



OPEN ACCESS

EDITED BY

Simon Jungblut,
University of Bremen, Germany

REVIEWED BY

Alexander Fedosov,
Severtsov Institute of Ecology and Evolution
(RAS), Russia
Rossanna Rodriguez-Canul,
Center for Research and Advanced Studies -
Mérida Unit, Mexico

*CORRESPONDENCE

Cátia Gonçalves

✉ cv.goncalves@campus.fct.unl.pt

Pedro M. Costa

✉ pmcosta@fct.unl.pt

RECEIVED 29 December 2023

ACCEPTED 20 May 2024

PUBLISHED 04 June 2024

CITATION

Gonçalves C, Moutinho Cabral I, Alves de
Matos AP, Grosso AR and Costa PM (2024)
Transcriptome profiling of the posterior
salivary glands of the cuttlefish *Sepia
officinalis* from the Portuguese West coast.
Front. Mar. Sci. 11:1362824.
doi: 10.3389/fmars.2024.1362824

COPYRIGHT

© 2024 Gonçalves, Moutinho Cabral, Alves de
Matos, Grosso and Costa. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Transcriptome profiling of the posterior salivary glands of the cuttlefish *Sepia officinalis* from the Portuguese West coast

Cátia Gonçalves^{1,2*}, Inês Moutinho Cabral^{1,2},
António P. Alves de Matos³, Ana R. Grosso^{1,2}
and Pedro M. Costa^{1,2*}

¹Associate Laboratory i4HB – Institute for Health and Bioeconomy, NOVA School of Science and Technology, NOVA University of Lisbon, Caparica, Portugal, ²UCIBIO – Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, NOVA University of Lisbon, Caparica, Portugal, ³Egas Moniz Centre for Interdisciplinary Research (CIEM); Egas Moniz School of Health & Science, Caparica, Portugal

Cephalopods like octopuses and cuttlefishes are known to secrete a 'toxic saliva' to inject into their prey, especially crustaceans since the XIX century. However, only in the mid-XX century were the first coleoid-specific toxins successfully isolated. Motivated by the growing interest on the global ocean as an almost inexhaustible source of novel bioactive compounds, we used RNA-Seq – based transcriptomics and *de novo* assembly of transcriptomes to screen the posterior salivary glands of *Sepia officinalis* (the common cuttlefish) from the Portuguese West coast for toxins and other bioactive proteins and peptides. Supported by microanatomical analyses, the posterior salivary glands constitute indeed the 'venom gland' whereas the more elusive anterior salivary glands (embedded in the buccal mass) are responsible for the production of mucin-rich saliva that is effectively the vehicle that transports the toxins as the venom is injected into the prey. Indeed, the transcriptomic profiling suggests that the cuttlefish venom is complex mixture of bioactive proteins, among which neurotoxins are major players, together with enzymes whose function is to digest the extracellular matrix to facilitate diffusion of the toxins. Nonetheless, by comparing with previous RNA-Seq data obtained from *S. officinalis* collected from other biogeographical areas, it may be suggested that significant inter-population variation in venom composition can occur, which may potentially increase the span of bioactives secreted by these animals. We isolated and validated the full coding sequences for three important toxins, a cysteine-rich venom protein (CRVP), a venom insulin (VIns) and a cephalotoxin (CTX). The toxins seem to be relatively conserved among coleoids but diverging from other venomous mollusks such as cone snails. Their properties as potent modulators of glucose (in the case of VIns) and as potential neurotoxins (like CRVP and CTX) can render them primer targets for drug development.

KEYWORDS

toxins, bioactives, cephalotoxins, cephalopoda, coleoidea, marine biotechnology, omics, histology

1 Introduction

The constant demand for novel effective bioproducts has been attracting growing enthusiasm on oceans as they harbor an immense biodiversity which is linked with a unique chemical diversity. Marine organisms that bear a chemical warfare are particularly interesting for ocean's bioprospection, since they secrete complex cocktails, as a strategy for predation or defense. These bioactive substances, especially toxins, are noteworthy for biotechnological applications since they interfere with specific molecular targets, such as ion channels, receptors and enzymes, and may consequently impair particular physiological pathways (Calvete et al., 2009). Altogether, the use of these natural bioactives can be a safer, efficient and sustainable therapeutic alternative to the risky and costly process of designing synthetic molecules (see for instance Fusetani, 2009; Rodrigo and Costa, 2019).

The reduced number of substances based on marine compounds translated into the market can be particularly explained by the tremendous marine biodiversity and the downsides associated with it, such as the scarcity of genomic resources. Here, not only can omics shed some light on these less-studied marine species, but they can also be a helpful tool for bioprospecting for novel bioactive compounds with biotechnological applications. Untargeted approaches, such as RNA-Seq, allow to identify multiple molecules without having a reference genome and transcriptome, which is the case of non-conventional model marine species (Martins et al., 2019; Rodrigo and Costa, 2019). Apart from unravelling and characterizing novel proteinaceous bioactives, transcriptomics also facilitates the prediction of their mode-of-action and possible targets. Peptides and proteins, especially toxins, are particular noteworthy as candidates for novel drugs as they can interact with specific molecular targets and can potentially be synthesized by heterologous expression.

Studies on marine invertebrates are showing their promising potential for producing molecules, including proteinaceous toxins, with putative biotechnological applications (Leal et al., 2012). The bioactive substances of these marine organisms are presenting interesting biomedical properties (e.g., painkiller, antitumoral, anticoagulant, anti-inflammatory, antimicrobial and antiviral) as well as cosmetical and nutritional applications or even as pesticides (Molinski et al., 2009; Martins et al., 2014; Rodrigo and Costa, 2019, and Cappello and Nieri, 2021). A group of organisms that emerged as a source of novel marine compounds for drug development are the Mollusca. Within this Phylum, cone snails are highlighted as pioneer organisms for marine biotechnologists due to the pharmaceutical value of peptidic bioactives such as conopeptides (Shen et al., 2000; Turner et al., 2020). Indeed, ziconotide (commercialized as prialt), a recombinant peptide designed from an ω -conotoxin secreted by the cone-snail *Conus magus*, became one of the first formally approved marine-inspired pharmaceuticals in Europe and the United States (since the early 2000s) for the management of severe chronic pain in patients insensitive or unable to take other treatments, being delivered intrathecally (Olivera et al., 1987; Bowersox et al., 1996; Williams et al., 2008). Adding to these gastropods, coleoid cephalopods are mollusks that have also been

suggested to secrete molecules with promising biotechnological proprieties. Indeed, their saliva can be venomous and contains multiple interesting bioactives, from toxins to permeabilizing agents, such as neurotoxins termed cephalotoxins (for a review, refer to Gonçalves and Costa, 2021). The secretion of these substances by the salivary glands can be fundamental for cephalopods to capture their preys (e.g., crabs) as they seemingly have paralyzing action. Cephalotoxins are secreted by the posterior salivary glands (PSG), as previously reported by Ghiretti (1959) and Ueda et al. (2008).

As cephalopods are becoming an auspicious group of marine invertebrates for biotechnologists, behavioral biologists and even neurobiologists, as examples, researchers are turning to state-of-the-art omics approaches for bioprospecting these animals for novel bioactives due to the ability to screen for multiple untargeted molecules in single runs (see Albertin and Simakov, 2020, as well as Baden et al., 2023). Despite the relatively scant number of omics studies, comparatively to more conventional model organisms, in marine invertebrates in general, including cephalopods, altogether leading to reduced molecular resources (such as reliable genomic annotation) compared to more conventional model organisms, there are already available literature suggesting molecular approaches as roadmaps towards bioprospecting cephalopods for proteinaceous bioproducts (Ruder et al., 2013; Caruana et al., 2016; Whitelaw et al., 2016). These works are accompanied by studies using proteomics and transcriptomics to describe the venom systems of marine invertebrates, from cone snails (Fassio et al., 2019; Fedosov et al., 2021) to annelids (von Reumont et al., 2014; Rodrigo et al., 2021; Moutinho Cabral et al., 2022) and even the first known poisonous crustacean, the remipede *Xibalbanus tulumensis* (von Reumont et al., 2017). These works show the potential of these methodologies and novel pipelines of research, from evolution to bioprospecting. In addition, previous analyses of the transcriptome and proteome of the PSG of the cuttlefish *Sepia officinalis* (Linnaeus, 1758) from the coast of France associated with bioassays suggested that these glands could have two functions linked with predation and immune defense as their substances and extracts have paralyzing and antimicrobial activities (Cornet et al., 2014).

The present work aims at exploring the PSG of the common cuttlefish, *Sepia officinalis* (Linnaeus, 1758) from the Portuguese West coast, at several levels: (i) comparatively characterize the microanatomy of the PSG and the anterior salivary glands (ASG) and associate their ultrastructure with specific functions, such as the secretion of toxins; (ii) identify and annotate transcripts putatively coding for toxins produced by this organ, in the case using a RNA-Seq – based transcriptomic approach, and (iii) isolate and analyze the full coding sequence of relevant toxins.

2 Materials and methods

2.1 Animal collection

Following the implementation of The Three Rs principle (first described by Russel and Burch and presently inserted in the Directive 2010/63/EU on the protection of animals used for

scientific purposes) the animals used in this study were provided by local fishermen. Animals were euthanized by sectioning the central brain and the tissues of interest were harvested immediately. Five *Sepia officinalis* (mantle length 14.5 ± 2.0 cm; weight 227.2 ± 68.9 g) were collected from two beaches of the Portuguese West coast: Praia do Norte ($38^{\circ}39'0.153''$ N, $9^{\circ}14'43.876''$ W) and Praia do Outão ($38^{\circ}29'19.422''$ N, $8^{\circ}56'0.455''$ W). The PSG were extracted and divided for both histological analyses and RNA extraction. The ASG were also excised and prepared for histology to add a structural comparison between the two salivary glands.

2.2 Sample preparation for microscopy analyses

Samples were immediately fixed after dissection. For optical microscopy, samples were fixed using Davidson's (9–10% v/v formalin, 10% v/v glacial acetic acid and 30% ethanol) and Zenker's (2.5% m/v potassium dichromate, 3% m/v mercury chloride, 1% sodium sulphate and 5% v/v glacial acetic acid) fixatives, both prepared in MilliQ-grade water (> 16 M Ω .cm). Fixation was done at room temperature for 24 h in case of Davidson's and Zenker's fixatives, followed by washing in MilliQ-grade water (4×15 min), dehydrated through a progressive series of (aqueous) ethanol: 70% (1×30 min), 95% (2×15 min) and 100% (3×30 min), following by intermediate infiltration and embedding with xylenes (3×15 min) and molten paraffin (overnight), respectively. Samples were then sectioned (5 μ m thickness) using a Leica RM2245 semi-automated rotary microtome. Histological sections were stained through a tetrachrome staining procedure that combines Alcian Blue for acid sugars; Periodic Acid–Schiff's for neutral polysaccharides; Weigert's Iron Hematoxylin for chromatin, and Picric Acid as counterstain for muscle fibers and cytoplasm, following Costa and Costa (2012). Previous to staining, slides were deparaffinated with xylenes (1 min) and rehydrated through a regressive ethanol series: 100% (1 min), 95% (1 min), 70% (30 s), finalized with MilliQ-grade water (6 min). Optical microscopy analyses were done with a DM 2500 LED model microscope with a MC 190 HD camera (all from Leica Microsystems). Image acquisition was made using the software Leica Application Suite Version 4.13.0 (Leica Microsystems). Image processing was done with ImageJ (Schneider et al., 2012) and GIMP (Montesanto, 2015).

Glutaraldehyde 2.5% v/v (in 0.1 M sodium cacodylate buffer, pH 7.4) or paraformaldehyde 4% m/v in phosphate-buffered saline (PBS), were used to fix samples for transmission electron microscopy (TEM). The fixation was performed over 2 h at room temperature, followed by washing in cacodylate buffer or PBS (3×15 min), post-fixation in 1% m/v osmium tetroxide (OsO₄) in 0.1 M sodium cacodylate buffer of PBS and washing in MilliQ-grade water (3×15 min). These samples were then embedded in Epoxy resin (Sigma-Aldrich, St. Louis, MO, USA), following Luft's (1961) methodology. In brief, samples were dehydrated through a progressive acetone series: 30%, 60%, 70%, 90% and 100% (15 min each), followed by intermediate infiltration with polypropylene oxide:Epoxy of 2:1, 1:1 and 1:2 (30 min each). A final infiltration was performed with Epoxy resin, in vacuum for 30 min. Resin

blocks were left to polymerize overnight at 60°C. Thin sections (50–60 nm) were obtained using a Leica Reichert-Jung Ultramicrotome Ultracut E. Sections were collected onto copper mesh grids and contrasted with 2% (w/v) aqueous Uranyl Acetate for 1 h 20 min and Reynold's lead citrate for 8 min (Venable and Coggeshall, 1965). Grids were analyzed using a 100-SX model TEM (JEOL, Tokyo, Japan) operated at 80 keV.

2.3 Total RNA extraction and RNA-seq

After excision, samples were immediately stabilized in RNAlater reagent (Sigma-Aldrich, St. Louis, MO, USA), where remained during 10 days at 4°C. An exchange to fresh RNAlater was made during that period. Afterwards, samples were archived at -80°C, until total RNA extraction. Total RNA was extracted from ≈ 20 mg of tissue per sample ($n = 3$ PSG from three different individuals) using the RNeasy Protect Mini Kit combined with the RNase-Free DNase Set (all from Qiagen, Hilden, Germany) for an efficient on-column digestion of DNA, following manufacturer instructions. Preliminary quantification of total RNA and initial quality assessment was performed using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The RNA integrity number (RIN) and quantification of total RNA were obtained using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and were found to range between 4.8. and 7.2, which is deemed adequate for marine invertebrates, with an input above 1 μ g uncontaminated RNA (refer to the guidelines suggested by Gayral et al., 2011; Gallego Romero et al., 2014; Puchta et al., 2020). The library construction of cDNA was carried out using a Stranded mRNA Library Preparation Kit. The generated DNA fragments were then sequenced on the Illumina NovaSeq platform (Illumina, San Diego, CA, USA), with a mean length of 150 bp paired-end reads within a coverage of 40 M reads.

2.4 Transcriptome *de novo* assembly and annotation

The quality control check of raw data obtained from RNA-Seq was pre-evaluated with FastQC v0.11.9 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The TrimGalore v0.6.6 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to discard low quality reads, i.e., reads under 20 bp of length. After applying the default parameters, less than 0.9% of reads were removed per sample. Filtered data was *de novo* assembled using Trinity v2.6.6 with default parameters (Grabherr et al., 2011; Haas et al., 2013). The quality of transcriptome assembly was evaluated through contigs N50, Ex90N50 and read content statistics, using TrinityStats, Bowtie2 v2.4.1 and Samtools v1.11 (Langmead and Salzberg, 2012; Haas et al., 2013; Langmead et al., 2019; Danecek et al., 2021). Transcript abundance was estimated for the novel assembled transcripts using Kallisto v0.43.0 (Bray et al., 2016). Subsequent analyses were performed using R 4.0.3 (Ihaka and Gentleman, 1996), where the

packages Timport v1.28.0 and SeqinR v4.2–30 were used to import data (Charif and Lobry, 2007; Sonesson et al., 2015). Normalized gene expression values were assessed through transcript per millions (TPMs). Coding regions were then predicted in the assembled transcriptome through TransDecoder v5.5.0, using the default parameters (Haas et al., 2013). The predicted open reading frames (ORFs) were functionally annotated by scanning for homology against Swiss-Prot (version 18–11-2020 13:41, downloaded from <https://ftp.ncbi.nlm.nih.gov/blast/db/swissprot.tar.gz>) and five customized UniProtKB databases (UniProt release 2023_03) using BLASTP from NCBI blast+ v2.13.0, with a cut off *e*-value of 1×10^{-5} (Altschul et al., 1990; Camacho et al., 2008; UniProt Consortium, 2021). The five customized databases were built from a subset of either UniProtKB or Swiss-Prot. The customized databases consisted of: (1) The ‘SeaTox’ database is mostly a customized subset of UniProtKB containing manually annotated toxin- or venom-related proteins from Eukaryota (excluding human), plus unreviewed proteins from Annelida, Cnidaria and Cephalopoda (this database can be freely accessed at <https://dx.doi.org/10.5281/zenodo.10418296>); (2) The ‘Toxin’ database is a subset of Swiss-Prot including toxin-related proteins and cephalotoxins; (3) ‘Enzyme’ database is a subset of Swiss-Prot including enzyme-related proteins; (4) ‘PTM’ database is a subset of Swiss-Prot including proteins associated with post-translational modifications (PTMs), according to the Gene Ontology annotation; (5) ‘Secreted’ database is a subset of Swiss-Prot including secreted proteins (UniProtKB, location_sl_0243). In

addition, ORFs were also functionally annotated by searching for reference protein domains (*e*-value < 0.05) against the Pfam database (Pfam-A.hmm version 35.0, Nov 2021) using HMMER v3.3.1 (Eddy, 2009; Mistry et al., 2021).

2.5 Validation by RT-qPCR

The RNA-Seq results were validated by reverse transcription quantitative polymerase chain reaction (RT-qPCR) of selected representative transcripts that encode for proteins with toxin or hormone activities or interfere with ion channels. The selected transcripts from RNA-Seq data encompassed a cysteine-rich venom protein (CRVP) as cysteine-rich secretory proteins (CRISPs) are often referred to as means to distinguish from their non-toxin counterparts, a cephalotoxin (CTX) and a venom insulin (VIns). The full coding sequence of these transcripts were firstly isolated and amplified by PCR. Primers (Table 1) were evaluated with *in silico* using PCR Primer Stats (Stothard, 2000) and OligoAnalyzer (<https://eu.idtdna.com/calc/analyzer>). Specific primers were designed to quantify expressed sequence tags (ESTs) in the coding region by RT-qPCR, avoiding conserved domains. The 18S gene was chosen as housekeeping. The cDNA was synthesized from total RNA samples using the NZY First-Strand cDNA Synthesis Kit (NZYTech, Lisbon, Portugal), following manufacturer instructions. Regions of interest were isolated by PCR in a Biometra TOne 96 gradient thermocycler (Analytik Jena, Jena, Germany) using a PCR kit from Invitrogen (Thermo

TABLE 1 Primer sequences used for sequence isolation by PCR and expression analysis by RT-qPCR in the posterior salivary glands of *Sepia officinalis*.

Gene	Primer	PCR/RT-qPCR	Primer sequence (5´-3´)	Amplicon size
18S	Forward	PCR/RT-qPCR	GTTCCGACCGTAAACG	140 bp
	Reverse	PCR/RT-qPCR	CCTCCGTCAATTCCT	
CRVP	Forward	PCR	AAGATGAAAGCTGTAGTATC	984 bp
	Reverse	PCR/RT-qPCR	AGGACACTTTTCTTTGC	
CTX	Forward_1	PCR	ATGGTTCTGTGGCAGAG	809 bp
	Reverse_1	PCR	GCCTTCGGTATCTGCAT	
	Forward_2	PCR	CGTCACTCCGGTCTT	815 bp
	Reverse_2	PCR	CGCTTGAGATCGAAGGA	
	Forward_3	PCR	GCTGTAGACACAAACATTGA	794 bp
	Reverse_3	PCR/RT-qPCR	GCAGTAGCATTCCATTCAA	
	Forward_4	PCR	TGAGTGCAGGTTATCGG	947 bp
	Reverse_4	PCR	AAGCAATATGAGGTCGGTA	
VIns	Forward	PCR/RT-qPCR	GGGACTATGGTCTTAAGTTTC	518 bp
	Reverse	PCR	GTATGCTTATGATGACGAGTC	
CRVP	Forward	RT-qPCR	GAAATCAGGTAAGTCTG	151 bp
CTX	Forward	RT-qPCR	CCGCCCTTGAATATCAC	161 bp
VIns	Reverse	RT-qPCR	CTCATTGCAGGTGTGTT	156 bp

The housekeeping gene encoding for the 18S ribosomal subunit was used for normalization (18S), whereas three full coding sequences of venom-related toxins for selected for transcriptome validation: Cysteine-rich venom protein (CRVP), SE-cephalotoxin (CTX) and Venom Insulin (VIns).

Fisher Scientific, Waltham, MA, USA). After resolving in a 1.2% agarose gel, PCR products were Sanger-sequenced for sequence confirmation. Gene expression analysis was performed in a Rotor-Gene Q (Qiagen, Hilden, Germany) using the NZY qPCR Green Master Mix (NZYTech, Lisbon, Portugal). The program comprised an initial denaturation procedure (95 °C, 10 min), followed by 40 cycles of denaturation (94 °C, 45 s), annealing (52 °C, 25 s) and extension (72 °C, 30 s). Expression was determined based on a simple ΔCt approach (Silver et al., 2006). Homoscedasticity and normality of data were assessed through Levene's and Shapiro-Wilk's tests, respectively. Differences in gene expression between the three genes were evaluated through the parametric Tukey's HSD test. A significance level $\alpha = 0.05$ was set for all analyses. Statistics were computed using R.

2.6 Phylogenetics of representative toxin transcripts

Three full coding sequences of the shortlisted transcripts, i.e., cysteine-rich venom protein latisemin (CRVP), a cephalotoxin (CTX) and a venom insulin (VIns) were analyzed. Each sequence was scanned for homology against NCBI RefSeq_protein database using BLASTP (Altschul et al., 1990; Camacho et al., 2008). The search was restricted to related taxonomic groups as 'Lophotrochozoa', phyla 'Arthropoda', 'Mollusca' and 'Cnidaria' as well as the specific classes: 'Cephalopoda', 'Arachnida', 'Gastropoda'. The most significant hits were chosen based on the best matches against venomous or toxin-bearing animals, e-value and percentage of identity retrieved from running BLASTP. Sequence alignment and dendrograms were made with MEGA X (Kumar et al., 2018). Phylogenetic consensus trees were produced for each selected sequence using the Maximum Likelihood method and Jones-Taylor-Thornton model, with 800 bootstrap replications (Jones et al., 1992). Tree branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (800 replicates) are shown next to the branches (Felsenstein, 1985). The trees were rooted on related sequences from Arthropoda. Sequence homology-based models of protein structure for each sequence of interest were constructed using Swiss-Model (Waterhouse et al., 2018), which takes AlphaFold Protein Structure Database (<https://alphafold.ebi.ac.uk/>) as reference, together with data from NMR, Cryo-EM, crystallography and other methods.

3 Results

3.1 The diverticular nature of the posterior salivary glands

The PSG or venom glands of *Sepia officinalis* are an exocrine paired organ located adjacently to the visceral mass, close to the digestive gland. This organ is connected to the buccal cavity by an excretory duct that runs through the salivary papilla.

Macroscopically, the PSG is homogeneous whitish in appearance, but its internal microanatomy reveals that it is formed by branched glandular tubules, seemingly blind-end diverticula (Figure 1A). The tubules are composed by columnar glandular epithelium whose cells are basally basophilic, easily recognized by its bluish or brown-black colors after staining with Harris's hematoxylin or Weigert's iron hematoxylin, respectively (Figures 1A, B). The organ also bears blood vessels and nervous tissue in the intertubular space, the latter of which is invariably formed chiefly by fibrocytes (see Figure 1B insets for examples). The basophilic nature of the basal cytoplasm of cells is mainly due to dense rough endoplasmic reticulum contiguous to nuclei (Figures 1C, D), which is conspicuous under TEM (Figure 2A). Apically, the columnar cells hold picric acid-positive secretory granules, which is indicative of proteinaceous content. Depending on their maturation stages, these dense protein-containing vesicles (secretory granules) can differ by their hue, which can range from yellowish to orange (visible in Figure 1C). This may be related to the cascade of events involving protein folding and other *post*-translational modifications as vesicles mature *a priori* to secretion. Protein biosynthesis and vesicle maturation is visible under TEM as well, resulting in abundant and often densely packed secretory granules (Figures 2B, C). No evidence was found for goblet (mucus-secreting) cells structurally as well as no evidence for the common histochemical signature of acidic or neutral mucins through Alcian Blue and PAS staining, respectively.

In comparison, the ASG of *Sepia officinalis* are also whitish structures formed by a pair of lobes like the PSG. However, ASG are embedded in the buccal mass, which is mostly comprised of the beak musculature, therefore being harder to localize and isolate. These glands are also microstructurally constant being formed chiefly by a network of branching glandular diverticula (blind-end 'tubules') interlinked by fibrous intertubular tissue and in this case, surrounded by muscular fibers (Figures 3A, B). The diverticula are also comprised of simple columnar secretory epithelium but histochemistry revealed a very distinct nature com PSG cells. The tubules of the ASG, which is often termed 'mucous gland' in coleoids, are mostly composed by serous and goblet (mucocytes) cells altogether expectedly responsible for the secretion of permeabilizing enzymes and copious amounts of mucins, respectively (Figures 3C, D). Serous cells bare also an abundant endoplasmic reticulum and should be responsible for the biogenesis of secretory granules, now dyed blueish-pink under Alcian Blue and PAS histochemical staining, respectively, revealing acidic and neutral sugars, respectively, presumably in glycoproteins like mucins (Figure 3C). In turn, goblet presented less abundance of endoplasmic reticulum and yielded a blue cytoplasm, following staining of mucus sacculi by Alcian Blue for acidic mucins (Figures 3D, E).

3.2 Transcriptome profiles of posterior salivary glands unveils several toxin-encoding transcripts

The assembled transcriptome (Figure 4) of the PSG of *Sepia officinalis* yielded a total of 171 318 transcripts, corresponding to predicted 124 209 genes. Of these, 27 753 resulted in potential open

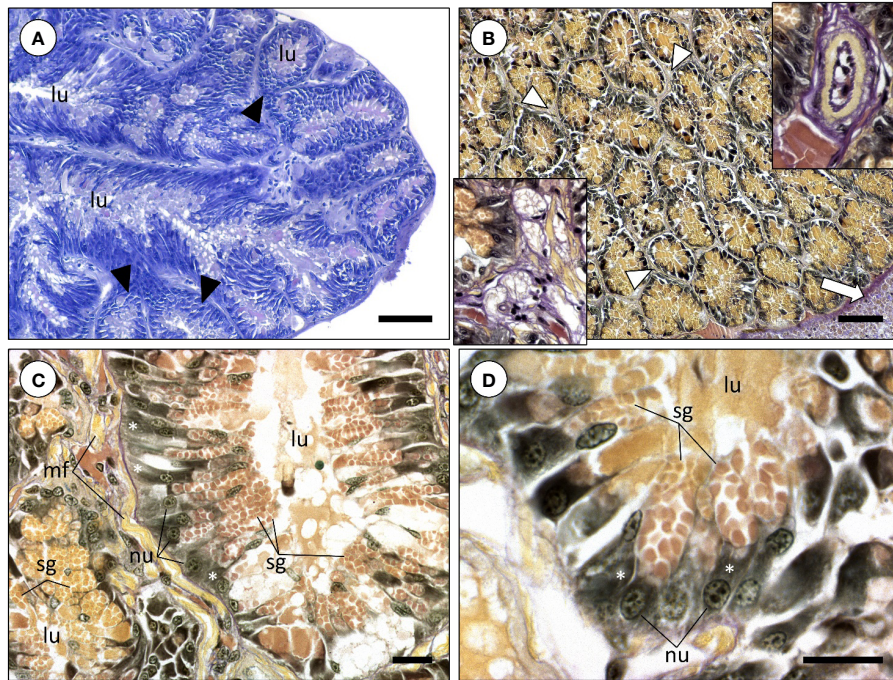


FIGURE 1

Histological photomicrographs of the posterior salivary glands of *Sepia officinalis*. **(A)** Overview of the gland structure stained with Hematoxylin & Eosin. Note the presence of the several different-shaped and sized blind-end diverticula, or 'tubules' (arrowheads), that are strongly basophilic basally, indicating dense rough endoplasmic reticulum. **(B)** General appearance of the posterior salivary glands revealed using a histochemical tetrachrome stain. The white arrow denotes the edge of the digestive gland. The secretory granules within glandular cells are stained by picric acid (yellowish), revealing the presence of proteinaceous material. Bottom inset: Detail of nervous tissue inserted within the glandular tubules. Upper inset: Detail of a blood vessel (arteriole-like) within the network of diverticula. **(C)** Diverticula formed by columnar glandular cells. The staining gradient among secretory granules most likely corresponds to different maturation stages. **(D)** Detail of glandular cells with strong basophilic cytoplasm (basal portion of cells), which indicates abundant rough endoplasmic reticulum (white asterisks). The apical portion of these cells are filled with secretory granules. Abbreviations: lu, lumen; mf, muscular fibers; nu, nucleus. Scale bars: **(A, B)** 100 μm ; **(C)** 25 μm ; **(D)** 20 μm .

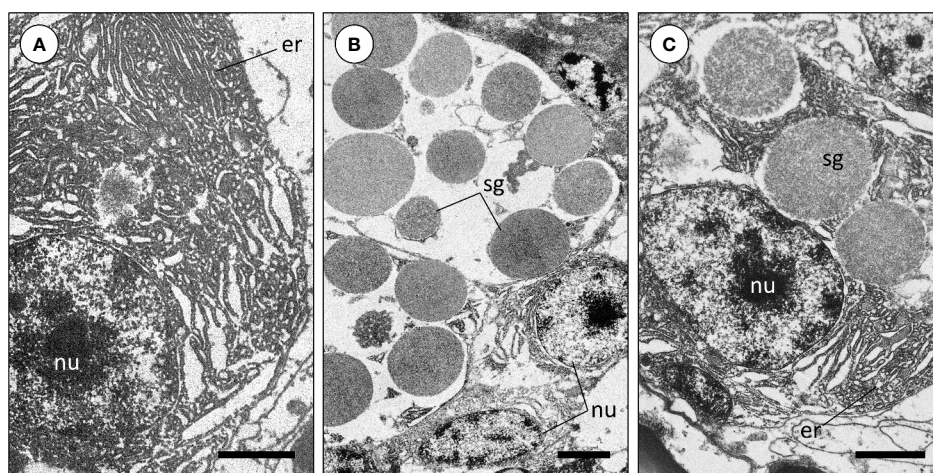


FIGURE 2

Transmission electron microscopy images of the posterior salivary gland cells of *Sepia officinalis*. **(A)** Abundant rough endoplasmic reticulum of a glandular cell. **(B)** Secretory granules (protein-dense secretory vesicles) of a glandular cell in different maturation stages. **(C)** Secretory granules located near the endoplasmic reticulum. Abbreviations: er, endoplasmic reticulum; nu, nucleus; sg, secretory granules. Scale bars: **(A, C)** 2 μm ; **(B)** 3 μm .

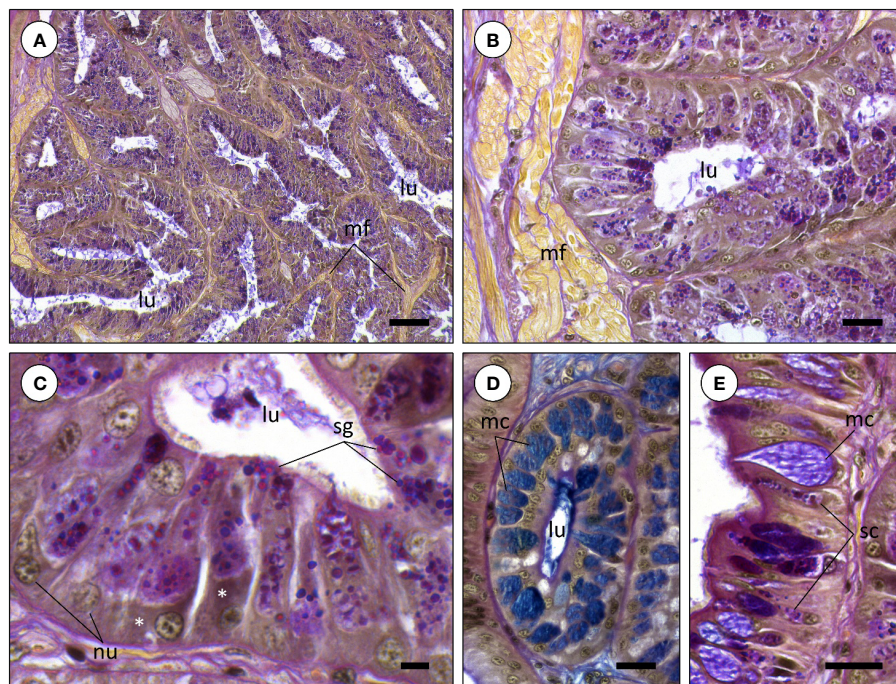


FIGURE 3

Histochemistry of the anterior salivary glands of *Sepia officinalis* (tetrachrome stain combining Alcian Blue, PAH, Wegert's iron Hematoxylin and Picric Acid). (A) General and homogeneous appearance of anterior salivary gland diverticula ('tubules'). (B) A diverticulum of an anterior salivary gland. Comparatively to the posterior salivary glands, diverticula are chiefly composed of mucin-secreting secretory cells. Note the presence of muscular fibers (mf) near the tubules resulting from the glands being embedded in the buccal mass. (C) Detail of the secretory granules being produced by tubule cells. (D) Diverticulum composed by mucus-secreting cells (mc), stained positive for Alcian blue, revealing acidic mucins. Note mucus, in blue, concentrated at the lumen (lu) of the diverticulum. (E) Glandular epithelium bearing serous (sc) and mucocytes (mc), also named goblet cells. Abbreviations: nu, nucleus; asterisk, (rough) endoplasmic reticulum; sg, secretory granules. Scale bars: (A) 100 μ m; (B, D, E) 25 μ m; (C) 10 μ m.

reading frames (ORFs), the shortest of which with 85 aa in length, from which 43.26% are full coding sequences. The initial homology-based search blast against Swiss-Prot (an expertly curated database) revealed about 19 157 annotated ORFs (Figure 4A). These were contrasted against subsets of either UniProtKB or Swiss-Prot databases to survey specifically for toxin-related proteins, enzyme-related proteins, proteins associated with post-translational modifications, secreted proteins or overall matches against eukaryote toxin or venom-related proteins (excluding humans), or other unreviewed proteins from Annelida, Cnidaria and Cephalopoda. This analysis resulted in a total of 16 158 annotated ORFs, which is equivalent to ~80% of ORFs annotated through Swiss-Prot. The results detailed in Figure 4B revealed that from those annotated ORFs, more than 1 200 ORFs had hits in all customized databases ('toxin', 'enzyme', 'secreted' and 'post-translational modifications'). 'Enzyme' detained the most hits, with about 93% of annotated ORFs (~15 000), with ~5 900 exclusive matches. About 45% of all annotated ORFs (~7 300) were significantly matched to proteins with known post-translational modifications. The annotated ORFs matched against 'Secreted' and 'Toxin', however with relatively lower percentages (about 27% and 21%, respectively, of total annotated ORFs). Since the PSG of *Sepia officinalis* is hypothesized to hold specific functions in secreting toxins and related bioactives, we contrasted the findings against a toxin-oriented customized database termed 'SeaTox' and

assessed how many annotated ORFs were in common with the preceding. The results (Figure 4C) showed that from ~1 500 annotated ORFs pertaining to the 'SeaTox' database, about 1 421 ORFs were detected in all databases, whereas 62 hits were exclusive to the 'SeaTox' database. The intersection of all databases revealed that cephalotoxins, cysteine-rich venom proteins and proteins with toxin activity or neurotoxins with metal ion binding activity yielded strongest homology matching, i.e., lowest *e*-value and highest percentage of identity (see chart in Figure 4C). The direct contrast with 'SeaTox' database also revealed high diversity of serine proteases (with 331 hits), followed by toxins, metalloproteases, chitinases, hyaluronidase and hormones (Figure 4D). Still, regardless of overall expression 'toxins' (inferred from annotation keywords) offered the highest number of hits (~750). *Sepia officinalis* transcripts encoded proteins that yielded a total of 203 292 domains, from which 19 873 were present in non-annotated predicted coding regions (5 275 ORFs). On the other hand, a total of 1 459 putative proteins that were annotated against the 'SeaTox' database also yielded 24 745 conserved protein domains. Within this group of ORFs, domains mostly linked associated with toxins, such as, CAP, Insulin, Kazal, Kunitz_BPTI and ShK were highlighted (Figure 4E), as well as, Ldl_recept_a and TSP_1, which are known to be present in cephalotoxins. Other domains (e.g., astacin, reprotolysin, trypsin) connected with venoms components, like permeabilizing and diffusing agents, were also

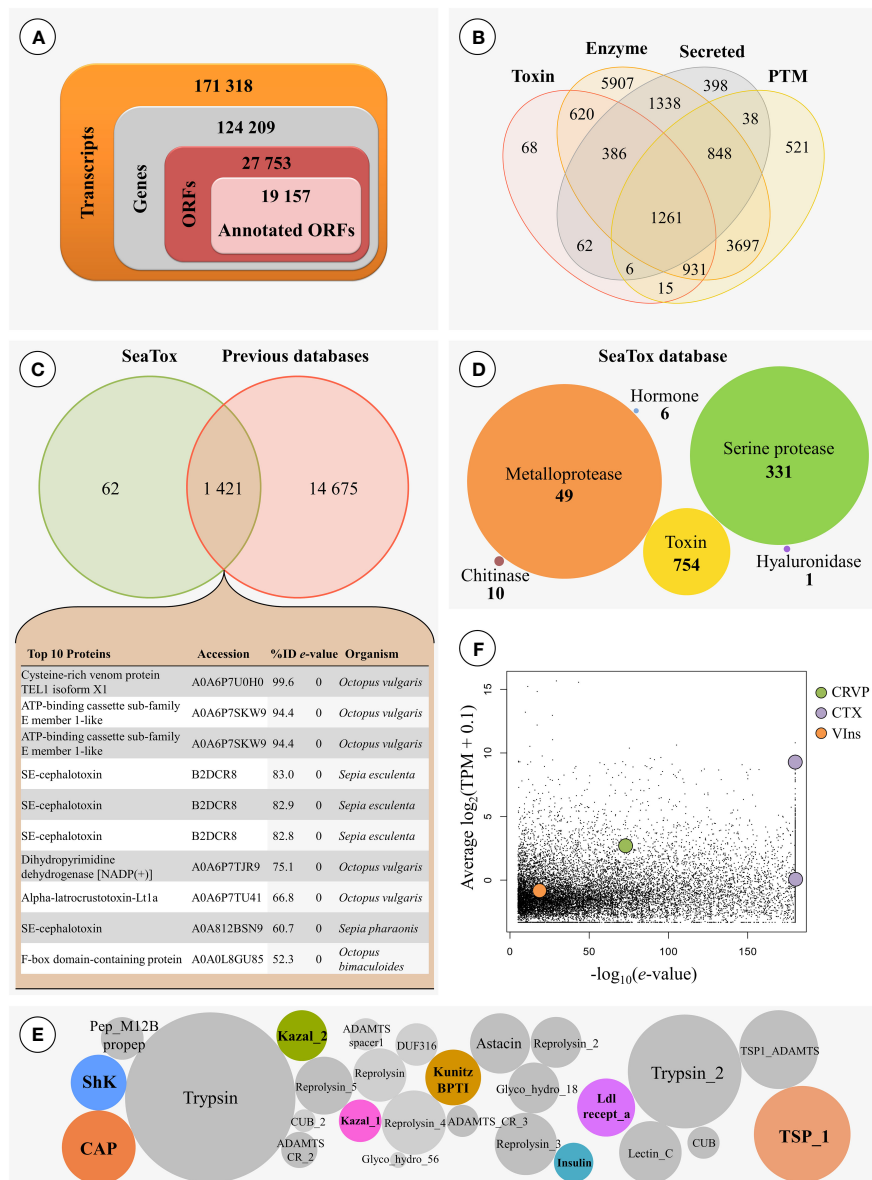


FIGURE 4

Global transcriptomics profiling of the posterior salivary glands of *Sepia officinalis*. (A) Refinement of the whole-transcriptome towards the selection of transcripts of interest. (B) Venn diagram showing the number of annotated open reading frames (ORFs) per comparison across four different customized databases ('toxin', 'enzyme', 'secreted' and 'post-translational modification'). Overlapping regions reveal the number of ORFs with homology matching against two or more databases. (C) Venn diagram representing the number of annotated ORFs with homology matching between all the previously-mentioned customized databases and the 'SeaTox' database. The top 10 hits of the common putative proteins conjoining contrasting against all customized databases are listed in the appended chart. (D) Categories (inferred from annotation) of proteins of interest with significant match against the 'SeaTox' database. The numbers within figures represent indicate the number of ORFs matching each category, while the radius of circles is representative of expression levels (TPMs). Note that ORFs may be classified in more than one category. (E) Conserved protein domains of interest in ORFs annotated against 'SeaTox' database. The colored circles represent domains most likely associated with toxins. (F) Expression levels and e-values retrieved from the BLAST analyses results for the annotated ORFs with match in all five databases, highlighting potential genes of interest within the venom secreted by all the annotated ORFs produced by *Sepia*: CRVP (Cysteine-rich venom protein), CTX (SE-cephalotoxin) and VIns (Venom Insulin) through expression levels and the quality of the annotation. The lowest e-values for the annotated ORF was represented within the homology-searches against all five customized databases.

detected. The expression of the annotated ORFs with match in all five databases is provided in Figure 4F.

From the mostly significantly matched transcripts, we selected three sequences of interest to perform validity expression analyses by RT-qPCR, based on several factors such as relevance in animal venom, reviewed annotation, expression, availability of

full coding sequence and even biotechnological potential. This short list thus includes transcripts that encode toxin-related proteins or venom components namely homologs for a cysteine-rich venom protein (CRVP) with 303 aa, a cephalotoxin (CTX) with 1042 aa and a venom insulin (VIns) with 161 aa. Overall, comparison of the expression levels and BLAST e-values revealed distinct expression

profiles these transcripts of interest, with CTX showing higher levels and similarity scores in comparison to the CRVP and VIns. Notably, the transcriptome profiles of the three independent biological replicates revealed high correlation between the annotated transcripts (Spearman's $R \sim 0.74$, $p < 0.05$), including for the relevant CRVP, CTX and VIns (Figures 5 and 6). Such findings, reinforce the biological relevance and consistency of the toxin-related proteins and venom components that were tentatively identified. Nonetheless, homology-matching analysis against the 'SeaTox' database yielded only eight ORFs that were homologous to proteins and peptides produced by *Conus*. Specifically, seven ORFs had homology-matching against a venom insulin, Con-Ins Im1 (e -value ranging from 4.4×10^{-19} to 7.51×10^{-19}) from *Conus imperialis*, whereas the remaining ORF was homologous to a conopressin/conophysin, isoform 2 fragment (e -value = 3.82×10^{-28}) from *Conus monile*.

3.3 *Sepia officinalis* express toxin-encoding transcripts conserved across venomous taxa

Given the potential biological relevance of the toxin related transcripts CRVP, CTX and VIns (for details on these proteins see

Ueda et al., 2008; De Meyts, 2016; Jiráček and Žáková, 2019; Zhang et al., 2022), these were shortlisted for detailed analysis and validation. Predicting ORFs revealed that all bear different well-conserved domains (Figure 7A). The cysteine-rich venom protein is a calcium channel regulator involved in impairing smooth muscle contraction and possesses a CAP (catapolite activator protein) domain. The toxin CTX has two specific well-conserved domains, namely the Thrombospondin type 1 domain and the Low-density lipoprotein receptor domain class A, the later containing several disulphide-bonding cysteines. The venom insulin acts on prey by inducing a hypoglycemic shock. This hormone presents the Insulin/IGF/Relaxin family domain that includes secreted regulatory hormones such as insulins, relaxins, insulin-like growth factor and bombyxin. Expression analyses by RT-qPCR involved consensus primer design for the two CTX transcript variants that were found in the assembled transcriptome. Assessing their expression by RT-qPCR confirmed their expression in the PSG (Figure 7B) and revealed comparative higher levels of expression of CTX (Tukey's HSD test, $p < 0.05$). The complete transcripts sequences were further corroborated by Sanger sequencing and then matched against their closest homologs amongst venomous or toxin-bearing animals from eumetazoan taxa namely Arthropoda, Mollusca and Cnidaria (Figure 7C). The phylogenetic trees showed closer similarity of target sequences to different species of

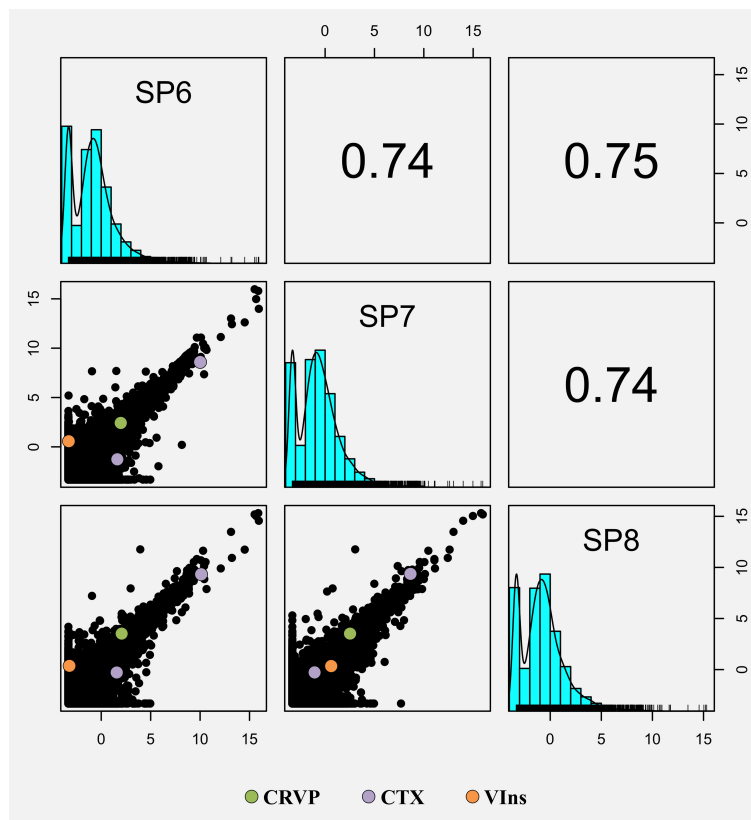


FIGURE 5

Generalized pair plot showing all the correlations given by the pairwise comparisons between the three RNA-Seq samples (SP6, SP7 and SP8). Below the diagonal are presented all the pairwise scatter plots comparing all the RNA-Seq samples, whereas Spearman correlation statistic values are displayed above the diagonal. Each point of the scatter plot represents the expression of a given gene (as \log_2 TPM). The genes of interest were highlighted in all scatter plots: CRVP (Cysteine-rich venom protein), CTX (SE-cephalotoxin) and VIns (Venom Insulin).

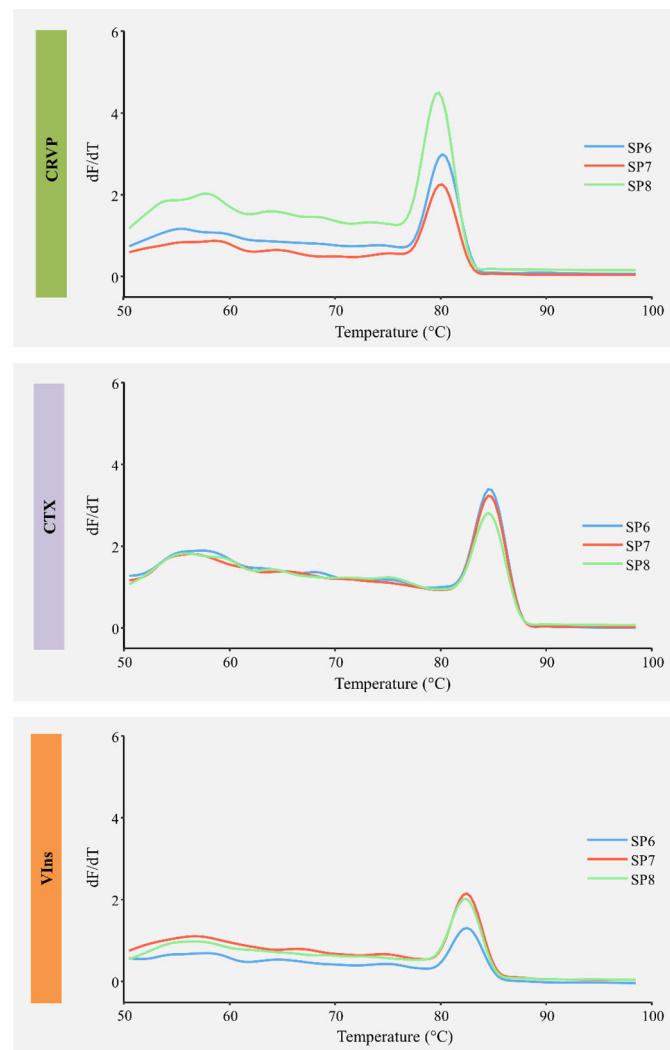


FIGURE 6

Melting curve analyses of the three primer pairs used in RT-qPCR for three shortlisted toxin full coding mRNAs isolated from the posterior salivary glands of *Sepia officinalis*. CRVP, Cysteine-rich venom protein; CTX, SE-Cephalotoxin; VIns, venom insulin. Individual replicates are identified as SP6 – SP8.

cephalopods (both cuttlefishes and octopuses). The findings allowed assuming that CRVP is similar to a cysteine-rich secretory protein (CRISP), probably a highly similar homolog to *S. pharaonis*. In turn, the CTX sequence had also displayed higher similarity with SE-cephalotoxins from other cuttlefishes but unclear positioning with other mollusks. The VIns sequence retrieved from *S. officinalis* was found to pertain to the cephalopod clade, as expected. In the case of VIns, there is a clear distinction between the cephalopod and the gastropod clades, the latter represented by *Conus* and *Aplysia*. However, this distinction is not evident for the two other proteins. The CRVP sequence holds greater similarities to CRISPs from some cnidarians (such as the bryozoan *Bugula*) than to *Aplysia*. In turn, the CTX from *S. officinalis* presented high similarities with other *Sepia* (namely *S. esculenta* and *S. pharaonis*) but higher similarities to some cnidarians than to *Octopus bimaculoides*, which yielded the closest match not only within octopodids but also the remaining mollusks.

Homology-based structural inference on the three proteins CRVP, CTX, VIns yielded global model quality estimate (GMQE, which ranges between 0 and 1) of 0.83, 0.34, and 0.67, respectively (Figure 7D). The most robust model can thus be considered for CRVP (75.83% sequence identity coverage A0A812CAC2.1.A from *Sepia pharaonis*); followed by VIns (97.52% against A0A6B9RMK7.1.A from *Sepia latimanus*) and finally CTX (73.66% against R4G2C8.1.A from *Sepioteuthis australis*).

4 Discussion

The common cuttlefish has two salivary glands, each bearing two lobes. Whereas the ASG is embedded into the buccal mass, and is therefore of difficult access and isolation, the PSG is located near the digestive gland and can be easily isolated and excised. Here we confirmed the very distinct microanatomy between the anterior and

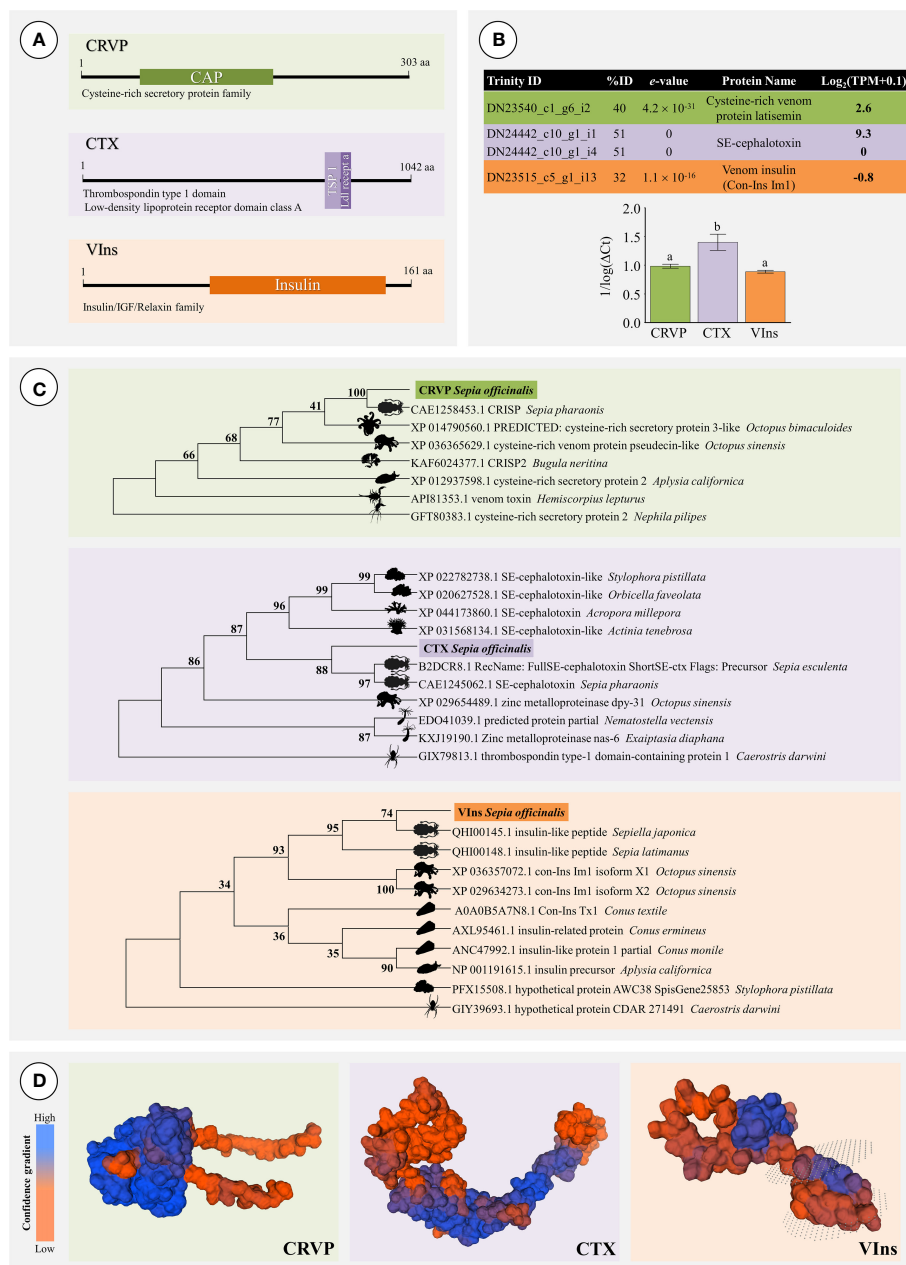


FIGURE 7

Validation of the transcriptome of the posterior salivary glands of *Sepia officinalis* by RT-qPCR of representative transcripts. **(A)** Representation of conserved domain locations for each putative protein selected for transcriptome validation: CRVP (Cysteine-rich venom protein), CTX (SE-cephalotoxin) and VIns (Venom Insulin). **(B)** Description of Expression analysis of selected transcripts (chart) by RT-qPCR (bar plot). Different letters indicate significant differences (Tukey's HSD test, $p < 0.05$). **(C)** Phylogenetic trees of CRVP, CTX and VIns. Each model includes the best matches from venomous or toxin-bearing animals for comparison. The phylogenetic models were produced using Maximum Likelihood and the Jones-Taylor-Thornton model, with 800 bootstrap replications. Bootstrap support values are given for all nodes. **(D)** Predicted homology (reference)-based models of the protein structure for each putative protein of interest: CRVP (Cysteine-rich venom protein), CTX (SE-cephalotoxin) and VIns (Venom Insulin). The latter was predicted to have a transmembrane domain (between dotted surfaces).

posterior salivary glands The dissimilarities have significant implications for their function, even though they share the general diverticula-based, well-irrigated structure that characterizes the annex glands of the cephalopod digestive tract (including the digestive gland), thus sharing the same endodermal origin (see Costa et al., 2014 and references therein). The ASG is chiefly involved in the secretion of saliva that constitute a mucin-rich vehicle for the bioactive compounds

(mostly proteinaceous) secreted *en masse* by the PSG, which can be inferred from the dense packing of secretory cells and secretory granules within. These are expelled apically into the glandular tubules' lumen through a merocrine (via exocytosis) or most likely, apocrine secretion (involving loss of parts of the cytoplasm) processes (see Rodrigo et al., 2018), even though the exact process could not be determined in the present work. The transcriptomic analyses disclosed

the secretion of a cocktail with toxins and matrix-directed proteinases such as metalloproteinases, chitinases and serine proteases that most likely act as permeabilizing agents to assist diffusion of toxins. Chitinases, in particular, bare particular relevance for *Sepia*, which is a known predator of decapod crustaceans. Most importantly, we isolated the full coding sequence of a cysteine-rich venom protein, venom insulin and a cephalotoxin, breaking way for future endeavors in recombinant expression and bioactivity testing for potential biotechnological purposes.

From animals collected in 2012 at Normandy, France, Cornet et al. (2014) disclosed that the PSG of *Sepia officinalis* holds a dual role in both predation and immune defense based on bioactivity assays with crabs and bacteria, respectively. Our findings are generally accordant with the identification of putative venom components inferred from the combination of proteomics and transcriptomics performed by these authors. However, the role of the PSG as an immune barrier, even though plausible, needs further investigation, as the vast majority of proteins that we found pertain to permeabilizing enzymes and toxins and not anti-microbial agents like bacterial lipopolysaccharide-binding peptides described by Cornet et al. (2014). In addition, besides its role in lubrication of the gut, the secretion of mucus by goblet cells (which comprise most of the tubule epithelia of the ASG) in the gut and annex glands is *per se* a critically important part of the innate immune system of eumetazoans. In addition, mucus is often the vehicle of antimicrobial and antiviral agents and other defensive molecules against pathogens (recently reviewed by Sheng and Hasnain, 2022). Altogether, our results, i.e., the combination between microanatomy and transcriptomics, suggest a separation of roles between the two salivary glands. Whereas the ASG is mostly involved in defense, the PSG is involved in predation. Interestingly, the microanatomy of gastropods is not as revealing of differences between ASG and PSG, which may be due to far less organ individualization in the Gastropoda than in the Cephalopoda. Still, our recent investigations on a predatory gastropod of the temperate intertidal that secretes cysteine-rich bioactives, likely toxins, *Nucella lapillus*, showed that its salivary gland blends goblet cell-rich tubules and tubules bearing acinar-like cells almost exclusively (D'Ambrosio et al., 2021), therefore resembling the cuttlefish's ASG and PSG tubules, respectively.

The differences between the present work and Cornet et al. (2014) are likely related to two factors: i) technical, as omics methods and associated databases and algorithms for annotation are constantly evolving, and ii) differences between the two populations of *S. officinalis*. Nonetheless, we highlight the presence of a CRVP and of CTX in common with the work by Cornet et al. (2014), which highlights the importance of these potential neurotoxins for the cuttlefish. In fact, high inter-population variability of venom composition has already been recorded in land animals, with emphasis on snakes, which is considered a serious problem for the treatment of envenomation (e.g., Casewell et al., 2020; Rashmi et al., 2021). Interestingly, Avella et al. (2022) used proteomics to address this issue in a species of viper from Portugal, which further highlights the pertinence of omics methods to study various aspects of venom composition. Even though intra-specific variation of venom composition is little investigated in marine animals besides cone snails (e.g., Fassio et al., 2019; Pardos-Blas et al., 2019), Smith et al. (2023) investigated venom variation in sea anemones combining

genomics, transcriptomics and proteomics and revealed the importance of gene duplication events and strong adaptive pressure to be major factors. Altogether, whether venom variation in various populations of *S. officinalis* results from selective pressure or phenotypic variation (such as caused by feeding) deserves further investigation. In addition, it also adds to increasing the span of bioactives of interest in this species and marine invertebrates in general.

The existence of a cephalopod 'toxic saliva' that is capable of paralyzing crabs is known since the work by Lo Bianco (1888) with octopus, but it was Ghiretti (1959) who first isolated a protein, hitherto named cephalotoxin, as the main causative agent, precisely from the PSG of *S. officinalis*. Since then, our understanding on the complex biochemistry of venoms has greatly refined. The term cysteine-rich venom proteins (CRVPs), which is occasionally used to identify cysteine-rich secretory proteins (CRISPs) in venoms and poisons, have been invariably found in the venoms of both land and marine animals. They are characterized by possessing CAP domains and high bioactivity. The CAP superfamily of proteins, among which CRISPs are one of multiple subfamilies (nine subfamilies are described in mammals alone), holds a wide span of known or predicted functions, from reproduction to immunity (Gibbs et al., 2008). Non-toxin CRISPs are commonly found in male testes of eumetazoans, mammals included, where they are believed to hold protective functions (see for instance the review by Gonzalez et al., 2021). In venoms, they may act as anti-microbials but also as neurotoxins, permeabilizing agents and regulators of inflammation (see Tadokoro et al., 2020, for a review). The latter aspect can hold particular importance for venom delivery as proteins that can promote blood flow and prevent immune function and perhaps even clotting. Interestingly, we recently found CRISP signatures in poisonous (non-injected) secretions from marine annelids as well, even in cases where low toxin-specificity is expected (Rodrigo et al., 2021; Moutinho Cabral et al., 2022), which further strengthens the ubiquitous roles of these proteinaceous bioactives in animal venoms.

Venom insulin, in turn, has not been found by Cornet et al. (2014) in *S. officinalis*. However, this hormone-like toxin has been described in cone snails, which release it to the water presumably to disorient schools of small prey fish by eliciting fast hypoglycemic shock (Safavi-Hemami et al., 2015). This recent discovery is leading to a renewed interest in hormone toxins by biotechnologists, especially because some insulin-like peptides from cone snails (Con-Ins) have been found to act much faster than human insulin (due to rapid activation without multiple cleaving, unlike native human insulin) and have high affinity to human insulin receptors, which may have very important implications for the treatment of diabetes (De Meyts, 2016; Jiráček and Žáková, 2019). The isolation of the full coding sequence for this protein can thus be a major achievement for the bioprospecting for novel peptidic bioactives in mollusks. However Swiss-Model retrieved no sufficient homology between any Con-Ins and the validated insulin-like peptide from *S. officinalis*, which indicates highly dissimilar structures. On the other hand, at this stage the structure of VIns is merely a prediction obtained against a different species of *Sepia*. Accordant with the data from structural predictions, phylogenetics (recall Figure 7) indicate that there can be two distinct clades of venom insulins based on sequence homology, one allocating cephalopods and other

the gastropods, including *Conus* and *Aplysia*. The tree thus suggests that the full sequences of these three main toxins present a significant degree of conservation with orthologs from other Cephalopoda, albeit noticeable variation within taxa and that the properties of VIns from these animals, which lacks experimental verification, may be distinct from *Conus*. Interestingly, the trend to separate between the Cephalopoda and Gastropoda clades does not extend to the CRVP/CRISPs or CTXs. The diversity and evolution of CRISPs in venomous animals is better described in snakes and terrestrial arthropods (like scorpions) than in mollusks. Even though they are believed to be relatively well conserved, only a few of these CAP domain-bearing proteins have been fully characterized (see Tadokoro et al., 2020; Zhang et al., 2022). It has been hypothesized that the evolution of CAP proteins (which started in bacteria) has been a multi-step process that potentially involved positive selection through non-toxin CRISPs and potentially from non-toxin to toxin (Vicens and Treviño, 2018; Tadokoro et al., 2020; Zhang et al., 2022). This may explain a consider variability in sequence, structure and function of these proteins even with the same taxon. In turn, cephalotoxin-like toxins (and potentially non-toxin proteins) have recently been found in very distinct animals using omics methods, including cnidarians and even fish, even though not all forms are validated (Moya et al., 2012; Domínguez-Pérez et al., 2018; Silva et al., 2018). Still, the evolution and function of these relatively large proteins (c.a. 500–1000 aa) is not yet understood, including ancestry and potential convergence. Interestingly, the validated cnidarian form (SwissProt accession B2DCR8) described by Moya et al. (2012) from the coral *Acropora millepora*, is not glycosylated and is devoid of the EGF (epidermal growth factor)-like domain present in *Sepia*, which is a disulphide bond-forming conserved domain common in venom proteins; but either form bear a peptide signal, which indicates they are actively secreted (see also Ueda et al., 2008).

To summarize, the relatively high number of overlapping hits against multiple databases (> 1 000) further sustains high biochemical diversity in the posterior salivary gland (PSG) of *Sepia officinalis*, including for aspects that have been little surveyed in cephalopods by marine toxinologists such as toxins and other secreted proteins with post-translation modifications that might be needed to assure function and specificity. Since protein folding and post-translation modifications of animal proteins can be cumbersome when promoting heterologous expression using prokaryote (e.g., *E. coli*) and even yeast models, this information can be paramount to develop adequate scale-up methods for the production, harvesting and purification for recombinant toxins and other bioactives. The present work confirmed the suspected high diversity of bioactives in the PSG of *S. officinalis*, an organ which should perhaps more correctly be named ‘venom gland’, as the ASG is chiefly responsible for the secretion of saliva vehicle of bioactives. Even though their specific bioactivity needs investigating, the toxins and accompanying peptides and proteins, of which permeabilizing agents such as metalloproteases and serine proteases are well-represented components that are likely able to affect a wide range of prey, especially crustaceans. From the annotated proteins or classes of proteins, we may highlight venom components with potential neurotoxic and endocrine-disrupting effects. The results also suggest, even though this issue

needs further enhancement, that can be considerable variability in venom composition between various populations of the common cuttlefish, as it is now known to occur in other venomous animals. However, *Sepia* venom proteins differ considerably from potential homologs in other molluscan taxa, especially the well-known *Conus* gastropods. Finally, it must be emphasized that, even in face of reduced genomic resources, modern omics can successfully be employed to screen for novel bioactives and produced validated results, here at least in the form for three full coding sequences for protein that bear biotechnological potential, a cysteine-rich venom protein (CRVP), a venom insulin (VIns) and a cephalotoxin (CTX).

Data availability statement

The datasets presented in this study can be found in online repositories. Transcriptome data is freely accessible at Gene Expression Omnibus (GEO) database, accession number GSE251667 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE251667>). The full coding sequences of shortlisted genes of interest are publicly available from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), under accession numbers OP198203 to OP198205. The ‘SeaTox’ toxin database in fasta format can be freely accessed at Zenodo (<https://doi.org/10.5281/zenodo.10418296>).

Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because animals were obtained from commercial sources.

Author contributions

CG: Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. IMC: Data curation, Formal analysis, Software, Writing – original draft, Writing – review & editing. APAM: Investigation, Methodology, Writing – original draft, Writing – review & editing. ARG: Methodology, Supervision, Writing – original draft, Writing – review & editing. PMC: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors acknowledge “Fundo azul” for the funding of research project MARVEN (Ref. FA_05_2017_007). This work is also financed by national funds from The Portuguese Foundation for Science and Technology (FCT), I.P., in the scope of project UIDP/04378/2020 and UIDB/04378/2020 of UCIBIO and project LA/P/0140/2020 of i4HB. FCT is also acknowledged for the grants SFRH/BD/144914/2019 to CG and 2022.11150.BD to IMC.

Acknowledgments

The author also thanks to Frederico Gimenez for the animal collection and the contribution of P. Henriques (CIEM) for the TEM sample preparation. Open-access publication fees were covered by the state of Bremen.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Albertin, C. B., and Simakov, O. (2020). Cephalopod biology: At the intersection between genomic and organismal novelties. *Annu. Rev. Anim. Biosci.* 8, 71–90. doi: 10.1146/annurev-animal-021419-083609
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Avella, I., Calvete, J. J., Sanz, L., Wüster, W., Licata, F., Quesada-Bernat, S., et al. (2022). Interpopulational variation and ontogenetic shift in the venom composition of Lataste's viper (*Vipera latastei*, Boscá 1878) from northern Portugal. *J. Proteomics* 263, 104613. doi: 10.1016/j.jprot.2022.104613
- Baden, T., Briseño, J., Coffing, G., Cohen-Bodénès, S., Courtney, A., Dickerson, D., et al. (2023). Cephalopod-omics: Emerging fields and technologies in cephalopod biology. *Integr. Comp. Biol.* 63, 1226–1239. doi: 10.1093/icb/icad087
- Bowersox, S. S., Gadbois, T., Singh, T., Pettus, M., Wang, Y. X., and Luther, R. R. (1996). Selective N-type neuronal voltage-sensitive calcium channel blocker, SNX-111, produces spinal antinociception in rat models of acute, persistent and neuropathic pain. *J. Pharmacol. Exp. Ther.* 279, 1243–1249.
- Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519
- Calvete, J. J., Sanz, L., Angulo, Y., Lomonte, B., and Gutiérrez, J. M. (2009). Venoms, venomomics, antivenomics. *FEBS Lett.* 583, 1736–1743. doi: 10.1016/j.febslet.2009.03.029
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2008). BLAST+: architecture and applications. *BMC Bioinform.* 10, 421. doi: 10.1186/1471-2105-10-421
- Cappello, E., and Nieri, P. (2021). From life in the sea to the clinic: The marine drugs approved and under clinical trial. *Life* 11, 1390. doi: 10.3390/life11121390
- Caruana, N. J., Cooke, I. R., Faou, P., Finn, J., Hall, N. E., Norman, M., et al. (2016). A combined proteomic and transcriptomic analysis of slime secreted by the southern bottletail squid, *Sepiadarium austrinum* (Cephalopoda). *J. Proteomics* 148, 170–182. doi: 10.1016/j.jprot.2016.07.026
- Caswell, N. R., Jackson, T. N. W., Laustsen, A. H., and Sunagar, K. (2020). Causes and consequences of snake venom variation. *Trends Pharmacol. Sci.* 41, 570–581. doi: 10.1016/j.tips.2020.05.006
- Charif, D., and Lobry, J. R. (2007). “SeqinR 1.0–2: A contributed package to the R Project for statistical computing devoted to biological sequences retrieval and analysis,” in *Structural approaches to sequence evolution: Molecules, networks, populations*. Eds. U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo (Springer-Verlag, Berlin Heidelberg), 207–232.
- Cornet, V., Henry, J., Corre, E., Le Corguille, G., Zanuttini, B., and Zatylny-Gaudin, C. (2014). Dual role of the cuttlefish salivary proteome in defense and predation. *J. Proteomics* 108, 209–222. doi: 10.1016/j.jprot.2014.05.019
- Costa, P. M., and Costa, M. H. (2012). Development and application of a novel histological multichrome technique for clam histopathology. *J. Invert. Pathol.* 110, 411–414. doi: 10.1016/j.jip.2012.04.013
- Costa, P. M., Rodrigo, A. P., and Costa, M. H. (2014). Microstructural and histochemical advances on the digestive gland of the common cuttlefish, *Sepia officinalis* L. *Zoomorphology* 133, 59–69. doi: 10.1007/s00435-013-0201-8
- D'Ambrosio, M., Gonçalves, C., Calmão, M., Rodrigues, M., and Costa, P. M. (2021). Localization and bioactivity of cysteine-rich secretions in the marine gastropod *Nucella lapillus*. *Mar. Drugs* 19, 276. doi: 10.3390/md19050276
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, 1–4. doi: 10.1093/gigascience/giab008
- De Meyts, P. (2016). Structural basis for the poisonous activity of a predator's venom insulin. *Nat. Struct. Mol. Biol.* 23, 872–874. doi: 10.1038/nsmb.3304
- Dominguez-Pérez, D., Campos, A., Alexei Rodríguez, A., Turkina, M. V., Ribeiro, T., Osorio, H., et al. (2018). Proteomic analyses of the unexplored sea anemone *Bunodactis verrucosa*. *Mar. Drugs* 16, 42. doi: 10.3390/md16020042
- Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211. doi: 10.1142/p715
- Fassio, G., Modica, M. V., Mary, L., Zaharias, P., Fedosov, A. E., Gorson, J., et al. (2019). Venom diversity and evolution in the most divergent cone snail genus *Profundiconus*. *Toxins* 11, 623. doi: 10.3390/toxins11110623
- Fedosov, A., Zaharias, P., and Puillandre, N. (2021). A phylogeny-aware approach reveals unexpected venom components in divergent lineages of cone snails. *Proc. Biol. Sci.* 288, 20211017. doi: 10.1098/rspb.2021.1017
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678
- Fusetani, N. (2009). “Marine toxins: An overview,” in *Marine Toxins as Research Tools. Progress in Molecular and Subcellular Biology*, vol. vol 46. Eds. N. Fusetani and W. Kem (Springer, Berlin, Heidelberg), 1–44. doi: 10.1007/978-3-540-87895-7_1
- Gallego Romero, I., Pai, A. A., Tung, J., and Gilad, Y. (2014). RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* 12, 42. doi: 10.1186/1741-7007-12-42
- Gayral, P., Weinert, L., Chiari, Y., Tsagkogeorga, G., Ballenghien, M., and Galtier, N. (2011). Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol. Ecol. Resour.* 11, 650–661. doi: 10.1111/j.1755-0998.2011.03010.x
- Ghiretti, F. (1959). Cephalotoxin: The crab-paralyzing agent of the posterior salivary glands of cephalopods. *Nature* 183, 1192–1193. doi: 10.1038/1831192b0
- Gibbs, G. M., Roelants, K., and O'Bryan, M. K. (2008). The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins – roles in reproduction, cancer, and immune defense. *Endocr. Rev.* 29, 865–897. doi: 10.1210/er.2008-0032
- Gonçalves, C., and Costa, P. M. (2021). Cephalotoxins: A hotspot for marine bioprospecting? *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.647344
- Gonzalez, S. N., Sulzyk, V., Weigel Muñoz, M., and Cuasnicu, P. S. (2021). Cysteine-rich secretory proteins (CRISP) are key players in mammalian fertilization and fertility. *Front. Cell. Dev. Biol.* 9. doi: 10.3389/fcell.2021.800351
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314. doi: 10.1080/10618600.1996.10474713
- Jiráček, J., and Žáková, L. (2019). From venom peptides to a potential diabetes treatment. *eLife* 8, e44829. doi: 10.7554/eLife.44829

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kumar, S., Stecher, G., Li, M., Nknyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Langmead, B., Wilks, C., Antonescu, V., and Charles, R. (2019). Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35, 421–432. doi: 10.1093/bioinformatics/bty648
- Leal, M. C., Puga, J., Seródio, J., Gomes, N. C., and Calado, R. (2012). Trends in the discovery of new marine natural products from invertebrates over the last two decades—where and what are we bioprospecting? *PLoS One* 7, e30580. doi: 10.1371/journal.pone.0030580
- Lo Bianco, S. (1888). Notizie biologiche riguardanti specialmente il periodo di maturità sessuale degli animali del Golfo di Napoli. *Mitth. Z. Stat. Neapel.* 8, 385.
- Luft, J. H. (1961). Improvements in epoxy resin embedding methods. *J. Biophys. Biochem. Cytol.* 9, 409–414. doi: 10.1083/jcb.9.2.409
- Martins, C., Dreij, K., and Costa, P. M. (2019). The state-of-the-art of environmental toxicogenomics: Challenges and perspectives of “omics” approaches directed to toxicant mixtures. *Int. J. Environ. Res. Public Health* 16, 4718. doi: 10.3390/ijerph16234718
- Martins, A., Vieira, H., Gaspar, H., and Santos, S. (2014). Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Mar. Drugs* 12, 1066–1101. doi: 10.3390/md12021066
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913
- Molinski, T. F., Dalisay, D. S., Lievens, S. L., and Saludes, J. P. (2009). Drug development from marine natural products. *Nat. Rev. Drug Discovery* 8, 69–85. doi: 10.1038/nrd2487
- Montesanto, G. (2015). A fast GNU method to draw accurate scientific illustrations for taxonomy. *Zookeys* 515, 191–206. doi: 10.3897/zookeys.515.9459
- Moutinho Cabral, I., Madeira, C., Grosso, A. R., and Costa, P. M. (2022). A drug discovery approach based on comparative transcriptomics between two toxin-secreting marine annelids: *Glycera alba* and *Hediste diversicolor*. *Mol. Omics* 18, 731–744. doi: 10.1039/D2MO00138A
- Moya, A., Huisman, L., Ball, E. E., Hayward, D. C., Grasso, L. C., Chua, C. M., et al. (2012). Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO₂-driven acidification during the initiation of calcification. *Mol. Ecol.* 21, 2440–2454. doi: 10.1111/j.1365-294X.2012.05554.x
- Olivera, B. M., Cruz, L. J., de Santos, V., LeCheminant, G. W., Griffin, D., Zeikus, R., et al. (1987). Neuronal calcium channel antagonists. Discrimination between calcium channel subtypes using ω -conotoxin from *Conus magus* venom. *Biochemistry* 26, 2086–2090. doi: 10.1021/bi00382a004
- Pardos-Blas, J. R., Irisarri, I., Abalde, S., Tenorio, M. J., and Zardoya, R. (2019). Conotoxin diversity in the venom gland transcriptome of the magician’s cone, *Pionoconus magus*. *Mar. Drugs* 17, 553. doi: 10.3390/md17100553
- Puchta, M., Boczowska, M., and Groszyk, J. (2020). Low RIN value for RNA-Seq library construction from long-term stored seeds: A case study of barley seeds. *Genes* 11, 1190. doi: 10.3390/genes11101190
- Rashmi, U., Khochare, S., Attarde, S., Laxme, R. R. S., Suranse, V., Martin, G., et al. (2021). Remarkable intrapopulation venom variability in the monocellate cobra (*Naja kaouthia*) unveils neglected aspects of India’s snakebite problem. *J. Proteomics* 242, 104256. doi: 10.1016/j.jprot.2021.104256
- Rodrigo, A. P., and Costa, P. M. (2019). The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. *Environ. Res.* 173, 270–280. doi: 10.1016/j.envres.2019.03.048
- Rodrigo, A. P., Grosso, A. R., Baptista, P. V., Fernandes, A. R., and Costa, P. M. (2021). A transcriptomic approach to the recruitment of venom proteins in a marine annelid. *Toxins* 13, 97. doi: 10.3390/toxins13020097
- Rodrigo, A. P., Martins, C., Costa, M. H., Alves de Matos, A. P., and Costa, P. M. (2018). A morphoanatomical approach to the adaptive features of the epidermis and proboscis of a marine Polychaeta: *Eulalia viridis* (Phyllodocida: Phyllodocidae). *J. Anat.* 233, 567–579. doi: 10.1111/joa.12870
- Ruder, T., Sunagar, K., Undheim, E. A., Ali, S. A., Wai, T. C., Low, D. H., et al. (2013). Molecular phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. *J. Mol. Evol.* 76, 192–204. doi: 10.1007/s00239-013-9552-5
- Safavi-Hemami, H., Gajewiak, J., Karanth, S., Robinson, S. D., Ueberheide, B., Douglass, A. D., et al. (2015). Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1743–1748. doi: 10.1073/pnas.1423857112
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Shen, G. S., Laver, R. T., and McCabe, R. T. (2000). Conopeptides: From deadly venoms to novel therapeutics. *Drug Discov. Today* 5, 98–106. doi: 10.1016/s1359-6446(99)01454-3
- Sheng, Y. H., and Hasnain, S. Z. (2022). Mucus and mucins: The underappreciated host defense system. *Front. Cell. Infect. Microbiol.* 12, 856962. doi: 10.3389/fcimb.2022.856962
- Silva, F., Huang, Y., Yang, V., Mu, X., Shi, Q., and Antunes, A. (2018). Transcriptomic characterization of the South American freshwater stingray *Potamotrygon motoro* venom apparatus. *Toxins* 10, 544. doi: 10.3390/toxins10120544
- Silver, N., Best, S., Jiang, J., and Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* 7, 33. doi: 10.1186/1471-2199-7-33
- Smith, E. G., Surm, J. M., Macrander, J., Simhi, A., Amir, G., Sachkova, M. Y., et al. (2023). Micro and macroevolution of sea anemone venom phenotype. *Nat. Commun.* 14, 249. doi: 10.1038/s41467-023-35794-9
- Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Res* 4, 1521. doi: 10.12688/f1000research.7563.2
- Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* 28, 1102–1104. doi: 10.2144/002861r01
- Tadokoro, T., Modahl, C. M., Maenaka, K., and Aoki-Shioi, N. (2020). Cysteine-rich secretory proteins (CRISPs) from venomous snakes: An overview of the functional diversity in a large and underappreciated superfamily. *Toxins* 12, 175. doi: 10.3390/toxins12030175
- Turner, A., Kaas, Q., and Craik, D. J. (2020). Hormone-like conopeptides – new tools for pharmaceutical design. *RSC Med. Chem.* 11, 1235–1251. doi: 10.1039/D0MD00173B
- Ueda, A., Nagai, H., Ishida, M., Nagashima, Y., and Shiomi, K. (2008). Purification and molecular cloning of SE-cephalotoxin, a novel proteinaceous toxin from the posterior salivary gland of cuttlefish *Sepia esculenta*. *Toxicon* 52, 574–581. doi: 10.1016/j.toxicon.2008.07.007
- UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100
- Venable, J. H., and Coggeshall, R. (1965). A simplified lead citrate stain for use in electron microscopy. *J. Cell Biol.* 25, 407–408. doi: 10.1083/jcb.25.2.407
- Vicens, A., and Treviño, C. L. (2018). Positive selection in the evolution of mammalian CRISPs. *J. Mol. Evol.* 86, 635–645. doi: 10.1007/s00239-018-9872-6
- von Reumont, B. M., Campbell, L. I., Richter, S., Hering, L., Sykes, D., Hetmank, J., et al. (2014). A Polychaete’s powerful punch: venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. *Genome Biol. Evol.* 6, 2406–2423. doi: 10.1093/gbe/evu190
- von Reumont, B. M., Undheim, E. A. B., Jauss, R. T., and Jenner, R. A. (2017). Venomics of remipede crustaceans reveals novel peptide diversity and illuminates the venom’s biological role. *Toxins* 9, 234. doi: 10.3390/toxins9080234
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi: 10.1093/nar/gky427
- Whitelaw, B. L., Strugnell, J. M., Faou, P., da Fonseca, R. R., Hall, N. E., Norman, M., et al. (2016). Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. *J. Proteome Res.* 15, 3284–3297. doi: 10.1021/acs.jproteome.6b00452
- Williams, J. A., Day, M., and Heavner, J. E. (2008). Ziconotide: an update and review. *Expert Opin. Pharmacother.* 9, 1575–1583. doi: 10.1517/14656566.9.9.1575
- Zhang, Q., Xu, J., Zhou, X., and Liu, Z. (2022). CAP superfamily proteins from venomous animals: Who we are and what to do? *Int. J. Biol. Macromolecules* 221, 691–702. doi: 10.1016/j.ijbiomac.2022.09.079