

To further investigate why our initial estimate of *Sepia officinalis's* chromosome numbers differred from those in other studies, we compared our assembly with the chromosome-scale genomes of two other decapod cephalopods, *Euprymna scolopes*[43] (a sepiolid) and *Doryteuthis pealeii*[49] (a loliginid), both described as having 46 chromosomes (Figure 3A). Using orthogroups to perform linkage analysis, we could detect a clear chromosome homology between *Doryteuthis* and *Sepia*, and to a lesser extent between *Euprymna* and *Sepia* (Figure 3B). We observed some small-scale rearrangements between *Euprymna* and *Doryteuthis*, such as a fusion of chromosomes 24 and 40 from *Doryteuthis* in chromosome 2 of *Euprymna*, which has been also observed in other Sepiolids[42,84]. Dotplots showing detailed pairwise synteny comparisons are shown in Supplementary Figure 4 (*Sepia* to *Doryteuthis*) and Supplementary Figure 5 (*Sepia* to *Euprymna*). The inferred species tree as part of this analysis places *Euprymna* as sister to *Sepia* and *Doryteuthis,* matching recent phylogenetic analyses[53,94] (note, however, that this tree was constructed without the inclusion of outgroup taxa, and therefore lacks a reliable root).

This comparison revealed that chromosomes 40 and 43 in *Sepia* are merged into one in *Euprymna* and in *Dorytheutis* (Figure 3C), as they were in the DToL *Sepia officinalis* assembly (Figure 2C). Together with the whole-genome alignments between the cuttlefish assemblies described earlier (Figure 2D), these results lead to the following parsimonious conclusion: that the karyotype for *S. officinalis* is 1n=46 chromosomes, as it is for other Decapodiformes.

We also performed a synteny comparison including the cuttlefish *A. esculentum* that was recently annotated by genome liftover from *Acanthosepion pharaonis*[95]. That assembly, constructed from a female individual, showed a low read coverage for a sex chromosome in a ZZ/Z0 sex determination system[96]. Our syntenic comparison indicated a strong homology between the inferred Z chromosome of *A. esculentum* and chromosome 46 of *S. officinalis* (Supplementary Figure 6A). As indicated above, we determined the sex of our *S. officinalis* specimen by replicating the analysis used to identify the Z chromosome in *A. esculentum*[96]. For this, we aligned short-reads (Illumina) from the same *S. officinalis* individual to the assembly and examined the normalized read coverage for each chromosome (Supplementary Figure 6B). In contrast to the low coverage observed in the female *A. esculentum* assembly (Supplementary Figure 6C), we observed no significant decrease in read coverage for chromosome 46, suggesting that our material came from a male animal. Additionally, we used a recently published genotyping protocol for cephalopods[76] and performed qPCR on extracted genomic DNA from tissue samples, confirming the male sex of our sequenced individual.

Genome repeat landscape

After creating a custom repeat library for *Sepia officinalis* using RepeatModeler[97], we masked the genome using RepeatMasker[93], resulting in 71.17% masked bases. The categories of repeats are shown in Figure 4A. Most repeats were not characterized (39.65% of total bp) and presumably represent ancient repeats that diverged beyond recognition[98]. Retroelements constituted the largest characterized repeat category (17.32%) followed by DNA transposons (5.92%); 5.83% were annotated as simple repeats. As observed in other

Decapodiformes, the *Sepia* genome contained almost no short interspersed nuclear elements (SINEs), supporting the hypothesis that the SINE expansion observed in octopuses occurred independently in their lineage[50].

Gene modeling and annotation

The genome was further annotated using BRAKER[80,99–114] combining short- and long-read RNA-seq data and publicly available protein data from multiple molluscan species (*Doryteuthis pealeii*[49], *Euprymna scolopes*[115], *Octopus bimaculoides*[49], *Octopus vulgaris*[47], *Nautilus pompilius*[116], and *Pecten maximus*[54]). A total of 18,663 gene models and 23,768 proteins were annotated.

The gene annotation was evaluated using BUSCO[80] v5.5.0 in protein mode, showing high completeness of the annotation (*metazoa_odb10*: C:98.2%[S:77.3%,D:20.9%], F:1.0%, M:0.8%, n:954 and *mollusca_odb10*: C:81.5%[S:59.6%,D:21.9%], F:0.9%, M:17.6%, n:5295). We further checked the completeness and consistency of our gene models using OMArk[117], and compared them to genome annotations for two other cuttlefish species, *Acanthosepion pharaonis* and *Acanthosepion lycidas* (Figure 4B). Completeness was assessed by comparing the annotated genes to conserved orthologs (present in 80% of extant species) in a given taxonomic clade (here the superorder Lophotrochozoa). The *S. officinalis* annotation missed 181 out of 2373 genes, which is higher than with *A. lycidas* (96 genes) but lower than with *A. pharaonis* (458 genes).

In a consistency assessment, where the presence of known gene families from the lineage is evaluated (Figure 4C), *S. officinalis* contains low proportions of taxonomically inconsistent or fragmented proteins (ca. 13%) similar to *A. lycidas* (9%), giving confidence in the annotation. In contrast, more than 50% of *A. pharaonis* proteins are labeled as inconsistent or fragmented, which could indicate annotation errors[117].

Overall, the annotations contain different numbers of predicted proteins which may reflect differences in the annotation method and the reference data used. The genome of *S. officinalis* contains fewer proteins (23,768) than those of *A. lycidas* (35,949) or *A. pharaonis* (53,515). In comparison, two octopus genomes contain similar numbers of proteins (*O. vulgaris*: 30,134; *O. bimaculoides*: 29,037) and were produced using NCBI's RefSeq[118] pipeline, suggesting that the differences observed across cuttlefish proteomes are probably of technical instead of biological origin.

Lastly, we assigned orthology information to the *S. officinalis* proteome using InterProScan[119] and eggNOG-mapper[120] to aid the interpretability of the resource for future transcriptomic or proteomic studies. Overall, 89% of proteins (21,204 out of 23,768) received an annotation from InterProScan from at least one of their databases. 59% of proteins (14,126 out of 23,768) were annotated by eggNOG-mapper, reflecting the more stringent orthology filters and prioritization of full-length matches implemented in the program[121].

Analysis of expanded gene families

We sought to investigate the *S. officinalis* gene annotation and place it in the context of gene repertoires from other cephalopod or molluscan species. First, we collected available

genome annotations from 12 other molluscan species (Table 2) and clustered them using OrthoFinder v.3.1.0[122], resulting in 23,658 orthogroups, hereafter named gene families.

Table 2: Overview of gene annotation of 13 molluscan species used for gene family analysis.

Organism Scientific Name	Accession	Source	URL	# of Proteins
Aplysia californica	GCF_000002075.1	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000002075.1/	21897
Crassostrea virginica	GCF_053477285.1	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/gene/GCF_053477285.1/	53819
Doryteuthis pealeii	GCA_023376005.1	custom[49]	https://metazoa.csb.univie.ac.at/CephData/dorPea.prot.gz	24931
Euprymna scolopes	GCA_024364805.1	Github	https://github.com/TheaFrances/E.scolopes-V2.2-BRAKER2-gene-annotation	31908
Gigantopelta aegis	GCF_016097555.1	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_016097555.1/	24904
Lottia gigantea	GCF_000327385.1	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000327385.1/	23822
Magallana gigas	GCF_963853765.1	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_963853765.1/	35231
Nautilus pompilius	GWHBECW00000000	GWH	https://ngdc.cncb.ac.cn/gwh/Assembly/21849/show	16536
Octopus bimaculoides	GCF_001194135.2	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001194135.2/	29037
Octopus vulgaris	GCA_951406725.2	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/gene/GCA_951406725.2/	30134
Pecten maximus	GCF_902652985.1	RefSeq	https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_902652985.1/	28975
Acanthosepion lycidas	GCA_963932145.1	Ensembl genebuild	https://ftp.ebi.ac.uk/pub/ensemblorganisms/Sepia_lycidas/GCA_963932145.1/ensembl/geneset/2024_05/	35949
Sepia officinalis	GCA_050097725.1	this study	https://doi.org/10.17617/1.5n7h-4385	23768

First, we investigated 36 of the gene families that contain more than 100 genes in any of the species, with 17 of these families containing at least one gene of *S. officinalis*, that reflect large-scale gene family expansions (Figure 4E). We used the InterProScan and eggNOG-mapper annotations to infer functional roles of these genes, selecting the most common gene annotation as the name of the gene family.

The zinc finger C2H2-type transcription factors (TFs) were grouped into three of the large gene families, with the largest family (OG0000000) only present in decapod cephalopods. This likely reflects the largely independent expansions in the octopod and decapod lineages that date back to a burst of transposon activity ca. 25 million years ago[46,48,49]. The largest expansion across mollusks occurs in the cadherin-like family (OG0000001): 310 in *S.*

officinalis, 283 in *D. pealeii*, 209 in *A. lycidas*, 102 in *O. vulgaris*, 55 in *O. bimaculoides*, with low but non-zero counts in bivalves (*C. virginica*, *M. gigas*). This profile is consistent with the protocadherin expansion first described in *O. bimaculoides*[46] and subsequently shown to be present across cephalopods[48,49,123].

HPGDS (OG0000005, hematopoietic prostaglandin D synthase) is a glutathione-S-transferase family member that catalyzes the conversion of prostaglandins, which have well-described roles in immune responses in vertebrates and insects[124,125]. This family shows a broad expansion in decapods, with a lesser expansion in octopods. Additionally, members of the glutathione-S-transferase families have been co-opted as S-crystallins, structural proteins found in the lens of cephalopods that may, or may not, retain enzymatic functions[126,127].

Two large families are mostly lineage-restricted. The RING-type zinc finger family (OG0000058) has 103 copies in *S. officinalis* and 26 in *A. lycidas* but is absent in all other species except for *E. scolopes*. Conversely, OG0000002 (unknown function) has 479 copies in *E. scolopes* and only a few copies in the other species. This interesting Sepiolid-specific expansion warrants further characterization.

We estimated gene family evolution rates using CAFE5[128] for all families with less than 100 copies in any species (this excludes the families described above, as very large copy-number differences between species preclude likelihood calculations under the applied birth-death model). After comparing different model parameters, we chose a gamma model with three rate categories, allowing for evolutionary rate variation among gene families. Out of the 12,895 gene families analyzed, 1,813 showed a significant ($p < 0.05$) expansion or contraction in at least one of the species. We focused our analysis on the 30 most significantly expanded families; among them were several retrotransposon-associated domains that have expanded specifically in *S. officinalis*: five families carrying Retrovirus-related Pol polyprotein domains, two Reverse transcriptase domain families, and four Ribonuclease H-like families (Supplementary Figure 7A). There was no coordinate-based overlap of the coding sequences with annotated TEs from the RepeatMasker output (Methods).

In addition to the three large gene families of C2H2 zinc finger expansions, 45 gene families containing this TF type showed a significant change in the CAFE5 analysis. Notably, eight of the significant gene families, as well as four of the largest gene families, were annotated as CCHC-type zinc fingers, which contain a “zinc knuckle” motif that is characteristic of retroviral nucleocapsid proteins[129] and is functionally integrated in the genomes of several species, including humans[130].

Some gene families without any relationship to retrotransposons were also expanded. For example, the UGT2A1-related family is a UDP-glucuronosyltransferase, a class of enzymes central to phase II detoxification and conjugation of metabolites, reported in other mollusks in the context of environmental chemical tolerance[131], and in insects in the context of pigmentation[132]. We also detected a family of homeodomain-like proteins, representing an expansion of this important TF family.

Tissue-specific expression of expanded gene families

To place the identified gene families in a functional context, we profiled their expression in the bulk RNA-seq data (taken from multiple tissues of *S. officinalis*) used originally for gene modeling (Figure 5A). Principal component analysis (PCA) revealed the largest axis of variation in gene expression to separate brain tissues from peripheral tissues, with skin being the most transcriptomically distinct (Figure 5A), consistent with the high number of tissue-specific differentially expressed (DE) genes identified in non-neural tissues (Figure 5B). We identified the genes belonging to expanded families that were differentially expressed across tissues and enriched gene ontology[133,134] (GO) terms for them to gain additional insight. The large families excluded from CAFE5 modelling and the significantly expanded families identified by CAFE5 were analyzed separately.

Eleven of the largest gene families were expressed in our data (Figure 5C) and five had enriched GO terms (Figure 5D,E). Among them, the cadherin family showed brain-restricted expression and GO terms related to cell–cell adhesion and calcium binding, consistent with their role in neuronal connectivity and circuit formation[46,135]. Two C2H2 zinc finger gene families were expressed in the optic and vertical/subvertical lobes of the brain and in the skin, with GO terms related to DNA-binding, transcriptional regulation or development. The RING-type zinc finger family was expressed specifically in the skin, with GO terms including zinc binding and ubiquitin protein ligase activity, the canonical function of RING-domain E3 ligases[136]. Genes of the HPGDS/S-crystallin family were expressed in the brain (basal and optic lobes and posterior subesophageal mass) and skin, with GO terms related to glutathione metabolism, matching their described enzymatic function. We did not find expression in the retina, which is expected given that S-crystallins are expressed in lentigenic cells of the eye[42,137] and these cells were not included during sampling.

Among the 30 most significantly expanded families examined (out of 1,813 total), expression was widespread (20/30) and tissue-specific differential expression was common (17/30), suggesting that a substantial proportion of expanded paralogs represent functional coding sequences with specialized spatial deployment (Supplementary Figure 7B). Ten of the retrotransposon-associated families were differentially expressed in the brain (optic and vertical/subvertical lobes) and skin, arguing against these loci being inactive repeat fragments and supporting their inclusion as transcribed gene models. Two significantly expanded families showed both differential expression and enriched GO terms (Supplementary Figure 7C). The first was the UGT2A1-related family, which had the largest number of differentially expressed genes overall, with expression concentrated in the skin, retina and posterior subesophageal mass of the brain. Enriched GO terms matched the described enzymatic function for this family, namely UDP-glycosyltransferase activity. The second gene family was the homeodomain-like family with enrichment for DNA binding terms consistent with their role as transcription factors, and was preferentially expressed in the vertical and subvertical brain lobes with weaker expression in other areas.

Collectively, many differentially expressed genes from expanded families were restricted to specific tissues or brain subregions (Figure 5F and Supplementary Figure 7D), indicating that paralogs within an expanded family have adopted distinct spatial expression domains and possibly, specialized functions.

Discussion

This study presents a detailed, chromosome-scale genome assembly and annotation for the common cuttlefish *Sepia officinalis*, providing an additional important resource for comparative genomics, neurobiology, and evolutionary studies within cephalopods and mollusks. The assembly was derived from a combination of PacBio HiFi long reads, Dovetail Omni-C chromatin conformation capture, and multiple rounds of scaffolding and manual curation. The resulting assembly has a haploid genome size of 5.68 Gb, and contains 47 chromosome-like clusters. This number prompted us to perform detailed analyses of chromosome structure and homology across various decapod genomes, leading to a likely consensus of 46 chromosomes. This resource should be valuable for future efforts in cephalopod genomics, especially within the Decapodiformes, where large and repetitive genomes have hindered previous efforts.

Chromosome Number and Karyotypic Variation

Our assembly contains an estimated haploid chromosome number of 47, a result produced by two independent scaffolding approaches (YAHS[82] and HapHiC[81]) and supported by Hi-C contact maps. This karyotype, however, differed from those reported for other decapod cephalopods, such as *Euprymna scolopes* and *Doryteuthis pealeii*[49], each with 46 chromosomes. Syntenic comparisons across datasets revealed that chromosomes 40 and 43 in *S. officinalis* correspond to a single chromosome in these other species, suggesting that the apparent split in *Sepia* may result from a technical artefact.

We also compared our results to the recently published *S. officinalis* genome from the Darwin Tree of Life project (DToL)[75], which proposed 49 chromosomes for that species. This difference was surprising: two genome assemblies from the same species are usually expected to have the same karyotype. The discrepancy is attributable to one chromosome split in our assembly and three splits in the DToL assembly, and could stem from technical differences of the Hi-C methods (Dovetail Omni-C vs. Arima v2), sequencing depth (11-fold vs. 83-fold coverage) or the tissue used (optic lobe vs. eye). We investigated these splits, or breakpoints in two ways: first, we reassembled both datasets with the same parameters and aligned both to the genome of another cuttlefish *A. esculentum* (1n=46). In this alignment, all four breakpoints were collinear with single *A. esculentum* chromosomes, providing a phylogenetically grounded prediction of 1n=46 for *Sepia officinalis.* Second, we investigated read depth at the breakpoint junctions, and saw a significant increase of Hi-C read pairs spanning the junctions compared to random chromosome pairings. Together with a higher repeat content at the breakpoints, these analyses suggest a technical cause for the differences between assemblies, caused by repeat-rich scaffold boundaries that impaired HiFi and Hi-C read alignment and in turn, made correct assembly challenging in these regions.

In conclusion, the two independent genome assemblies and our analysis suggest three possible karyotypes for *Sepia officinalis*: 46, 47, or 49. The most likely explanation based on chromosome synteny is that the true number is 46, matching the numbers established for other cuttlefish and decapod species[42,49,84]. 47 chromosomes (our initial estimate) and 49 chromosomes (DToL estimate) would therefore be erroneous values explained by technical factors related to repeat content and low alignment rates preventing complete assembly.

Incorporating ultra-long read data may help to correctly assemble these problematic regions, as is now common for telomere-to-telomere assemblies[138].

Another, but less likely, explanation for the observed differences could be variations in chromosomal architecture of different *Sepia officinalis* populations, due to idiosyncratic individual chromosome fusion. Intraspecific chromosome number variation is rare but not unprecedented; for instance, chromosomal polymorphism has been described in the butterfly *Leptidea sinapis* across different populations[139,140]. Notably, the two *S. officinalis* individuals used for sequencing originated from different regions - the Portuguese Atlantic (this study) and the French Mediterranean (DToL project), raising the possibility of geographic variation in chromosome structure. Future population-level analyses will hopefully determine whether *S. officinalis* exhibits karyotypic polymorphism.

Additional methods such as cytogenetic karyotyping or optical mapping such as BioNano[141] (imaging of fluorescently tagged, linearized DNA) could be used to validate chromosome numbers. However, whereas karyotypes of octopuses have been consistent throughout the literature (1n=30)[142,143], those measured in decapods vary greatly. For example, 1n=46 chromosomes have been reported for two species of cuttlefish (*A. esculentum* and *A. lycidas*) and three loliginid squids[85]; 1n=36 has been reported for *A. arabica*[86] and 1n=24 in *A. pharaonis*[87]. In *S. officinalis*, a karyotype of 1n=52 is reported for testis samples[88]. Combining cytogenetic preparations with fluorescent labeling of centromeric or telomeric sequences, as demonstrated in the octopus *A. aerolatus*[143] could help resolve these issues. Establishing a routine staining protocol would enable comprehensive tests at the species- and population-level.

Taken together, our results illustrate the difficulty of assembling large genomes with high repeat content and large karyotypes, at least from sequencing data alone. Internal validation methods and genome comparisons across species are therefore important. Convergence of reliable estimates will, in turn, help identify chromosomal fusion-with-mixing events (FWM; fusion of two ancestral chromosomes followed by extensive shuffling of their gene content) that are clade specific. Early branching order in Decapodiformes has been notoriously unstable[53,84,94,144–147]; thus, such rare and irreversible FWM characters could be useful in further phylogenetic analysis of this clade[51,148].

In addition to studying chromosomal topology in phylogenetic reconstructions, some of the most interesting aspects of these rearrangements relate to changes of and innovation in regulatory elements that underlie phenotypic diversity. In coleoid cephalopods, it is thought that an ancient large-scale genome rearrangement was combined with lineage-specific changes and repeat expansions[48–50]. This restructuring gave rise to hundreds of tightly linked, evolutionarily unique microsyntenies, corresponding to distinct topological compartments with specialized regulatory architectures that contribute to complex, tissue-specific expression patterns in the nervous system and elsewhere[43]. Extending this, chromosomal conformation analyses in *E. scolopes* revealed that co-regulated eye and light-organ genes cluster at topologically associating domain (TAD) boundaries, and that an evolutionarily recent rearrangement at the dachshund (DAC) locus may have been instrumental in the emergence of the symbiotic light organ in *Euprymna* - directly linking specific chromosomal topology to morphological innovation[44].

To understand the broader functional impact of these changes across coleoids, a recent study investigating Micro-C, RNA-seq, and ATAC-seq data from multiple species revealed broadly conserved chromatin domains, but also many lineage-specific chromatin loops that form novel regulatory signatures and impact expression profiles across species and tissues[149].

Despite the observed small-scale regulatory changes, the chromosomes of decapods are considered to be more closely related to the ancestral coleoid karyotype than those of octopods. The derived octopod karyotype becomes apparent when comparing it to the genome of the vampire squid, an early-branching octopodiform (sister to all octopods) which retained features of the decapod, ancestral karyotype[150]. Taken together, the conserved karyotype of decapods accommodates fine-scale regulatory diversity that might underlie morphological diversity among species, which suggests that many regulatory innovations are still being evolutionarily explored through rearrangements within the existing chromosomes.

Genome Size and Repeat Landscape

The size of the genome of *S. officinalis* is estimated to be 5.68 Gb - comparable to those of other Decapodiformes and roughly twice the size of typical Octopodiformes genomes. We found that over 71% of the genome is repetitive, the repeats being dominated by unclassified elements and retrotransposons. The near absence of SINEs reinforces the idea that their expansion in octopuses was a lineage-specific event[50]. These results underscore the evolutionary complexity and dynamism of cephalopod genomes, shaped by waves of transposon activity that likely played a role in the evolution of novel traits[43,49].

Gene Annotation and Comparative Assessment

We annotated 18,663 gene models and 23,768 proteins using a combination of short- and long-read RNA-seq data and molluscan protein references. Compared with other annotated cuttlefish genomes (*A. lycidas*, *A. pharaonis*), our annotation is conservative concerning gene numbers but stands out in terms of completeness and taxonomic consistency. OMArk-based evaluations confirmed that our gene models have low levels of fragmentation and contamination, and high lineage-specific coherence. Striking differences across species in the number of predicted genes, duplications, and missing orthologs likely reflect variation in the annotation pipelines and reference data rather than true biological differences. Note that the comprehensive annotation of protein isoforms remains challenging even for model organisms[151,152]. Still, our dataset provides a solid foundation that may be refined with additional (long-read) transcriptomic data taken from diverse tissues and developmental stages.

We characterized the *S. officinalis* gene models further by clustering them into families together with 12 other cephalopod and non-cephalopod mollusks. As in other coleoid genomes, we observed large-scale expansions of gene families such as C2H2 zinc finger transcription factors and protocadherins, genes implicated in regulatory innovation, neural development and plasticity[46,48,49,123]. These conserved expansions may be linked to the unusual cognitive complexity and behavioral sophistication of these invertebrates.

We profiled the expression of expanded gene families in bulk RNA-seq data from multiple *S. officinalis* tissues and found that many genes were differentially expressed, frequently in a tissue-restricted manner. This spatial partitioning of paralogs across tissues validates many expanded families as likely functional and biologically relevant, and suggests that subfunctionalization or neofunctionalization may have accompanied gene family expansion in *S. officinalis*, with individual family members acquiring specialized roles[153,154]. Note that we only investigated the expression of the largest copy number families, and the most significantly expanded families. Whether this pattern generalizes to the broader set of expanded families remains to be determined.

Among the large families, the cadherin and C2H2 zinc finger families showed preferential expression in neural tissues, consistent with known roles in synaptic organization and transcriptional regulation of neural development, respectively. The skin-specific expression of the RING-type zinc finger family is notable given its restricted expansion in decapods. RING domains are characteristic of E3 ubiquitin ligases that catalyze protein ubiquitination and target substrates for proteasomal degradation[155,156]. Their expression could hint at unique demands in protein turnover in the skin, but the specific function of this gene family in cuttlefish skin remains to be determined.

Many of the smaller but significantly expanded gene families, which were predominantly found in decapod genomes, were linked to retrotransposons, like retrovirus-related Pol polyprotein, reverse transcriptase domain, and ribonuclease H-like families. We confirmed that these genes were genuine coding sequences and not artifacts of TE repeats, and that they were expressed in *S. officinalis* tissues. The combination of coding-sequence annotation, expression, and tissue specificity is consistent with at least some of these loci having been retained as functional genes after retroelement-related origins.

Notably, we observed many expanded families containing CCHC-type zinc fingers, which contain a “zinc knuckle” motif, characteristic of retroviral nucleocapsid proteins[129]. In addition to the retroviral function, some eukaryotic proteins containing CCHC-domains have been described, which likely originate from domesticated nucleocapsid sequences and play important roles in RNA metabolism[130,157]. Taken together, these gene families may represent a recent wave of retroelement activity, and the variation in copy number across different Decapodiformes is consistent with repeated recruitment and/or independent retention of retroelement-derived sequences in different decapod lineages.

Two non-retroviral gene families were also differentially expressed in *S. officinalis* tissues: a UDP-glucuronosyltransferase (UGT) family and a homeodomain-like family. The expression of the UGT family, particularly in skin and retina is interesting: enzymes of the UGT family are involved in pigment metabolism in insects[132], but have also been reported in mollusks as playing roles in environmental chemical tolerance[131]. The homeodomain-like family contains transcription factors with important roles in body patterning and brain regionalization[158]. We found members of this family to be expressed in the vertical and subvertical lobes of the brain, but also more weakly in other brain areas, suggesting novel regulatory roles of this transcription factor class in higher-order neural circuits in brain areas associated with learning and memory in cephalopods[18].

A potential caveat of this analysis is that the gene modeling approaches used for the species were different: while the majority of genomes were annotated with the RefSeq pipeline[118], others used Ensembl genebuild[159], BRAKER[80,99–114] or other custom approaches. This may have introduced artificially inflated gene counts due to insufficient isoform resolution, or missing gene models due to the absence of reference data. In the future, as genome annotations become progressively refined (as for *D. pealeii*[160]), gene family-level analyses will become more powerful and enable the identification of additional lineage- or species-specific innovations.

Conclusions and Outlook

This study provides a new genomic resource for *Sepia officinalis*. The chromosome-scale assembly and annotation open the door to in-depth studies of gene regulation, neural circuit development, and genome evolution across coleoid cephalopods.

As we move toward a more complete picture of cephalopod genome evolution, the integration of chromosomal synteny, regulatory architecture, and transcriptomic diversity across species will be especially important. The *S. officinalis* genome represents an important step on this path, enabling high-resolution comparative and functional analyses of one of the most enigmatic and evolutionarily successful invertebrate lineages.

Materials & Methods

Animal husbandry

All research and animal care procedures were carried out following the institutional guidelines that comply with national and international laws and policies (DIRECTIVE 2010/63/EU; German animal welfare act; FELASA guidelines). European cuttlefish *Sepia officinalis* were hatched from eggs collected in the Portuguese Atlantic and reared in a seawater system, at 20°C. The closed system contains 4,000 L of artificial seawater (ASW; Instant Ocean) with a salinity of 33‰ and pH of 8–8.5. Water quality was tested daily and adjusted as required. Trace elements and amino acids were supplied weekly. Marine LED lights above each tank provided a 12/12-h light/dark cycle with gradual on- and off-sets at 07:00 and 19:00. The animals were fed live food (either *Hemimysis* spp. or small *Palaemonetes* spp.) *ad libitum* twice per day. The animals were housed together in 120 L glass tanks with a constant water through-flow resulting in five complete water exchanges per hour. Enrichment consisted of natural fine-grained sand substrate, seaweed (*Caulerpa prolifera*), rocks of different sizes and various natural and man-made three-dimensional objects.

Tissue preparation

Animals underwent terminal anesthesia to prevent animal suffering or distress, following the Guidelines for the Care and Welfare of Cephalopods in Research[161,162]. Animals were transferred to a bucket containing ethanol 2% (v/v) in ASW. After 5 minutes, animals were

gently probed for simple avoidance reflexes, and the ethanol concentration in the ASW was gradually increased to a maximum of 5%. Sufficient depth of anesthesia was reached when the animal no longer reacted even to stronger stimuli (touching and pinching with tweezers), and no reaction was observed when a hand was moved in the visual area and the cornea was touched[163]. In this deep anesthesia, the animals were decapitated and the tissue was rapidly transferred into ice-cold, calcium-free Ringer solution (460 mM NaCl, 10 mM KCl, 51 mM $MgCl_2$, 10 mM glucose, 2 mM Glutamine, 10 mM HEPES, pH 7.4) bubbled with oxygen. Brain and body tissue was dissected under a stereoscope and flash-frozen in liquid nitrogen before storage at -80ºC until further use.

Unless stated otherwise, sequencing libraries were prepared from the same individual (6-month-old adult *Sepia officinalis*, F1 from eggs collected in Portugal).

Long-read Whole Genome Library Preparation and Sequencing

The long-read library was prepared and sequenced at the MPI for Plant Research in Cologne, Germany. Genomic DNA was extracted from flash-frozen brain tissue with the MagAttract HMW DNA kit (Qiagen, Cat. no. 67563) and the sequencing library was prepared with the SMRTbell express template prep kit 2.0 (PacBio, Cat. no. 100-938-900). The library was sequenced on 5 SMRT cells on the PacBio Sequel II with the Sequel II binding kit 2.2 (102-089-000), Sequel II sequencing kit 2.0 (101-820-200) and SMRT Cell 8M tray (101-389-001).

Long-read RNA library preparation and sequencing

RNA was isolated from various flash-frozen tissues (different brain areas, mantle/epidermis, arm/tentacle; 5-10 mg each). First, tissue samples were homogenized in TRIzol using a homogenizer (VDI12/S12N-5S, VWR, Germany). RNA was isolated and DNAseI treated according to the Direct-zol RNA Mini prep kit from ZymoResearch (R2050). RNA was quantified by spectrophotometry (Nanodrop, Thermo Scientific, USA) and its quality was assessed with a RNA 6000 Nanochip on Bioanalyzer (Agilent Technologies, Germany).
The Iso-Seq libraries were prepared and sequenced at the Sequencing facility of the MPI for Plant Research, using a method targeting the 5' cap and poly-A tail of mRNAs, described in detail in[164]. Briefly, mRNA was pooled from the 8 tissue samples and cDNA was synthesized using the TeloPrime Full-Length cDNA Amplification Kit V2 (Lexogen, Cat Nr: 013.08 or 013.24). The sequencing library was prepared with the SMRTbell express template prep kit 2.0 (PacBio, Cat. no. 100-938-900). The pooled library was sequenced on 1 SMRT cell on the PacBio Sequel II system using the Sequel II binding kit 2.1 (101-843-000), Sequel II sequencing kit 2.0 (101-820-200) and SMRT Cell 8M tray (101-389-001).
Further, a separate library was prepared from optic lobe cDNA and sequenced on a second SMRT cell using the same reagents described above.

Omni-C Library Preparation and Sequencing

An Omni-C library was prepared and sequenced at the MPI for Plant Research in Cologne, Germany. The library was prepared from brain tissue using the Dovetail Omni-C™ Kit (Dovetail Genomics, Cat. No. 21005) according to the manufacturer's protocol. Briefly, chromatin was fixed in place in the nucleus. Fixed chromatin was digested with DNaseI then

extracted. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed and the DNA was purified from proteins. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The sequencing library was generated using Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of the library. The library was sequenced on an Illumina NextSeq 2000 platform to generate 400 million 2 x 150 bp read pairs.

Short-read DNA sequencing

High molecular weight genomic DNA was isolated from brain tissue (optic lobe) using NEB Monarch gDNA Purification Kit (T3010S). First, 10 mg of flash-frozen tissue were submerged in 100 µl of tissue lysis buffer, cut into small pieces and transferred into a 1.5 ml tube containing an additional 100 µl of lysis buffer. Then, 3 µl of 20 U/µl Proteinase K was added to the tissue sample and incubated for 2 h at 56°C in a thermal mixer with agitation until tissue pieces were completely dissolved. After incubation, tissue lysate was purified from the debris by centrifugation for 3 min at maximal speed (Eppendorf, Germany). Clean supernatant was transferred to a new 1.5 ml tube and incubated with 3 µl RNase A for 5 min at 56°C with agitation at full speed. After RNase A incubation tissue lysate was column purified and eluted with 80 µl of elution buffer according to the manufacturer protocol. The purity of gDNA was assessed using A260/280 and A260/230 absorbance ratios measured by NanoDrop spectrophotometer (ThermoFisher Scientific). The integrity of the obtained gDNA was checked using 0,75% agarose gel electrophoresis by loading 100 ng of gDNA sample and using HindIII digested Lambda DNA (NEB#3012) as a marker.
DNA sequencing libraries were prepared from 500 ng of high MW gDNA using Illumina DNA PCR-Free Tagmentation Library Prep Kit (20041794) and UD Indexes Set A (20026121). Obtained dual-indexed single-stranded libraries were quantified using the Qubit single-stranded DNA (ssDNA) assay kit (Q10212, ThermoFisher Scientific). 1000 pM of the final library were run on P3 reagents and Illumina NextSeq2000 using Illumina DNA PCR-Free Sequencing and Indexing primer (20041797) with 151 bp paired-end reads.

Short-read RNA library preparation and sequencing

For short-read RNA sequencing, tissue from another animal (8-month-old adult, F0 from eggs collected in Normandie, France) was used. RNA was isolated from various flash-frozen tissues (different brain areas, skin and retina; 5 mg each). First, tissue samples were homogenized in TRIzol using a homogenizer (VDI12/S12N-5S, VWR, Germany) following DNAseI treatment and column purification according to the Direct-zol RNA Micro prep kit from ZymoResearch (R2062). The RNA integrity and quantity were measured with the Qubit fluorometer (Invitrogen, Q33216) and the 2100 Bioanalyzer (Agilent Technologies, Germany).
RNA-seq libraries were prepared from 300 ng of total RNA, using the Illumina TruSeq stranded mRNA library prep kit (Cat:20020594) and IDT for Illumina xGen UDI-UMI Adapters (Cat:10005903). Libraries were sequenced on an Illumina NextSeq500 (Mid output, 300 cyc) and NextSeq2000 (P3, 300 cyc) with 145 bp paired-end reads.

Nuclear genome assembly

Before assembly, PacBio HiFi reads were subjected to additional trimming to remove residual adapter sequences using NCBI's VecScreen tool. The k-mer distribution was estimated using Meryl[165] within the Merfin[166] package with a k-mer size of 21, and genome size was estimated using GenomeScope[77] from Illumina short reads and PacBio HiFi data. The primary assembly was generated with hifiasm[167] using a combination of HiFi and Hi-C reads. Scaffolding was performed iteratively with YAHS[82] on the phased haplotype 1. We adjusted the run parameters for scaffolding resolutions (-r 1000,2000,5000,10000,20000,30000,50000,70000,100000,150000,200000,250000,300000,350000,400000,500000,700000,1000000,2000000,5000000,10000000,15000000,20000000,30000000,40000000,50000000,60000000,70000000,80000000,100000000,110000000,120000000,150000000,170000000,200000000,500000000), repetitions per resolution (-R3), minimum read mapping quality (-q1) and telomeric sequence (--telo-motif "TTAGGG") from the only experimentially determined telomere motif for cephalopods (*Amphioctopus areolatus*[143], accessed via TeloBase[168]) to optimize the tool's performance for a large, fragmented assembly. Finally, the assembly was manually curated using JBAT[83] to place residual scaffolds into chromosome scaffolds. Different versions of the assembly were evaluated based on Hi-C coverage, mRNA alignment and completeness based on BUSCO[80] v5.5.0 (*metazoa_odb10* and *mollusca_odb10*).

Mitochondrial genome assembly

To assemble the mitochondrial genome, we aligned a published *S. officinalis* mitochondrial genome[169] (NC_007895.1) to the PacBio Hi-Fi and Omni-C reads using minimap2[170]. The hits were extracted using seqtk[171] subseq and assembled using hifiasm[167]. The resulting contigs contained multiple repeats of the circular mitochondrial genome, and were aligned back to NC_007895.1 to extract the final sequence.

Nuclear genome annotation

Repetitive elements in the genome were softmasked using RepeatMasker v4.1.7-p1[93] (with rmblast v2.14.1+) (-xsmall and -gff options) and after creating a custom repeat library with RepeatModeler v.2.0.6[97] (without LTRstruct option).

Gene models were created using BRAKER3[80,99–114], run via the Docker container, on the softmasked genome using both RNA-seq and protein data. We used long-read Iso-Seq data and short-read Illumina RNA-seq data from various tissues (brain, skin, mantle, retina) generated for this study (see above). For protein data, publicly available proteomes for *Doryteuthis pealeii*[49], *Euprymna scolopes*[115], *Octopus bimaculoides*[49], *Octopus vulgaris*[47], *Nautilus pompilius*[116], *Pecten maximus*[54] were used for training, as well as a previously generated proteome for *Sepia officinalis* from StringTie gene models (this study, described below). RNA-seq data was input into BRAKER3 using the --bam option, protein files were specified with the --prot_seq option. Untranslated regions (UTRs) were added by the -addUTR=on parameter. The BUSCO completeness of the resulting gene set was maximized using TSEBRA within BRAKER on the BUSCO lineage *metazoa_odb10*.

For the initial protein set used as input for BRAKER, the short and long RNA reads were aligned to the genome using *minimap2*[170]. Transcript models were predicted with StringTie

v3.0.0[172] using the *--conservative* and *--mix* options. The resulting GTF files were combined using the transcript merge mode, resulting in a set of non-redundant transcripts. Finally, TransDecoder v5.7.0[173] was run with default parameters to translate coding regions in the transcripts.

The final gene annotations for *S. officinalis* were assessed for completeness using OMArk v.0.3.0[117] with the ancestral clade Lophotrochozoa and without including splice information, accessed via their webserver. Proteins were annotated with orthology information using InterProScan[119] v5.73-104 including lookup of annotations and GO terms (options -iprlookup -goterms). Further orthology information was added using the eggNOG-mapper v2.1.12[120] webserver with the eggNOG v5.0 database and default parameters.

Whole genome alignment and synteny analysis

For whole genome alignments, the assembly produced for *Sepia officinalis* in this study and the published assembly from the Darwin Tree of Life project (GCA_964300435.1) or the *Acanthosepion esculentum* genome (GCA_964036315.1) were aligned using Winnowmap2[89] and visualized with a custom script in R v4.4.2[174].

For synteny analyses, proteome and gtf files were downloaded for *Doryteuthis pealeii* and *Euprymna scolopes.* Annotation files for *Acanthosepion esculentum* were recently generated by liftover annotation from *Acanthosepion pharaonis*[96] and downloaded from Zenodo[95].

Synteny analyses between all chromosomes of the compared species were performed using the R package GENESPACE v.1.2.3[175] with default parameters, described briefly below. Protein sequence similarity was first estimated using DIAMOND2[109] in fast mode, and orthogroups and pairwise orthologues were inferred using OrthoFinder v2.5[176] with hierarchical orthogroups (HOGs) enabled. Prior to synteny inference, tandem arrays were condensed to their most central representative gene, and gene rank order was recalculated on these array-representative genes to reduce confounding effects of tandem duplication on collinearity detection.
Syntenic blocks were identified pairwise between all genome combinations using MCScanX[177], constrained to DIAMOND hits where both query and target genes belonged to the same orthogroup (onlyOgAnchors = TRUE). Initial anchor hits were clustered into large syntenic regions using a density-based spatial clustering approach (dbscan[178]), with a minimum block size of five anchor genes (blkSize = 5) and a maximum of five intervening non-anchor genes permitted within a block (nGaps = 5). Anchor clustering used a search radius of 25 gene-rank positions (blkRadius = 25). All hits falling within a syntenic buffer of 100 gene-rank positions around confirmed block anchors (synBuff = 100) were retained as syntenic. No secondary syntenic hits were included (nSecondaryHits = 0). Syntenic orthogroups were integrated across all pairwise comparisons and collapsed into a pan-genome annotation anchored to *S. officinalis* as the reference genome.
Syntenic relationships were visualized as riparian plots and pairwise dotplots using the built-in plotting functions of GENESPACE v1.2.3. Riparian plots were constructed using physical chromosomal coordinates (useOrder = FALSE) with *S. officinalis* as the reference, displaying all three genomes. A second riparian plot was generated highlighting a region of interest. Pairwise dotplots were produced for the *S. officinalis–D. pealeii* and *S. officinalis–E. scolopes* genome comparisons, displaying only synteny-validated hits (type = "syntenic")

with a minimum synteny score of 10 (minScore = 10) and a minimum of 10 genes per chromosome pair required for display (minGenes2plot = 10).

Coverage analysis

For the analysis of breakpoints between *S. officinalis* assemblies, Hi-C data was aligned using bwa-mem2 v2.3[91] and quantified using pairtools v1.1.0[92]. Contacts between scaffolds (trans pairs) were extracted from deduplicated read pairs (pair type UU). For each of the four breakpoints, all trans pairs between the two flanking scaffold halves were extracted using awk. As controls, the corresponding joined scaffold from the opposing assembly was paired with the nearest-length uninvolved scaffold from the same assembly, and trans pairs were extracted identically.

Genome-wide background and intra-scaffold contact distributions were computed in a single pass over each pairs file using awk, recording the total number of trans pairs and intra-scaffold pairs (>1 kb separation) for all scaffold pairs across the genome. Trans contact rates were normalized by the product of the two scaffold lengths (pairs per Mb^2) to correct for length bias. For visualization, genome-wide trans pair positions were binned into 500 kb windows along each scaffold to produce contact density tracks. To test whether trans contact rates at breakpoints were elevated above background, empirical p-values were computed as the fraction of all genome-wide scaffold pair rates equal to or exceeding the observed breakpoint rate. For the three DToL breakpoints jointly, a one-tailed Wilcoxon rank-sum test was applied against the background distribution.

HiFi reads were aligned to the scaffolded assemblies using minimap2[170] and duplicate-marked alignments were removed. HiFi read depth was computed over the terminal 200 kb at each scaffold end using pysam v0.22.1[179] count_coverage with a mapping quality threshold of ≥10, binned at 1 kb resolution. Spanning reads - defined as reads with supplementary alignments crossing a scaffold boundary - were identified by querying split alignments at each junction using pysam.

Repeat content at scaffold junctions was characterized by extracting 200 kb windows flanking each breakpoint and the corresponding correct scaffold termini, and running RepeatMasker v4.1.7-p1[93] (with rmblast v2.14.1+) against a *de novo* repeat library generated by RepeatModeler v.2.0.6[97] from the same assembly (described above). Repeat annotations were parsed and collapsed into eight classes (SINE, LINE, DNA transposons, LTR elements, simple repeats, low-complexity regions, unknown repeats, and other), and the masked fraction per class was quantified for each window.

For sex chromosome analysis, read coverage across chromosomes was analyzed as described recently[96]. Short reads were aligned using STAR v2.7.11b[180] to our chromosome-scale assembly. For *A. esculentum*, short reads (ERR12945500) and assembly (GCA_964036315.1) were downloaded from NCBI. The sequencing depth was calculated using mosdepth[181] by a window size of 500,000 bp and normalized to the median coverage of the first chromosome. Chromosomes with significantly reduced read coverage were identified by a one-sided Wilcoxon rank-sum test of each chromosome's normalized depth windows against all remaining chromosomes. P-values were corrected using the

Benjamini-Hochberg method and considered only for chromosomes with at least 10% decrease in median normalized depth.

Gene family expansion analysis

Orthogroups were inferred across 13 molluscan species (Table 2), including *S. officinalis*, using OrthoFinder v3.1.0[122] with default parameters. The input proteomes included the longest protein isoform per gene for each species. The rooted species tree from OrthoFinder[182,184] was converted to an ultrametric tree using the R package ape[183] v5.8.1.

Gene families were filtered by removing orthogroups present in only a single species, and by separating orthogroups containing 100 or more gene copies in any species, as extreme copy-number differences in gene families prevent likelihood calculation under the applied birth-death model.

Gene family evolution rates were estimated using CAFE5[128] v5.1.1 on the filtered orthogroups, using the ultrametric species tree as input. Four models were evaluated: the base model (single global lambda), and Gamma models with k = 2, 3, and 4 rate categories, which allow evolutionary rate variation among gene families. The Gamma k = 3 model was selected based on the best (lowest) final log-likelihood score. All subsequent statistical inferences were performed under this model.

For families showing statistically significant expansion or contraction ($p < 0.05$ after Bonferroni correction), branch-specific copy-number changes were extracted from the CAFE5 output. Families were categorized as *S. officinalis*-specific, coleoid-specific, or broad expansions based on the distribution of significant changes across the phylogeny.

To assess whether expanded gene families in *S. officinalis* contained genes derived from or embedded within repetitive elements, a coordinate-based overlap analysis was performed. For each gene in an expanded orthogroup, the overlap between its coding sequence (CDS) coordinates and RepeatMasker annotations was computed using *bedtools intersect* v2.30[185]. To avoid double-counting when multiple repeat annotations overlapped the same coding bases, overlapping repeat intervals were merged per gene prior to summing covered bases, and the overlap fraction was computed as merged covered bases divided by total CDS length.

Bulk RNA-seq analysis

Quality-filtered paired-end RNA-seq reads were aligned to the *S. officinalis* genome assembly using STAR[180] (v2.7.11b) with the following parameters: --outSAMmultNmax 1 (retaining only uniquely mapping reads), --outFilterIntronMotifs RemoveNoncanonical (removing reads with non-canonical splice junctions), and --outSAMtype BAM SortedByCoordinate.
Gene-level read counts were quantified using featureCounts[186] from the Subread package (v2.0.8) with the gene annotation described above (GTF format). Counting was performed at the exon level (-t exon) and summarized by gene (-g gene_id), using paired-end mode (-p --countReadPairs) to count fragments rather than individual reads. Only reads with mapping quality ≥ 255 (-Q 255), corresponding to uniquely mapped reads in the STAR output, were included in the count matrix. The resulting gene count matrix was used as input for differential expression (DE) analysis.

Tissue-specific marker genes were identified using DESeq2[187] (v1.42.0) with a one-vs-all comparison strategy. For each tissue, samples were grouped into a binary condition ("target" vs. "other"), where "target" represented the tissue of interest and "other" comprised all remaining tissues. Raw count matrices were filtered to retain genes with ≥10 counts in at least 3 samples prior to analysis. DE testing was performed using the Wald test, log2 fold changes were shrunk using the apeglm[188] method. Genes were classified as tissue markers if they met the following criteria: adjusted p-value (Benjamini-Hochberg FDR) < 0.05, log2 fold change > 1, and mean expression in the target tissue > 5.

To generate gene-level GO[133,134] annotations for the *S. officinalis* genome, GO term assignments from the InterProScan output (detailed earlier) were used. First, transcript IDs were collapsed to gene IDs and for genes with multiple transcripts, GO terms were aggregated using a union strategy, collecting all unique GO terms across all isoforms and annotation sources to maximize coverage. GO terms present in fewer than 5 genes or more than 500 genes in the expressed gene universe were excluded from enrichment testing to reduce noise from overly specific or generic terms.

GO enrichment analysis was performed using the clusterProfiler[189,190] (v4.12.6) enricher function with custom GO annotations generated as described above. For each gene set (DE genes for each tissue), enrichment was tested against the background universe of all expressed genes (genes passing the low-count filter in DESeq2). Over-representation was assessed using the hypergeometric test, with p-values adjusted for multiple testing using the Benjamini-Hochberg procedure. GO terms with adjusted p-value (FDR) < 0.05 were considered significantly enriched.

S. officinalis members of each expanded gene family were cross-referenced against tissue-specific marker genes identified by DESeq2 (as described earlier). For families with at least one differentially expressed member, GO enrichment was performed using enricher() from clusterProfiler as described for tissue marker genes. For families with at least one enriched GO term, expression patterns of differentially expressed members were visualized as a heatmap of z-scored VST values, with samples grouped by tissue.

Data availability

The genome assembly and raw data can be found at the BioProject PRJNA1091451 on NCBI. Raw sequencing reads are deposited at SRA (study accession SRP570862). The code for the genome assembly and annotation is available at https://gitlab.mpcdf.mpg.de/mpibr/laur/cuttlefishomics/soffgenome.
Genome annotation files are deposited at https://public.brain.mpg.de/Laurent/sepoff/annotation/.

Author contributions

O.S., and G.L. conceived of the study design. S.R., O.S. and G.L. wrote the first draft of the manuscript. D.H. collected tissue for sequencing. E.C. and S.R. extracted DNA and RNA

from tissue, E.C. performed Illumina sequencing. G.T. performed genome assembly. G.T. and S.R. manually curated and annotated the genome. S.R., G.T., and O.S. performed genomic analyses. S.R. created the figures. S.R., O.S. and G.L. coordinated and led the project. G.L. acquired funding for this study.
All authors contributed to the review and editing of the manuscript.

Acknowledgments

We thank Bruno Huettel for PacBio and Omni-C sequencing. We thank Xitong Liang, Theodosia Woo and Mathieu Renard for help with tissue dissection. We thank Darrin Schultz, Dalila Destanović and Thea Rogers for helpful discussions and advice on manual curation and gene modeling. We thank Victor Nieto Caballero for initial discussions on gene family expansion analysis.

This work was funded by the Max Planck Society (GL) and the European Research Council (GL; ERC grant CAMOUFLAGE, 101141501). O.S. was supported by the ERC's Horizon 2020: European Union Research and Innovation Programme, grant No. 945026.

All authors declare no conflict of interest. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Figures

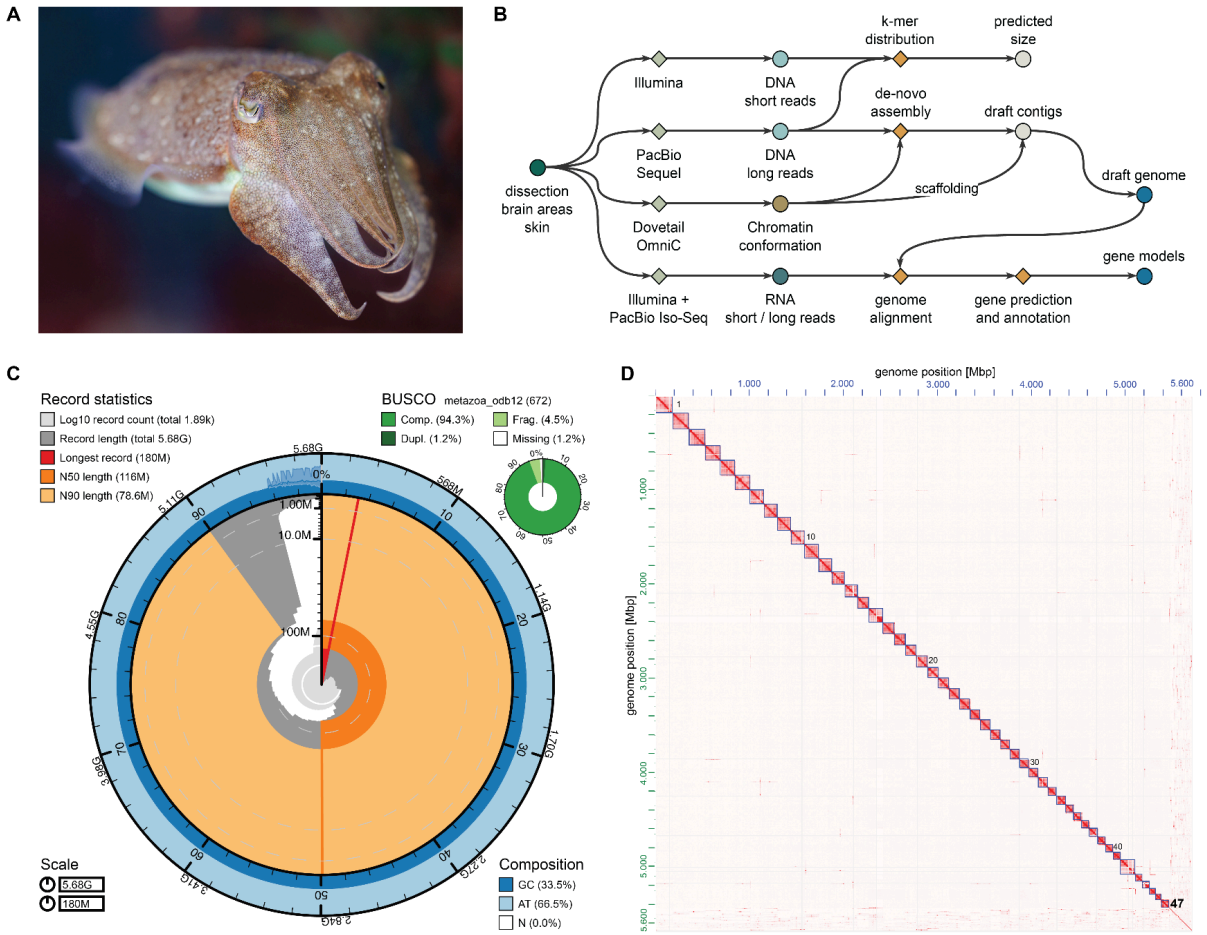


Figure 1: *Sepia officinalis* assembly statistics and quality control. A) Specimen of *S. officinalis* (credit: Stephan Junek, MPI for Brain Research). **B)** Overview of the genome assembly workflow. Genome size was estimated from short DNA reads (Illumina) using GenomeScope[77,78]. The primary assembly was generated from long DNA reads (PacBio Sequel II) and chromatin conformation capture (Hi-C) reads (Dovetail OmniC) with hifiasm[167]. Assembly was scaffolded with YAHS[82] and residual small scaffolds were manually placed in chromosomes. **C)** Snail plot of chromosome-scale *S. officinalis* assembly generated using blobtools2[191] showing scaffold statistics (e.g. number of scaffolds, median scaffold length N50), base composition and completeness measured using Benchmarking Universal Single-Copy Orthologs (BUSCO)[80] against the *metazoa_odb12* database. **D)** Hi-C heatmap showing the 47 chromosome-scale scaffolds with few sequences remaining in unplaced scaffolds. X and y-axes show the genome position in Mbp. The heatmap was generated using juicebox[83], 0-7039 observed counts (balanced) are shown.

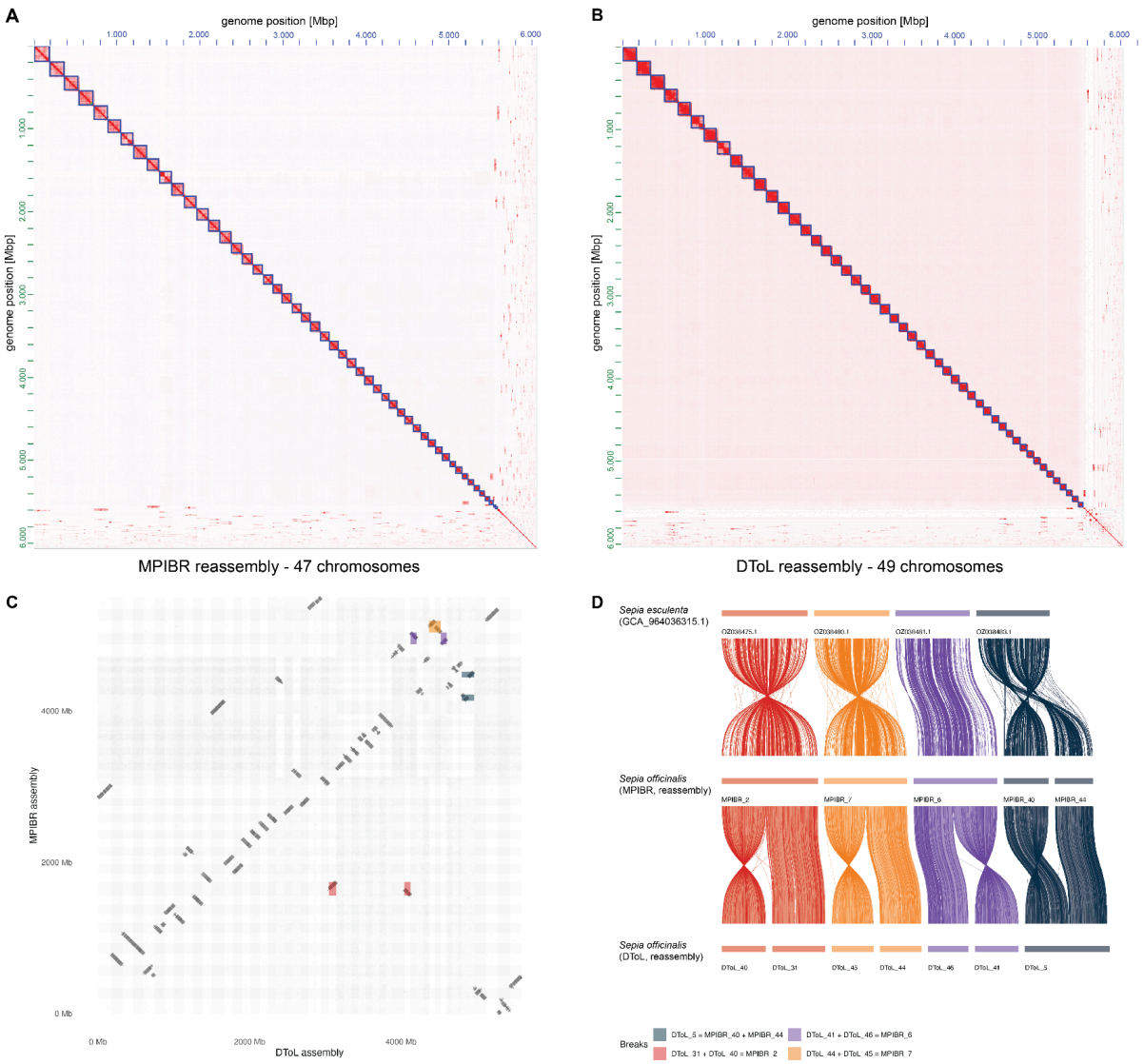


Figure 2: Comparison of two *Sepia officinalis* chromosome-scale assemblies indicates chromosome number of 1n=46. Datasets were collected from two *S. officinalis* animals, one as described in this study (MPIBR), the second by the Darwin Tree of Life consortium (DToL)[75]. Both datasets were assembled using a common pipeline (hifiasm and YAHS). **A)** Hi-C contact map of the MPIBR primary assembly, scaffolded using YAHS without manual curation. Assembled 47 chromosome scaffolds are shown as blue boxes. **B)** Hi-C contact map of DToL primary assembly, scaffolded using YAHS without manual curation, showing 49 assembled chromosome scaffolds as blue boxes. **C)** Whole-genome alignment of both scaffolded assemblies using Winnowmap2[89], showing DToL on x-axis and MPIBR on the y-axis. The 4 “breakpoints” of chromosomes in either of the assemblies (3 breaks in DToL chromosomes compared to MPIBR, 1 break in MPIBR compared to DToL) are highlighted in different colors. **D)** Ribbon diagram showing the four breakpoints from C) compared to the chromosome-scale assembly from another cuttlefish, *Acanthosepion esculentum* (1n=46). The color of breakpoints are the same in panels C+D.

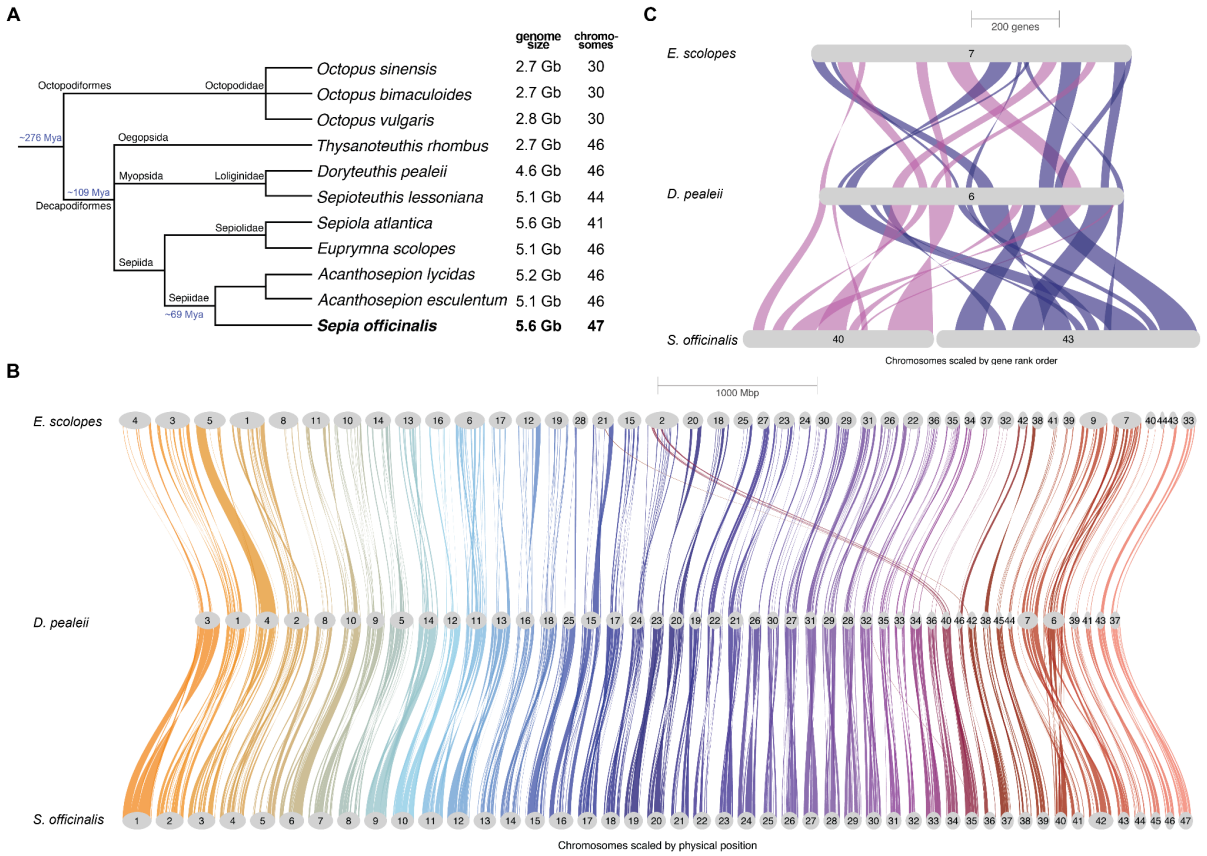


Figure 3: Syntenic comparison of three decapod species. A) Taxonomy of selected cephalopod species showing their genome size (in gigabases, Gb) and haploid chromosome numbers. Taxonomy information was downloaded from NCBI taxonomy browser, divergence times for Coleoidea and Decapodiformes from[192] and for Sepiidae from[94]. **B)** Genome-wide syntenic relationship between chromosomes of *E. scolopes*[49] (top), *D. pealeii*[49] (middle) and *S. officinalis* (bottom). Colored braids connect syntenic regions across genomes, with chromosomes drawn to physical scale. *Euprymna* chromosomes 45 and 46 are not shown because they contain too few orthogroups. **C)** Detailed synteny of *Sepia* chromosomes 40 (magenta) and 43 (darkblue) shown, that are joined in the other species and cause the different haploid chromosome number in *Sepia*. Riparian plots were generated using GENESPACE v1.2.3[175].

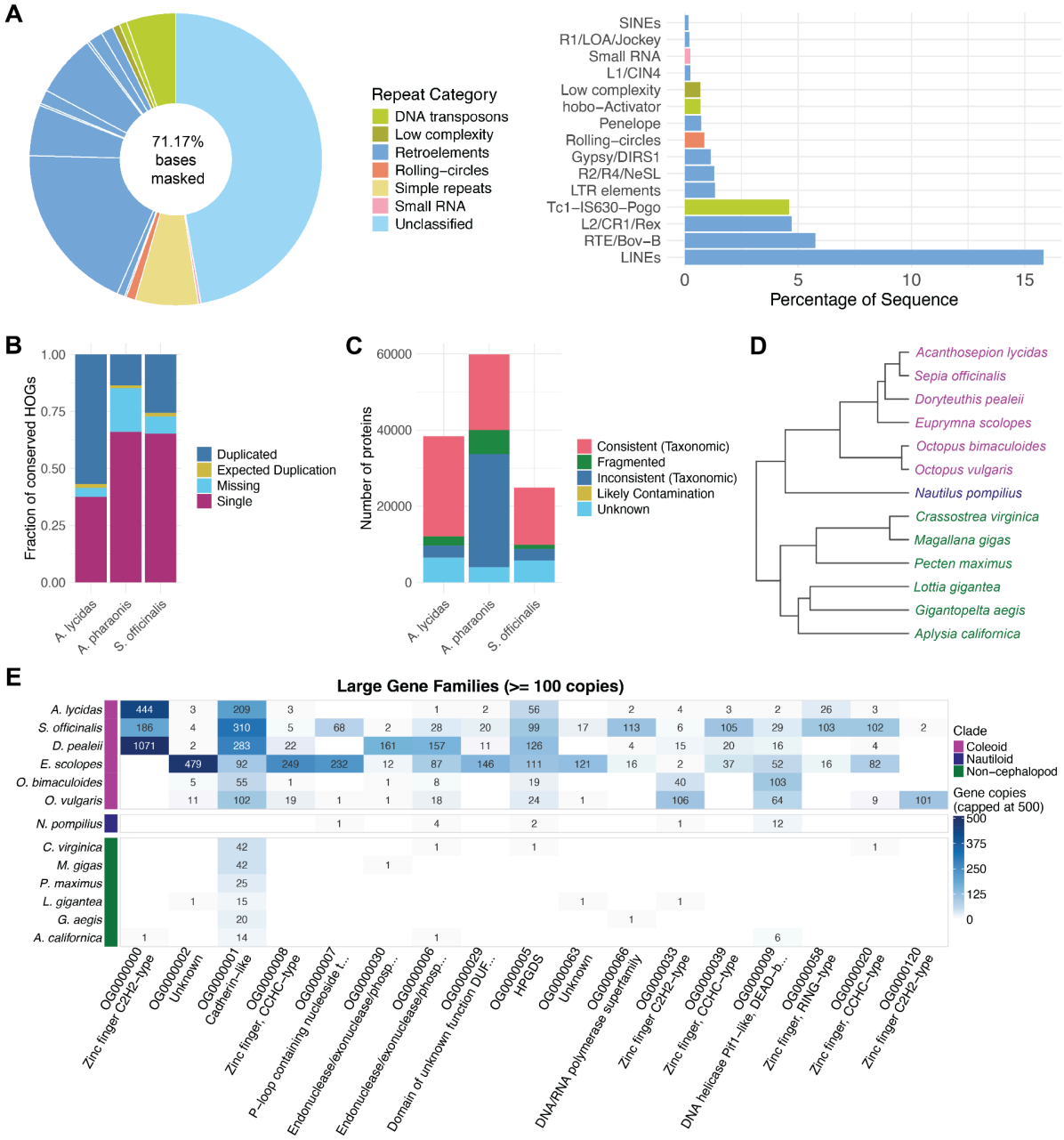


Figure 4: Genome annotation for *Sepia officinalis*. A) Annotation of repeat landscape of the *S. officinalis* genome, annotated using RepeatModeler[97]. Full repeat landscape is shown on the left, annotated repeats (excluding unclassified or simple repeats) are shown on the right. **B-C)** Quality control of gene annotation and comparison to two other cuttlefish species using OMArk[117]. Results shown for *Acanthosepion lycidas* (GCA_963932145.1, Ensembl Genebuild), *Sepia officinalis* (BRAKER, this study) and *Acanthosepion pharaonis*[193] (BRAKER). Lophotrochozoa was used as the ancestral clade. **B)** Completeness assessed by the presence of genes conserved in the clade, classified as *single* or multiple copies (*duplicated*), or *missing*. **C)** Consistency assessed by the proportion of proteins placed in the correct lineage (*consistent*); placement in incorrect lineages randomly (*inconsistent*) or to specific species (*contamination*), or no placement in known gene families (*unknown*). **D)** Phylogenetic tree of 13 molluscan species used for analysis of gene families with Orthofinder[122]. Species are colored by clade: purple = coleoid cephalopods, blue = nautiloid (non-coleoid cephalopod), green = non-cephalopod mollusk. **E)** Heatmap of largest gene

families (orthogroups from Orthofinder, with more than 100 genes in any species), ordered from largest gene count across all species on the left. Families with at least one gene in *S. officinalis* are depicted. Rows show gene counts for each species (color capped at 500 genes), columns show orthogroups and their annotation by eggNOG mapper[120,121] or InterProScan[119], if available. Clade colors match D).

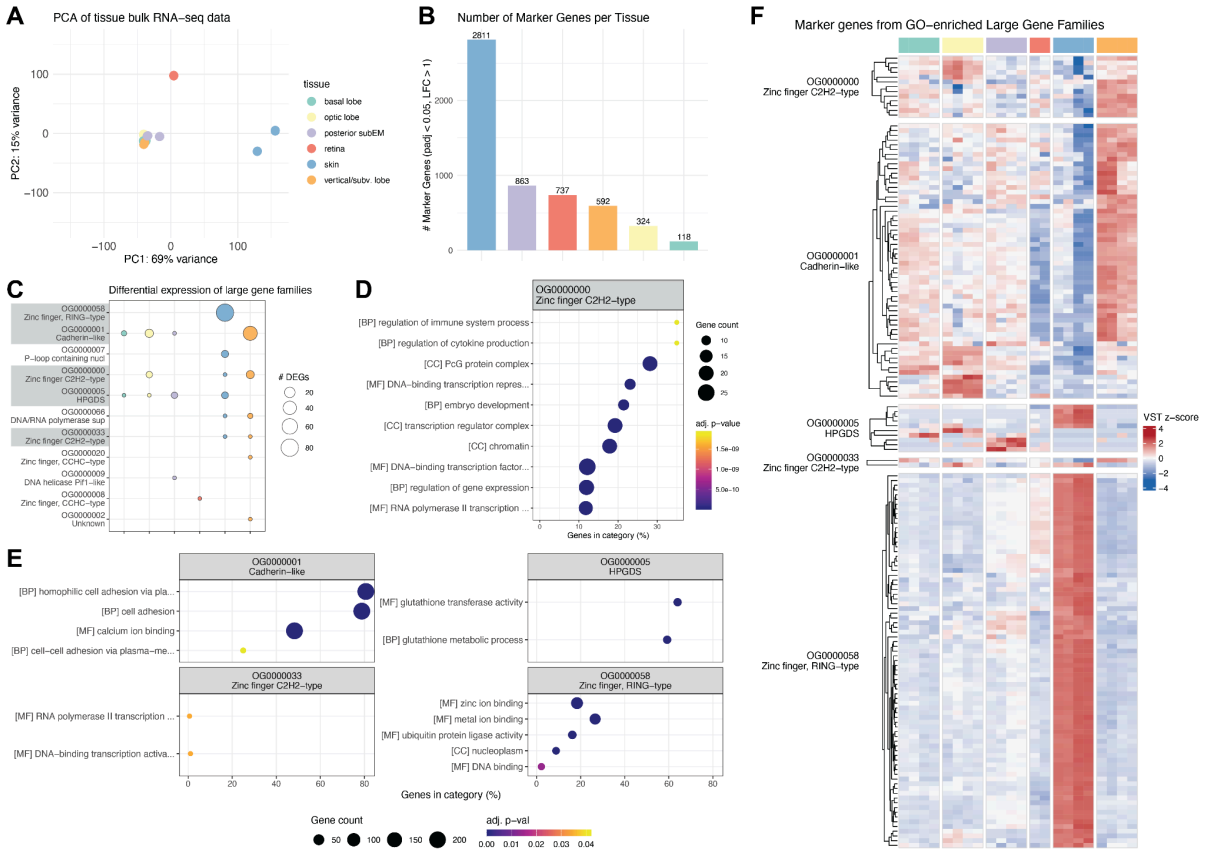
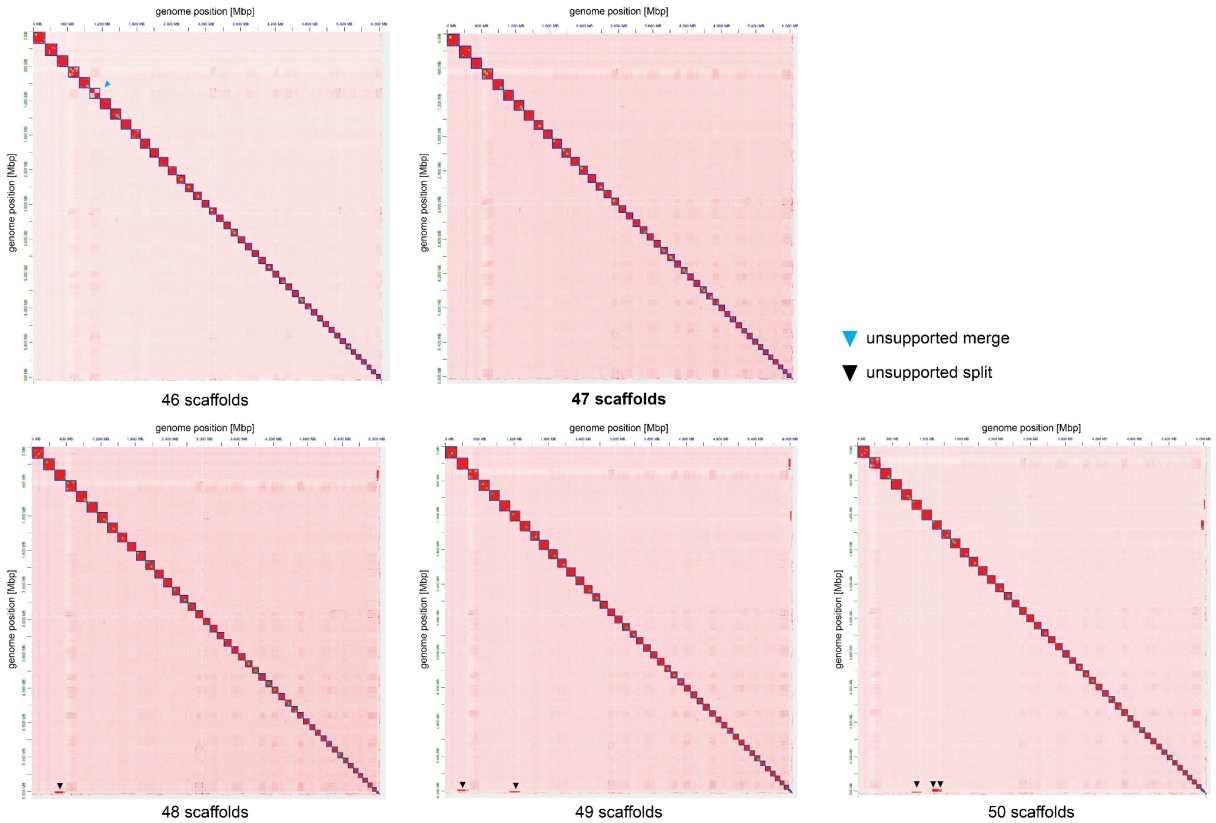
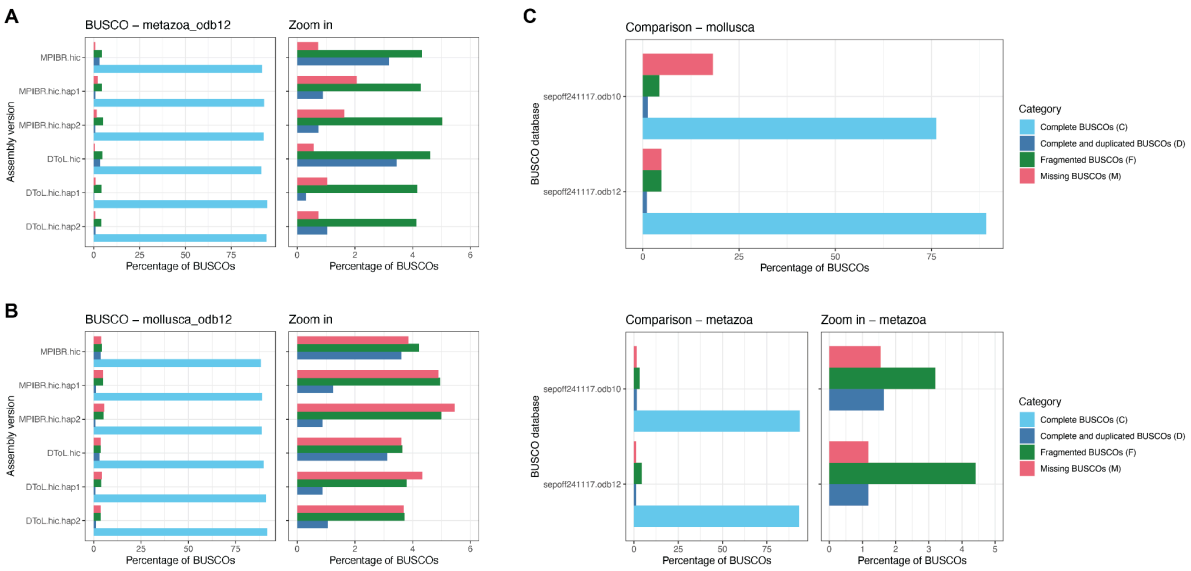


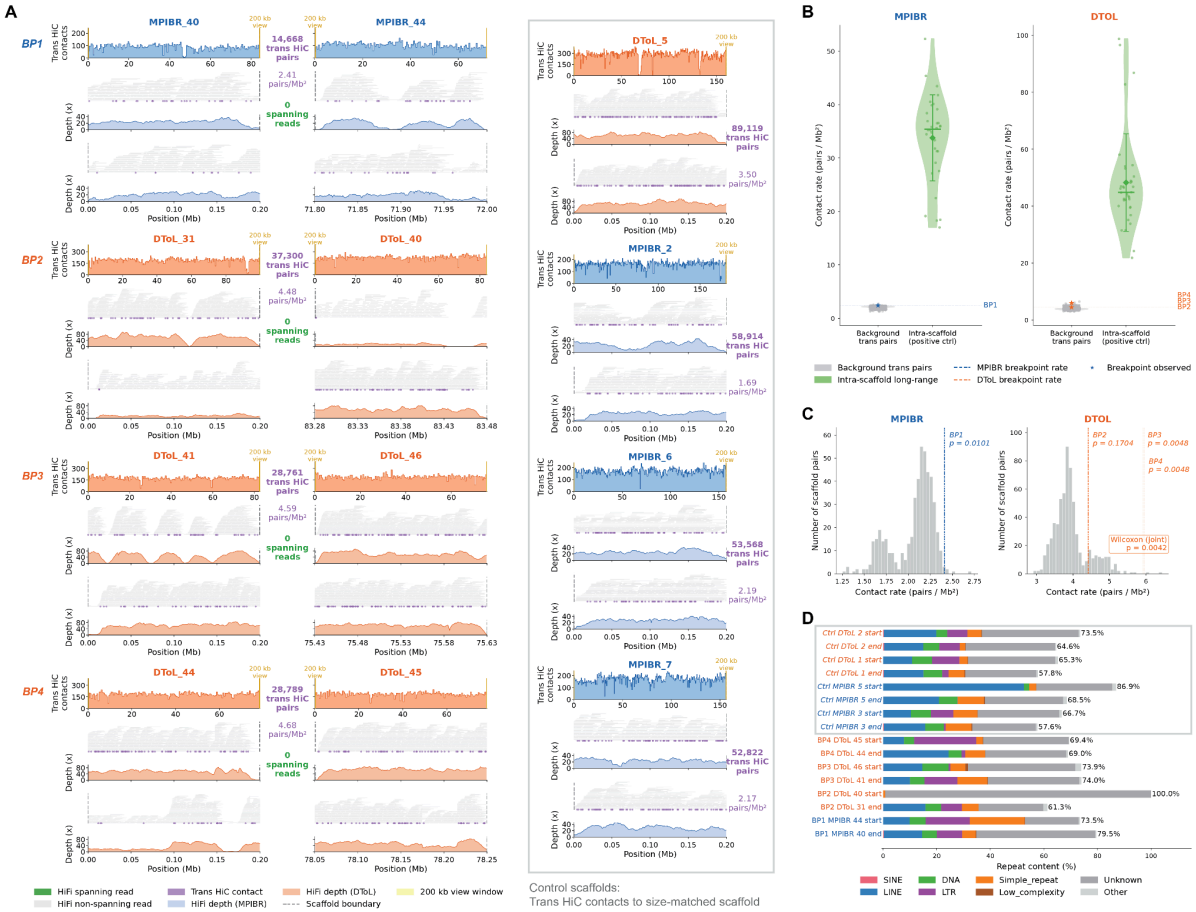
Figure 5: Expression of expanded gene families in tissue bulk RNA-seq data. Bulk RNA-seq data collected from one adult *S. officinalis* from different brain tissues (optic lobes - yellow, basal lobes - turquoise, vertical and subvertical lobes - orange, posterior subesophageal mass - purple), retina (red) and skin (blue, from the dorsal mantle). Tissue color code is identical throughout the figure. **A)** Principal component analysis (PCA) of the data, showing the first 2 PCs, colored by tissue. **B)** Barplot showing number of differentially expressed (DE) genes (i.e. marker genes) for each tissue, calculated against all other tissues using DESeq2[187]. **C)** Largest gene families (orthogroups) with differential expression in bulk RNA-seq data. Dot size shows the number of DE genes for each tissue. Families with enriched GO terms are highlighted in grey. **D)+E)** Dotplots of enriched gene ontology[133,134] (GO) terms for large gene families, enriched using clusterProfiler[189] using a hypergeometric test. Dot size shows the number of expressed genes per family with this GO term, x-axis shows the percentage of expressed genes from all genes with this GO term. Dot color shows the adjusted p-value after Benjamini-Hochberg false discovery rate (FDR) correction. CC: cellular component, MF: molecular function, BP: biological process. **F)** Heatmap of z-scored expression of all DE genes from the largest gene families with enriched GO terms.



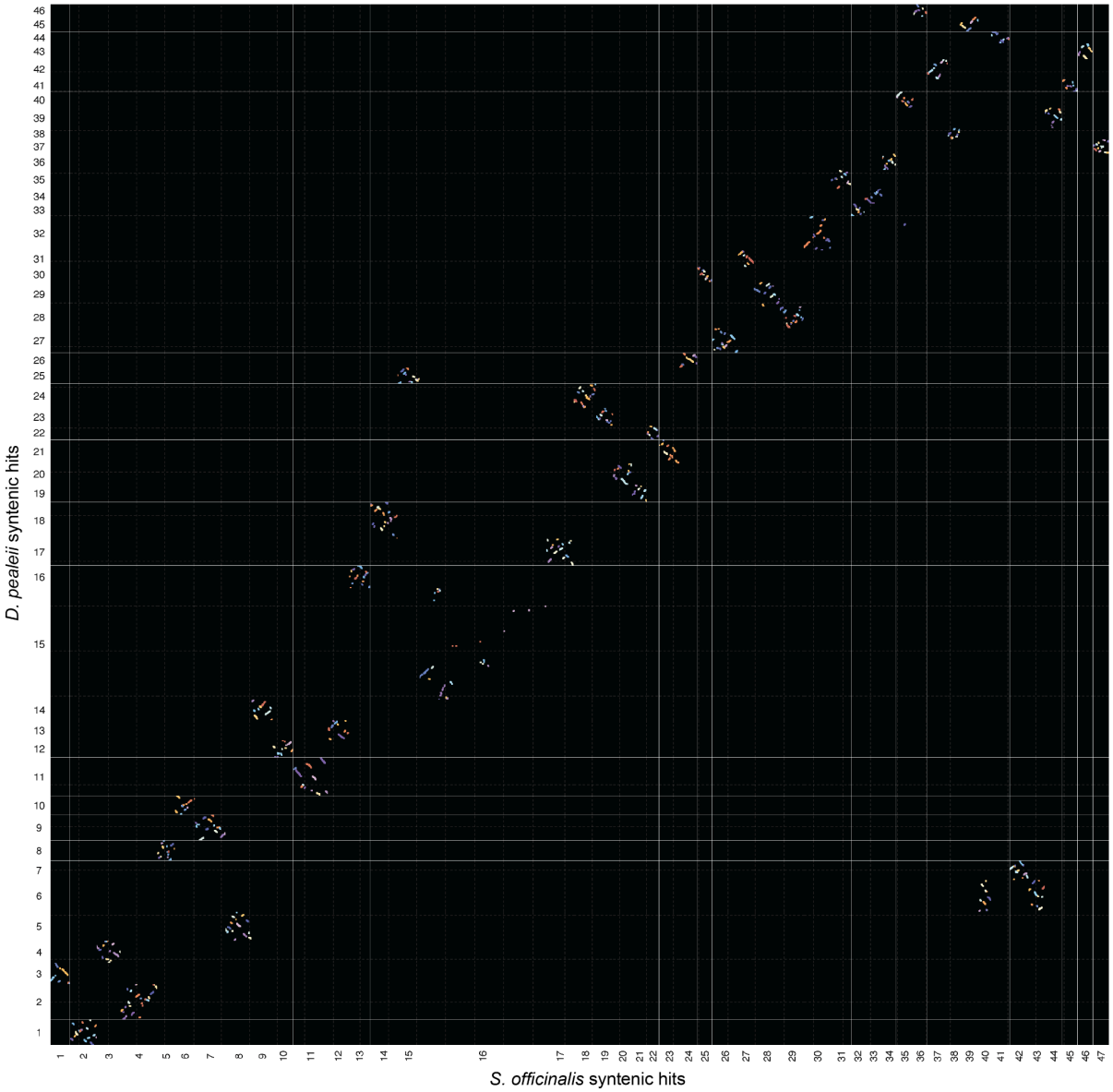
Supplementary Figure 1: HapHiC scaffolding for different numbers of expected chromosome scaffolds show 47 chromosomes as most supported. Hi-C contact maps from HapHiC[81] are shown for 46, 47, 48, 49 and 50 expected chromosome scaffolds. Assembled chromosomes are shown as blue boxes, Hi-C signal indicating a false (unsupported) merger is shown by cyan arrow, false splits are shown by black arrows. The contact maps differ from the map shown in Figure 1, which was created using YAHS and manual curation.



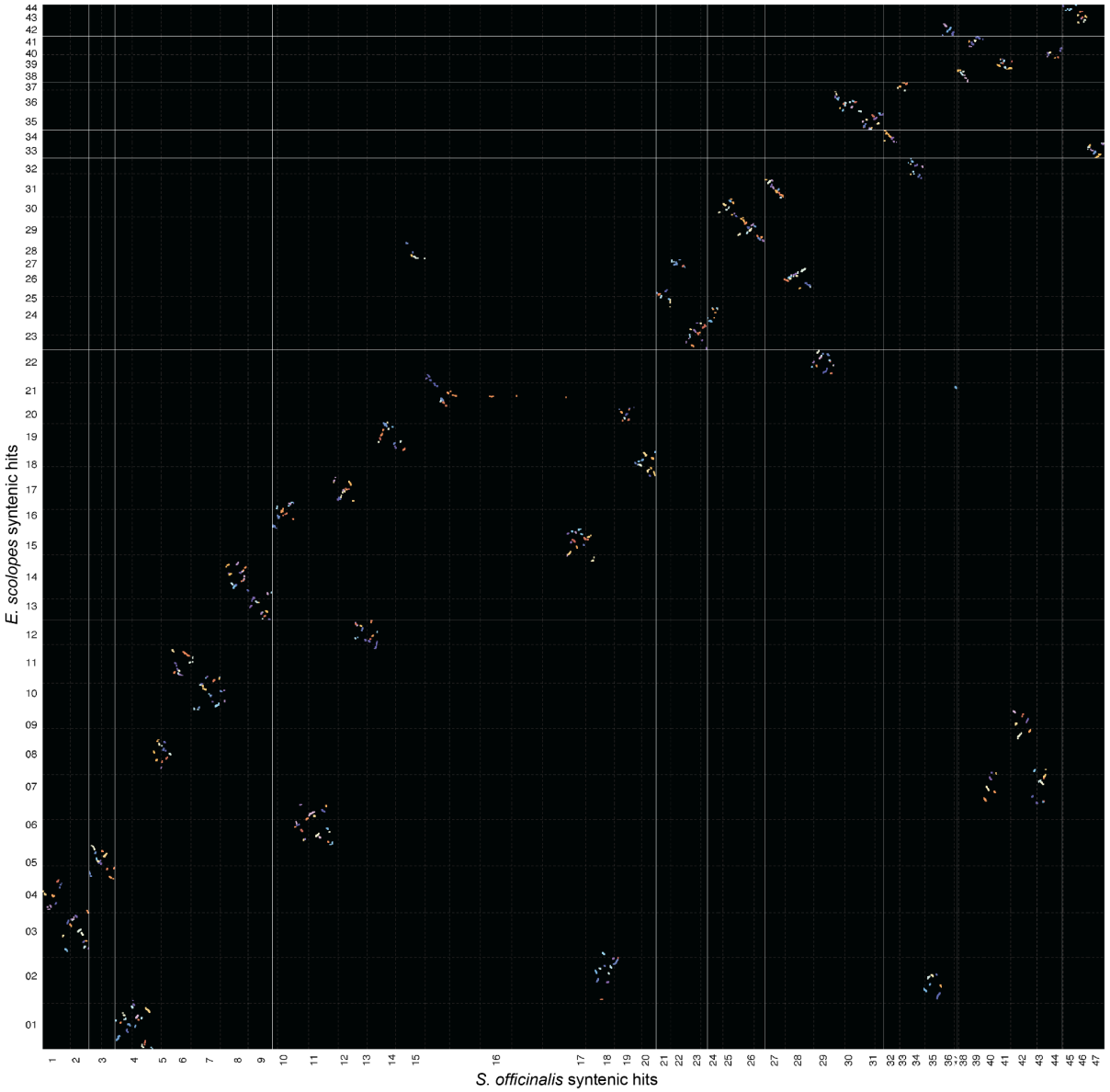
Supplementary Figure 2: BUSCO completeness results. A) Comparison of two *S. officinalis* chromosome-scale assemblies, which were constructed from two independent datasets (this study: MPIBR, Darwin Tree of Life project: DToL), assembled using a common pipeline (hifiasm[167] with PacBio HiFi and Hi-C reads). Results for the database *metazoa_odb12*, the zoom in shows only duplicated, fragmented and missing fractions to improve readability. The DToL assemblies have slightly higher completeness than MPIBR, due to the higher sequencing coverage used as input. In both datasets, compared to the primary assembly (".hic"), the phased haplotypes (".hic.hap1" and "hic.hap2") have less duplicated but more missing genes. **B)** BUSCO results for the *mollusca_odb12* database, showing the same trend as in A). **C)** Comparison of different BUSCO databases *odb10* and *odb12* on the manually curated assembly ("sepoff241117"). For the *mollusca* gene sets (top), a strong improvement in completeness was observed between *odb10* and *odb12*, reflecting that the updated gene set is more concise and conserved across species. For the *metazoa* gene sets (bottom) the completeness was marginally increased for *odb12* compared to *odb10*.



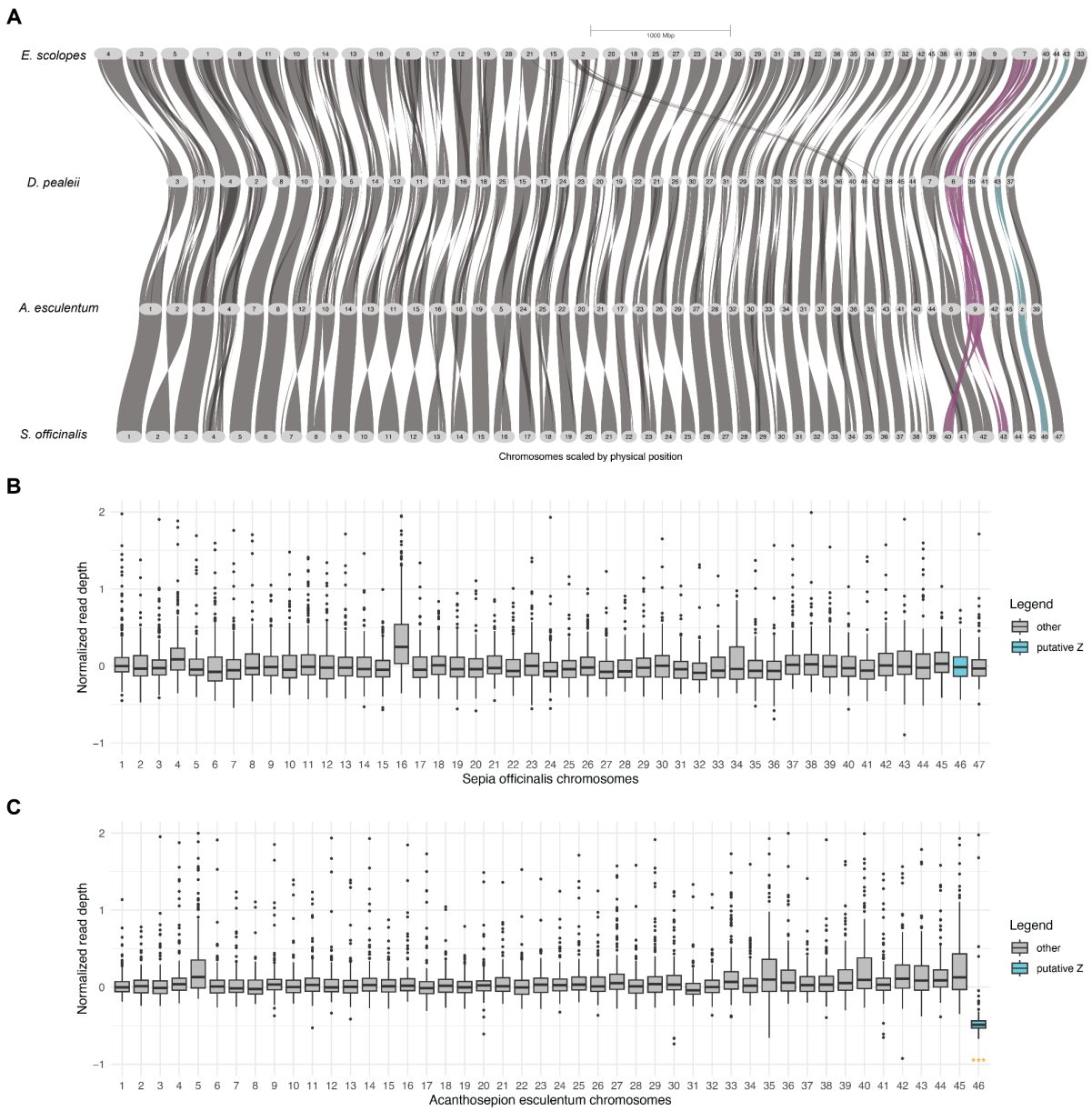
Supplementary Figure 3: Analysis of raw data at breakpoints between *S. officinalis* assemblies hints at a technical cause of breakpoints. A) Coverage of HiC and HiFi data shown for pairs of scaffolds exhibiting breakpoints. Blue shows MPIBR data, orange shows DToL data. For each breakpoint, trans HiC contacts are shown on top across the full scaffold, with terminal 200 kb windows highlighted in yellow. Both terminal windows are shown below with aligned HiFi reads (grey horizontal bars) and normalized HiFi read density. Trans HiC contacts are shown as purple dots. Right grey box: same data shown for the complete breakpoint scaffold of the other assembly, with trans HiC contacts calculated to a size-matched scaffold. **B)** Distribution of normalized trans HiC contact rate (pairs per Mb^2) for random scaffold pairs ("background pairs", grey) and within scaffolds ("intra scaffold", green) for MPIBR (left) and DToL (right) data. Values for scaffolds with breakpoints are indicated in blue and orange, respectively. **C)** Histogram of contact rates from B) shown for random scaffold pairs and breakpoint pairs. Contact rates and empirical p-values of breakpoint pairs are indicated in blue (left, MPIBR) and orange (right, DToL). Joint p-value for three rates for DToL breakpoints is indicated in box (Wilcoxon rank-sum, one-tailed). **D)** Repeat analysis of 200 kb scaffold ends at breakpoints and control scaffolds (grey box). Overall repeat content (% of base pairs) and type are shown.



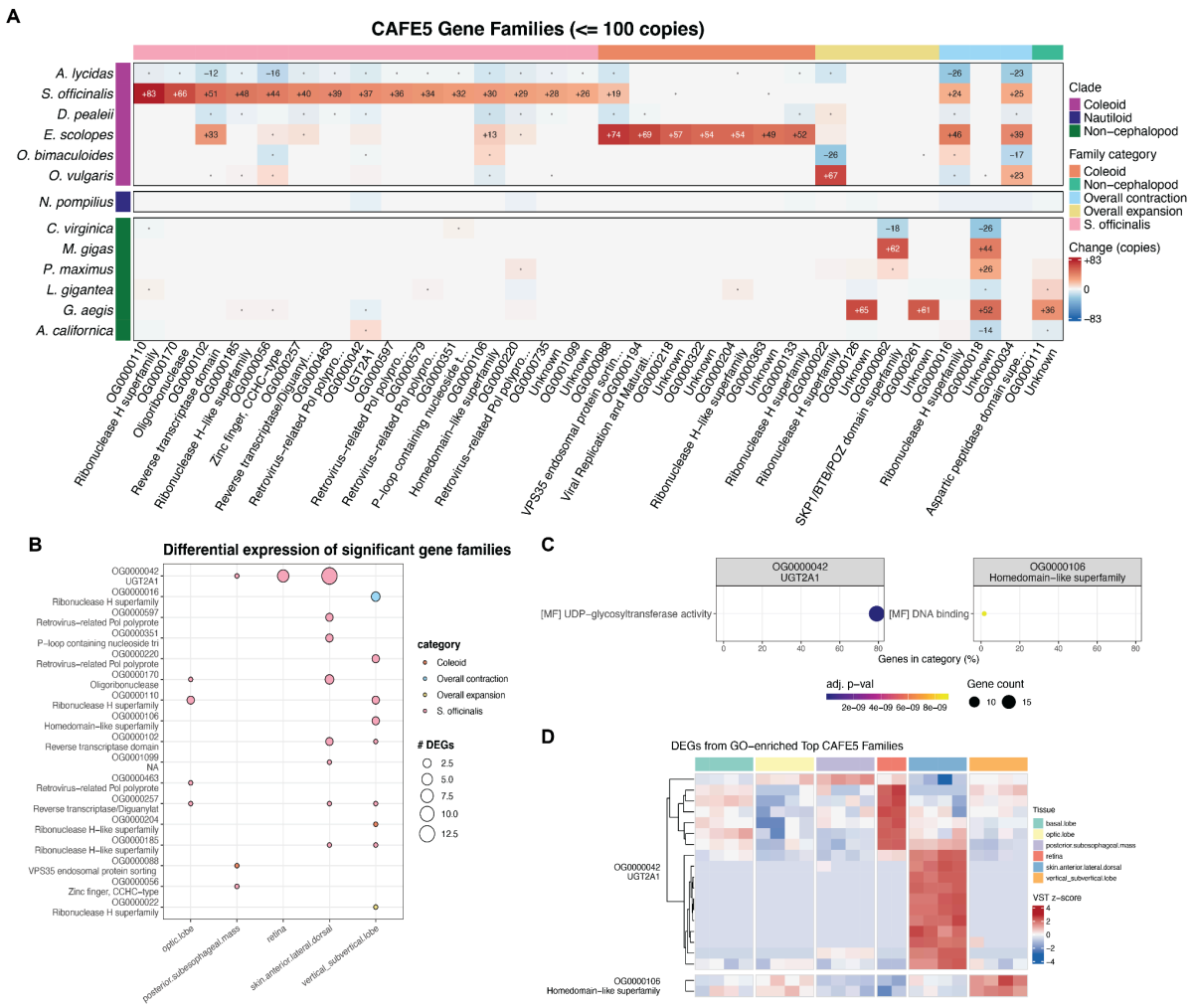
Supplementary Figure 4: Syntenic relationship between *S. officinalis* and *D. pealeii* chromosomes. Dot plot showing finer-resolution syntenic anchor hits (perfectly collinear blast hits within the same orthogroup). Genes are ordered along the chromosomes, gridlines are shown every 1000 genes. Only chromosome pairs with a minimum synteny score of 10 and at least 10 syntenic genes are shown. Synteny analysis and visualization were performed using GENESPACE v1.2.3[175].



Supplementary Figure 5: Syntenic relationship between *S. officinalis* and *E. scolopes* chromosomes. Dot plot showing finer-resolution syntenic anchor hits (perfectly collinear blast hits within the same orthogroup). Genes are ordered along the chromosomes, gridlines are shown every 1000 genes. Only chromosome pairs with a minimum synteny score of 10 and at least 10 syntenic genes are shown. *E. scolopes* chromosomes 45 and 46 are not shown because they contain too few orthogroups. Synteny analysis and visualization were performed using GENESPACE v1.2.3[175].



Supplementary Figure 6: Syntenic comparison of four decapod species hints at a cephalopod sex chromosome. A) Riparian plot showing synteny relationships of chromosomes from four decapod species, generated using GENESPACE[175] with orthogroups. *Euprymna* chromosomes 45 and 46 are not shown because they contain too few orthogroups. Chromosome split in *S. officinalis* compared to other species is shown in purple, putative sex chromosome as identified recently[96] is shown in cyan. **B)** Normalized coverage of sequencing data in *S. officinalis* chromosomes. **C)** Normalized coverage of short reads to female *A. esculentum* genome, reproduced from[96]. Decrease in read coverage for chromosome 46 is visible, the putative Z sex chromosome. Read depth was calculated from Illumina gDNA reads in windows of 500,000 bp and normalized to the median coverage of chromosome 1. Box plots showing median divergence (box dividing line), interquartile range (box), and 1.5 times the interquartile range (whiskers). The putative Z chromosome is highlighted in cyan. Chromosomes with significantly reduced read coverage (orange label) were identified by a one-sided Wilcoxon rank-sum test of each chromosome's normalized depth windows against all remaining chromosomes (Benjamini-Hochberg-corrected, at least 10% decrease in median normalized depth, * $p < 0.5$, ** $p < 0.01$, *** $p < 0.001$).



Supplementary Figure 7: Gene family expansion analysis. A) Gene family expansion analysis using CAFE5[128] with a gamma model (k=3) on all smaller gene families (less than 100 genes in any species). 30 families with the most change in different categories are shown (expanded only in *S. officinalis* (pink), in all coleoids (orange), in all species (yellow), in non-cephalopod mollusks (green) or overall contraction (blue)). Rows show change (expansion or contraction) of gene families in any species, columns show orthogroups and annotation, if available. Dots show significant change ($p < 0.05$), gene counts are shown for at any orthogroup with at least 12 genes in any species. **B)** Gene families with differential expression in bulk RNA-seq data. Dot size shows the number of DE genes for each tissue. **C)** Dotplots of enriched (GO) terms for large gene families, enriched using clusterProfiler using a hypergeometric test. Dot size shows the number of expressed genes per family with this GO term, x-axis shows percentage of expressed genes from all genes with this GO term. Dot color shows adjusted p-value after Benjamini-Hochberg false discovery rate (FDR) correction. CC: cellular component, MF: molecular function, BP: biological process. **D)** Heatmap of z-scored expression of all DE genes from the gene families with enriched GO terms.

References

1. Giuditta, A., Libonati, M., Packard, A., and Prozzo, N. (1971). Nuclear counts in the brain lobes of octopus vulgaris as a function of body size. Brain Res. *25*, 55–62.

2. Young, J. (1963). The number and sizes of nerve cells in Octopus. In Proceedings of the Zoological Society of London (Wiley Online Library), pp. 229–254.

3. Dorkenwald, S., Matsliah, A., Sterling, A.R., Schlegel, P., Yu, S., McKellar, C.E., Lin, A., Costa, M., Eichler, K., Yin, Y., et al. (2024). Neuronal wiring diagram of an adult brain. Nature *634*, 124–138. https://doi.org/10.1038/s41586-024-07558-y.

4. Herculano-Houzel, S., Mota, B., and Lent, R. (2006). Cellular scaling rules for rodent brains. Proc. Natl. Acad. Sci. *103*, 12138–12143. https://doi.org/10.1073/pnas.0604911103.

5. Hodgkin, A.L., Huxley, A.F., and Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of Loligo. J. Physiol. *116*, 424–448. https://doi.org/10.1113/jphysiol.1952.sp004716.

6. Llinás, R., Steinberg, I.Z., and Walton, K. (1981). Presynaptic calcium currents in squid giant synapse. Biophys. J. *33*, 289–321. https://doi.org/10.1016/S0006-3495(81)84898-9.

7. Llinás, R., Steinberg, I.Z., and Walton, K. (1981). Relationship between presynaptic calcium current and postsynaptic potential in squid giant synapse. Biophys. J. *33*, 323–351. https://doi.org/10.1016/S0006-3495(81)84899-0.

8. Llinas, R.R. (1984). The Squid Giant Synapse. In Current Topics in Membranes and Transport The Squid Axon., A. Kleinzeller, ed. (Academic Press), pp. 519–546. https://doi.org/10.1016/S0070-2161(08)60483-9.

9. Hanlon, R.T., and Messenger, J.B. (2018). Cephalopod Behaviour 2nd ed. (Cambridge University Press) https://doi.org/10.1017/9780511843600.

10. Josef, N., Berenshtein, I., Rousseau, M., Scata, G., Fiorito, G., and Shashar, N. (2017). Size Matters: Observed and Modeled Camouflage Response of European Cuttlefish (Sepia officinalis) to Different Substrate Patch Sizes during Movement. Front. Physiol. *7*.

11. Osorio, D., Ménager, F., Tyler, C.W., and Darmaillacq, A.-S. (2022). Multi-level control of adaptive camouflage by European cuttlefish. Curr. Biol. *32*, 2556-2562.e2. https://doi.org/10.1016/j.cub.2022.04.030.

12. Marshall, N.J., and Messenger, J.B. (1996). Colour-blind camouflage. Nature *382*, 408–409. https://doi.org/10.1038/382408b0.

13. Zylinski, S., Osorio, D., and Shohet, A. j (2009). Perception of edges and visual texture in the camouflage of the common cuttlefish, Sepia officinalis. Philos. Trans. R. Soc. B Biol. Sci. *364*, 439–448. https://doi.org/10.1098/rstb.2008.0264.

14. Zylinski, S., Darmaillacq, A.-S., and Shashar, N. (2012). Visual interpolation for contour completion by the European cuttlefish (Sepia officinalis) and its use in dynamic camouflage. Proc. R. Soc. B Biol. Sci. *279*, 2386–2390. https://doi.org/10.1098/rspb.2012.0026.

15. Zylinski, S., Osorio, D., and Shohet, A.J. (2009). Cuttlefish camouflage: context-dependent body pattern use during motion. Proc. R. Soc. B Biol. Sci. *276*, 3963–3969. https://doi.org/10.1098/rspb.2009.1083.

16. Jozet-Alves, C., Bertin, M., and Clayton, N.S. (2013). Evidence of episodic-like memory in cuttlefish. Curr. Biol. *23*, R1033–R1035. https://doi.org/10.1016/j.cub.2013.10.021.

17. Schnell, A.K., Boeckle, M., Rivera, M., Clayton, N.S., and Hanlon, R.T. (2021). Cuttlefish exert self-control in a delay of gratification task. Proc. R. Soc. B Biol. Sci. *288*, 20203161. https://doi.org/10.1098/rspb.2020.3161.

18. Boycott, B.B., and Young, J.Z. (1955). A memory system in Octopus vulgaris Lamarck. Proc. R. Soc. Lond. Ser. B - Biol. Sci. *143*, 449–480. https://doi.org/10.1098/rspb.1955.0024.

19. Mather, J.A. (2006). Behaviour Development: A Cephalopod Perspective. Int. J. Comp. Psychol. *19*. https://doi.org/10.46867/ijcp.2006.19.01.02.

20. Hall, K., and Hanlon, R. (2002). Principal features of the mating system of a large spawning aggregation of the giant Australian cuttlefish Sepia apama (Mollusca: Cephalopoda). Mar. Biol. *140*, 533–545.

21. Norman, M.D., Finn, J., and Tregenza, T. (1999). Female impersonation as an alternative reproductive strategy in giant cuttlefish. Proc. R. Soc. Lond. B Biol. Sci. *266*, 1347–1349. https://doi.org/10.1098/rspb.1999.0786.

22. Sampaio, E., Seco, M.C., Rosa, R., and Gingins, S. (2021). Octopuses punch fishes during collaborative interspecific hunting events. Ecology *102*, e03266. https://doi.org/10.1002/ecy.3266.

23. Sampaio, E., Sridhar, V.H., Francisco, F.A., Nagy, M., Sacchi, A., Strandburg-Peshkin, A., Nührenberg, P., Rosa, R., Couzin, I.D., and Gingins, S. (2024). Multidimensional social influence drives leadership and composition-dependent success in octopus–fish hunting groups. Nat. Ecol. Evol. *8*, 2072–2084. https://doi.org/10.1038/s41559-024-02525-2.

24. Feord, R.C., Sumner, M.E., Pusdekar, S., Kalra, L., Gonzalez-Bellido, P.T., and Wardill, T.J. (2020). Cuttlefish use stereopsis to strike at prey. Sci. Adv. *6*, eaay6036. https://doi.org/10.1126/sciadv.aay6036.

25. Medeiros, S.L. de S., Paiva, M.M.M. de, Lopes, P.H., Blanco, W., Lima, F.D. de, Oliveira, J.B.C. de, Medeiros, I.G., Sequerra, E.B., de Souza, S., Leite, T.S., et al. (2021). Cyclic alternation of quiet and active sleep states in the octopus. iScience *24*, 102223. https://doi.org/10.1016/j.isci.2021.102223.

26. Iglesias, T.L., Boal, J.G., Frank, M.G., Zeil, J., and Hanlon, R.T. (2019). Cyclic nature of the REM sleep-like state in the cuttlefish Sepia officinalis. J. Exp. Biol. *222*, jeb174862. https://doi.org/10.1242/jeb.174862.

27. Pophale, A., Shimizu, K., Mano, T., Iglesias, T.L., Martin, K., Hiroi, M., Asada, K., Andaluz, P.G., Van Dinh, T.T., Meshulam, L., et al. (2023). Wake-like skin patterning and neural activity during octopus sleep. Nature *619*, 129–134. https://doi.org/10.1038/s41586-023-06203-4.

28. Young, J.Z. (1974). The central nervous system of Loligo I. The optic lobe. Philos. Trans.

R. Soc. Lond. B Biol. Sci. *267*, 263–302.

29. Young, J.Z. (1962). The optic lobes of Octopus vulgaris. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *245*, 19–58. https://doi.org/10.1098/rstb.1962.0005.

30. Young, J. (1960). The Visual System of Octopus:(1) Regularities in the Retina and Optic Lobes of Octopus in Relation to Form Discrimination. Nature *186*, 836–839.

31. Young, J.Z. (1991). Computation in the Learning System of Cephalopods. Biol. Bull. *180*, 200–208. https://doi.org/10.2307/1542389.

32. Young, J.Z. (1963). Light-and Dark-Adaptation in the Eyes of Some Cephalopods. Proc. Zool. Soc. Lond. *140*, 255–272. https://doi.org/10.1111/j.1469-7998.1963.tb01863.x.

33. Young, J.Z. (1962). The retina of cephalopods and its degeneration after optic nerve section. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *245*, 1–18.

34. Nixon, M., and Young, J.Z. (2003). The brains and lives of cephalopods (Oxford University Press).

35. Case, N.M., Gray, E., and Young, J. (1972). Ultrastructure and synaptic relations in the optic lobe of the brain of Eledone and Octopus. J. Ultrastruct. Res. *39*, 115–123.

36. Dilly, P., Gray, E.G., and Young, J.Z. (1963). Electron microscopy of optic nerves and optic lobes of Octopus and Eledone. Proc. R. Soc. Lond. B Biol. Sci. *158*, 446–456.

37. Stephens, P., and Young, J.Z. (1969). The glio-vascular system of cephalopods. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *255*, 1–12.

38. Baden, T., Briseño, J., Coffing, G., Cohen-Bodénès, S., Courtney, A., Dickerson, D., Dölen, G., Fiorito, G., Gestal, C., Gustafson, T., et al. (2023). Cephalopod-omics: Emerging Fields and Technologies in Cephalopod Biology. Integr. Comp. Biol. *63*, 1226–1239. https://doi.org/10.1093/icb/icad087.

39. Styfhals, R., Zolotarov, G., Hulselmans, G., Spanier, K.I., Poovathingal, S., Elagoz, A.M., De Winter, S., Deryckere, A., Rajewsky, N., Ponte, G., et al. (2022). Cell type diversity in a developing octopus brain. Nat. Commun. *13*, 7392. https://doi.org/10.1038/s41467-022-35198-1.

40. Songco-Casey, J.O., Coffing, G.C., Piscopo, D.M., Pungor, J.R., Kern, A.D., Miller, A.C., and Niell, C.M. (2022). Cell types and molecular architecture of the Octopus bimaculoides visual system. Curr. Biol. *0*. https://doi.org/10.1016/j.cub.2022.10.015.

41. Duruz, J., Sprecher, M., Kaldun, J.C., Al-Soudy, A.-S., Lischer, H.E., van Geest, G., Nicholson, P., Bruggmann, R., and Sprecher, S.G. (2023). Molecular characterization of cell types in the squid Loligo vulgaris. eLife *12*, e80670. https://doi.org/10.7554/eLife.80670.

42. Gavriouchkina, D., Tan, Y., Parey, E., Ziadi-Künzli, F., Hasegawa, Y., Piovani, L., Zhang, L., Sugimoto, C., Luscombe, N., Marlétaz, F., et al. (2025). A single-cell atlas of the bobtail squid visual and nervous system highlights molecular principles of convergent evolution. Nat. Ecol. Evol. *9*, 1245–1262. https://doi.org/10.1038/s41559-025-02720-9.

43. Schmidbaur, H., Kawaguchi, A., Clarence, T., Fu, X., Hoang, O.P., Zimmermann, B., Ritschard, E.A., Weissenbacher, A., Foster, J.S., Nyholm, S.V., et al. (2022). Emergence

of novel cephalopod gene regulation and expression through large-scale genome reorganization. Nat. Commun. *13*, 2172. https://doi.org/10.1038/s41467-022-29694-7.

44. Rouressol, L., Briseno, J., Vijayan, N., Chen, G.Y., Ritschard, E.A., Sanchez, G., Nyholm, S.V., McFall-Ngai, M.J., and Simakov, O. (2023). Emergence of novel genomic regulatory regions associated with light-organ development in the bobtail squid. iScience *26*, 107091. https://doi.org/10.1016/j.isci.2023.107091.

45. Yoshida, M., Hirota, K., Imoto, J., Okuno, M., Tanaka, H., Kajitani, R., Toyoda, A., Itoh, T., Ikeo, K., Sasaki, T., et al. (2022). Gene Recruitments and Dismissals in the Argonaut Genome Provide Insights into Pelagic Lifestyle Adaptation and Shell-like Eggcase Reacquisition. Genome Biol. Evol. *14*, evac140. https://doi.org/10.1093/gbe/evac140.

46. Albertin, C.B., Simakov, O., Mitros, T., Wang, Z.Y., Pungor, J.R., Edsinger-Gonzales, E., Brenner, S., Ragsdale, C.W., and 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.

47. Destanović, D., Schultz, D.T., Styfhals, R., Cruz, F., Gómez-Garrido, J., Gut, M., Gut, I., Fiorito, G., Simakov, O., Alioto, T.S., et al. (2023). A chromosome-level reference genome for the common octopus, Octopus vulgaris (Cuvier, 1797). G3 GenesGenomesGenetics, jkad220. https://doi.org/10.1093/g3journal/jkad220.

48. 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., et al. (2019). Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. Proc. Natl. Acad. Sci. *116*, 3030–3035. https://doi.org/10.1073/pnas.1817322116.

49. Albertin, C.B., Medina-Ruiz, S., Mitros, T., Schmidbaur, H., Sanchez, G., Wang, Z.Y., Grimwood, J., Rosenthal, J.J.C., Ragsdale, C.W., Simakov, O., et al. (2022). Genome and transcriptome mechanisms driving cephalopod evolution. Nat. Commun. *13*. https://doi.org/10.1038/s41467-022-29748-w.

50. Marino, A., Kizenko, A., Wong, W.Y., Ghiselli, F., and Simakov, O. (2022). Repeat Age Decomposition Informs an Ancient Set of Repeats Associated With Coleoid Cephalopod Divergence. Front. Genet. *13*.

51. Simakov, O., Bredeson, J., Berkoff, K., Marletaz, F., Mitros, T., Schultz, D.T., O'Connell, B.L., Dear, P., Martinez, D.E., Steele, R.E., et al. (2022). Deeply conserved synteny and the evolution of metazoan chromosomes. Sci. Adv. *8*, eabi5884. https://doi.org/10.1126/sciadv.abi5884.

52. Zou, Y., Fu, J., Liang, Y., Luo, X., Shen, M., Huang, M., Chen, Y., You, W., and Ke, C. (2024). Chromosome-level genome assembly of the ivory shell Babylonia areolata. Sci. Data *11*, 1201. https://doi.org/10.1038/s41597-024-04001-9.

53. Chen, Z., Baeza, J.A., Chen, C., Gonzalez, M.T., González, V.L., Greve, C., Kocot, K.M., Arbizu, P.M., Moles, J., Schell, T., et al. (2025). A genome-based phylogeny for Mollusca is concordant with fossils and morphology. Science *387*, 1001–1007. https://doi.org/10.1126/science.ads0215.

54. Zeng, Q., Liu, J., Wang, C., Wang, H., Zhang, L., Hu, J., Bao, L., and Wang, S. (2021). High-quality reannotation of the king scallop genome reveals no 'gene-rich' feature and evolution of toxin resistance. Comput. Struct. Biotechnol. J. *19*, 4954–4960. https://doi.org/10.1016/j.csbj.2021.08.038.

55. Sun, S., Han, X., Han, Z., and Liu, Q. (2024). Chromosomal-scale genome assembly and annotation of the land slug (Meghimatium bilineatum). Sci. Data *11*, 35. https://doi.org/10.1038/s41597-023-02893-7.

56. Männer, L., Schell, T., Spies, J., Galià-Camps, C., Baranski, D., Ben Hamadou, A., Gerheim, C., Neveling, K., Helfrich, E.J.N., and Greve, C. (2024). Chromosome-level genome assembly of the sacoglossan sea slug Elysia timida (Risso, 1818). BMC Genomics *25*, 941. https://doi.org/10.1186/s12864-024-10829-7.

57. Ma, B., Jin, W., Fu, H., Sun, B., Yang, S., Ma, X., Wen, H., Wu, X., Wang, H., and Cao, X. (2023). A High-Quality Chromosome-Level Genome Assembly of a Snail Cipangopaludina cathayensis (Gastropoda: Viviparidae). Genes *14*. https://doi.org/10.3390/genes14071365.

58. Liu, Z., Huang, Y., Chen, H., Liu, C., Wang, M., Bian, C., Wang, L., and Song, L. (2023). Chromosome-level genome assembly of the deep-sea snail Phymorhynchus buccinoides provides insights into the adaptation to the cold seep habitat. BMC Genomics *24*, 679. https://doi.org/10.1186/s12864-023-09760-0.

59. Peñaloza, C., Gutierrez, A.P., Eöry, L., Wang, S., Guo, X., Archibald, A.L., Bean, T.P., and Houston, R.D. (2021). A chromosome-level genome assembly for the Pacific oyster Crassostrea gigas. GigaScience *10*, giab020. https://doi.org/10.1093/gigascience/giab020.

60. Arias-Montecino, A., Sykes, A., Álvarez-Hernán, G., de Mera-Rodríguez, J.A., Calle-Guisado, V., Martín-Partido, G., Rodríguez-León, J., and Francisco-Morcillo, J. (2024). Histological and scanning electron microscope observations on the developing retina of the cuttlefish (Sepia officinalis Linnaeus, 1758). Tissue Cell *88*, 102417. https://doi.org/10.1016/j.tice.2024.102417.

61. O'Brien, C.E., Mezrai, N., Darmaillacq, A.-S., and Dickel, L. (2017). Behavioral development in embryonic and early juvenile cuttlefish (Sepia officinalis). Dev. Psychobiol. *59*, 145–160. https://doi.org/10.1002/dev.21476.

62. Bellingham, J., Morris, A.G., and Hunt, D.M. (1998). The Rhodopsin Gene of the Cuttlefish Sepia Officinalis: Sequence and Spectral Tuning. J. Exp. Biol. *201*, 2299–2306. https://doi.org/10.1242/jeb.201.15.2299.

63. Chemello, G., Faraoni, V., Notarstefano, V., Maradonna, F., Carnevali, O., and Gioacchini, G. (2022). First Evidence of Microplastics in the Yolk and Embryos of Common Cuttlefish (Sepia officinalis) from the Central Adriatic Sea: Evaluation of Embryo and Hatchling Structural Integrity and Development. Anim. Open Access J. MDPI *13*. https://doi.org/10.3390/ani13010095.

64. Court, M., Macau, M., Ranucci, M., Marquês, T., Repolho, T., Lopes, V.M., Rosa, R., and Paula, J.R. (2024). Oxygen loss compromises growth and cognition of cuttlefish newborns. Proc. Biol. Sci. *291*, 20241291. https://doi.org/10.1098/rspb.2024.1291.

65. Gladman, N.W., and Askew, G.N. (2023). The hydrodynamics of jet propulsion swimming in hatchling and juvenile European common cuttlefish, Sepia officinalis. J. Exp. Biol. *226*. https://doi.org/10.1242/jeb.246225.

66. Yang, T., Jia, Z., Chen, H., Deng, Z., Liu, W., Chen, L., and Li, L. (2020). Mechanical design of the highly porous cuttlebone: A bioceramic hard buoyancy tank for cuttlefish. Proc. Natl. Acad. Sci. U. S. A. *117*, 23450–23459.

https://doi.org/10.1073/pnas.2009531117.

67. Von Boletzky, S. (1987). Fecundity variation in relation to intermittent or chronic spawning in the cuttlefish, Sepia officinalis L. (Mollusca, Cephalopoda). Bull. Mar. Sci. *40*.

68. Laptikhovsky, V., Salman, A., Onsoy, B., and Katagan, T. (2003). Fecundity of the common cuttlefish, Sepia officinalis L. (Cephalopoda, Sepiidae): A new look at an old problem. Sci. Mar. *67*, 279–284.

69. Reiter, S., Hülsdunk, P., Woo, T., Lauterbach, M.A., Eberle, J.S., Akay, L.A., Longo, A., Meier-Credo, J., Kretschmer, F., Langer, J.D., et al. (2018). Elucidating the control and development of skin patterning in cuttlefish. Nature *562*, 361–366. https://doi.org/10.1038/s41586-018-0591-3.

70. Woo, T., Liang, X., Evans, D.A., Fernandez, O., Kretschmer, F., Reiter, S., and Laurent, G. (2023). The dynamics of pattern matching in camouflaging cuttlefish. Nature *619*, 122–128. https://doi.org/10.1038/s41586-023-06259-2.

71. Imarazene, B., Andouche, A., Bassaglia, Y., Lopez, P.-J., and Bonnaud-Ponticelli, L. (2017). Eye Development in Sepia officinalis Embryo: What the Uncommon Gene Expression Profiles Tell Us about Eye Evolution. Front. Physiol. *8*, 613. https://doi.org/10.3389/fphys.2017.00613.

72. Cocci, P., Mosconi, G., and Palermo, F.A. (2023). Effect of polycyclic aromatic hydrocarbons on homeobox gene expression during embryonic development of cuttlefish, Sepia officinalis. Chemosphere *325*, 138315. https://doi.org/10.1016/j.chemosphere.2023.138315.

73. Andouche, A., Bassaglia, Y., Baratte, S., and Bonnaud, L. (2013). Reflectin genes and development of iridophore patterns in Sepia officinalis embryos (Mollusca, Cephalopoda). Dev. Dyn. *242*, 560–571. https://doi.org/10.1002/dvdy.23938.

74. McKenna, V., Archibald, J., Beinart, R., Dawson, M., Hentschel, U., Keeling, P., Lopez, J., Martin-Duran, J., Petersen, J., Sigwart, J., et al. (2024). The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the Tree of Life [version 2; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Res. *6*. https://doi.org/10.12688/wellcomeopenres.17222.2.

75. The Darwin Tree of Life Project Consortium (2022). Sequence locally, think globally: The Darwin Tree of Life Project. Proc. Natl. Acad. Sci. *119*, e2115642118. https://doi.org/10.1073/pnas.2115642118.

76. Rubino, F.A., Coffing, G.C., Gibbons, C.J., Small, S.T., Desvignes, T., Pessutti, J., Petersen, A.M., Arkhipkin, A., Shcherbich, Z., Postlethwait, J.H., et al. (2025). A non-invasive method to genotype cephalopod sex by quantitative PCR. Preprint at bioRxiv, https://doi.org/10.1101/2025.10.28.685099 https://doi.org/10.1101/2025.10.28.685099.

77. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. *11*, 1432. https://doi.org/10.1038/s41467-020-14998-3.

78. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short

reads. Bioinformatics *33*, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

79. Hjelmen, C.E. (2024). Genome size and chromosome number are critical metrics for accurate genome assembly assessment in Eukaryota. Genetics *227*, iyae099. https://doi.org/10.1093/genetics/iyae099.

80. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

81. Zeng, X., Yi, Z., Zhang, X., Du, Y., Li, Y., Zhou, Z., Chen, S., Zhao, H., Yang, S., Wang, Y., et al. (2024). Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat. Plants *10*, 1184–1200. https://doi.org/10.1038/s41477-024-01755-3.

82. Zhou, C., McCarthy, S.A., and Durbin, R. (2023). YaHS: yet another Hi-C scaffolding tool. Bioinformatics *39*, btac808. https://doi.org/10.1093/bioinformatics/btac808.

83. Dudchenko, O., Shamim, M.S., Batra, S.S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., Hilaire, B.G.S., Yao, W., Stamenova, E., et al. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. Preprint at bioRxiv, https://doi.org/10.1101/254797 https://doi.org/10.1101/254797.

84. Sanchez, G., Fernández-Álvarez, F.Á., Bernal, A., Heath-Heckman, E., Lami, R., McFall-Ngai, M., Nishiguchi, M., Nyholm, S., Simakov, O., Allcock, A.L., et al. (2026). Rapid mid-Cretaceous diversification of squid and cuttlefish preceded radiation into coastal niches. Nat. Ecol. Evol., 1–15. https://doi.org/10.1038/s41559-026-03009-1.

85. Gao, Y.M., and Natsukari, Y. (1990). Karyological studies on seven cephalopods. Venus Jpn. J. Malacol. *49*, 126–145.

86. Jazayeri, A., Papan, F., Motamedi, H., and Mahmoudi, S. (2011). Karyological investigation of Persian Gulf cuttle fish (sepia arabica) in the coasts of Khuzestan province. Life Sci. J. *8*.

87. Ebrahimi Pour, M. (2009). The Study of Persian Gulf Cuttlefish (Sepia pharaonis) Chromosome Via Incubation of Blood Cells. Appl. Biol. *22*, 1–8. https://doi.org/10.22051/jab.2009.3307.

88. Vitturi, R., Rasotto, M.B., and Farinella-Ferruzza, N. (1982). The chromosomes of 16 molluscan species. Ital. J. Zool. *49*, 61–71.

89. Jain, C., Rhie, A., Hansen, N.F., Koren, S., and Phillippy, A.M. (2022). Long-read mapping to repetitive reference sequences using Winnowmap2. Nat. Methods *19*, 705–710. https://doi.org/10.1038/s41592-022-01457-8.

90. Jain, C., Rhie, A., Zhang, H., Chu, C., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Weighted minimizer sampling improves long read mapping. Bioinformatics *36*, i111–i118. https://doi.org/10.1093/bioinformatics/btaa435.

91. M. Vasimuddin, S. Misra, H. Li, and S. Aluru (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 314–324.

https://doi.org/10.1109/IPDPS.2019.00041.

92. Open2C, Abdennur, N., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., and Venev, S.V. (2023). Pairtools: from sequencing data to chromosome contacts. Preprint at bioRxiv, https://doi.org/10.1101/2023.02.13.528389 https://doi.org/10.1101/2023.02.13.528389.

93. Smit, A., Hubley, R., and Green, P. RepeatMasker http://www.repeatmasker.org.

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

95. Coffing, G., TITTES, S., Small, S., Songco-Casey, J.O., Pungor, J., Piscopo, D., Miller, A., Niell, C., and Kern, A. (2024). Data for: Cephalopod Sex Determination and its Ancient Evolutionary Origin. (Zenodo). https://doi.org/10.5281/zenodo.14010217 https://doi.org/10.5281/zenodo.14010217.

96. Coffing, G.C., Tittes, S., Small, S.T., Songco-Casey, J.O., Piscopo, D.M., Pungor, J.R., Miller, A.C., Niell, C.M., and Kern, A.D. (2025). Cephalopod sex determination and its ancient evolutionary origin. Curr. Biol. *35*, 931-939.e4. https://doi.org/10.1016/j.cub.2025.01.005.

97. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

98. Feschotte, C. (2008). The contribution of transposable elements to the evolution of regulatory networks. Nat. Rev. Genet. *9*, 397–405. https://doi.org/10.1038/nrg2337.

99. Hoff, K.J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-Genome Annotation with BRAKER. Methods Mol. Biol. Clifton NJ *1962*, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.

100. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics Bioinforma. *3*, lqaa108. https://doi.org/10.1093/nargab/lqaa108.

101. Brůna, T., Gabriel, L., and Hoff, K.J. (2024). Navigating Eukaryotic Genome Annotation Pipelines: A Route Map to BRAKER, Galba, and TSEBRA. Preprint at arXiv, https://doi.org/10.48550/arXiv.2403.19416 https://doi.org/10.48550/arXiv.2403.19416.

102. Gabriel, L., Hoff, K.J., Brůna, T., Borodovsky, M., and Stanke, M. (2021). TSEBRA: transcript selector for BRAKER. BMC Bioinformatics *22*, 566. https://doi.org/10.1186/s12859-021-04482-0.

103. Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. Bioinforma. Oxf. Engl. *32*, 767–769. https://doi.org/10.1093/bioinformatics/btv661.

104. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction

in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics *7*, 62. https://doi.org/10.1186/1471-2105-7-62.

105. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013.

106. Li, H. (2023). Protein-to-genome alignment with miniprot. Bioinformatics *39*, btad014. https://doi.org/10.1093/bioinformatics/btad014.

107. Iwata, H., and Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. Nucleic Acids Res. *40*, e161. https://doi.org/10.1093/nar/gks708.

108. Gotoh, O. (2008). A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. Nucleic Acids Res. *36*, 2630–2638. https://doi.org/10.1093/nar/gkn105.

109. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

110. Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. *20*, 278. https://doi.org/10.1186/s13059-019-1910-1.

111. Brůna, T., Lomsadze, A., and Borodovsky, M. (2024). GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. Genome Res. *34*, 757–768. https://doi.org/10.1101/gr.278373.123.

112. Huang, N., and Li, H. (2023). compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics *39*, btad595. https://doi.org/10.1093/bioinformatics/btad595.

113. Pertea, G., and Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. Preprint at F1000Research, https://doi.org/10.12688/f1000research.23297.2 https://doi.org/10.12688/f1000research.23297.2.

114. Gabriel, L., Brůna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M., and Stanke, M. (2024). BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. *34*, 769–777. https://doi.org/10.1101/gr.278090.123.

115. Rogers, T.F., Yalçın, G., Briseno, J., Vijayan, N., Nyholm, S.V., and Simakov, O. (2024). Gene modelling and annotation for the Hawaiian bobtail squid, Euprymna scolopes. Sci. Data *11*, 40. https://doi.org/10.1038/s41597-023-02903-8.

116. Zhang, Y., Mao, F., Mu, H., Huang, M., Bao, Y., Wang, L., Wong, N.-K., Xiao, S., Dai, H., Xiang, Z., et al. (2021). The genome of Nautilus pompilius illuminates eye evolution and biomineralization. Nat. Ecol. Evol. *5*, 927–938. https://doi.org/10.1038/s41559-021-01448-6.

117. Nevers, Y., Rossier, V., Train, C.M., Altenhoff, A., Dessimoz, C., and Glover, N. (2022). Multifaceted quality assessment of gene repertoire annotation with OMArk. bioRxiv, 2022.11.25.517970. https://doi.org/10.1101/2022.11.25.517970.

118. Goldfarb, T., Kodali, V.K., Pujar, S., Brover, V., Robbertse, B., Farrell, C.M., Oh, D.-H., Astashyn, A., Ermolaeva, O., Haddad, D., et al. (2025). NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. Nucleic Acids Res. *53*, D243–D257. https://doi.org/10.1093/nar/gkae1038.

119. Blum, M., Andreeva, A., Florentino, L.C., Chuguransky, S.R., Grego, T., Hobbs, E., Pinto, B.L., Orr, A., Paysan-Lafosse, T., Ponamareva, I., et al. (2025). InterPro: the protein sequence classification resource in 2025. Nucleic Acids Res. *53*, D444–D456. https://doi.org/10.1093/nar/gkae1082.

120. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.

121. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. *34*, 2115–2122. https://doi.org/10.1093/molbev/msx148.

122. Emms, D.M., Liu, Y., Belcher, L., Holmes, J., and Kelly, S. (2025). OrthoFinder: scalable phylogenetic orthology inference for comparative genomics. bioRxiv, 2025.07.15.664860. https://doi.org/10.1101/2025.07.15.664860.

123. Styfhals, R., Seuntjens, E., Simakov, O., Sanges, R., and Fiorito, G. (2019). In silico Identification and Expression of Protocadherin Gene Family in Octopus vulgaris. Front. Physiol. *Volume 9-2018*.

124. Ahmed, S., Stanley, D., and Kim, Y. (2018). An Insect Prostaglandin E2 Synthase Acts in Immunity and Reproduction. Front. Physiol. *9*, 1231. https://doi.org/10.3389/fphys.2018.01231.

125. Kanaoka, Y., and Urade, Y. (2003). Hematopoietic prostaglandin D synthase. Prostaglandins Leukot. Essent. Fatty Acids *69*, 163–167. https://doi.org/10.1016/s0952-3278(03)00077-2.

126. Tomarev, S.I., Chung, S., and Piatigorsky, J. (1995). Glutathione S-transferase and S-crystallins of cephalopods: evolution from active enzyme to lens-refractive proteins. J. Mol. Evol. *41*, 1048–1056. https://doi.org/10.1007/BF00173186.

127. Tan, W.-H., Cheng, S.-C., Liu, Y.-T., Wu, C.-G., Lin, M.-H., Chen, C.-C., Lin, C.-H., and Chou, C.-Y. (2016). Structure of a Highly Active Cephalopod S-crystallin Mutant: New Molecular Evidence for Evolution from an Active Enzyme into Lens-Refractive Protein. Sci. Rep. *6*, 31176. https://doi.org/10.1038/srep31176.

128. Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2021). CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics *36*, 5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.

129. Summers, M.F., Henderson, L.E., Chance, M.R., South, T.L., Blake, P.R., Perez-Alvarado, G., Bess Jr., J.W., Sowder III, R.C., Arthur, L.O., Sagi, I., et al. (1992). Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1. Protein Sci. *1*, 563–574. https://doi.org/10.1002/pro.5560010502.

130. Aceituno-Valenzuela, U., Micol-Ponce, R., and Ponce, M.R. (2020). Genome-wide analysis of CCHC-type zinc finger (ZCCHC) proteins in yeast, Arabidopsis, and humans. Cell. Mol. Life Sci. *77*, 3991–4014. https://doi.org/10.1007/s00018-020-03518-7.

131. Qi, R., Li, P., Wang, Q., Miao, J., Li, Z., and Pan, L. (2026). Integrated analysis of detoxification pathways and hepatotoxicity of butylated hydroxytoluene (BHT) in the Manila clam (Ruditapes philippinarum). Ecotoxicol. Environ. Saf. *309*, 119677. https://doi.org/10.1016/j.ecoenv.2026.119677.

132. Ahn, S.-J., and Marygold, S.J. (2021). The UDP-Glycosyltransferase Family in Drosophila melanogaster: Nomenclature Update, Gene Expression and Phylogenetic Analysis. Front. Physiol. *12*. https://doi.org/10.3389/fphys.2021.648481.

133. Gene Ontology Consortium (2026). The Gene Ontology knowledgebase in 2026. Nucleic Acids Res. *54*, D1779–D1792. https://doi.org/10.1093/nar/gkaf1292.

134. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

135. Redies, C. (1997). Cadherins and the formation of neural circuitry in the vertebrate CNS. Cell Tissue Res. *290*, 405–413. https://doi.org/10.1007/s004410050947.

136. Joazeiro, C.A.P., Wing, S.S., Huang, H., Leverson, J.D., Hunter, T., and Liu, Y.-C. (1999). The Tyrosine Kinase Negative Regulator c-Cbl as a RING-Type, E2-Dependent Ubiquitin-Protein Ligase. Science *286*, 309–312. https://doi.org/10.1126/science.286.5438.309.

137. Ryu, K.-B., Jo, G.-H., Gil, Y.-C., Jeon, D., Choi, N.-R., Jung, S.-H., Jo, S., An, H.S., Lee, H.-Y., Eyun, S., et al. (2023). Eye development and developmental expression of crystallin genes in the long arm octopus, Octopus minor. Front. Mar. Sci. *10*. https://doi.org/10.3389/fmars.2023.1136602.

138. Li, H., and Durbin, R. (2024). Genome assembly in the telomere-to-telomere era. Nat. Rev. Genet. *25*, 658–670. https://doi.org/10.1038/s41576-024-00718-w.

139. Lukhtanov, V.A., Dincă, V., Talavera, G., and Vila, R. (2011). Unprecedented within-species chromosome number cline in the Wood White butterfly Leptidea sinapis and its significance for karyotype evolution and speciation. BMC Evol. Biol. *11*, 109. https://doi.org/10.1186/1471-2148-11-109.

140. Lukhtanov, V.A., Dincă, V., Friberg, M., Šíchová, J., Olofsson, M., Vila, R., Marec, F., and Wiklund, C. (2018). Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. Proc. Natl. Acad. Sci. *115*, E9610–E9619. https://doi.org/10.1073/pnas.1802610115.

141. Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., et al. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat. Biotechnol. *30*, 771–776. https://doi.org/10.1038/nbt.2303.

142. Wang, J., and Zheng, X. (2017). Comparison of the genetic relationship between nine Cephalopod species based on cluster analysis of karyotype evolutionary distance. Comp. Cytogenet. *11*, 477–494. https://doi.org/10.3897/compcytogen.v11i3.12752.

143. Adachi, K., Ohnishi, K., Kuramochi, T., Yoshinaga, T., and Okumura, S.-I. (2014). Molecular cytogenetic study in Octopus (Amphioctopus) areolatus from Japan. Fish. Sci. *80*, 445–450. https://doi.org/10.1007/s12562-014-0703-4.

144. 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., et al. (2017). Molecular clocks indicate turnover and diversification of modern coleoid cephalopods during the Mesozoic Marine Revolution. Proc. R. Soc. B Biol. Sci. *284*, 20162818. https://doi.org/10.1098/rspb.2016.2818.

145. Anderson, F.E., and Lindgren, A.R. (2021). Phylogenomic analyses recover a clade of large-bodied decapodiform cephalopods. Mol. Phylogenet. Evol. *156*, 107038. https://doi.org/10.1016/j.ympev.2020.107038.

146. Lindgren, A.R., Pankey, M.S., Hochberg, F.G., and 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.

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

148. Schultz, D.T., Haddock, S.H.D., Bredeson, J.V., Green, R.E., Simakov, O., and Rokhsar, D.S. (2023). Ancient gene linkages support ctenophores as sister to other animals. Nature *618*, 110–117. https://doi.org/10.1038/s41586-023-05936-6.

149. Rogers, T.F., Stock, J., Schulz, N.G., Yalçin, G., Rencken, S., Weissenbacher, A., Clarence, T., Schultz, D.T., Ragsdale, C.W., Albertin, C.B., et al. (2025). Genome reorganisation and expansion shape 3D genome architecture and define a distinct regulatory landscape in coleoid cephalopods. bioRxiv, 2025.08.29.672809. https://doi.org/10.1101/2025.08.29.672809.

150. Yoshida, M., Tóth, E., Kon-Nanjo, K., Kon, T., Hirota, K., Toyoda, A., Toh, H., Miyazawa, H., Terauchi, M., Noguchi, H., et al. (2025). Giant genome of the vampire squid reveals the derived state of modern octopod karyotypes. iScience *28*. https://doi.org/10.1016/j.isci.2025.113832.

151. Amaral, P., Carbonell-Sala, S., De La Vega, F.M., Faial, T., Frankish, A., Gingeras, T., Guigo, R., Harrow, J.L., Hatzigeorgiou, A.G., Johnson, R., et al. (2023). The status of the human gene catalogue. Nature *622*, 41–47. https://doi.org/10.1038/s41586-023-06490-x.

152. Frankish, A., Carbonell-Sala, S., Diekhans, M., Jungreis, I., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Arnan, C., Barnes, I., et al. (2023). GENCODE: reference annotation for the human and mouse genomes in 2023. Nucleic Acids Res. *51*, D942–D949. https://doi.org/10.1093/nar/gkac1071.

153. Lynch, M., and Force, A. (2000). The probability of duplicate gene preservation by subfunctionalization. Genetics *154*, 459–473. https://doi.org/10.1093/genetics/154.1.459.

154. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics *151*, 1531–1545. https://doi.org/10.1093/genetics/151.4.1531.

155. Freemont, P.S. (2000). Ubiquitination: RING for destruction? Curr. Biol. *10*, R84–R87. https://doi.org/10.1016/S0960-9822(00)00287-6.

156. Gamsjaeger, R., Liew, C.K., Loughlin, F.E., Crossley, M., and Mackay, J.P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. Trends Biochem. Sci. *32*, 63–70. https://doi.org/10.1016/j.tibs.2006.12.007.

157. Wang, Y., Yu, Y., Pang, Y., Yu, H., Zhang, W., Zhao, X., and Yu, J. (2021). The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism. RNA Biol. *18*, 2107–2126. https://doi.org/10.1080/15476286.2021.1909320.

158. Focareta, L., Sesso, S., and Cole, A.G. (2014). Characterization of Homeobox Genes Reveals Sophisticated Regionalization of the Central Nervous System in the European Cuttlefish Sepia officinalis. PLOS ONE *9*, e109627. https://doi.org/10.1371/journal.pone.0109627.

159. Aken, B.L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., et al. (2016). The Ensembl gene annotation system. Database J. Biol. Databases Curation *2016*, baw093. https://doi.org/10.1093/database/baw093.

160. Veenstra, G.J.C. (2025). pita_genes_v0.3: Doryteuthis pealeii gene models. https://doi.org/10.5281/zenodo.15222686.

161. Ponte, G., Roumbedakis, K., COST Action FA1301, Galligioni, V., Dickel, L., Bellanger, C., Pereira, J., Vidal, E.A., Grigoriou, P., Alleva, E., et al. (2023). General and species-specific recommendations for minimal requirements for the use of cephalopods in scientific research. Lab. Anim. *57*, 26–39. https://doi.org/10.1177/00236772221111261.

162. Fiorito, G., Affuso, A., Basil, J., Cole, A., de Girolamo, P., D'Angelo, L., Dickel, L., Gestal, C., Grasso, F., Kuba, M., et al. (2015). Guidelines for the Care and Welfare of Cephalopods in Research –A consensus based on an initiative by CephRes, FELASA and the Boyd Group. Lab. Anim. *49*, 1–90. https://doi.org/10.1177/0023677215580006.

163. Andrews, P.L.R., Darmaillacq, A.-S., Dennison, N., Gleadall, I.G., Hawkins, P., Messenger, J.B., Osorio, D., Smith, V.J., and Smith, J.A. (2013). The identification and management of pain, suffering and distress in cephalopods, including anaesthesia, analgesia and humane killing. J. Exp. Mar. Biol. Ecol. *447*, 46–64. https://doi.org/10.1016/j.jembe.2013.02.010.

164. Cartolano, M., Huettel, B., Hartwig, B., Reinhardt, R., and Schneeberger, K. (2016). cDNA Library Enrichment of Full Length Transcripts for SMRT Long Read Sequencing. PLOS ONE *11*, e0157779. https://doi.org/10.1371/journal.pone.0157779.

165. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. *21*, 245. https://doi.org/10.1186/s13059-020-02134-9.

166. Formenti, G., Rhie, A., Walenz, B.P., Thibaud-Nissen, F., Shafin, K., Koren, S., Myers, E.W., Jarvis, E.D., and Phillippy, A.M. (2022). Merfin: improved variant filtering, assembly evaluation and polishing via k-mer validation. Nat. Methods *19*, 696–704. https://doi.org/10.1038/s41592-022-01445-y.

167. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

168. Lyčka, M., Bubeník, M., Závodník, M., Peska, V., Fajkus, P., Demko, M., Fajkus, J., and Fojtová, M. (2024). TeloBase: a community-curated database of telomere sequences across the tree of life. Nucleic Acids Res. *52*, D311–D321. https://doi.org/10.1093/nar/gkad672.

169. Akasaki, T., Nikaido, M., Tsuchiya, K., Segawa, S., Hasegawa, M., and Okada, N. (2006). Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. Mol. Phylogenet. Evol. *38*, 648–658. https://doi.org/10.1016/j.ympev.2005.10.018.

170. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

171. Li, H. (2013). Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences.

172. Shumate, A., Wong, B., Pertea, G., and Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. PLOS Comput. Biol. *18*, e1009730. https://doi.org/10.1371/journal.pcbi.1009730.

173. Haas, B.J. GitHub - TransDecoder/TransDecoder: TransDecoder source. https://github.com/TransDecoder/TransDecoder.

174. R Core Team (2024). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

175. Lovell, J.T., Sreedasyam, A., Schranz, M.E., Wilson, M., Carlson, J.W., Harkess, A., Emms, D., Goodstein, D.M., and Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife *11*, e78526. https://doi.org/10.7554/eLife.78526.

176. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

177. Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49. https://doi.org/10.1093/nar/gkr1293.

178. Hahsler, M., Piekenbrock, M., and Doran, D. (2019). dbscan: Fast Density-Based Clustering with R. J. Stat. Softw. *91*, 1–30. https://doi.org/10.18637/jss.v091.i01.

179. pysam-developers/pysam (2026). (pysam-developers).

180. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

181. Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics *34*, 867–868.

https://doi.org/10.1093/bioinformatics/btx699.

182. Emms, D.M., and Kelly, S. (2018). STAG: Species Tree Inference from All Genes. Preprint at bioRxiv, https://doi.org/10.1101/267914 https://doi.org/10.1101/267914.

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

184. Emms, D.M., and Kelly, S. (2017). STRIDE: Species Tree Root Inference from Gene Duplication Events. Mol. Biol. Evol. *34*, 3267–3278. https://doi.org/10.1093/molbev/msx259.

185. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

186. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

187. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

188. Zhu, A., Ibrahim, J.G., and Love, M.I. (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics *35*, 2084–2092. https://doi.org/10.1093/bioinformatics/bty895.

189. Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., Tang, W., Wang, Q., Liu, B., Wang, R., et al. (2024). Using clusterProfiler to characterize multiomics data. Nat. Protoc. *19*, 3292–3320. https://doi.org/10.1038/s41596-024-01020-z.

190. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

191. Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. (2020). BlobToolKit – Interactive Quality Assessment of Genome Assemblies. G3 GenesGenomesGenetics *10*, 1361–1374. https://doi.org/10.1534/g3.119.400908.

192. Kröger, B., Vinther, J., and Fuchs, D. (2011). Cephalopod origin and evolution: A congruent picture emerging from fossils, development and molecules. BioEssays *33*, 602–613. https://doi.org/10.1002/bies.201100001.

193. Song, W., Li, R., Zhao, Y., Migaud, H., Wang, C., and Bekaert, M. (2021). Pharaoh Cuttlefish, Sepia pharaonis, Genome Reveals Unique Reflectin Camouflage Gene Set. Front. Mar. Sci. *8*.