Peer-reviewed & recommended by PCI Genomics

Two *de novo* transcriptome assemblies and functional annotations from juvenile cuttlefish (*Sepia officinalis*) under various metal and *p*CO2 exposure conditions

Sol Dourdin Thomas*[1], Minet Antoine[1], Pante Eric**[2] & Lacoue-Labarthe Thomas**[1]

[1] Littoral Environnement et Sociétés (LIENSs), UMR 7266 CNRS – La Rochelle Université, 2 rue Olympe de Gouges, 17000 La Rochelle, France.
[2] CNRS, IRD, Ifremer, UMR 6539, Laboratory of Environmental Marine Sciences (LEMAR), Université de Bretagne, rue Dumont D'Urville, 29280 Plouzané, France.

*Corresponding author
Correspondence: t.soldourdin@gmail.com
**Authors contributed equally

ABSTRACT

The cuttlefish *Sepia officinalis* is a precious model in behavioural and neurobiology studies. It is currently facing combined environmental changes related to the anthropogenic global change. However, genomic resources available to support investigations tackling this issue are still scarce. Therefore, we present two annotated *de novo* transcriptome assemblies from recently hatched (whole body) and one-month-old (head) *Sepia officinalis* juveniles. Both assemblies rely on an important read depth validated by a pseudo-rarefaction analysis, and gathered several individuals from various metal and pCO_2 exposure conditions. After redundancy reduction, assemblies from newly hatched and one-month-old individuals comprised 230,672 and 370,613 transcripts with 35,590 and 44,233 putative ORFs, respectively, and an annotation rate arounf 70%. Assemblies were compared to each other, revealing age-specific transcriptomic landscapes. These two assemblies constitute highly valuable genomic resources complementing reference genome assembly and facilitating the investigation of transcriptomic endpoints in environmental studies considering coleoid cephalopods.

Keywords: Bioinformatics, Transcriptome announcement, Cephalopods, Ecotoxicology, Non-model species

Introduction

Cephalopods constitute a remarkable class within molluscs, exhibiting an outstanding diversity of morphologies, habitats and ecology. The class is divided in two subclasses: Nautiloidea Blainville, 1825 and Coleoidea Bather, 1888. While the former contains only a few living fossil species, the latter, counting more than 800 species, represents the large majority of extant cephalopods (*c.a.* octopuses, cuttlefishes, squids). Coleoid cephalopods also display some traits thought to be the result of convergent evolution with vertebrates such as camera-like eye, closed circulatory system and complex central nervous system (Packard, 1972).

So far mostly considered in behaviour and neurobiology studies, cephalopods are becoming promising model systems in ecotoxicology, for biomonitoring (Ajala et al., 2022) as well as experimental studies (García-Flores et al., 2025; Gouveneaux et al., 2023; Penicaud et al., 2017). This growing interest has been supported by the recent development of cephalopod genomic studies (Albertin & Simakov, 2019; Baden et al., 2023) producing several transcriptomes from diverse tissue types (*i.e. Octopus vulgaris* whole body, early life stages (Prado-Álvarez et al., 2022); *Idiosepius pygmaeus* central nervous system and eyes (Thomas et al., 2024)) along with genome assemblies (*i.e. O. vulgaris* (Destanović et al., 2023), *Amphioctopus fangsiao* (Jiang et al., 2022)) during the last years.

Among coleoids, the cuttlefish *Sepia officinalis* is of great economic interest in the northern east Atlantic as one of the main halieutic resource (Arkhipkin et al., 2021). Besides, *S. officinalis* from different areas were shown to be chronically contaminated by various metals such as silver, lead, cadmium or mercury (Ahmed et al., 2022; Ajala et al., 2022; Ihya et al., 2021; Minet et al., 2021; Miramand et al., 2006), as well as anti-depressant compounds like venlafaxine or fluoxetine (Chabenat, Bidel, et al., 2021; Chabenat, Knigge, et al., 2021) leading to behavioural alterations and neurotoxic effects (see Lacoue-Labarthe et al. (2016) for a focus on early-life stages) potentially amplified by ocean acidification (see Minet (2022) for the example of combined exposure with mercury). Yet, despite scientific and ecological relevance, genomic resources for *S. officinalis* remain limited: only one tissue specific transcriptome assembly was released (Gonçalves et al., 2024) and the first chromosome-level reference genome is about to be published (Rencken et al., 2025). High-quality transcriptomes provide essential complement to genome assemblies (Elliott, 2014; Martin & Wang, 2011) by capturing tissue- and condition-specific expression patterns, as well as providing qualitative diversity of isoforms resulting from alternative splicing or RNA editing – particularly abundant in cephalopods and promoting proteome plasticity (Rosenthal & Eisenberg, 2025; Voss & Rosenthal, 2023).

In this context, we present two *de novo* transcriptome assemblies built from the whole body of newly hatched juveniles (project “rnasep1”) and head tissues of one-month-old juveniles (project “rnasep2”) *S. officinalis* reared in various metal and pCO_2 conditions. The gathering of individuals from different experimental conditions enabled the provision of valuable ready-made genomic resources, facilitating future gene expression analyses and comparative transcriptomic studies in coleoid cephalopods.

Methods

Description of the biological material

Freshly laid eggs of common cuttlefish (*Sepia officinalis*) were collected in the Minimes Port, Bay of Biscay, France (46°08’44.7”N; 1°10’16.4”W) during the Spring of 2016 (rnasep1) and 2021 (rnasep2). For both experiments, eggs were gathered from multiple independent clutches, and each experimental replicate contained eggs originating from different parental pairs, thereby minimizing potential clutch- or family-specific effects.

Newly hatched juveniles

Upon field collection, eggs were immediately transported to the laboratory. They were individualized from their clutches and then placed in eighteen 9 L glass aquaria (n=30 eggs per aquarium; salinity: 34.4 ppt; temperature: 18.4 ± 0.6°C; 12 h light:12 h dark cycle, with aeration, recirculation pumps). After 5 days of acclimation, nine aquaria were maintained at ambient seawater pH (pHT = 7.95 ± 0.03) and the nine others were gradually acidified by bubbling CO_2 until reaching a pHT of 7.70 ± 0.08. The pCO_2 regulation was handled by an IKS system from Aquastar© and weekly calibration of IKS probes with NBS standard pH solutions (pH 4.0 and 7.0), a glass electrode and using TRIS buffer solutions (salinity 35, provided by A. Dickson, Scripps Institution of Oceanography, San Diego, USA). Among each pH condition, two set of three aquaria were spiked with $AgNO_3$ or $HgCl_2$ solutions in order to reach 1 $\mu g.L^{-1}$ of metal in seawater, respectively. The last three aquaria were kept metal-free for control conditions. This resulted in 6 rnasep1 experimental conditions: Control (hereon "CT"), acidification ("pCO_2"), exposed to $HgCl_2$ ("$HgCl_2$"), exposed to $AgNO_3$ ("$AgNO_3$"), exposed to both acidification and $HgCl_2$ ("pCO_2 + $HgCl_2$") or exposed to acidification and $AgNO_3$ ("pCO_2+ $AgNO_3$") (Figure 1a). Water renewal was done every two days and spikes of metals were renewed accordingly to maintain the metal concentrations as constant as possible.

Eggs were maintained in these conditions for 40 days, until hatching. When hatched (considering a daily survey), juveniles were immediately collected, anaesthetized with 17.5 $mg.L^{-1}$ $MgCl_2$ and eventually put to death using 2% ethanol in seawater. Then, individuals (n = 6 per condition) were frozen in liquid nitrogen and stored at -80°C before RNA extraction that was performed straight away. Additionally, fourteen hatchlings were sampled for measurement of Ag and Hg concentrations in whole body to validate the exposure conditions (data available in sample description at ENA PRJEB89715, as well as in table S1a).

One-month-old juveniles

Eggs were acclimated in the laboratory and raised until one-month old juvenile stage to investigate the effects of pCO_2 and MeHg on behaviour of cuttlefish early life stages (Minet, 2022). Briefly, eggs were individualized from their clutches and then acclimated for two weeks in four black plastic basins (*i.e.* batches) in control conditions (n=70 eggs per batch; salinity: 35.0 ppt; temperature: 18.3 ± 0.2°C; pH_{NBS}: 8.0 ± 0.1; 12 h light:12 h dark cycle, with aeration, recirculation pumps and daily water renewal). After the acclimation period, two batches remained in control conditions (pH_{NBS}: 8.03 ± 0.05 in 2020 and 8.01 ± 0.05 in 2021, equivalent to ~440 µatm), while the two others were gradually acclimated to acidified conditions by reducing pH_{NBS} (CO_2 bubbling) from 8.01 to 7.54 ± 0.06 (equivalent to ~1600 µatm). The pCO_2 regulation was handled by an IKS system from Aquastar© (pH measurements every 20 min and weekly calibration of IKS probes with NBS standard pH 4.0 and 7.0 solutions and a glass electrode.

Immediately after hatching, 27 individuals from each pH condition were transferred into six 20 L black plastic basins and individualized in cylindrical baskets of 60 mm width (*c.a.* 9 indiv. per batch in triplicate). Thereafter, juvenile cuttlefishes were fed *ad libitum* with control (half of the individuals) or MeHg-contaminated (other half of the individuals) live ditch shrimp (*Palaemon varians*). For the MeHg-contaminated condition, only the first shrimp of each day was contaminated to ensure the number of contaminated shrimps eaten to be the same for all MeHg-exposed individuals. This experiment resulted in 4 distinct conditions: Control (hereon "CT"), Acidification ("pCO_2"), contaminated by methyl-mercury ("MeHg") or a combination of acidification and methyl-mercury contamination ("pCO_2+MeHg") (Figure 1b).

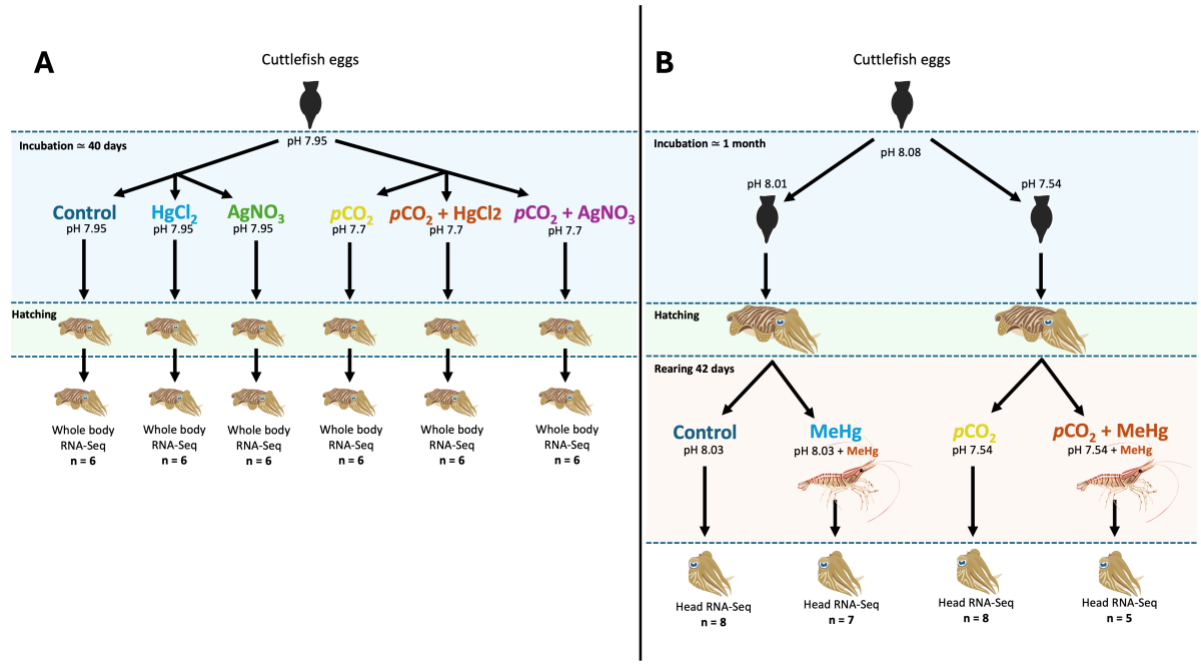

Figure 1 - Experimental design for A) newly hatched ("rnasep1"; carried out in 2016) and B) one-month-old juveniles ("rnasep2"; carried out in 2021) exposure to heavy metals and hypercapnia.

Hatchlings were raised for 42 days before being anaesthetized with 17.5 mg.L^{-1} $MgCl_2$ and eventually sacrificed using 2% ethanol in seawater. In order to investigate the transcriptomic responses in the whole central nervous system to the experimental conditions, the heads of half of the individuals (CT: n=8; pCO_2: n=8; MeHg: n=7; pCO_2+MeHg: n=5) were immediately sampled and freeze dried in liquid nitrogen before storage at -80°C until RNA extraction. Total mercury concentrations were measured in the muscle of mantle to validate the dietary exposure (data available in sample description at ENA PRJEB88225, as well as in Table S1b).

The experiments were performed in accordance with the French and the European regulations (Directive 2010/63/EU) for the use of animals for scientific purposes. The experimental procedures were referenced and authorized by the French Ministry of Higher Education and Research for the use of animals for scientific purposes (APAFIS # 20520-2019050614554709).

RNA extraction and Illumina sequencing

Newly hatched juveniles

RNA extraction was performed using the Nucleospin for NucleoZOL (Macherey Nagel©), using manufacturer instructions. Whole individuals were first ground in liquid nitrogen. The resulting power was incubated in NucleoZOL + RNase-free water (>700 µL/50 mg of tissue powder) for 15 min after 15 s of vigorous vortexing. A 700 µL aliquot was then used for downstream RNA extraction steps. RNA was eluted in 60 µL RNase-free H_2O. A 5 µL aliquot was sampled for quantification on Qubit (using Thermo Fisher Scientific© HS and BR quantification kits) and 1% gel electrophoresis; the rest of the RNA extraction was immediately placed at -80°C until next steps. Further quality control on a Fragment Analyser (Agilent©), library construction and Illumina sequencing were outsourced to GeT-PlaGe platform from GenoToul, using the Illumina TruSeq Stranded mRNA Library Prep kit and an Illumina HiSeq 3000 platform (2x150bp).

Table 1 - Sequencing statistics for newly hatched juveniles. GeT-PlaGe sample code is used for the labelling of reads in the European Nucleotide Archive repository. R1 GC% and R2 GC% are percent GC content on the forward and reverse reads, respectively.

GeT-PlaGe Sample Code	Sample Name	Number of Paired-end Reads	R1 GC%	R2 GC%
Ag771	SEP1-Ag7.7_1	21,283,196	40	41
Ag8-1-102	SEP1-Ag8.1_102	21,080,270	41	41
Ag7-7-29	SEP1-Ag7.7_29	17,316,330	41	41
Ag772	SEP1-Ag7.7_2	19,177,260	41	41
Ag7-7-31	SEP1-Ag7.7_31	22,963,224	41	41
Ag7-7-34	SEP1-Ag7.7_34	20,399,022	41	41
Ag773	SEP1-Ag7.7_3	17,805,326	41	41
Ag8-1-21	SEP1-Ag8.1_21	23,068,454	41	41
Ag8-1-27	SEP1-Ag8.1_27	21,202,072	40	41
Ag812	SEP1-Ag8.1_2	19,480,720	41	41
Ag813	SEP1-Ag8.1_3	20,526,026	41	41
Ag816	SEP1-Ag8.1_6	21,223,080	41	41
Hg7-7-125	SEP1-Hg7.7_125	24,256,850	41	41
Hg771	SEP1-Hg7.7_1	22,357,978	40	41
Hg772	SEP1-Hg7.7_2	19,811,618	40	41
Hg7-7-48	SEP1-Hg7.7_48	20,937,382	41	41
Hg7-7-49	SEP1-Hg7.7_49	19,721,682	41	41
Hg7-7-52	SEP1-Hg7.7_52	19,410,228	41	41
Hg8-1-111	SEP1-Hg8.1_111	21,700,826	41	42
Hg811	SEP1-Hg8.1_1	17,635,966	40	41
Hg812	SEP1-Hg8.1_2	24,195,958	40	40
Hg8-1-35	SEP1-Hg8.1_35	20,576,074	41	41
Hg8-1-36	SEP1-Hg8.1_36	19,962,954	40	40
Hg816	SEP1-Hg8.1_6	20,235,544	41	41
CT7-7-15	SEP1-CT7.7_15	18,380,114	40	41
CT7-7-18	SEP1-CT7.7_18	18,430,516	41	42
CT772	SEP1-CT7.7_2	17,710,208	40	40
CT773	SEP1-CT7.7_3	16,961,648	40	40
CT774	SEP1-CT7.7_4	15,762,346	40	40
CT7-7-93	SEP1-CT7.7_93	18,755,500	41	41
CT811	SEP1-CT8.1_1	20,027,126	40	41
CT812	SEP1-CT8.1_2	18,261,584	41	41
CT813	SEP1-CT8.1_3	18,859,762	41	41
CT8-1-7	SEP1-CT8.1_7	19,050,674	41	41
CT8-1-82	SEP1-CT8.1_82	20,522,436	41	41
CT8-1-9	SEP1-CT8.1_9	17,506,642	40	41

One-month-old juveniles

The RNA extraction and library preparation were outsourced to Fasteris©. Briefly, 28 heads underwent manual grinding in liquid nitrogen and shredded tissues were then preserved in RLT and beta-mercaptoethanol. Total RNA was isolated using the RNeasy 96 kit (Qiagen©, ref 74181) with final elution in 50 µL of nuclease-free water. After the extraction step, samples were quantified by Quant-iT (Thermo Fisher Scientific©) and the quality of RNA was checked by loading 2 µL of a dilution on a Fragment Analyzer (Agilent©). All samples reached the required concentration of 250-500 ng in a total volume of 50 µL. After poly(A) tail selection, libraries were constructed using the Illumina TruSeq Stranded mRNA Library Prep kit with NEXTFLEX unique dual index barcodes (Revvity©). The libraries were sequenced on an Illumina NovaSeq 6000 platform (2x150bp).

Table 2 - Sequencing statistics for one-month-old juveniles. Fasteris sample code is used for the labelling of reads in the European Nucleotide Archive repository. R1 GC% and R2 GC% are percent GC content on the forward and reverse reads, respectively.

Fasteris Sample Code	Sample Name	Number of Paired-end Reads	R1 GC%	R2 GC%
AVRX-1	SEP2-CT8.1_5	31,819,911	42	42
AVRX-2	SEP2-CT8.1_6	37,005,211	43	43
AVRX-3	SEP2-CT8.1_8	29,634,371	42	43
AVRX-4	SEP2-CT8.1_9	33,148,390	42	43
AVRX-5	SEP2-CT8.1_37	25,239,430	42	42
AVRX-6	SEP2-CT8.1_38	31,105,749	43	43
AVRX-7	SEP2-CT8.1_39	33,593,329	41	41

Fasteris Sample Code	Sample Name	Number of Paired-end Reads	R1 GC%	R2 GC%
AVRX-8	SEP2-CT8.1_80	28,294,978	42	43
AVRX-9	SEP2-CT7.7_10	29,927,066	43	43
AVRX-10	SEP2-CT7.7_11	27,023,641	43	43
AVRX-11	SEP2-CT7.7_12	31,302,866	42	42
AVRX-12	SEP2-CT7.7_13	26,791,740	43	44
AVRX-13	SEP2-CT7.7_46	33,106,734	42	42
AVRX-14	SEP2-CT7.7_48	27,834,839	45	46
AVRX-15	SEP2-CT7.7_85	27,648,758	43	43
AVRX-16	SEP2-CT7.7_86	30,032,397	43	43
AVRX-17	SEP2-Hg8.1_19	29,059,690	43	43
AVRX-18	SEP2-Hg8.1_20	32,381,027	44	44
AVRX-19	SEP2-Hg8.1_21	24,445,510	43	45
AVRX-20	SEP2-Hg8.1_24	30,206,664	42	43
AVRX-21	SEP2-Hg8.1_25	31,415,272	43	43
AVRX-22	SEP2-Hg8.1_55	24,646,853	43	43
AVRX-23	SEP2-Hg8.1_56	30,177,602	42	42
AVRX-24	SEP2-Hg7.7_28	28,964,924	42	45
AVRX-25	SEP2-Hg7.7_30	39,767,058	42	41
AVRX-26	SEP2-Hg7.7_31	26,843,136	41	42
AVRX-27	SEP2-Hg7.7_32	44,104,080	41	41
AVRX-28	SEP2-Hg7.7_68	25,479,196	41	41

Pre-assembly processing

The same pipeline of bioinformatic analyses (Figure 2) was carried out for rnasep1 and rnasep2, using the IFB-core high-performance computing cluster of the Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013). The raw data quality was assessed using FastQC (fastqc_v0.12.1) (Andrews, 2010) in combination with MultiQC (multiqc_v1.14) (Ewels et al., 2016). Fastp (fastp_v0.23.1) (Chen, 2023) was used to filter out adapter sequences as well as low-quality reads (PHRED score ≤ 30), and retrieve only post-filtering reads longer than 120 bp. Kraken2 (kraken2_v2.14) was used to identify and remove any prokaryote (bacteria and archaea) contamination.

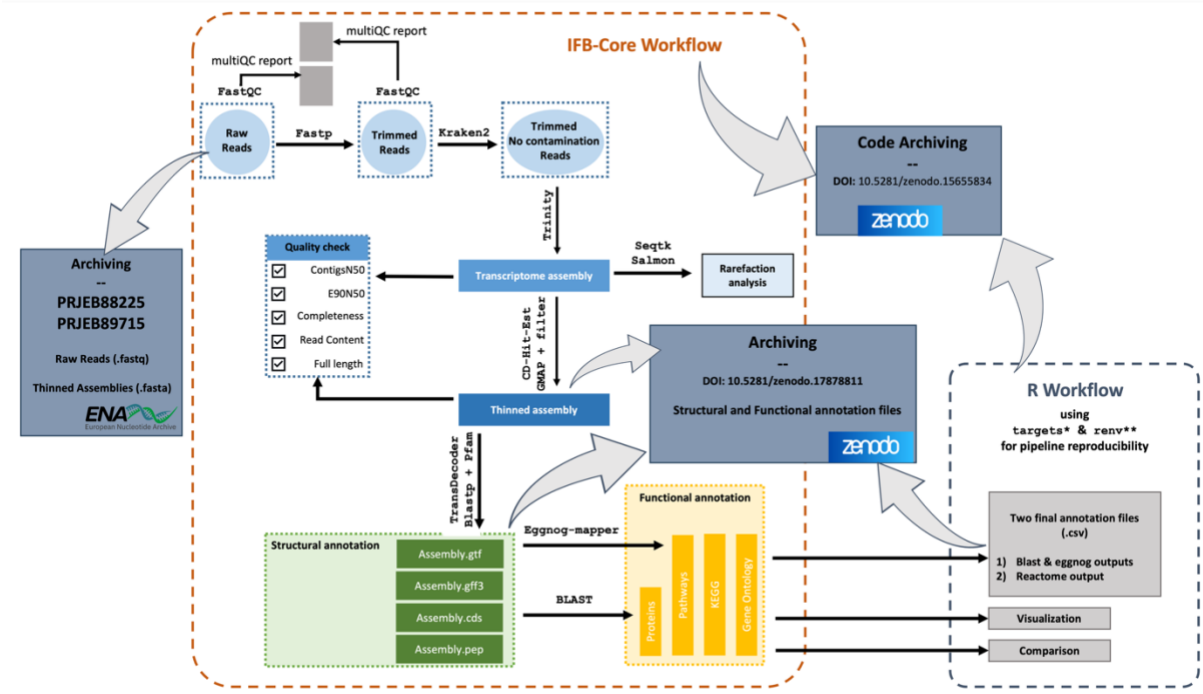

Figure 2 - Bioinformatic workflow from raw data to final assembly and annotation files. *(Landau, 2021); **(Ushey & Wickham, 2025)

De novo transcriptome assembly and thinning

Eighteen samples (n=3 per condition) and 16 samples (n=4 per condition) were randomly selected to be included in the rnasep1 and rnasep2 assemblies, respectively, to reduce the time and resource consuming. The *de novo* transcriptome assembly was performed using Trinity (trinity_v2.15.1) (Grabherr et al., 2011) with default parameters and *in-silico* normalization.

Assembled isoforms were clustered based on their similarities (default percentage identity threshold used: 95%) using the CD-HIT tool (cd-hit_v4.8.1) (Li & Godzik, 2006) to retain only the most representative sequences, thereby reducing transcriptome redundancy likely to result from the Trinity method. The resulting assemblies were mapped onto the genomic sequence GCA_050097725.1 (Rencken et al., 2025) using GMAP (gmap_v2020.06.01) (Wu & Watanabe, 2005). From the gff3 file produced, we identified, then filtered out chimeras transcripts (mapping on multiple chromosomes or opposite strands, with too long introns and rearranged exons), as well as transcripts with alignments < 100 bp or < 20% of their total length. Transcripts smaller than 200 bp were removed as well to comply with the European Nucleotide Archive standards.

Assessment of the transcriptome assemblies

Several metrics were used to fully assess the quality of the assembly. First, TransRate (transrate_v1.0.3) (Smith-Unna et al., 2016) retrieved general statistics from the raw and thinned assemblies. Then, the contigs N50 statistic (at least 50% of the assembled transcript nucleotides are found in contigs that are at least of N50 length), based on the longest isoform per gene and the E90N50 statistic (N50 for top 90% most highly expressed genes) were computed using dedicated Trinity utilities and Kallisto (kallisto_v0.46.2) (Bray et al., 2016). BUSCO (Busco_v5.5.0) (Manni et al., 2021) was used to evaluate the completeness of the transcriptome against two ortholog databases: Metazoa (metazoan_odb10) and Mollusca (mollusca_odb10). The full-length transcript analysis provided by Trinity was also performed using BLAST (blast_v2.16.0) (Camacho et al., 2009) and STAR (star_v2.7.11a) (Dobin et al., 2013) was used to align back the reads against the assembly to assess its read content.

To confirm that the reduced number of samples still ensure a sufficient read depth to catch all available transcripts, we performed an *ad hoc* rarefaction analysis as follows. For each assembly, all the input R1 and R2 fastq files were concatenated, then reads were sub-sampled (keeping pair information) using seqtk (seqtk_v1.3). (https://github.com/lh3/seqtk). For newly-hatched individuals, 1 to 80 million read pairs were sub-sampled. For one-month-old juveniles, 1 to 300 million read pairs were sub-sampled. Then, Salmon (salmon_v1.10.2) (Patro et al., 2017) was used to index Trinity assemblies and quantify the number of reads from each corresponding sub-sample that matches each Trinity transcript. Finally, the number of Trinity transcripts matched by at least 1 read was reported for each sub-sample.
(Haas, n.d.) (Camacho et al., 2009)

Transcriptome annotation

Transcriptome annotation was performed stepwise. First, the putative coding regions (open reading frames [ORFs] ≥ 100 amino acids) were identified by Transdecoder (transdecoder_v5.7.0) (Haas, n.d.), and the prediction was enhanced by homology evidence from BLASTp (blast_v2.16.0) (Camacho et al., 2009) and domain detection using Pfam (Mistry et al., 2021) *via* HMMER/hmmscan (hmmer_v3.3.2) (Eddy, 2011). To generate a representative protein set for downstream annotation, we retained only the longest ORF for each transcript isoform. Then, predicted protein-coding sequences were annotated using a two-step BLASTp strategy. We first performed similarity searches against a curated molluscan UniProt/SwissProt (Bateman et al., 2023) database to obtain taxonomically relevant annotations. Sequences with no significant hits were subsequently queried against the full UniProt/SwissProt database to obtain additional functional annotations. The two BLASTp steps were run using an e-value threshold of 1e-5,

retaining only the best hit per query (Bateman et al., 2023). The best hit was retained for each query and its associated Gene Ontology (GO) terms were extracted from the UniProt GO annotation file (*goa_uniprot_all.gaf.gz*). EggNOG-mapper v2 (eggnog-mapper_v2.1.12) (Cantalapiedra et al., 2021) tool was used to complement homology-based annotations by performing functional annotations based on precomputed orthologous groups and complete the Blast annotation with hits from the PFAM (Mistry et al., 2021), COG (Tatusov et al., 2000), KEGG (Kanehisa et al., 2017) and GO (Ashburner et al., 2000) databases. Searches were performed using DIAMOND, restricting the taxonomic scope to bilateria (NCBI TaxID 33213) and retaining only the one-to-one orthologs with an e-value cutoff of 0.001.

Comparison between newly hatched and one-month-old juveniles

The predicted protein sequences from both assemblies were compared to each other through a reciprocal BLASTp (blast_v2.16.0) (Camacho et al., 2009) to identify reciprocal best hits (RBH) considered to be proteins shared between both assemblies, as well as sequences exhibiting no hit considered as putatively specific. The number of RBH and no-RBH sequences between the two assemblies was represented using the ggvenn R package (ggvenn_v0.1.19). The running of GO enrichment analysis on the lists of specific sequences using GO_MWU script (Fisher's exact test, FDR: 0.1) did not identified significant enrichment. Therefore, the top 20 specific GO terms from 'biological_process' namespace were displayed in scatter plots and summarised by their level 6 ontology using the ontologyIndex (ontologyindex_v2.12) (Greene et al., 2017) R package to facilitate interpretation.

Results

Illumina sequencing and pre-processing

The Illumina sequencing produced a total of 358,400,00 paired-end reads for newly hatched and 830,584,708 paired-end reads for one-month-old juveniles (see Tables 1 and 2 for sequencing statistics). The pre-processing resulted in 350,000,000 and 1,108,000,000 cleaned reads accounting for 48.8% and 66.7% of the raw data for newly hatched and one-month-old juveniles, respectively. Then, taxonomic classification of reads did not identify any prokaryote contamination.

De novo transcriptome assembly quality assessment

For newly hatched juveniles, the thinned transcriptome assembly contained 230,572 transcripts with an average length of 677 bp (Table 3). The quality of the assembly was validated thanks to several metrics. About 83% and 95% of the orthologs from the BUSCO Mollusca and Metazoa databases respectively matched the assembly. In addition, 1,301 proteins matched Trinity transcripts from 80% to 90% of their protein length and 4,892 proteins were represented by nearly full-length transcripts. Besides, 91,3% of the reads used in the assembly process mapped back to the transcriptome. A more detailed alignment report is in Table S2a.

The one-month-old juveniles transcriptome assembly obtained after redundancy reduction comprised 370,613 transcripts with a mean length of 569 bp (Table 4). A high proportion of conserved orthologs from the BUSCO Mollusca (83%) and Metazoa (95%) databases were found in the assembly. Moreover, 1,352 proteins matched Trinity transcripts from 80% to 90% of their protein length and 4,927 proteins were represented by nearly full-length transcripts. Finally, 93.4% of the reads used to build the assembly mapped back to it. A more detailed alignment report is in Table S2b.

Rarefaction analysis showed that, for both dataset, the subset of individuals used for assembly is an acceptable compromise between exhaustiveness of transcript inventory and computational effort (Figure 3).

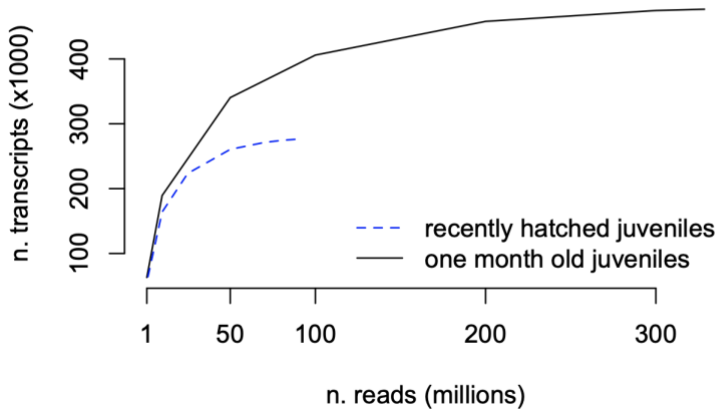

Figure 3 - Rarefaction curves performed on the raw Trinity assemblies.

Based on this assessment, we can assume that both assemblies are of good quality in terms of transcripts length and representation of conserved transcripts.

Table 3 - Transcriptome assembly and assessment metrics for Raw and Thinned assemblies for newly hatched juveniles

Type	Metric	Raw	Thin
Baseline			
	Number of Trinity "genes"	207,672	196,944
	Number of transcripts	282,660	230,572
	GC content (%)	36.4	36.01
	Smallest transcript	174	200
	Largest transcript	36,551	36,551
	Mean transcript length	741	677
	Bases with N (ambiguity)	0	0
	Proportion of N bases	0	0
	N50 (longest isoform per gene)	758	752
	E90N50	1,932	1,864
	TransRate predicted number of transcripts with ORF	35,681	24,804
	Mean ORF (%)	49.43	47.99
Completeness			
	Mollusca_odb Complete and Single-copy (%)	43.9	53.8
	Mollusca_odb Complete and duplicated (%)	43.0	28.8
	Metazoa_odb Complete and Single-copy (%)	61.7	72.1
	Metazoa_odb Complete and duplicated (%)	36.2	23.1
FullLength			
	Number of proteins matching Trinity transcript from 80% to 90% of their protein length	1295	1,301
	Number of proteins reprented by nearly full-length (>80%) transcripts	4901	4,892
ReadContent			
	Aligned reads (%)	97.8	91.3

Table 4 - Transcriptome assembly and assessment metrics for Raw and Thinned assemblies for one-month-old juveniles

Type	Metric	Raw	Thin
Baseline			
	Number of Trinity "genes"	331,430	311,810
	Number of transcripts	483,699	370,613
	GC content (%)	36.94	36.49

Type	Metric	Raw	Thin
	Smallest transcript	179	200
	Largest transcript	33,040	32,845
	Mean transcript length	621	569
	Bases with N (ambiguity)	0	0
	Proportion of N bases	0	0
	N50 (longest isoform per gene)	594	591
	E90N50	2,005	1,977
	TransRate predicted number of transcripts with ORF	43,409	28,159
	Mean ORF (%)	49.5	48.5
Completeness			
	Mollusca_odb Complete and Single-copy (%)	38.9	50.5
	Mollusca_odb Complete and duplicated (%)	48.6	32.3
	Metazoa_odb Complete and Single-copy (%)	59.5	70.1
	Metazoa_odb Complete and duplicated (%)	38.8	24.9
FullLength			
	Number of proteins matching Trinity transcript from 80% to 90% of their protein length	1467	1352
	Number of proteins represented by nearly full-length (>80%) transcripts	5388	4927
ReadContent			
	Aligned reads (%)	98.6	93.4

Transcriptome annotation

The structural annotation performed with Transdecoder on the thinned assemblies identified 42,959 and 53,992 ORFs for rnasep1 and rnasep2, respectively, which is consistent with the number of proteins reported from the reference genome (n = 50,979) (Rencken et al., 2025). After the filtering of longest ORF per transcript isoform, we obtained 35,590 ORFs for rnasep1 and 44,233 ORFs for rnasep2. These ORFs were submitted to functional annotation.

In the assembly of newly hatched individuals, 26,290 ORFs were annotated at least once. BLASTp annotated 24,299 transcripts against the SwissProt/Uniprot database (i.e. 4,958 from the curated database and 19,341 from the whole database). In addition, 21,351 orthologous hits were found by eggNOG-mapper with 14,709 transcripts annotated with gene ontology terms (for the occurrence of the top 20 GO terms from each namespace: Figure 4a). Moreover, 12,967 transcripts were associated to KEGG KO and 6,947 to KEGG pathways (Figure 5a).

In the one-month-old assembly, 29,474 ORFs were annotated at least once. BLASTp annotated 21,014 transcripts against the SwissProt/Uniprot database (i.e. 6,370 from the curated database and 14,644 from the whole database). This annotation was completed using EggNOG-mapper. Overall, 24,196 orthologous hits were reported and 16,204 transcripts matched gene ontology terms (Figure 4b), 14,269 transcripts matched KEGG KO and 7,444 matched KEGG pathways (Figure 5b).

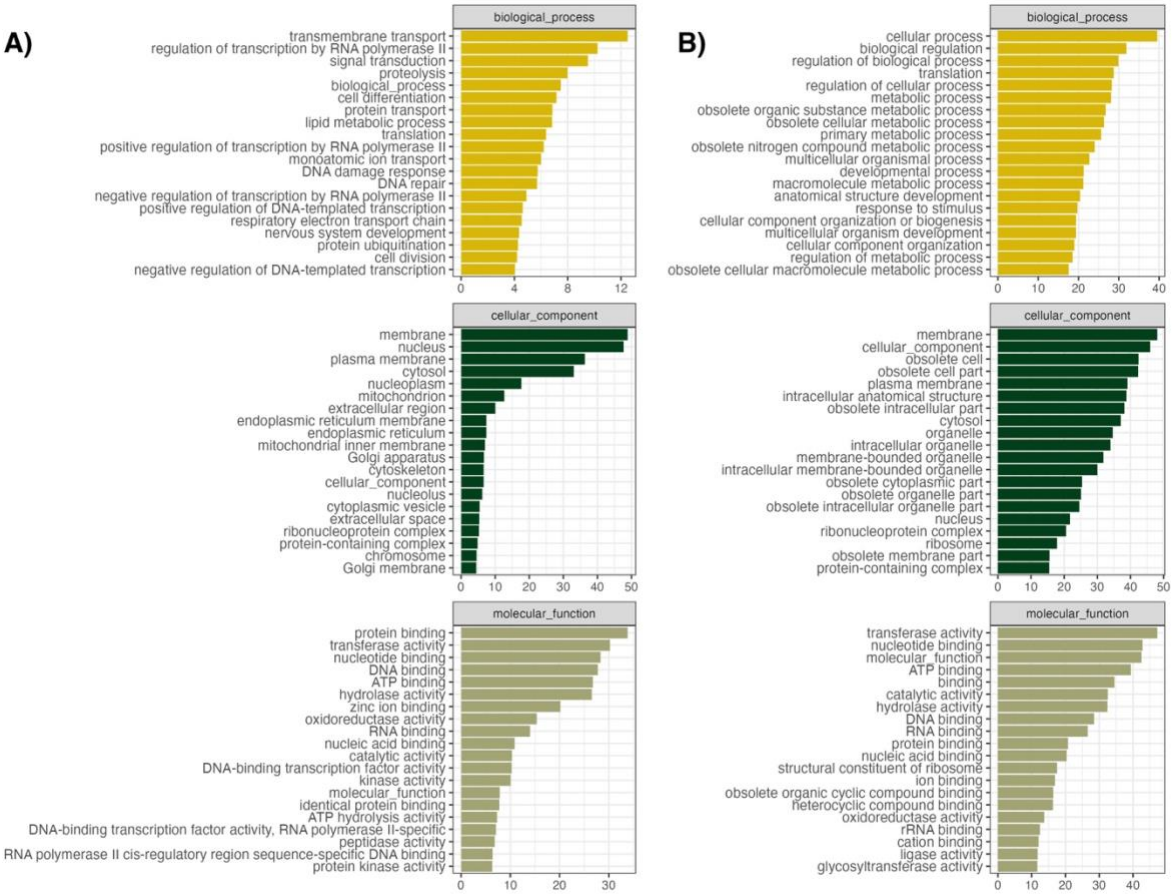

Figure 4 - Occurrence of the 20 top GO terms related to the putative ORFs by namespace for A) newly hatched and B) one-month-old juveniles.

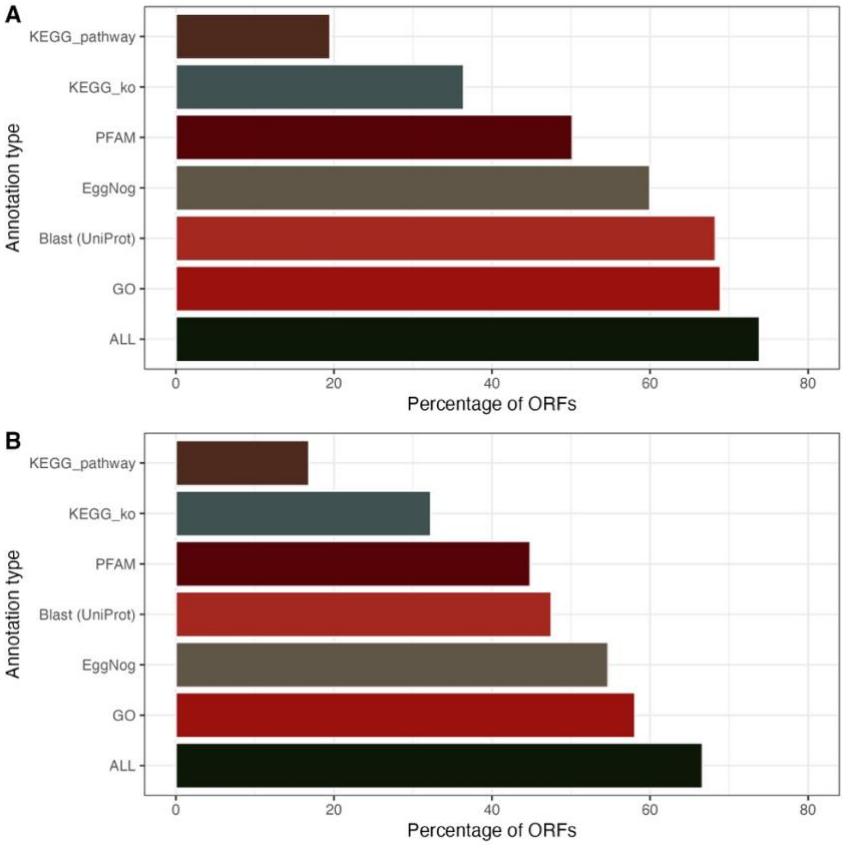

Figure 5 - Percentage of ORFs matching at least one entry from each annotation type for A) whole body hatchlings and B) head of one-month-old juveniles.

Newly hatched vs one-month-old juveniles

Overall, 17,526 transcripts were shared between rnasep1 and rnasep2 assemblies (RBH), representing 49% and 40% of each assembly, respectively (Figure 6). Consequently, 18,064 and 26,707 transcripts did not match RBH requirements in rnasep1 and rnasep2, respectively. Among them, ~40% (rnasep1) and ~20% (rnasep2) were classified as “putative specific” (i.e. no hit from the reciprocal BLASTp). The visualization of the top 20 unique GO categories from the “putative specific” sequences revealed consistent patterns (Figure 6). Three main functional themes are highlighted from newly hatched juveniles (whole body): early structuring of neural network (e.g. *neural crest formation*, *dendrite arborization* and *presynaptic modulation of chemical synaptic transmission*), response to cellular stress and protein quality (e.g. *negative regulation of PERK-mediated unfolded protein response*, *protein branched polyubiquitination*) and metabolism/transport/homeostasis (e.g. *glycine betaine transport*, *amino acid transport*, *plasma membrane phospholipid scrambling*). The presence of *neural crest formation* term echoes the hypothesis of cordal pattern of neurogenesis in cephalopods (Shigeno et al., 2015), and other neural structuring-related terms are consistent with the active neural development at this stage (Nixon & Mangold, 1998). The main unique GOs from one-month-old juveniles (head tissues) displayed clear pattern as well. There is a strong signal of lipid processing (e.g. *sphinganine metabolic process*, *unsaturated fatty acid metabolic process*, *acylglycerol acyl-chain remodeling*), neural signalling (e.g. *neuronal ion channel clustering*, *catecholamine transport*, *spinal cord association neuron specification*), secondary metabolism and post-translational modifications (e.g. *C-terminal protein amino acid modification*, *RNA 5’-end processing*).

Overall, specific transcripts from rnasep1 are dominated by developmental processes such as nervous system structuring, cellular stress control and some metabolic functions potentially implicated in the post-hatching transition (Minet et al., 2025). On the other side, specific transcripts from rnasep2 are more clearly associated with a mature neural system functioning, including lipidic

remodelling of membranes, biosynthesis and metabolism of sphingolipids as well as neurotransmission. These main functions are consistent with the lipid-rich cephalopod brain (Navarro et al., 2000), which specifically contains sphingolipids (Kreps et al., 1968), and the role catecholamines (mostly dopamine and noradrenaline) play as neurotransmitter in *Sepia officinalis* nervous system (Juorio, 1971). Therefore, Figure 6 shows that specific transcripts are likely to results from biological signal related to developmental stage and tissue diversity.

In addition, the difference in the number of transcripts detected between newly hatched and one-month-old juveniles is more likely to result from biological signal than technical artefact. Indeed, from the Figure 3, we can argue that the sampling depth was enough in both assemblies to capture all the information contained in the sequencing data. Moreover, the brief GO analysis demonstrated that both assemblies encompass biological specificities resulting from the collected tissues (whole body vs head) and the developmental stage of the individuals. In this sense, cephalopod brains are known to produce a high quantity of transcripts, in particular thanks to an extraordinary rate of mRNA editing (~60% of brain transcripts) (Rosenthal & Eisenberg, 2025; Voss & Rosenthal, 2023) resulting in a unique diversification of isoforms. And this diversification is likely to occur with the acquisition of new behaviours triggered by environmental context experienced throughout life (Fischer et al., 2021), explaining the higher number of isoforms found in one-month-old juveniles while newly hatched individuals probably express only innate behaviour-related transcripts. Moreover, rnasep1 individuals may still be under transcriptional canalization due to their temporal proximity with hatchling transition, that is embryonic stage, resulting in an overall less variable transcriptomic activity (Irvine, 2020). Finally, the predominance of brain tissues in rnasep2 while rnasep1 contains a lot of tissues producing poorly diversified transcripts (i.e. mantle, muscles) is likely to dilute signal from more plastic tissues (neural tissues), thereby lower the overall variability as well.

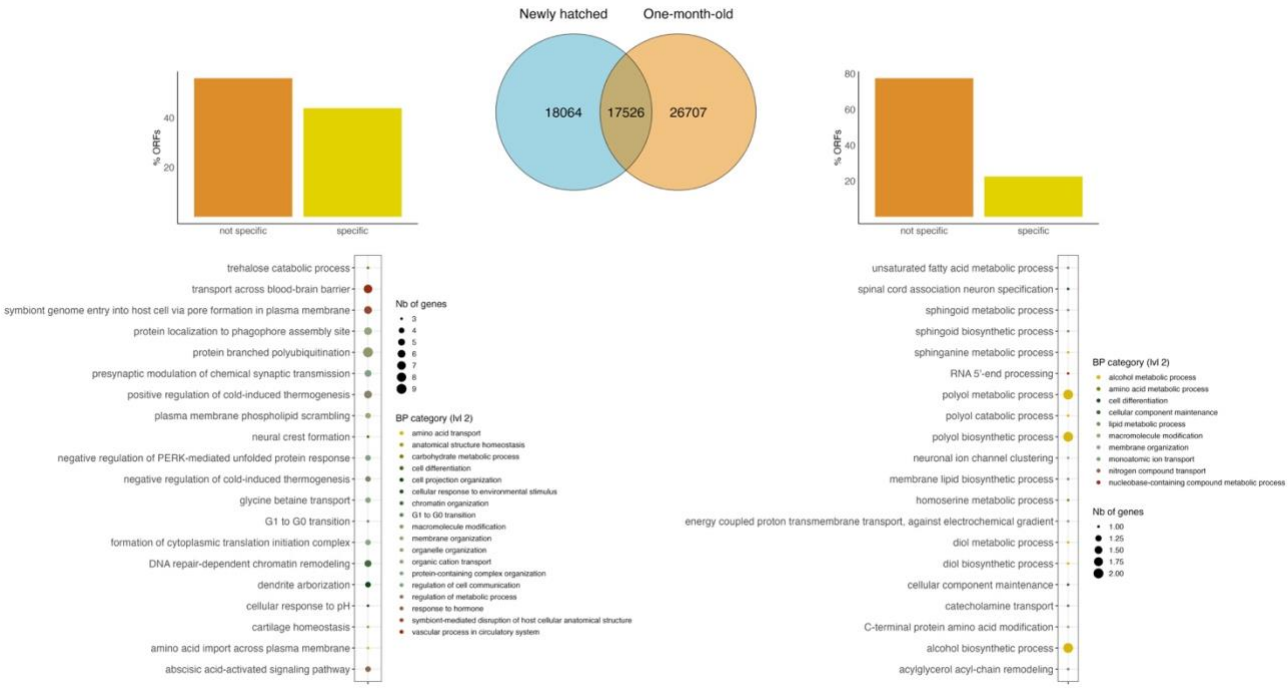

Figure 6 - Venn diagram displaying the number of RBH and non-RBH protein sequences. The bar plots represent the proportion non-RBH sequences specific to each assembly (i.e. no Blastp hit; left: newly hatched juveniles; right: one-month-old juveniles). The top 20 (based on occurrrence) level 6 GO terms are displayed in the scatter plots (left: newly hatched juveniles; right: one-month-old juveniles).

Conclusion

Both transcriptome assemblies significantly contribute to improve molecular resources for conducting genomic studies on Cephalopods, particularly *S. officinalis*. The overall overlap rate (49% and 40% for rnasep1 and rnasep2, respectively) between transcripts of whole body hatchlings and head of one-month-old juveniles, as well as the functional specificity of unique sequences attest to the importance of considering developmental stages and tissues in transcriptomic studies. In this sense, the use of multiple tissues (whole body in newly hatched and heads in one-month-old juveniles), although it may obscure some tissue-specific patterns, provides valuable resources for organism-level investigations. Moreover, the inclusion of individuals from various metal and pCO_2 exposure conditions diversified the transcriptomic landscapes included in the assemblies, broadening their range of relevance. This is particularly critical in the context of global change which requires to consider the influence of combined environmental stressors.

Acknowledgement

We are grateful to the *Institut Français de Bioinformatique* and the IFB-Core cluster platform, financed under the *Programme d'Investissements d'Avenir* funded by the *Agence Nationale de la Recherche* (RENABI-IFB ANR-11-INBS-0013 and MUDIS4LS ANR-21-ESRE-0048), for providing computing resources

Funding

The transciptomic analyses of the rnasep 1 samples were supported by the EC2CO ECODYN funding and by the Fédération de Recherche en Environnement pour le Développement Durable (FREDD) of La Rochelle University. The transcriptomic analyses and the post-doc grant of TSD were supported by the RNA-Sep2 project funded by *Agence National de la Recherche - France 2030* plan (ANR-21-EXES-0010). Biological materials were sampled from the MERCy project funded by *la Fondation pour la Recherche sur la Biodiversité* and the *Ministère de la Transition Ecologique et Solidaire*. The *Région Nouvelle Aquitaine* is acknowledged for its support to the PhD grant to AM through the EXPO project.

Conflict of interest disclosure

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data, scripts, code and supplementary information availability

All the transcriptomic data (raw reads) are available at the European Nucleotide Archive under the accession numbers: PRJEB89715 (rnasep1) and PRJEB88225 (rnasep2), with the two assemblies as "related ENA records": HCHA01230572 (rnasep1) and HCHB01370613 (rnasep2). The two assemblies were also deposited with their related annotation files on Zenodo with DOI: https://doi.org/10.5281/zenodo.17878811. The raw annotation files produced by the annotation process and used as input data in the R pipeline of the project are available on Zenodo with DOI: https://doi.org/10.5281/zenodo.17896603. The Github repository containing the codes used for bioinformatic analyses was released on Zenodo, accessible under the DOI: https://doi.org/10.5281/zenodo.15655834.

References

Ahmed, H. O., Moustafa, A. Y., Abd El-Wakeil, K. F., & Omer, M. Y. (2022). Heavy metals distribution in the body parts of the cephalopods (Sepia officinalis and Octopus vulgaris) collected

from the Mediterranean Sea, Egypt. *Egyptian Journal of Aquatic Biology and Fisheries*, *26*(2), 339. https://doi.org/10.21608/EJABF.2022.229841

Ajala, M., Ameur, W. Ben, & Annabi, A. (2022). First evidence of the utility of cephalopods for biomonitoring program in the field: case of Sepia officinalis south west of Mediterranean Sea (Gulf of Gabes, Tunisia). *Environmental Science and Pollution Research*, *29*(19), 28675–28687. https://doi.org/10.1007/s11356-021-17804-9

Albertin, C. B., & Simakov, O. (2019). Cephalopod Biology: At the Intersection Between Genomic and Organismal Novelties. *Downloaded from Www.AnnualReviews.Org Guest*. https://doi.org/10.1146/annurev-animal-021419-083609

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/.

Arkhipkin, A. I., Hendrickson, L. C., Payá, I., Pierce, G. J., Roa-Ureta, R. H., Robin, J. P., & Winter, A. (2021). Stock assessment and management of cephalopods: advances and challenges for short-lived fishery resources. *ICES Journal of Marine Science*, *78*(2), 714–730. https://doi.org/10.1093/ICESJMS/FSAA038

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., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Baden, T., Briseño, J., Coffing, G., Cohen-Bodenès, S., Courtney, A., Dickerson, D., Dölen, G., Fiorito, G., Gestal, C., Gustafson, T., Heath-Heckman, E., Hua, Q., Imperadore, P., Kimbara, R., Król, M., Lajbner, Z., Lichilín, N., MacChi, F., McCoy, M. J., … Albertin, C. B. (2023). Cephalopod-omics: Emerging Fields and Technologies in Cephalopod Biology. In *Integrative and Comparative Biology* (Vol. 63, Issue 6, pp. 1226–1239). Oxford University Press. https://doi.org/10.1093/icb/icad087

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Bye-A-Jee, H., Cukura, A., Denny, P., Dogan, T., Ebenezer, T. G., Fan, J., Garmiri, P., da Costa Gonzales, L. J., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., … Zhang, J. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/NAR/GKAC1052

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/NBT.3519

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 1–9. https://doi.org/10.1186/1471-2105-10-421

Cantalapiedra, C. P., Herṇandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Chabenat, A., Bidel, F., Knigge, T., & Bellanger, C. (2021). Alteration of predatory behaviour and growth in juvenile cuttlefish by fluoxetine and venlafaxine. *Chemosphere*, *277*, 130169. https://doi.org/10.1016/J.CHEMOSPHERE.2021.130169

Chabenat, A., Knigge, T., & Bellanger, C. (2021). Antidepressants Modify Cryptic Behavior in Juvenile Cuttlefish at Environmentally Realistic Concentrations. *Environmental Toxicology and Chemistry*, *40*(9), 2571–2577. https://doi.org/10.1002/ETC.5149,

Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *IMeta*, *2*(2), e107. https://doi.org/10.1002/IMT2.107

Destanović, D., Schultz, D. T., Styfhals, R., Cruz, F., Gómez-Garrido, J., Gut, M., Gut, I., Fiorito, G., Simakov, O., Alioto, T. S., Ponte, G., & Seuntjens, E. (2023). A chromosome-level reference genome for the common octopus, Octopus vulgaris (Cuvier, 1797). *G3 Genes|Genomes|Genetics*, *13*(12). https://doi.org/10.1093/G3JOURNAL/JKAD220

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Computational Biology*, *7*(10). https://doi.org/10.1371/journal.pcbi.1002195

Elliott, D. J. (2014). Illuminating the transcriptome through the genome. In *Genes* (Vol. 5, Issue 1, pp. 235–253). MDPI AG. https://doi.org/10.3390/genes5010235

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/BIOINFORMATICS/BTW354

Fischer, E. K., Hauber, M. E., & Bell, A. M. (2021). Back to the basics? Transcriptomics offers integrative insights into the role of space, time and the environment for gene expression and behaviour. In *Biology Letters* (Vol. 17, Issue 9). Royal Society Publishing. https://doi.org/10.1098/rsbl.2021.0293

García-Flores, M., De Silva-Dávila, R., Rodríguez-Figueroa, G. M., Moncayo-Estrada, R., Hernández-Guerrero, C. J., & Arellano-Martínez, M. (2025). Ecotoxicology of cephalopod early life phases: review and perspectives. *Environmental Science and Pollution Research 2024 32:4*, *32*(4), 1693–1709. https://doi.org/10.1007/S11356-024-35829-8

Gonçalves, C., Moutinho Cabral, I., Alves de Matos, A. P., Grosso, A. R., & Costa, P. M. (2024). Transcriptome profiling of the posterior salivary glands of the cuttlefish Sepia officinalis from the Portuguese West coast. *Frontiers in Marine Science*, *11*, 1362824. https://doi.org/10.3389/FMARS.2024.1362824

Gouveneaux, A., Minet, A., Jozet-Alves, C., Knigge, T., Bustamante, P., Lacoue-Labarthe, T., & Bellanger, C. (2023). Cuttlefish color change as an emerging proxy for ecotoxicology. *Frontiers in Physiology*, *14*, 1162709. https://doi.org/10.3389/FPHYS.2023.1162709

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. https://doi.org/10.1038/nbt.1883

Greene, D., Richardson, S., & Turro, E. (2017). OntologyX: A suite of R packages for working with ontological data. *Bioinformatics*, *33*(7), 1104–1106. https://doi.org/10.1093/bioinformatics/btw763

Haas, B. (n.d.). *TransDecoder v_5.7.0*. Https://Transdecoder.Github.Io/.

Ihya, A., Aicha, A. A., Sara, H., Mustapha, A., & Abdellatif, M. (2021). Assessment of Trace Metal Concentrations in Three Seafood Deep-Sea Fished Caught in the Moroccan Atlantic. *Journal of Ecological Engineering*, *22*(9), 87–98. https://doi.org/10.12911/22998993/141480

Irvine, S. Q. (2020). Embryonic canalization and its limits—A view from temperature. In *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* (Vol. 334, Issue 2, pp. 128–144). John Wiley and Sons Inc. https://doi.org/10.1002/jez.b.22930

Jiang, D., Liu, Q., Sun, J., Liu, S., Fan, G., Wang, L., Zhang, Y., Seim, I., An, S., Liu, X., Li, Q., & Zheng, X. (2022). The gold-ringed octopus (Amphioctopus fangsiao) genome and cerebral single-nucleus transcriptomes provide insights into the evolution of karyotype and neural novelties. *BMC Biology*, *20*(1), 1–15. https://doi.org/10.1186/S12915-022-01500-2

Juorio, A. V. (1971). Catecholamines and 5-Hydroxytryptamine in Nervous Tissue of Cephalopods. *The Journal of Physiology*, *216*, 213–226. https://doi.org/10.1113/jphysiol.1971.sp009519

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. https://doi.org/10.1093/NAR/GKW1092

Kreps, E., Smirnov, A., Chirkovskaya, E., Patrikeeva, M., & Krasilnikova, V. (1968). Phospholipids in the nervous sytem of molluscs. *Journal of Neurochemistry*, *15*(4), 285–291. https://doi.org/10.1111/J.1471-4159.1968.TB11612.X

Lacoue-Labarthe, T., Le Pabic, C., & Bustamante, P. (2016). Ecotoxicology of early-life stages in the common cuttlefish Sepia officinalis: review and perspectives. *Vie et Milieu - Life & Environment*, *66*(1), 65–79. https://hal.science/hal-01330645

Landau, W. M. (2021). The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. https://doi.org/10.21105/joss.02959

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/BIOINFORMATICS/BTL158

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/MOLBEV/MSAB199

Martin, J. A., & Wang, Z. (2011). Next-generation transcriptome assembly. In *Nature Reviews Genetics* (Vol. 12, Issue 10, pp. 671–682). https://doi.org/10.1038/nrg3068

Minet, A. (2022). *Bioaccumulation and effects of mercury in cuttlefish in the context of ocean acidification.* Université de La Rochelle.

Minet, A., Manceau, A., Valada-Mennuni, A., Brault-Favrou, M., Churlaud, C., Fort, J., Nguyen, T., Spitz, J., Bustamante, P., & Lacoue-Labarthe, T. (2021). Mercury in the tissues of five cephalopods species: First data on the nervous system. *Science of The Total Environment*, *759*, 143907. https://doi.org/10.1016/J.SCITOTENV.2020.143907

Minet, A., Melvin, S., Metian, M., Taylor, A., Oberhänsli, F., Lefrançois, C., Swarzenski, P., Bustamante, P., & Lacoue-Labarthe, T. (2025). Metabolomic and phenotypic effects of ocean acidification on cuttlefish differ across early life stages. *Marine Environmental Research*, *205*. https://doi.org/10.1016/j.marenvres.2025.107013

Miramand, P., Bustamante, P., Bentley, D., & Kouéta, N. (2006). Variation of heavy metal concentrations (Ag, Cd, Co, Cu, Fe, Pb, V, and Zn) during the life cycle of the common cuttlefish Sepia officinalis. *Science of The Total Environment*, *361*(1–3), 132–143. https://doi.org/10.1016/J.SCITOTENV.2005.10.018

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1), D412–D419. https://doi.org/10.1093/NAR/GKAA913

Navarro, J. C., Villanueva, R., & Spain´b, S. (2000). Lipid and fatty acid composition of early stages of cephalopods: an approach to their lipid requirements. *Aquaculture*, *183*, 161–177. https://doi.org/10.1016/S0044-8486(99)00290-2

Nixon, M., & Mangold, K. (1998). The early life of Sepia officinalis , and the contrast with that of Octopus vulgaris (Cephalopoda) . *Journal of Zoology*, *245*(4), 407–421. https://doi.org/10.1111/j.1469-7998.1998.tb00115.x

Packard, A. (1972). Cephalopods and Fish : The Limits of Convergence. *Biological Reviews*, *47*(2), 241–307. https://doi.org/10.1111/J.1469-185X.1972.TB00975.X

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/NMETH.4197

Penicaud, V., Lacoue-Labarthe, T., & Bustamante, P. (2017). Metal bioaccumulation and detoxification processes in cephalopods: A review. *Environmental Research*, *155*, 123–133. https://doi.org/10.1016/J.ENVRES.2017.02.003

Prado-Álvarez, M., Dios, S., García-Fernández, P., Tur, R., Hachero-Cruzado, I., Domingues, P., Almansa, E., Varó, I., & Gestal, C. (2022). De novo transcriptome reconstruction in aquacultured early life stages of the cephalopod Octopus vulgaris. *Scientific Data 2022 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41597-022-01735-2

Rencken, S., Tushev, G., Hain, D., Ciirdaeva, E., Simakov, O., & Laurent, G. (2025). Chromosome-scale genome assembly of the European common cuttlefish Sepia officinalis. *BioRxiv*, 2025.04.22.649952. https://doi.org/10.1101/2025.04.22.649952

Rosenthal, J. J. C., & Eisenberg, E. (2025). Extensive Recoding of the Neural Proteome in Cephalopods by RNA Editing. *Annual Review of Animal Biosciences*, *41*, 25. https://doi.org/10.1146/annurev-animal-060322

Shigeno, S., Parnaik, R., Albertin, C. B., & Ragsdale, C. W. (2015). Evidence for a cordal, not ganglionic, pattern of cephalopod brain neurogenesis. *Zoological Letters*, *1*(1). https://doi.org/10.1186/s40851-015-0026-z

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M., & Kelly, S. (2016). TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research*, *26*(8), 1134–1144. https://doi.org/10.1101/GR.196469.115

Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, *28*(1), 33. https://doi.org/10.1093/NAR/28.1.33

Thomas, J. T., Huerlimann, R., Schunter, C., Watson, S. A., Munday, P. L., & Ravasi, T. (2024). Transcriptomic responses in the nervous system and correlated behavioural changes of a cephalopod exposed to ocean acidification. *BMC Genomics*, *25*(1). https://doi.org/10.1186/s12864-024-10542-5

Ushey, K., & Wickham, H. (2025). *Renv: Project environments*. 2025. https://rstudio.github.io/renv/

Voss, G., & Rosenthal, J. J. C. (2023). High-level RNA editing diversifies the coleoid cephalopod brain proteome. In *Briefings in Functional Genomics* (Vol. 22, Issue 6, pp. 525–532). Oxford University Press. https://doi.org/10.1093/bfgp/elad034

Wu, T. D., & Watanabe, C. K. (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, *21*(9), 1859–1875. https://doi.org/10.1093/bioinformatics/bti310