

Article

Transcriptomic Profile of the Cockle *Cerastoderma edule* Exposed to Seasonal Diarrhetic Shellfish Toxin Contamination

Dany Domínguez-Pérez ^{1,*}, José Carlos Martins ¹, Daniela Almeida ², Pedro Reis Costa ³, Vitor Vasconcelos ^{1,4} and Alexandre Campos ^{1,*}

¹ CIIMAR—Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Rua General Norton de Matos s/n, Terminal de Cruzeiros do Porto de Leixões, 4450208 Matosinhos, Portugal; jmartins@ciimar.up.pt (J.C.M.); vvascon@fc.up.pt (V.V.)

² Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA; danielaalmeida23@gmail.com

³ IPMA—Instituto Português do Mar e da Atmosfera, Rua Alfredo Magalhães Ramalho, 6, 1495006 Lisbon, Portugal; prcosta@ipma.pt

⁴ Biology Department, Faculty of Sciences, University of Porto, Rua do Campo Alegre, s/n, 4169007 Porto, Portugal

* Correspondence: dany.perez@ciimar.up.pt (D.D.-P.); acampos@ciimar.up.pt (A.C.); Tel.: +351-22-340-1800 (A.C.); Fax: +351-22-339-0608 (A.C.)

Abstract: Bivalves constitute an important source of proteins for human consumption, but some accumulate biotoxins such as diarrhetic shellfish toxins (DSTs), constituting a risk to human health. The cockle *Cerastoderma edule* is one of the most important species harvested in the Portuguese coast but also one of the most affected species due to recurrent DSTs exposure. However, little is known regarding the effects of the toxins produced by blooming dinoflagellates on *C. edule*. Herein, we explore the Differentially Expressed Genes (DEGs) of two tissues (gills and digestive gland) from wild cockles sampled in Portugal, through their whole transcriptomic response in two different seasons (exposed and not exposed to DSTs). The de novo transcriptome assembly returned 684,723 contigs, N50 of 1049, and 98.53% completeness. Altogether, 1098 DEGs were identified, of which 353 DEGs were exclusive for the digestive gland, 536 unique for the gills and 209 DEGs were common. Among DEGs were identified known DSTs-biomarkers including glutathione peroxidase, glutathione S-transferase, superoxide dismutase, cytochrome P450, ABC transporters, actin and tubulin-related proteins, Heat shock proteins and complement C1Q-like proteins. This study provides the first transcriptomic profile of *C. edule*, giving new insights about its molecular responses under different environmental conditions of DSTs exposure.

Keywords: DSTs; bivalves; RNAseq; biomarkers; Differentially Expressed Genes (DEGs); Portugal

Key Contribution: This study provides the first transcriptomic profile of two tissues from *C. edule* (digestive gland and gills), in two different environmental conditions. Our findings give new insights about its molecular responses towards DSTs exposure; including known biomarkers involved in antioxidant system; metabolic detoxification and immune responses. The data generated constitute a valuable genomic resource to develop more specific experiments; serving as template for evolutionary analyses allowing the specific-primer design for the in vitro assessment of the DEGs/biomarkers responses in controlled conditions. Besides, the provided transcriptomic profile can be used to create a custom protein database for further proteogenomic analyses to evaluate the DEGs/DSTs-biomarker expressions at the proteomic level.

1. Introduction

Seasonal microalgae blooms constitute an ongoing concern due to their potential impacts on fisheries, seafood safety, tourism and other ecosystem services [1–5]. The high



Citation: Domínguez-Pérez, D.; Martins, J.C.; Almeida, D.; Costa, P.R.; Vasconcelos, V.; Campos, A. Transcriptomic Profile of the Cockle *Cerastoderma edule* Exposed to Seasonal Diarrhetic Shellfish Toxin Contamination. *Toxins* **2021**, *13*, 784. <https://doi.org/10.3390/toxins13110784>

Received: 10 October 2021
Accepted: 2 November 2021
Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

level of some marine biotoxins produced during these events, known as harmful algal blooms (HABs), can kill fish and other sea organisms on a large scale [6–8]. Currently, one of the most impactful marine biotoxins is the group of Diarrhetic Shellfish Toxins (DSTs) [6], comprising okadaic acid (OA) and analogous molecules named dinophysistoxins (DTX–1, –2 and –3) [9]. Indeed, DSTs outbreaks are the most abundant from recurrent HABs in Europe (North Atlantic) [8,10], as well as other temperate regions of the world such as Asia or South America [6]. The main source of DSTs in the marine trophic chain are species of the genus *Dinophysis*, but they can be produced by dinoflagellates belonging to the genera *Prorocentrum* as well [9].

DSTs are lipophilic toxins that can be filtered and accumulated in the adipose tissues by some marine organisms [11], being shellfish the main route of transmission to humans [12,13]. Shellfish constitutes an important source of protein for humans, but their consumption represents a food safety threat when contaminated with DSTs [13–15]. The ingestion of contaminated shellfish with DSTs can produce poisoning symptoms such as diarrhea, nausea, vomiting and abdominal pain [12,16]. In Portugal, for instance, there is an old tradition of eating shellfish, including some commercial species of bivalves like mussels, clams, scallops, oysters, and cockles [13]. However, DSTs shellfish contamination occurs every year, between April and October [2], resulting in the temporary banning of shellfish harvesting and trade [10].

The toxic effects of DSTs in humans mainly involve the inhibition of serine/threonine protein phosphatases 1 and 2A (PP1 and PP2A) [15,17], causing deregulation of several intracellular processes [13]. However, shellfish can accumulate a high amount of DSTs, mainly in the digestive gland [18,19], without known records of mortality associated with blooms of DST-producing dinoflagellates [13]. The mechanism underlying filter-feeding bivalves' resistance to DSTs includes the removal of the toxins contained in non-digested microalgae cells and food materials through the production of feces and pseudo-feces, as well as activation of biotransformation and cellular excretion [9,18]. The toxin resistance at the molecular level has been associated with some antioxidant enzymes such as superoxide dismutase (SOD), catalase (CAT), and glutathione peroxidase (GPx) [20], phase-I and -II drug-metabolizing enzymes cytochrome P450 (CYP450) and glutathione s-transferase (GST) [21,22], and ATP-binding cassette (ABC) transporters (e.g., P-glycoprotein) [23–25].

Advances in OMICs approaches have corroborated these biomarkers in some bivalves contaminated with DSTs [9,13]. OMICs could contribute to identify the repertoire of the proteins involved in the dynamics of DSTs in shellfish, specificity, and selectivity of transforming enzymes to OA and DTX–1,–2 [13,20,26–28]. Nonetheless, a wide range of characterized transcripts, proteins, or metabolites in public databases are only available for some species, such as mussels and oysters. Indeed, the marine mussel *Mytilus galloprovincialis* [29] or oyster *Magallana gigas* (previously known as *Crassostrea gigas*) [30] are better represented, with fully sequenced genomes, transcriptome, and proteome characterization, while other commercial species like clams *Ruditapes decussatus* [31], the marine gastropod *Lobatus gigas* [32], tellin *Donax trunculus* and the cockle *C. edule*, remain poorly studied [13].

It has been demonstrated that depuration rates of DSTs vary among species (cockles > oysters > scallops > clams > mussels) [9,18,33–36]. The cockle *C. edule* (Figure 1) is, within the group of bivalves, the species with the highest clearance rates. This unique depuration capacity is, however, poorly understood. *C. edule* constitutes an important commercial species in Portugal, which is still harvested and collected by the Portuguese population, using traditional techniques. Nonetheless, this is one of the most seasonally affected bivalve species by DST outbreaks in Portugal, mainly in the summer (Figure 1).

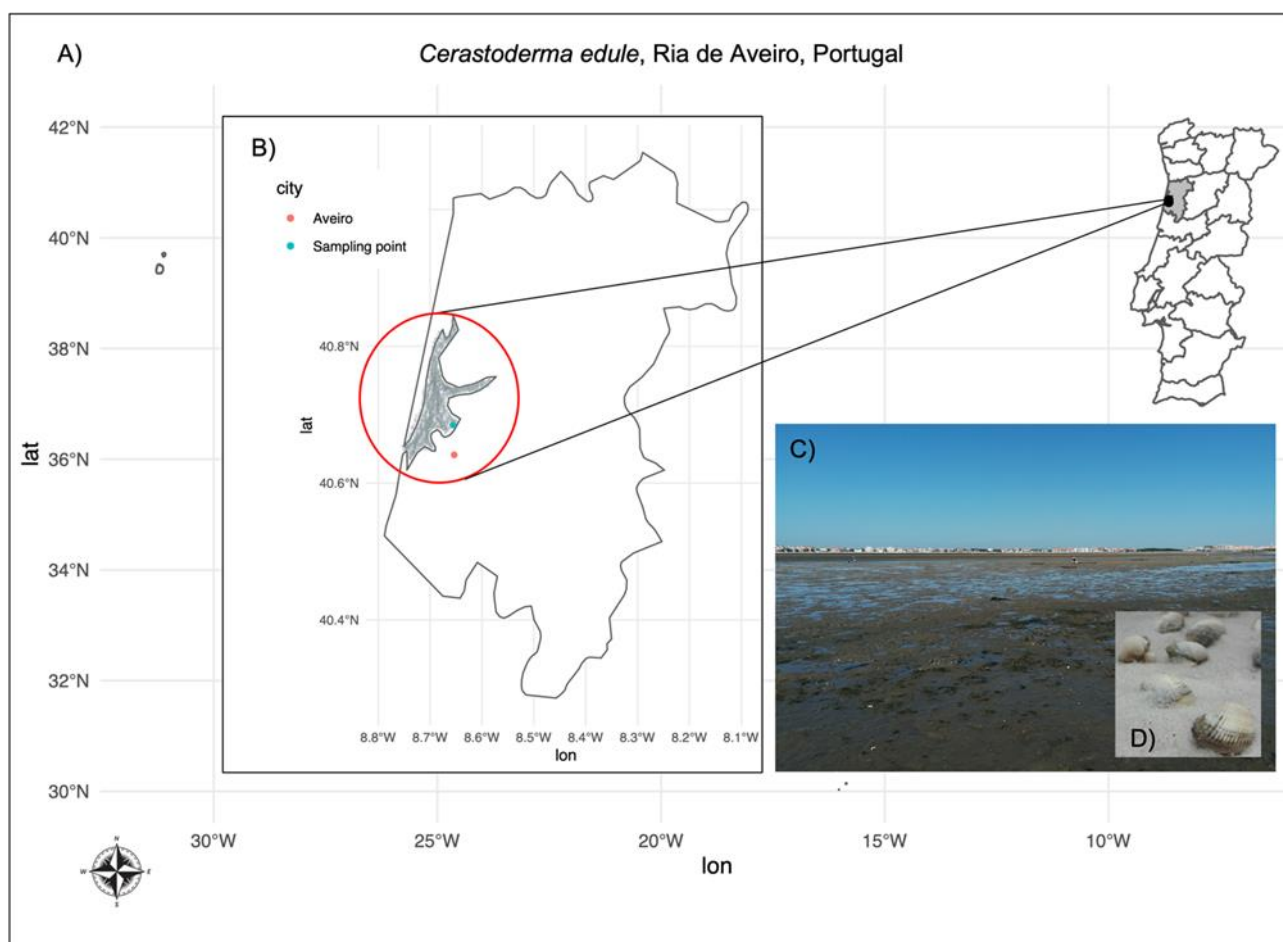


Figure 1. Shellfish harvesting area in Ria de Aveiro, Portugal. The composite figure displays in the background (A) the map of Portugal (x -axis: longitude, y -axis: latitude), highlighting the geographical localization of Ria de Aveiro, which is enlarged in panel (B), showing the coordinates of Aveiro city (red) and the sampling point (green). In the panel (C) is shown a picture of the harvesting area in Ria de Aveiro, during the low tide in the sampling point of (D) the cockles *Cerastoderma edule*.

In this work we performed for the first time a transcriptomic analysis of two tissues from wild cockles sampled in Ria de Aveiro, Portugal (Figure 1), during two different seasons, namely during blooms of *Dinophysis* spp when high levels of DSTs are reached in cockles and in periods of no shellfish toxicity (exposed and not exposed to DSTs). Ria de Aveiro is a shallow lagoon located on the Portuguese Northwest Atlantic coast (Figure 1), and one of the major shellfish harvesting sites in this country [37]. This lagoon is repeatedly exposed to several HABs, including DSTs-producing dinoflagellates, that significantly impact the commercial harvesting of bivalves [37,38]. Thus, we sequenced the total RNA of two tissues: gills and digestive gland, using a paired-end RNA sequencing approach in an Illumina NovaSeq 6000 platform. These two organs are highly important to consider when examining the impacts of toxic algae exposure. The digestive gland is where the accumulation of biotoxins and their biotransformation processes mainly occurs, while the gills are in direct contact with water and toxic cells [18,39]. Some of these DEGs identified were consistent with known molecular biomarkers reported in the DSTs depuration process in bivalves. This study constitutes the first transcriptomic profile of *C. edule* exposure shedding new lights of the molecular response of this species in different environmental scenarios of DST exposure.

2. Results

2.1. DSTs Determination by LC-MS/MS in *C. edule* Sampled in Ria de Aveiro, Portugal

Herein, we performed a transcriptomic analysis of two tissues from wild cockles sampled in Ria de Aveiro (Figure 1), Portugal, in two different seasons of the year 2019. The DSTs content in the homogenate of samples from both seasons studied was determined via liquid chromatography with tandem mass spectrometry (LC-MS/MS). The quantity of DSTs was below the detection limit in the samples used as control (March 2019), which confirmed that such samples were not exposed to DSTs, whereas the exposed samples (May 2019) contained amounts over the limits permitted for human consumption, being 218.3 µg/kg for OA, 124.0 µg/kg for DTX2, which in terms of total OA equivalents corresponds to 292.8 µg/kg of total OA equivalent.

2.2. Illumina Sequencing, Transcriptome Assembly and Completeness

The total RNA of two tissues: gills and digestive gland, was extracted using an RNeasy Mini kit (Qiagen, Venlo, The Netherlands). RNA samples were conducted for paired-end RNA sequencing using an Illumina NovaSeq 6000 platform (Illumina, Inc., San Diego, CA, USA) at Macrogen (Macrogen, Inc., Seoul, Korea). The total RNA sequencing delivered around 41.5 reads per sample with 151 bp of read length, accounting for more than 6.2 billion bases (Table 1). Most of the sequenced reads were identified with high accuracy, with approximately 98% showing a Phred quality score (i.e., ratio of bases that expresses the accuracy of each nucleotide) over 20 (Q20: corresponding to 99% base call accuracy), and around 94% of reads with Phred score over 30 (Q30: corresponding to 99.9% base call accuracy) (Table 1). After adapter trimming, high quality reads were conducted for de novo transcriptome assembly using Trinity v2.10.0.

Table 1. Statistics summary of transcriptomic analyses performed with wild cockles sampled in Ria de Aveiro, Portugal, exposed to Diarrhetic Shellfish Toxins (DSTs), and without DST. The table includes Raw Data Statistics from sequencing process, the de novo Assembly and Annotation metrics.

Summary of Analyses per Sample/Condition ¹	Statistics	Ce_DGc	Ce_DGe	Ce_Gc	Ce_Ge
Raw Data Statistics	Total Bases ²	6,248,071,054	6,402,268,630	6,258,018,330	6,333,078,920
	Read Count ³	41,377,954	42,399,130	41,443,830	41,940,920
	GC (%) ⁴	42.33	41.63	41.09	41.05
	AT (%) ⁵	57.67	58.37	58.91	58.95
	Q20 (%) ⁶	97.95	97.89	98.02	97.9
	Q30 (%) ⁷	94.06	93.83	93.98	93.8
de novo Transcriptome Assembly (Trinity v2.10.0)	Assembled transcripts			684,723	
	Total genes			361,460.0	
	Minimum/Maximum Contig length (bp: base pair) ⁸			200/42,164 bp	
	Contig length (N50) ⁸			1094 bp	
	Contig length (Average) ⁹			736.09 bp	
	GC (%)			37.39	
	Total assembled bases (all transcripts)			504,019,879	
	Total assembled bases (longest isoform per gene)			199,701,724	
Assembly Completeness (BUSCO v5) ¹¹	Percentage of all read aligned / not aligned ¹⁰	92.1%/7.8%	92.4%/7.5%	93.1%/6.8%	93.3%/6.6%
	number of BUSCOs/species			954/65	
	Complete Single-Copy			282/29.56%	
	Complete duplicated			658/69%	
Assembly Clustering ¹² (CD-HIT 4.8.1) and Completeness ¹¹	Fragmented			8/0.84%	
	Missing			6/0.63%	
	number of clusters at 90% identity			402,961	
	Complete Single-Copy			771/80.82%	
	Complete Duplicated			170/17.82%	
Fragmented			8/0.84%		
Missing			5/0.52%		

Table 1. Cont.

Summary of Analyses per Sample/Condition ¹	Statistics	Ce_DGc	Ce_DGe	Ce_Gc	Ce_Ge
ORFs prediction ¹³ (TransDecoder v5.5.0)	Transcripts with ORF (best isoform per gene)		48,107		
	Blast Hits nr/NCBI (Metazoa/non-Metazoa) ¹⁴		37,184/362		
	Blast Hits SwissProt (Metazoa/non-Metazoa) ¹⁵		26,968/22		
ORFs Annotation	Averall ORFs with Blast hits/ no hits ¹⁶		37,262 (77.5%)/10,845 (22.5%)		
	EggNOG ¹⁷		31,848 (66.2%)		
	KEGG Pathways/KO ¹⁸		12,760 (25.9%)/20,621 (42.8%)		
	IPS/GO annotation (%) ¹⁹		48,107 (100%)/19,398 (40.3%)		
	Gene Ontology (ORFs with GO annotation) ²⁰		21,799 (45.3%)		

¹ Sample ID/Condition: **Ce_DGc**: *Cerastodesma edule* digestive gland (DG) without dinophysistoxins (DSTs), **Ce_DGe**: *C. edule* DG exposed to DSTs, **Ce_Gc**: *C. edule* gills (G) without DSTs, **Ce_Ge**: *C. edule* gills exposed to DSTs. ² Total read bases: Total number of bases sequenced. ³ Total reads: Total number of reads, corresponding to the sum of read 1 (Upstream Files Pattern: _1) and read 2 (Downstream Files Pattern: _2), obtained from Illumina paired-end sequencing. ⁴ GC(%): GC content. ⁵ AT(%): AT content. ⁶ Q20(%): Ratio of bases that have Phred quality score (expresses the accuracy of each nucleotide) of over 20 (99% base call accuracy). ⁷ Q30(%): Ratio of bases that have Phred quality score (expresses the accuracy of each nucleotide) of over 30 (99.9% Base call accuracy). ⁸ N50: is defined as the sequence length of the shortest contig at 50% of the total transcriptome/genome length. ⁹ Average contigs length. ¹⁰ RNA-Seq Read Representation corresponding to the percentage of the overall alignment (reads aligned/not aligned). ¹¹ The de novo assembly quality and completeness assessment with the Benchmarking Universal Single-Copy Orthologs (BUSCOs), showing the number of BUSCOs/relative representation (percentage) of Complete Single-Copy, Complete Duplicated, Fragmented and Missing. BUSCO v5 run in mode: Transcriptome, lineage Metazoa, e-value 1.0E−3. ¹² Sequence redundancy of the de novo assembled transcriptome was reduced using CD-HIT 4.8.1, setting 0.9 as the Sequence Identity Threshold. ¹³ Protein Coding Sequences (CDS) obtained by six-frame translation with TransDecoder v5.5.0., considering a minimum length of 100 amino acids for open reading frames (ORFs), homology to known proteins via Pfam searches, and the best/longest isoform per gene. ¹⁴ Blast hits of the 48,107 CDS (best isoforms) obtained from the transcriptome assembly searched against the non-redundant (nr) protein database from NCBI, using BlastX program, filtering by the taxonomic section Metazoa/non-Metazoans (accessed on 11 May 2021), setting a cut-off e-value 1.0E−3. ¹⁵ Blast hits of the 48,107 CDS (best isoforms) obtained from the transcriptome assembly searched against the protein database UniProtKB/Swiss-Prot, using BlastX program, filtering by the taxonomic section Metazoa/non-Metazoans (accessed on 12 May 2021), setting a cut-off e-value 1.0E−3. ¹⁶ Summary of Blast hits/no hits found against both databases searched, and their relative representation (percentage) within all (48,107) predicted ORFs. ¹⁷ Functionally annotated ORFs according to orthologous groups (OGs: i.e., COGs, arCOGs, ENOGs, KOGs), using the eggNOG Mapper 1.0.3 with EggNOG 5.0.0, and its relative representation (percentage) within all (48,107) predicted ORFs. ¹⁸ Functionally annotated ORFs according to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and KEGG Orthology (KO), and its relative representation (percentage) within all (48,107) predicted ORFs. ¹⁹ Functionally annotated ORFs according to Gene Ontology (GO) and known proteins domains using InterProScan (IPS) v5.51–85.0, searched against InterPro member databases (i.e., Pfam, PROSITE, PRINTS, ProDom, SMART, TIGRFAMs, PIRSE, SUPERFAMILY, Gene3D, and PANTHER), as well as its relative representation (percentage) within all (48,107) pred. ²⁰ Overall functionally annotated ORFs according to Gene Ontology (GO) and its relative representation (percentage) within all (48,107) predicted ORFs.

Around 93% of the pre-processed reads comprising 504,019,879 of bases, were aligned in the de novo assembly. Altogether, we obtained an assembled transcriptome with 684,723 transcripts and 361,460 genes by combining all four samples into a single transcriptome, of which 684,576 transcripts showed length over 200 bp (Table 1). The de novo assembly generated showed a contig length (N50) of 1094 bp, average contig length of 736.09 bp, with a minimum and maximum contig length of 200 bp and 42,164 bp, respectively (Table 1). The assembly completeness revealed that 98.53% of the BUSCO groups have complete gene representation (single-copy or duplicated), while 0.84% are only partially recovered, and 0.63% are missing (Table 1). The number of duplicated BUSCOs were reduced from 68.97% to 29.56% after clustering analyses (Figure S1), resulting in 402,961 number of clusters at 90% identity (Table 1). The transcriptomic project has been registered with the BioProject database under BioProject ID: PRJNA739261 (<http://www.ncbi.nlm.nih.gov/bioproject/739261>). It is noteworthy, that 147 remaining transcripts with length below 200 bp, and additional 4930 transcripts, detected as putative contaminants were removed during NCBI transcriptome submission. This filtered Transcriptome Shotgun Assembly project containing 679,646 contigs has been deposited at DDBJ/EMBL/GenBank under the accession GJGL00000000. All statistics and additional data of the total RNA sequencing, assembly, and clustering are summarized in Table 1, DATASET S1, as well as in the data availability section.

2.3. Open Reading Frame Prediction and Annotation

The 402,961 clusters obtained with 0.9 of sequence identity were used for the open reading frames (ORFs) prediction using the TransDecoder v5.5.0., considering a minimum length of 100 amino acids. Overall, 48,107 protein coding sequences (CDS) were identified with homology to known proteins via Pfam searches as the best/longest isoform per gene (Table 1). A total of 21,788 CDS were complete proteins, corresponding to 45%, whereas around 33% were partial sequences, and 21% internal sequences (Figure S2, DATASET S2).

These 48,107 ORFs were CloudBlasted against the metazoans section of the non-redundant (nr) protein database from NCBI and UniProtKB/Swiss-Prot protein database [40], using CloudBlast option with the BlastX program setting a cut-off E-Value $1.0E-3$ [41]. A total of 37,184 transcripts had hits against metazoans and 362 against non-metazoans species within the nr database. On the other hand, 26,968 transcripts showed significant homology to metazoans species in UniProtKB/Swiss-Prot protein database, and 22 transcripts with non-metazoans species, of which 20 hits were common to non-metazoan matches found in the nr database. Overall, 37,262 transcript sequences corresponding to 77.5% had blast hits, whereas 10,845 corresponding to 22.5% (Table 1) remained without hits.

Besides, 31,848 (66.2%) of these 48,107 ORFs were also annotated with eggNOG-Mapper 1.0.3 with EggNOG 5.0.0 [42]. Moreover, the corresponding GO annotation was performed by combining the existing GO terms (Blast hits mapping/GO annotation, EggNOG, InterProScan) and KEGG pathways [43]. In total, 21,799 (45.3%) transcripts were functionally annotated by Gene Ontology (Table 1).

It is noteworthy that our samples were retrieved from the wild, thus, putative contaminants were detected (Figure S3) and carefully filtered before pairwise differential expression analyses. Indeed, 611 transcripts out of the 48,107 predicted ORFs (as best isoforms), were filtered out considering Blast hits description against non-metazoans section, eggNOG best Taxa Level, and putative contaminants automatically detected by the NCBI/TSA submission system. The remaining 47,496 ORFs encoded by 33,665 genes were then used for differential expression analyses. The list of these 47,496 ORFs and the list of putative contaminants with their corresponding annotation (e.g., Blast Top-Hits, GO annotation and eggNOG annotation) can be found in Tables S1 and S2, respectively.

2.4. Differential Expression Analyses and Enrichment

All transcripts encoded by the 33,665 *C. edule* coding genes (related with the 47,496 ORFs) were used as the template to calculate gene expression. In total, 141,638 transcripts were retrieved from the de novo transcriptome assembly and then used to map the pre-processed raw data to create the count tables for each pairwise comparison. Considering the distribution of counts across libraries (Figure S4), some feature/isoforms were filtered on a count-per-million basis (CPM). Altogether, 31,772 genes were retained after counts filtering and then considered for pairwise differential expression analyses in the digestive gland, while in the gills were used 30,861 features. The differential expression analyses for the digestive gland resulted in 562 DEGs, of which 202 genes were up-regulated and 360 down-regulated, whereas in the gills were found 745 DEGs, being 99 up-regulated and 646 genes down-regulated (Figure 2).

In total, 1098 unique/non-redundant DEGs were obtained in both pairwise comparisons, 209 DEGs were common in both tissues, being 353 DEGs exclusive for the digestive gland comparison, and 536 unique DEGs found in the gills (Figure 3). The pattern of DEGs expression is very similar among samples from different tissues of the same conditions. However, some genes show a tissue-dependent expression pattern (Figure 3). Indeed, among those 209 overlapped DEGs, 16 DEGs showed a similar up-regulated pattern in both exposed conditions, 189 DEGs showed a similar down-regulated pattern, and four DEGs were putatively non-condition-dependent (Figure 2, Tables S3 and S4).

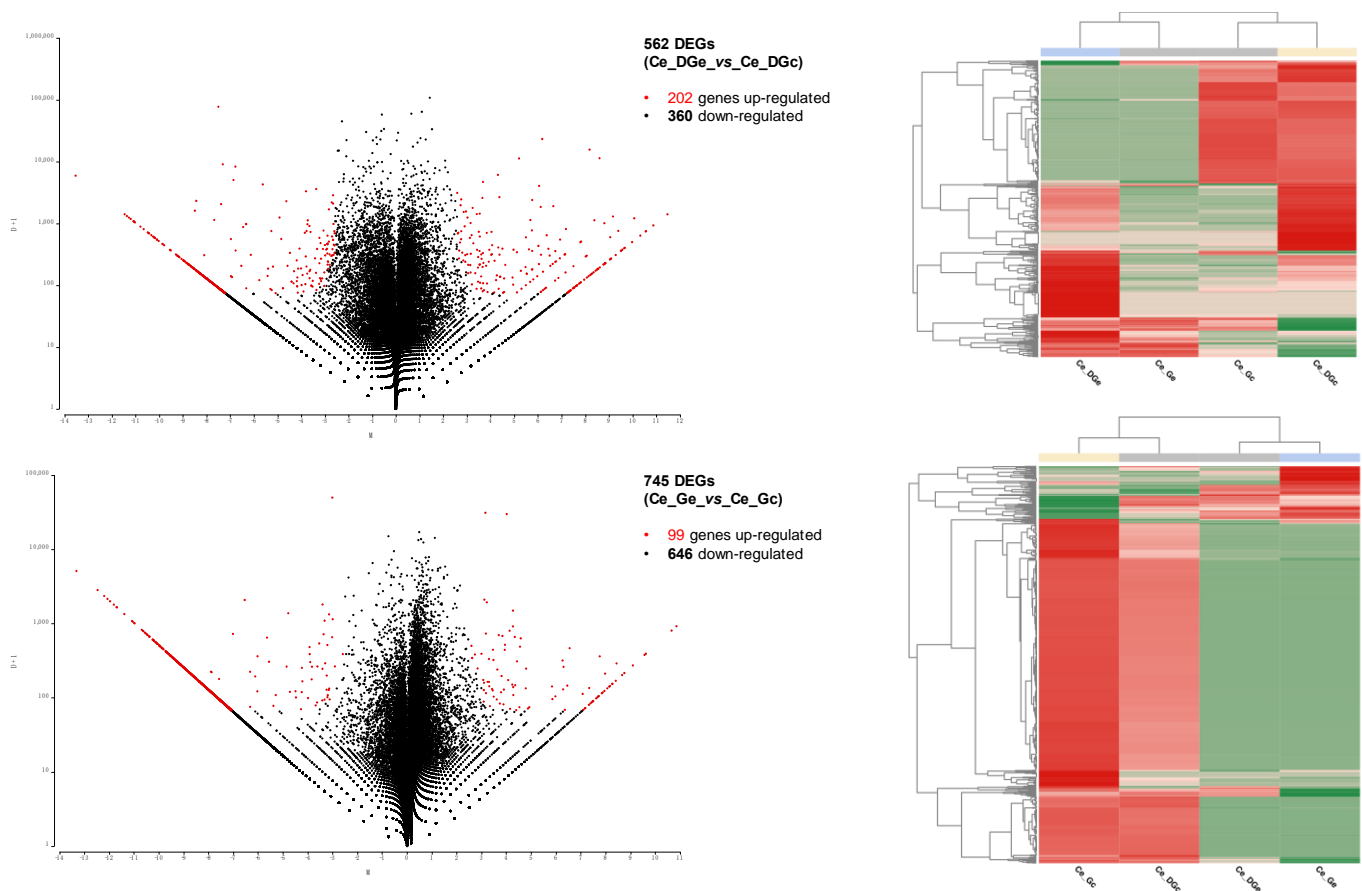


Figure 2. Pairwise Differential Expression results using NOISeq R/Bioc package ($p < 0.01$). The figure shows a scatter plot showing the log-fold change (M) and the absolute value of the difference in expression between conditions (D), where D values are displayed in a log scale. The Differentially Expressed Genes (DEGs) are depicted as red dots. The corresponding heat map of DEGs (Z-score values) from each condition tested (in columns) are depicted: Ce_DGe_vs_Ce_DGc, in the top panel corresponding to Digestive Gland (DG) samples, and Gills (G) in the bottom (Ce_Ge_vs_Ce_Gc). The contrast condition and reference/control samples are shown as Ce DGe/Ce Ge and Ce DGc/Ce Gc, respectively.

Although some known stress and detoxing-DSTs related-genes were detected (Table 2), those 16 common up-regulated DEGs in the exposed conditions are related to immune response (immunoglobulin-like and macro domain-like), enzymes, integral components of the membrane, unknown and/or poorly characterized elements (Figure 2, Tables S3 and S4). Thus, this subset did not enrich any GO term, as well as another subset of up-regulated DEGs (Table S3), by applying the hypergeometric distribution significance test (one-tailed Fisher's exact test, $FDR < 0.05$). On the contrary, down-regulated DEGs showed significant differences in the enrichment analyses (Figure 4, Table S5), resulting in 110 enriched GO terms for the subset of 189 common DEGs with a similar down-regulated pattern identified in both conditions exposed to DSTs (189_common_DEGs), only one for those unique down-regulated DEGs identified in the digestive gland (168_unique_DEGs_DG), and 176 enriched GO terms for the 456 unique down-regulated DEGs identified in the gills (456_unique_DEGs_G).

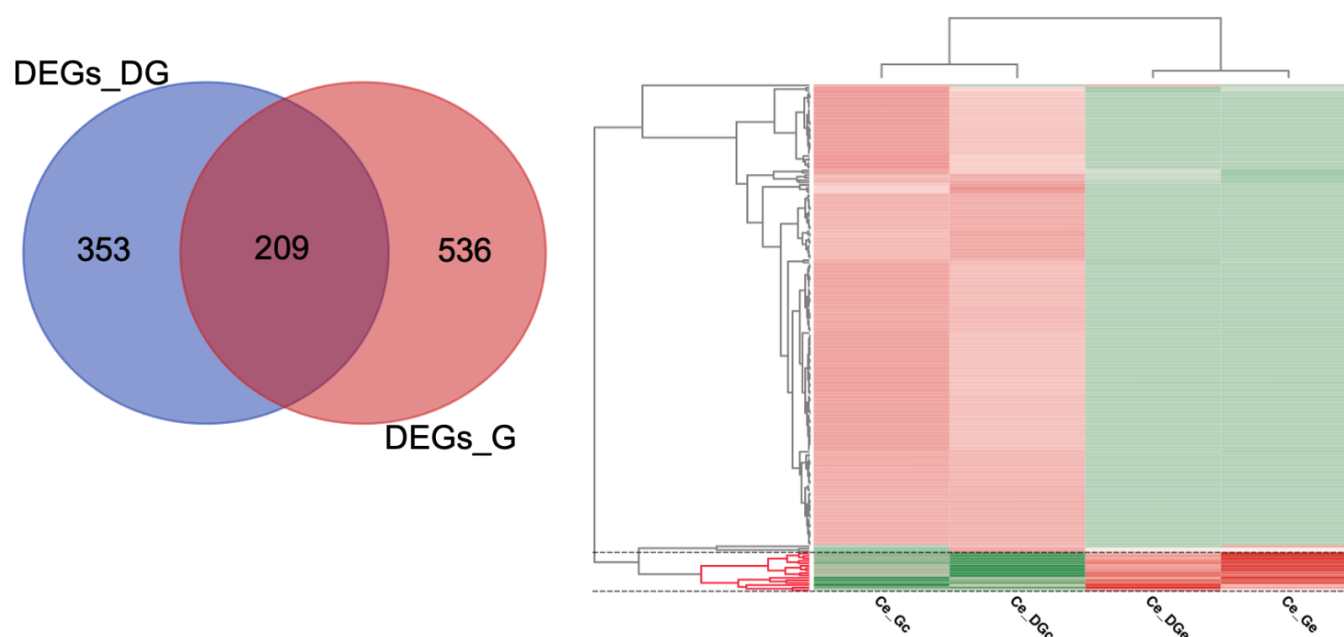


Figure 3. Venn Diagram of the Differentially Expressed Genes (DEGs) and heat map of the shared DEGs between pairwise comparisons. The figure shows in the left panel unique and overlapped DEGs between the tested conditions *Ce_DGe_vs_Ce_DGc*, and *Ce_Ge_vs_Ce_Gc*, corresponding to Digestive Gland (DEGs_DG) samples, and Gills (DEGs_G). The contrast condition and reference/control sample are shown in the columns of the heat map (right panel) as *Ce_DGe/Ce_Ge* and *Ce_DGc/Ce_Gc*, respectively. The heat map depicts the corresponding expression (Z-score values) of the 209 common DEGs found, of which a cluster of up-regulated genes in the exposed condition, comprising 16 transcripts highlighted (in the bottom between dashed lines). Venn diagram was drawn with an online free tool, available at the webserver of the Bioinformatics and Evolutionary Genomics Center (BEG/Van de Peer Lab site, Ghent University, Belgium, <http://bioinformatics.psb.ugent.be/webtools/Venn/> accessed on 15 July 2021).

Among the top-thirty enriched GO terms highlights: binding, cellular process, metabolic process, catalytic activity, intracellular anatomical, intracellular organelle, and biosynthetic process as common GO terms. Some terms related to non-membrane-bounded organelle and macromolecule biosynthetic process, are enriched in the subset of 189 common DEGs. In contrast, small molecule binding, nucleotide binding, nucleoside phosphate binding and anion binding, are exclusive for the subset of 456 unique DEGs found in the gill's comparisons (Figure 4). The complete list of DEGs and the results of the Venn diagram, listing unique and shared DEGs among samples are provided in Tables S3 and S4, respectively. The enriched GO-terms found for each subset of DEGs, and the names of the related genes, can be found in Table S5. Additional files related to the gene expression and enrichment analyses can be found in the data availability section (DATASET S3–S5).

Table 2. List of Differentially Expressed Genes (DEGs) related to known biomarkers involved in the shellfish molecular response after DSTs exposure.

Functional ¹ Categories	Gene/ Transcript ²	Gene/Transcript Description ³	Gene/Transcript ⁴ GO/IP Description	Organism ⁵ (Accession)	DEGs Subset ⁶ (UP-Regulated/Down-Regulated) LFC ⁷					
					Unique (DG) ^{6.1} 185 UR	Unique (DG) ^{6.2} 168 DR	Common 16 UR ^{6.3}	Common 189 DR ^{6.4}	Unique (G) ^{6.5} 80 UR	Unique (G) ^{6.6} 456 DR
Antioxidant system	TRINITY_DN158 227_c0_g1	Extracellular superoxide dismutase [Cu-Zn]	Superoxide metabolic process; metal ion binding	<i>Fasciolopsis buski</i> (KAA019 7766.1)		-8.15				
	TRINITY_DN250 845_c0_g1	Extracellular superoxide dismutase [Cu-Zn]		<i>Fasciolopsis buski</i> (KAA019 7766.1)		-8.15				
	TRINITY_DN57 639_c0_g2	Superoxide dismutase [Cu-Zn]		<i>Elysia chlorotica</i> (RUS8 3892.1)						-9.14
	TRINITY_DN54 684_c0_g1	Glutathione peroxidase	Response to oxidative stress; glutathione peroxidase activity	<i>Clonorchis sinen- sis</i> (GAA50772.1)						-7.91
	TRINITY_DN1 482_c1_g1	Glutathione hydrolase 1 proenzyme-like	Glutathione catabolic process glutathione hydrolase activity	<i>Pecten max- imus</i> (XP_03375 2752.1)	2.84					
	TRINITY_DN7 416_c0_g1	Glutathione S-transferase Mu 3	Glutathione metabolic process; protein binding	<i>Alligator mississip- piensis</i> (XP_00626 0118.1)						-8.29
	TRINITY_DN8 990_c0_g1	Putative glutathione- specific γ-glutamyl cyclotransferase 2	Glutathione catabolic process; γ-glutamyl cyclotransferase activity	<i>Octopus vulgaris</i> (XP_02963 7948.1)	3.46					
	TRINITY_DN9 032_c0_g3	Cytochrome P450 2U1-like	Monoxygenase activity; iron ion binding;	<i>Mizuhopecten yessoen- sis</i> (XP_02137 3153.1)	3.66					
Metabolic detoxification	TRINITY_DN2 028_c0_g1	Steroid 17-α-hydroxylase /17,20 lyase-like	oxidoreductase activity Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen; heme binding	<i>Pecten maximus</i> (XP_03376 3830.1)			5.2			
	TRINITY_DN12 127_c0_g1	ABC transporter G family member 21-like isoform X1	ATPase-coupled transmembrane transporter activity	<i>Pomacea canalicu- late</i> (XP_02508 9510.1)	6.64					
	TRINITY_DN 86_c1_g1	Phosphatidylcholine translocator ABCB4-like	ATP binding; xenobiotic transmembrane transporter activity	<i>Mizuhopecten yessoen- sis</i> (XP_02136 7043.1)	3.18					

Table 2. Cont.

Functional Categories ¹	Gene/Transcript ²	Gene/Transcript Description ³	Gene/Transcript GO/IP Description ⁴	Organism ⁵ (Accession)	DEGs Subset ⁶ (UP-Regulated/Down-Regulated) LFC ⁷					
					Unique (DG) ^{6.1} 185 UR	Unique (DG) ^{6.2} 168 DR	Common 16 UR ^{6.3}	Common 189 DR ^{6.4}	Unique (G) ^{6.5} 80 UR	Unique (G) ^{6.6} 456 DR
Transcription	TRINITY_DN70033_c0_g1	KH domain protein	RNA processing; DNA binding; RNA binding	<i>Opisthorchis felineus</i> (TGZ67600.1)				-8.01		
	TRINITY_DN22122_c0_g1	Cell division cycle and apoptosis regulator protein 1	regulation of transcription, DNA-templated	<i>Paragonimus westermani</i> (KAA3674315.1)				-9.21		
	TRINITY_DN54844_c0_g2	CREB-binding protein	regulation of transcription, DNA-templated; protein binding	<i>Schistosoma japonicum</i> (TNN09875.1)						-7.39
	TRINITY_DN116751_c0_g1	Transcription elongationregulator 1	FF domain; protein-protein interaction	<i>Paragonimus westermani</i> (KAA3677206.1)				-8.26		
	TRINITY_DN9687_c0_g1	Tubulin α chain testis-specific	structural constituent of cytoskeleton; GTP binding; microtubule	<i>Fasciola hepatica</i> (KAA3677206.1)				-8.65		
cytoskeleton, actin related proteins	TRINITY_DN9687_c1_g2	Tubulin α -chain	microtubule	<i>Fasciola hepatica</i> (THD25403.1)						-8.48
	TRINITY_DN181044_c0_g1	Actin depolymerizing factor	actin filament depolymerization; actin binding	<i>Opisthorchis viverrine</i> (XP_009175913.1)						-7.5
	TRINITY_DN34498_c0_g1	Actin-1 isoform X2	actin family	<i>Rattus norvegicus</i> (XP_002726247.1)						-7.08
	TRINITY_DN613_c1_g1	VChain V, beta-actin	ATP binding; cytoplasm; cytoskeleton	<i>Sus scrofa</i> (pdb 5NW4)				-11.08		
	TRINITY_DN6932_c0_g1	Actin β/γ 1	Actin family	<i>Clonorchis sinensis</i> (GAA57823.1)				-11.81		
Immune response	TRINITY_DN7730_c0_g1	Actin β/γ 1		<i>Opisthorchis felineus</i> (TGZ74009.1)				-5.42		
	TRINITY_DN10313_c0_g2	Complement C1q-like protein		<i>Sinonovacula constricta</i> (QCO31675.1)		-3.83				
	TRINITY_DN1050_c0_g1	Complement C1q tumor necrosis factor-related protein 4-like Putative C1q domain	Protein binding	<i>Octopus vulgaris</i> (XP_029647271.1)	8.06					
	TRINITY_DN14010_c0_g1	containing protein MgC1q83		<i>Mytilus galloprovincialis</i> (CBX41732.1)						-8.38
	TRINITY_DN1909_c0_g2	Complement C1q-like protein 2		<i>Lates calcarifer</i> (XP_018543622.1)	3.91					
	TRINITY_DN45102_c0_g1	Complement C1Q-like protein		<i>Pecten maximus</i> (XP_033749093.1)		-5.41				

Table 2. Cont.

Functional Categories ¹	Gene/Transcript ²	Gene/Transcript Description ³	Gene/Transcript GO/IP Description ⁴	Organism ⁵ (Accession)	DEGs Subset ⁶ (UP-Regulated/Down-Regulated) LFC ⁷					
					Unique (DG) ^{6.1} 185 UR	Unique (DG) ^{6.2} 168 DR	Common 16 UR ^{6.3}	Common 189 DR ^{6.4}	Unique (G) ^{6.5} 80 UR	Unique (G) ^{6.6} 456 DR
Heat shock protein HSP 90 (HSP90)	TRINITY_DN16801_c0_g1	Hsp90 protein	Protein folding; ATP binding; ATP hydrolysis activity; unfolded protein binding	<i>Opisthorchis viverrine</i> (XP_009172981.1)				−8.31		
	TRINITY_DN260239_c0_g1	Hsp90 co-chaperone Cdc37 isoform 1	Protein kinase binding	<i>Schistosoma japonicum</i> (AAP06244.1)				−10.24		−7.99

¹ Major Functional Categories related to known proteins/biomarkers involved in dinophysistoxins (DSTs) metabolism and detoxification. ² Sequence name of the DEGs identified within the corresponding Major Functional Category. ³ Blast top-hit description of the corresponding DEGs. ⁴ Gene Ontology (GO)/InterProScan (IPS) description of the corresponding DEG (Gene/transcript). ⁵ Organism and accession number of the corresponding Blast top-hit obtained for each DEGs. ⁶ Subsets of unique and overlapped DEGs (used as test set in the enrichment analyses), identified in the digestive gland and gills of the cockle *C. edule*, exposed to DSTs, where contrast condition and reference/control sample were Ce_DGe/Ce_Ge and Ce_DGc/Ce_Gc, respectively. Subsets: ^{6.1} 185 unique up-regulated DEGs identified in the digestive gland (DG) (unique DG 185 UR, Ce_DGe_vs_Ce_DGc); ^{6.2} 168 unique down-regulated DEGs identified in the digestive gland (unique DG 168 DR, Ce_DGe_vs_Ce_DGc); ^{6.3} 16 common DEGs with a similar up-regulated pattern identified in both conditions exposed to DSTs (common 16 UR); ^{6.4} 189 common DEGs with a similar down-regulated pattern identified in both conditions exposed to DSTs (common 189 DR); ^{6.5} 80 unique up-regulated DEGs identified in the gills (Ce_Ge_vs_Ce_Gc, unique (G) 80 up-regulated); ^{6.6} 456 unique down-regulated DEGs as test set (file: test_set_456_unique_DEGs_down_Ce_Ge_vs_Ce_Gc). ⁷ The log-fold change (M) and the absolute value of the difference in expression obtained with NOISeq R/Bioc package between conditions: Ce_DGe_vs_Ce_DGc, in the Digestive Gland (DG), and Ce_Ge_vs_Ce_Gc in the Gills (G).

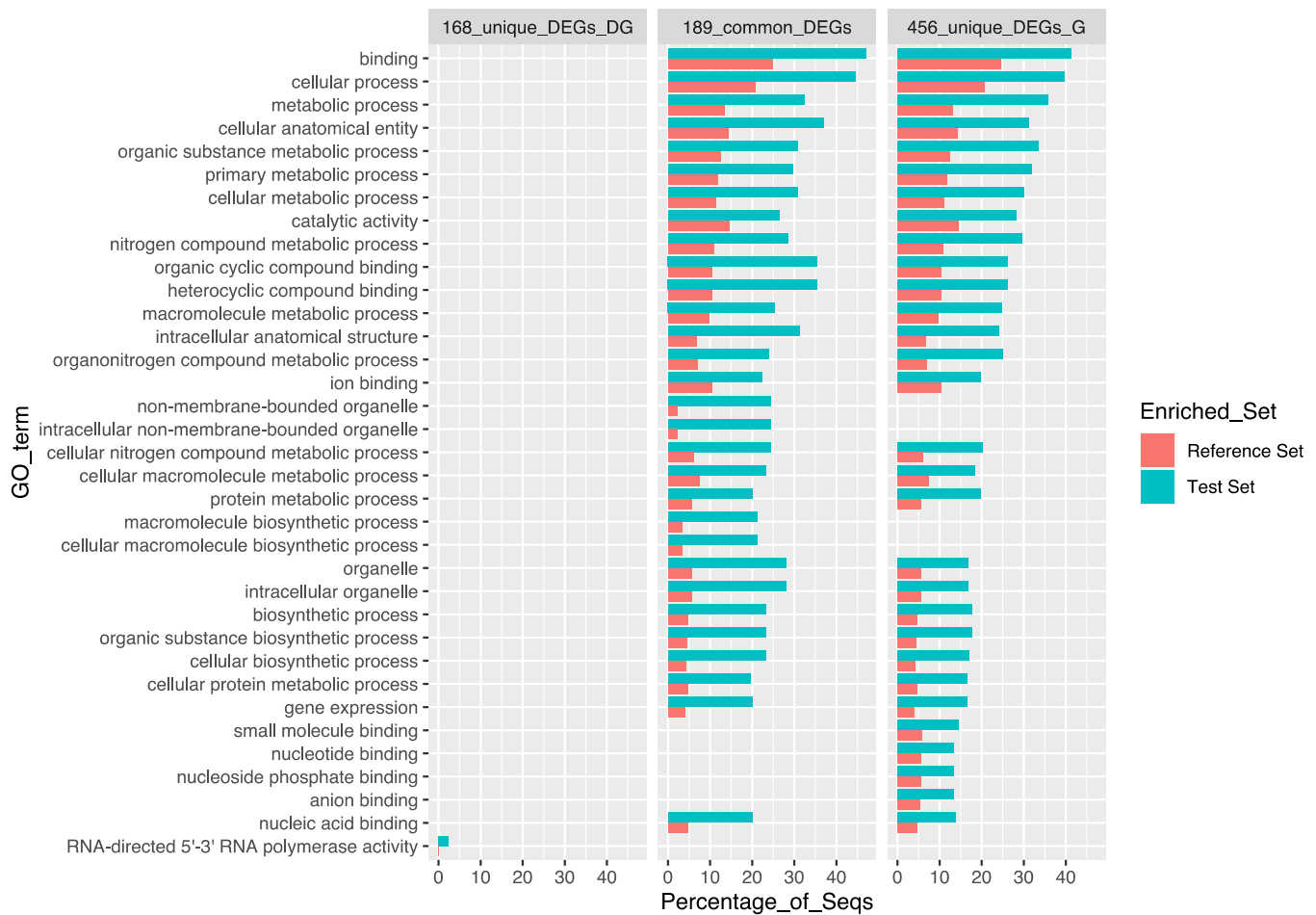


Figure 4. Enrichment analyses performed for the Differentially Expressed Genes (DEGs). The top-thirty enriched GO terms obtained with one-tailed Fisher's exact test ($FDR < 0.05$) are represented by a stacked bar (in green) as the percentage of the GO terms found in the Reference Set (in red), for each subset of DEGs used as Test Set: 189 common DEGs with a similar down-regulated pattern identified in both condition exposed to DSTs (189_common_DEGs), unique down-regulated DEGs identified in the digestive gland (168_unique_DEGs_DG), and 456 unique down-regulated DEGs identified in the gills (456_unique_DEGs_G). The tested conditions in the pairwise analyses were Ce_DGe_vs_Ce_DGc and Ce_Ge_vs_Ce_Gc, corresponding to Digestive Gland (DEGs_DG) samples, and Gills (DEGs_G), where the contrast condition and reference/control samples were Ce_DGe/Ce_Ge and Ce_DGc/Ce_Gc, respectively.

3. Discussion

3.1. Common DEGs and Known Biomarkers Facing DSTs

Herein, we identified 1098 unique/non-redundant DEGs, with 209 DEGs shared between comparisons (Figure 3), with a higher number of 745 DEGs in the gills (Ce_Ge_vs_Ce_Gc) compared to the 562 DEGs found in the digestive gland (Ce_DGe_vs_Ce_DGc) (Figure 2, Tables S3 and S4). A higher number of DEGs have been previously reported in the gills in similar studies assessed in the mussel *Mytilus galloprovincialis* [44]. However, this behavior is likely non-tissue dependent, since the number of DEGs and the Differentially Abundant/Expressed Proteins found have been variable at transcriptomic [28] and proteomic [26] levels, after mussels' exposure to DSTs and their derivatives (i.e., OA, DTX1, DTX2 and DTX3). Nonetheless, some genes expression could be tissue-dependent once 353 DEGs were exclusive for the digestive gland comparisons, whereas 536 unique DEGs were found in the gills (Figure 3).

Although 98% of the common DEGs showed a similar expression pattern, only 7.65% (16 DEGs) were up-regulated in the exposed condition of both comparisons (Ce_DGe_vs_Ce_DGc and Ce_Ge_vs_Ce_Gc) (Figure 3, Table 2). This subset includes DEGs involved in immune response, enzymes, and poorly characterized genes/transcripts. Among the 16 common up-regulated DEGs in the exposed conditions of both comparisons (Figure 3), highlights the transcript/gene TRINITY_DN2028_c0_g1 with homology to the CYP450 (Table 2, Table S3). The family of CYP450 enzymes with oxidoreductase/monooxygenase activity has been recognized as an important biomarker involved in metabolic detoxification, which has been found as up-regulated after DSTs exposure in shellfish [9,13,21,45,46]. In addition, another transcripts/gene TRINITY_DN9032_c0_g3 with homology to a CYP450 2U1-like (XP_021373153.1) from *Mizuhopecten yessoensis* was also found as up-regulated in the digestive gland (Table 2).

Likewise, DEGs involved in immune response were relatively abundant in the subset of 16 common up-regulated DEGs, denoting a possible immunological response of the clam to the exposure and ingestion of toxic alga and/or parasites (Table 2). The transcript/genes TRINITY_DN909_c0_g1, showed homology to neural cell adhesion molecule 1-like from the oyster *Crassostrea gigas* (XP_034299592.1) and TRINITY_DN802_c0_g1 with nephrin-like from *M. yessoensis* (XP_021372895.1), which contain immunoglobulin-like domains. Besides, another DEGs (TRINITY_DN3944_c0_g2), showed an Immunoglobulin-like domain, identified throughout IPS accession (IPR007110). These domains are often involved in interactions, commonly with other Ig-like domains via their β -sheets and can be found in different types of receptors like T-cell antigen receptors, cell adhesion molecules, MHC class I and II antigens, as well as the hemolymph protein hemolin [47–51]. Noteworthy, one protein can contain more than one of these types of Ig-like domains, for instance: Immunoglobulin C-2 Type, Ig_3 domain and cytokine receptor motif NF3 (fibronectin type 3 domain), C1-set (constant-1; IPR003597), C2-set (constant-2; IPR008424) and I-set (intermediate; IPR013098) [52].

Interestingly, hemolin, formerly known as bacteria-induced protein P4 [53], belongs to the immunoglobulin superfamily, considered one of the first hemolymph components to bind to the bacterial surface [53]. This protein has been mainly characterized in insects [54–57] within the major bacteria-inducible hemolymph protein [57], but hemolin also occurs in crustaceans and mollusks [58,59] as part of the host defense mechanisms [60]. An hemolin-related molluscan defense molecule, with five Ig-like domains was identified as down-regulated in *Lymnaea stugnalis*, when infected by the schistosome parasites [59]. However, this report is inconsistent with our findings, since most of the DEGs related with parasites were down-regulated (Figure 2, Tables 2 and S3), suggesting less abundance in the exposed condition, where hemolin showed higher expression (Figure 3, Table S3). Only the DEGs TRINITY_DN842_c2_g1 within this subset, with homology to a C-Jun-amino-terminal kinase-interacting 4-like isoform X6 from *Schistosoma japonicum* (TNN04577.1) was up-regulated. Although the population of parasites detected in *C. edule* is variable among seasons in Ria de Aveiro, their abundances lack significant differences [61]. Likewise, a higher mean temperature of previous months can negatively affect the prevalence of parasites in *C. edule* [62]. In addition, two transcripts/genes matching to heat shock proteins HSP 90 (HSP90) from parasites, were down-regulated, being one of them commonly found within the 189 shared DEGs in both tissues (Table 2). In general, the probability of detecting a trematode-related genes as down-regulated within DEGs suggests a less infestation of those parasites in the exposed condition (sampled in May), than control (March).

In addition, other elements related to immune response were identified among DEGs (Table 2). Although some of them have been previously identified in response to DSTs in bivalves [13], others could potentially belong to parasites, as well as DEGs involved in the transcription regulation (Table 2). However, any of the matched parasites have been reported within the list of species that use *C. edule* as their first and/or second intermediate host [61,62]. Likewise, the best-hit species distribution could be biased by the annotation method using the blastX program (homology-based), and the availability of homologues

sequences in the public database. Some of these DEGs involved in immune response include five transcript/genes related to complement C1q (Table 2). The complement components C1q, such as mgC1q83 and mgC1q29 were significantly down-regulated in the transcriptomic analyses of the digestive gland of the mussel *Perna viridis*, after 6-h exposure to DSTs-producing dinoflagellate *P. lima* [63]. In contrast, we found four transcripts differentially expressed in the digestive gland, of which two DEGs (TRINITY_DN10313_c0_g2 and TRINITY_DN45102_c0_g1) were down-regulated, and two (TRINITY_DN1050_c0_g1 and TRINITY_DN1909_c0_g2) up-regulated (Table 2). Besides, we also found a gene/transcript (TRINITY_DN14010_c0_g1) with homology to a C1q domain-containing protein MgC1q83 from *M. galloprovincialis* (CBX41732.1) in the gills (Table 2). This DEG was significantly down-regulated in the gills, having the same blast-hit (CBX41732.1) of the DEGs reported as a C1q genes down-regulated in the digestive gland of *P. viridis* [63].

Herein, we found that both organs studied can produce these components under immune stress conditions, but with some differences regarding the isoform distribution and the up-down-regulation pattern among DEGs between conditions/tissues (Table 2). The C1q family is involved in various known immune responses, such as pathogen recognition, activation of the complement system, mediating cell migration, and apoptotic cell clearance [64,65]. In the Mediterranean mussel *M. galloprovincialis*, MgC1q plays a role in the recognition of pathogens as part of their innate immune response [66]. Thus, considering their important role and versatility [63,65,66], the expression of some transcripts/isoforms could be tissue-dependent, since any of the DEG related to C1q complement were shared among conditions, being negatively affected in some cases such as MgC1q83 (CBX41732.1), but also activated in the digestive gland (Table 2).

In addition, other biomarkers involved in the antioxidant system and metabolic detoxification were detected (Table 2). Three down-regulated DEGs (two in the digestive gland and one in the gills) showed homology to SOD (Table 2). The transcripts/gene TRINITY_DN7416_c0_g1 with homology to a GST Mu 3 (XP_006260118.1) was down-regulated in the gills, as well as the DEGs TRINITY_DN54684_c0_g1 matching to a GPx enzyme (Table 2). On the contrary, two DEGs (TRINITY_DN1482_c1_g1 and TRINITY_DN8990_c0_g1) functionally related to the glutathione catabolic process, glutathione hydrolase activity and γ -glutamylcyclotransferase activity, were up-regulated in the digestive gland. Similarly, two DEGs related to ABC transporters' family genes were up-regulated in the digestive gland (Table 2). Altogether, these antioxidant elements (such as SOD, CAT, GST, GPx), as well as detoxifying enzymes like CYP450, and the ABC transporters family, play an active role in the first line of defense against DSTs in bivalves' mollusks [20–24].

Previous studies also found GST as down-regulated, suggesting a main role during short-term exposure of DSTs, as proposed for ABC transporters [63]. However, other GST-related/Phase II detoxification enzymes were up-regulated in the digestive gland as well as some members of the ABC transporters family genes. These group of genes seem to remain up-regulated/activated in a longer exposure period to catalyze the conjugation of glutathione to the DSTs and derivatives for its further elimination through the ABC transporter (ABCB10) system and/or multidrug resistance-associated protein (ABCC1, ABCG) [63,67]. Nonetheless, the up-regulation of some ATP-binding cassette (ABC) transporters family genes could be relevant in the mechanism of bivalve's resistance in a longer exposure period, since they can play a variety of roles in xenobiotic transmembrane transporter activity and detoxification.

Components of the cytoskeleton, microtubule constituents, and actin-related genes were also affected (Table 2, Table S3). Indeed, a total of seven DEGs were found as down-regulated in both tissues, comprising actin-1 isoform X2, tubulin α chain, actin depolymerizing factor, VChain V/ β -actin, and actin β/γ 1 (Table 2). Damage in the cytoskeleton and/or destabilization in its elements like actin and tubulin have been corroborated as truthful markers of DST effects in bivalves [23,63,67]. Indeed, TUBA1C and TUBB1 were significantly down-regulated in the gills of the mussel *P. viridis* after a short-time exposure of

P. lima [67]. The histological study revealed the branchial filament structure completely deformed, and the cells atrophied into strips with irregular arrangement during the short-term exposure. However, the negative effect to the gill decreased, and the gill was restored after longer exposure (96 h), associated with a significant increase in the expression of TUBA1C and TUBB1 transcripts [67]. This behavior was not observed in this study since those genes TUBA1C and TUBB1 were not detected within our list of DEGs. However, the down-regulation of cytoskeleton-related genes in the exposed condition can be explained by the recovery after the long-term DSTs exposure.

3.2. Enriched Pathways and Degs Involved

The lack of GO terms associated with these 16 DEGs up-regulated genes may difficult the unbiased detection of enriched pathways in this subset. In fact, only seven genes/transcript (43.75%) had GO annotations, and four DEGs (TRINITY_DN961_c0_g1, TRINITY_DN9102_c0_g1, TRINITY_DN768_c3_g1, TRINITY_DN13635_c0_g1) lacked any Blast hits, IPS results, and Pfam/conserved domains (Figure 2, Tables S3 and S4).

Similarly, only one GO term RNA-directed 5'-3' RNA polymerase activity (GO:0003968) was enriched in the subset of the 168 unique down-regulated DEGs found in the digestive gland (Figure 4, Table S5). Four transcript/genes TRINITY_DN71766_c0_g1, TRINITY_DN126419_c0_g1, TRINITY_DN7379_c0_g1 and TRINITY_DN18481_c0_g1, included within the test set of DEGs are related to flavivirus-like protein and/or polyprotein/polyprotein-4 (Table S3). These transcripts/genes show domain of enzymes (RNA-dependent RNA polymerase, RdRp), which are essential protein encoded in the genomes of all RNA containing viruses with no DNA stage [68,69]. These transcripts/genes are putatively derived from the non-exposed samples (used as control, sampled in March), containing retroviruses, seemingly single stranded RNA viruses (Figure S5).

Apart from this unique over-represented GO term, the other 189 shared down-regulated DEGs from the digestive gland, enriched most of the GO terms found in the subset of 456 unique down-regulated DEGs from the gills (Figure 4, Table S5). Indeed, many overrepresented GO terms were found in both subsets of shared DEGs (168_unique_DEGs_DG), and unique down-regulated DEGs from the gills (456_unique_DEGs_G) (Figure 4, Table S5). Although most of the over-represented pathways are common in the abovementioned subsets of DEGs (189 shared down-regulated DEGs and 456 unique down-regulated DEGs in the gills), some GO terms are exclusive for each group (Figure 4).

For instance, some enriched GO terms like binding (GO:0005488), cellular process (GO:0009987), metabolic process (GO:0008152), among others, display a similar enrichment pattern in the subsets of the common down-regulated DEGs and unique from the gills (Figure 4). Their absent in the exclusive down-regulated DEGs in the digestive gland (168_unique_DEGs_DG), suggest that both tissues can play similar function in non-exposed/exposed DSTs condition (without excluding other environmental factors (e.g., light, temperatures, parasites). Nonetheless, the gills possess some additional DEGs that can play similar roles related to those GO terms, except for non-membrane-bounded organelle (GO:0043228), intracellular non-membrane-bounded organelle (GO:0043232), macromolecule biosynthetic process (GO:0034645) and cellular macromolecule biosynthetic process (GO:0034645). Noteworthy, the GO terms small molecule binding (GO:0036094), nucleotide binding (GO:0000166), nucleoside phosphate binding (GO:1901265) and anion binding (GO:0043168), are exclusively over-represented in the subset of the 456 down-regulated DEGs in the gills (456_unique_DEGs_G) (Figure 4).

As expected, a high percentage of the down-regulated DEGs contained in those common 23 GO terms (Figure 4, Table S6), are redundant among those GO terms (Table S6). Besides, this redundancy is also observed when overlap the GO terms associated with the resulting DEGs groups from the Venn Diagram (Table S6). The “pleiotropic” behavior of the common DEGs may be related with the active role played in several basic function of great relevance to response the environmental changes such as DSTs blooms. Indeed, some terms like ribosome, transcriptions, HSPs, cytoskeleton, histones and glycolysis were more

frequent among the DEGs in this subset (Table S6). Similar redundancy and functional roles were observed with the GO terms found in the subsets of common downregulated DEGs, which are also absent in the subset of unique DEGs in the digestive gland and gills (Figure 4). Likewise, the uniquely over-represented GO terms in the subset of the 456 down-regulated DEGs in the gills (small molecule binding, nucleotide binding, nucleoside phosphate binding and anion binding), are mainly involved in similar functions and putatively related to parasites as well (Tables S3 and S5).

4. Remarks

Herein, we performed a transcriptomic analysis of two tissues (gills and digestive gland) from wild cockles sampled in Ria de Aveiro, Portugal, in two different seasons (with and without DSTs). We found that some biomarkers previously reported among known stress and detoxing-DSTs related-genes were identified within DEGs such as: GPx, GST Mu 3, SOD and CYP450 (involved in antioxidant system), ABC transporter family members (metabolic detoxification/xenobiotic transmembrane transporter activity), apoptosis regulator protein 1 and CREB-binding protein (transcription regulators), actin 1 isoform X2, actin β/γ 1 and tubulin α chain (cytoskeleton, actin-related proteins), HSP90, and complement C1Q-like proteins (immune responses).

The detection of some known biomarkers involved in metabolic detoxification was consistent with previous studies, being CYP540 up-regulated in the exposed conditions of both tissues analyzed, and two members of the ABC transporters family genes up-regulated in the digestive gland. Other biomarkers, such as enzymes, cytoskeleton components, and C1q component mgC1q83 showed a similar pattern as previously described. However, we found that C1q family gene expression could be more complex than previously known, seemingly tissue-dependent considering the isoforms diversity, distribution, and their down and up-regulation pattern. The isoforms of the C1q components found within DEGs should be particularly further studied since this C1q is recognized as the target recognition protein of the classical complement pathway concerning various immune responses. Although some of the DEGs found have been previously reported in shellfish after DST exposures, appearing as consistent biomarkers, their expression should not be only associated with DSTs, since we analyzed environmental samples. For instance, some enzymes in the oxidative stress response can be affected by high temperatures in mussels [70,71]. Similarly, the differential expression of the genes involved in the immune response can be induced by the DSTs concentration, but also influenced by the presence of parasites, bacteria, symbionts, and viruses. Noteworthy, some of these biomarkers ascribed to bivalves' responses facing DST exposures involved in immune responses, antioxidant system, regulation of transcription, detoxification could also be found differentially expressed in both *C. edule* and putative parasites. This behavior in which parasites have been well represented within the annotation of mussels' transcriptome was previously observed [63]. Hence, some genes should be carefully reviewed to avoid false conclusions whether some DEGs/known biomarkers belong to parasites or to the host.

To the best of our knowledge, this study provides the first transcriptomic profile of *C. edule*, giving new insights about its molecular responses facing DSTs exposure in a different environmental condition. Besides, the data generated and publicly shared constitute a valuable resource for bioinformatics analyses to identify sequences signatures in bio-transforming enzymes/proteins, allowing the design of specific primers for further in vitro assays such as RT-PCR targeting known and putative novel DSTs-biomarkers. Moreover, the transcriptomic profile can be used to create a custom protein database for high-throughput proteogenomic analyses to evaluate and/or validate the DEGs/DSTs-biomarkers expressions at the proteomic level. Altogether, our findings and the data provided could help to improve or address new depuration in the shellfish industry.

5. Materials and Methods

5.1. Sampling, RNA Extraction and RNASeq

Specimens of *C. edule* were retrieved from the wild in Ria de Aveiro (Canal de Mira, coordinates: 40.62713495410355, -8.739629051658452) in two different seasons of the year 2019 for comparison purposes. Conditions were established as without DSTs (control) and contaminated (exposed to DSTs) samples considering the warnings reports released by the Instituto Português do Mar e da Atmosfera (IPMA) (available at <http://www.ipma.pt/pt/bivalves/index.jsp> accessed on 20 March 2019, and 18 May 2019). The control samples corresponding to the season without risk of contamination with DSTs were collected on 22 March 2019. On this first sampling day, and according to IPMA phytoplankton records, the abundance (cells/L) of HAB species was below detection limits. Contaminated samples were sampled on 20 May 2019, when DSP toxin-producing species abundance levels attained 360 cells/L. Amnesic shellfish poisoning (ASP) toxin producers and yessotoxin-producing organisms were also detected, although in a much lower relative amount (taking into account regulated values).

Firstly, cockles were hand dredged during low tide at the harvest location in Ria de Aveiro, and then the specimens were transported in ice to the lab for subsequent tissues sampling. Small pieces (1 cm³) of gills (G) and digestive gland (DG) from five individuals of both conditions (control and treated) were pooled per tissue and condition, stabilized in RNAlater solution, and stored at 20 °C until used. The soft resting parts of the specimens were pooled and conducted to LC-MS/MS analysis for dinophysistoxins quantification.

5.2. DSTs Quantification by LC-MS/MS Analyses

To confirm the amount of DSTs such as okadaic acid (OA) and its analogs dinophysistoxins (DTXs) reported by IPMA in Ria de Aveiro (2019), the remaining soft-parts from the same *C. edule* specimens used for RNAseq, were pooled, and conducted to liquid chromatography with tandem mass spectrometry (LC-MS/MS), following the Standardized Operating Procedure (SOP) of the European Reference Laboratory for Marine Biotoxins (EURLMB) for the determination of marine lipophilic biotoxins in bivalve mollusks [72].

As previously described [26], a 2 g aliquot of cockle tissue homogenate was extracted with 9.0 mL MeOH 100% and homogenized with a Polytron mixer (Kinematica, Maltes, Switzerland). The mixture was centrifuged at 2000 × *g* for 10 min. The supernatant was then collected into a 30 mL tube and the pellet was re-extracted with 9.0 mL MeOH by vortexing for 2 min. After centrifugation, supernatants were combined and made up to 20 mL with MeOH. A 2.5 mL aliquot was the raw methanolic extract was transferred to 10 mL glass tube for alkaline hydrolysis to converting the 7-O-acyl ester derivatives (DTX3) into their parent toxin. The hydrolysis was started by adding 313 µL of 2.5 M NaOH and heated at 76 °C for 40 min in a heating block. The reaction was then neutralized with 313 µL of 2.5 M HCl. After filtered through 0.2 µm syringe filter 5 µL were injected to the LC-MS/MS system.

The LC-MS/MS analysis was performed using on in 1290 Infinity chromatograph (Agilent Technologies, Waldbronn, Germany) coupled to an Agilent 6470 triple quadrupole mass spectrometer. The chromatographic separation was conducted with a Zorbax SB-C8 RRHT column (2.1 × 50 mm, 1.8 µm), protected with a guard column (2.1 × 5 mm, 1.8 µm). The mobile phase A was water with 2 mM ammonium formate and 50 mM formic acid, and the mobile phase B was 95% acetonitrile with 2 mM ammonium formate and 50 mM formic acid. An elution gradient at a flow rate of 0.4 mL/minute was used as follows: 0–3 min, gradient from 88 to 50% eluent A; 3–6.5 min gradient 50 to 10% eluent A; 6.5–8.9 min 10% eluent A; 8.9–10 min, gradient 10 to 88% eluent A. The detection of the toxins was carried out in multiple reaction monitoring (MRM) acquisition mode. Two MRM transitions were monitored in negative polarity, one for quantification and the other for confirmation. OA and DTX2: *m/z* 803 > 255 and *m/z* 803 > 563; DTX1: *m/z* 817 > 255 and *m/z* 817 > 563. A matrix match calibration with five standard solutions ranging in concentration levels from

2.0 to 22.0 ng/mL with a correlation > 0.990 was set up for quantification using certified OA, DTX1 and DTX2 reference standards purchased from CIFGA (Lugo, Spain).

5.3. RNA Extraction and Illumina Sequencing

Preserved samples in RNAlater solution were transposed to 2 mL tubes pre-filled with ceramic beads (Precellys[®] Ceramic kit 1.4 mm, Bertin Technologies, Montigny-le-Bretonneux, France) and then homogenized in homogenization buffer (RLT buffer) using an automatic bead-based homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France) with a custom program (speed: 5500 RPM at 0 °C; 2×20 s cycles; 30 s pause). The supernatant of the two tissues (G and DG) and both conditions, containing a pool from the five biological replicates, were recovered, and conducted for total RNA extraction using Qiagen's RNeasy Mini kit (Venlo, The Netherlands) according to the manufacturer's instructions.

Total RNA concentration and relative quality were determined photometrically with a DeNovix DS-11 spectrophotometer (DeNovix Technologies, Wilmington, DE, USA), and RNA integrity was evaluated with the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Four samples, corresponding to control samples (without toxins)—gills (Gc), digestive gland (DGc)—and contaminated samples (exposed or contaminated with shellfish toxins)—gills (Ge) and digestive gland (DGe)—were submitted to Macrogen, Inc. (Seoul, South Korea) for total RNA sequencing using Illumina TruSeq Stranded Total RNA Library Construction with Ribo-Zero + NovaSeq 6000.

5.4. Transcriptome Assembly and Completeness Assessment

The RNA sequencing on the aforementioned Illumina platform 150 bp (base-pair) paired-end (PE) resulted in around 41 million reads for each sample transcriptome. Transcriptomic and downstream analyses were performed with a set of bioinformatics tools incorporated in the OmicsBox v1.4.11 [41]. Firstly, the quality of the raw reads was assessed using the FastQC software v0.11.8 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> accessed on 13 January 2021) [73]. Then, the original FASTQ files were pre-processed using Trimmomatic v0.38 [74]. The adapters were trimmed for each sample individually, low-quality bases and low-quality reads removed, and only reads longer than 36 bp were retained for further analyses, as previously described [75]. Likewise, the filtered reads across all four samples were combined into a single dataset [75], and de novo assembled with Trinity v2.10.0 [76], using three as the minimum kmer coverage in the strand-specific mode. The assembly quality and completeness were assessed with the Benchmarking Universal Single-Copy Orthologs (BUSCOs, BUSCO v5) [77], searched for metazoans orthologues in transcriptome mode with a cut-off e-value $1.0E-3$ on OrthoDB v10 [78].

5.5. Open Reading Frame Prediction and Transcript Abundance Estimation

Due to the lack of a reference genome for *C. edule*, the resulting de novo transcriptome assembly was used as a reference to perform the transcript expression analysis [75], using individual tools embedded in the Trinity differential expression module [79], and already included in OmicsBox [41]. In order to reduce sequence redundancy, the de novo assembled transcriptome was first submitted to clustering analyses using CD-HIT 4.8.1 [80,81], setting 0.9 as the sequence identity threshold. Then, the clustered transcriptome was scanned for coding regions by TransDecoder v5.5.0 (<http://transdecoder.sf.net> accessed on 3 February 2021) [79,82], and all six-frame translations were filtered for a minimum length of 100 amino acids for open reading frames (ORFs). Simultaneously, the resulting ORFs were scanned for homology to known proteins via Pfam searches (PFAM 32, HMMER 3.2.1) [83,84], and only the best isoform per gene was retained, considering as priority the homology to known proteins domain versus the transcript length [79,82]. The corresponding coding genes for the best isoforms and all their transcript isoforms were used as reference transcriptome templates for the expression analyses. The original pre-processed reads (clean reads without adapters) were mapped back against the reference transcriptome on a per-sample

basis using Bowtie2 v2.3.4.1 [85], followed by calculation of abundance estimates using RSEM [86]. The resulting count tables obtained were used for the Pairwise Differential Expression Analysis (No Replicates) tool, based on the Bioconductor software package NOISeq v2.30.0 [87,88], suited to compare samples from two experimental conditions without biological replicates. The genes without counts across libraries were filtered on a count-per-million basis (CPM), and five technical replicates were simulated per sample. The data were normalized using the trimmed mean of M-values (TMM) normalization method, which was corrected for library size and reduced RNA compositional effect [89]. Finally, the Differentially Expressed Genes (DEGs) were determined (p -value 0.01) between samples exposed to DST (contrast condition) versus samples without toxins (control conditions) from both tissues studied (digestive gland: Ce_DGe_versus_Ce_DGc and gills: Ce_Ge_versus_Ce_Gc).

5.6. Functional Annotation and Enrichment

The best isoforms per coding gene obtained from the clustered transcriptome, previously analyzed against Pfam, were then characterized based on sequence similarity against public databases. Firstly, these transcripts (best isoforms) were searched against the metazoans section of the non-redundant (nr) protein database from NCBI (nr database: <ftp://ftp.ncbi.nlm.nih.gov/blast/db>; accessed on 11 May 2021) and UniProtKB/Swiss-Protproteindatabase (accessed on 12 May 2021) [40], using CloudBlast option with the BlastX 2.10.0+ program setting a cut-off e -value $1.0E-3$ [41]. The transcripts without hits were retrieved and CloudBlasted against the same protein database but excluding metazoans. Protein signatures were also obtained from the InterPro member databases (i.e., Pfam, PROSITE, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, Gene3D, and PANTHER) using the InterProScan (v5.51–85.0) software [43]. All annotation and the corresponding gene ontology (GO) terms were then transferred to the sequences and eventually merged with existing GO terms resulting from mapping, InterProScan, EggNOG (evolutionary genealogy of genes: non-supervised orthologous groups) annotation and KEGG pathways [43], obtained with eggNOG-Mapper 1.0.3 with EggNOG 5.0.0 [42]. The remaining transcripts without annotation were further conducted to Rfam blast search against Xfam servers (<rfam.xfam.org>) [90]. Enrichment analyses to determine over/underexpressed pathways were performed to the categorical GO terms corresponding to the DEGs, excluding those transcripts corresponding to non-metazoans hits, and filtering the remaining GO terms of Bacteria, Virus, and other putative contaminants. The enriched genes/pathways were statistically determined with the hypergeometric distribution significance test (Fisher's exact test, $FDR < 0.05$).

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/toxins13110784/s1>, Figure S1: The de novo assembly quality and completeness assessment with the Benchmarking Universal Single-Copy Orthologs (BUSCOs), Figure S2: The Protein Coding Sequences (CDS) obtained by six-frame translations with TransDecoder v5.5.0., considering a minimum length of 100 amino acids for open reading frames (ORFs) with homology to known proteins via Pfam searches, and the best/longest isoform per gene, Figure S3: Species distribution according to Orthologous Groups (OGs) categories obtained from the eggnog database using the eggNOG-Mapper 1.0.3 with Egg-NOG 5.0.0, Figure S4: Distribution of counts per gene across libraries (based on a count-per-million basis (CPM), Figure S5: Multi-dimensional scaling (MDS) plot of the Differentially Expressed Genes (DEGs), Table S1: Transcripts used for pairwise differential expression analyses, Table S2: Putative contaminants filtered before pairwise differential expression analyses, Table S3: Differentially Expressed Genes (DEGs) obtained from the pairwise comparison between the conditions studied in the transcriptomic profile of the cockle *Cerastoderma edule* exposed to Diarrhetic Shellfish Toxins (DSTs), using NOISeq v2.30.0, with 5 technical replicates simulated per sample and p -value 0.01, Table S4: Venn Diagram results between the Differentially Expressed Genes (DEGs) from the pairwise comparison between the tested condition Ce_DGe_vs_Ce_DGc and Ce_Ge_vs_Ce_Gc, obtained in the transcriptomic analyse of the cockles *Cerastoderma edule*, Table S5: Enrichment analyses performed to the

Differentially Expressed Genes (DEGs) using the hypergeometric distribution significance test (Fisher's exact test, $FDR < 0.05$), Table S6: results of the Venn Diagram performed to 23 common enriched GO terms in the 189 shared downregulated DEGs from the digestive gland, and the subset of 456 unique DEGs downregulated in the gills.

Author Contributions: Conceptualization, D.D.-P., J.C.M. and A.C.; Identification, Sampling and sample preparation, J.C.M.; D.A. and D.D.-P.; writing—original draft preparation, D.D.-P.; J.C.M. and A.C.; P.R.C.; V.V.; review and editing, D.A.; P.R.C. and V.V.; LC-MS/MS analyses, P.R.C.; Bioinformatics analyses, D.D.-P.; funding acquisition, A.C. and V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Portuguese Science Foundation (Fundação para a Ciência e a Tecnologia, FCT) and under the Projects MOREBIVALVES (PTDC/ASP-PES/31762/2017) and by the Strategic Funding UIDB/04423/2020 and UIDP/04423/2020 through national funds provided by FCT and European Regional Development Fund (ERDF), in the framework of the program PT2020.

Institutional Review Board Statement: This study involved bivalve mollusks which, according to the Portuguese and European laws, do not require previous approval by Institutional Review Board or Ethics Committee.

Informed Consent Statement: Not applicable.

Data Availability Statement: The transcriptomic project has been registered with the BioProject database under BioProject ID: PRJNA739261 (<http://www.ncbi.nlm.nih.gov/bioproject/739261>). The pre-processed raw data have been deposited at the NCBI Sequence Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/Traces/sra>), under accessions SAMN19778280 (Ce_DGc, <https://www.ncbi.nlm.nih.gov/biosample/19778280>), SAMN19778281 (Ce_DGe, <https://www.ncbi.nlm.nih.gov/biosample/19778281>), SAMN19778282 (Ce_Gc, <https://www.ncbi.nlm.nih.gov/biosample/19778282>), SAMN19778283 (Ce_Ge, <https://www.ncbi.nlm.nih.gov/biosample/19778283>). The Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GJGL00000000. DATASET S1: The resulting original files obtained from de novo transcriptome assembly and clustering analyses (e.g., .fasta and .gff files, transcript_to_gene_map.txt, summary files), have been deposited at the Mendeley Data repository as DATASET S1 (<http://dx.doi.org/10.17632/vkw2jcmmt.1>). DATASET S2: The resulting .fasta and .gff files corresponding to the open reading frames (ORFs) prediction obtained, as well as summary files have been deposited at the Mendeley Data repository as DATASET S2 (<http://dx.doi.org/10.17632/bgps766mxp.1>). DATASET S3: All data used for the gene expression calculation comprising the count tables, as well as the resulting files obtained from the differential expression analyses, have been deposited at the Mendeley Data repository as DATASET S3 (<http://dx.doi.org/10.17632/6z6rpf9w8z.1>). DATASET S4: The additional files used for enrichment analyses have been deposited at the Mendeley Data repository as DATASET S4 (<http://dx.doi.org/10.17632/5yp5ncm2vp.1>).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Anderson, D.M.; Fensin, E.; Gobler, C.J.; Hoeglund, A.E.; Hubbard, K.A.; Kulis, D.M.; Landsberg, J.H.; Lefebvre, K.A.; Provoost, P.; Richlen, M.L. Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. *Harmful Algae* **2021**, *102*, 101975. [[CrossRef](#)] [[PubMed](#)]
2. Berdalet, E.; Fleming, L.E.; Gowen, R.; Davidson, K.; Hess, P.; Backer, L.C.; Moore, S.K.; Hoagland, P.; Enevoldsen, H. Marine harmful algal blooms, human health and wellbeing: Challenges and opportunities in the 21st century. *J. Mar. Biol. Assoc. UK* **2016**, *96*, 61–91. [[CrossRef](#)] [[PubMed](#)]
3. Bresnan, E.; Arévalo, F.; Belin, C.; Branco, M.A.; Cembella, A.D.; Clarke, D.; Correa, J.; Davidson, K.; Dhanji-Rapkova, M.; Lozano, R.F. Diversity and regional distribution of harmful algal events along the Atlantic margin of Europe. *Harmful Algae* **2021**, *102*, 101976. [[CrossRef](#)] [[PubMed](#)]
4. Lewitus, A.J.; Horner, R.A.; Caron, D.A.; Garcia-Mendoza, E.; Hickey, B.M.; Hunter, M.; Huppert, D.D.; Kudela, R.M.; Langlois, G.W.; Largier, J.L. Harmful algal blooms along the North American west coast region: History, trends, causes, and impacts. *Harmful Algae* **2012**, *19*, 133–159. [[CrossRef](#)]
5. Yñiguez, A.T.; Lim, P.T.; Leaw, C.P.; Jipanin, S.J.; Iwataki, M.; Benico, G.; Azanza, R.V. Over 30 years of HABs in the Philippines and Malaysia: What have we learned? *Harmful Algae* **2021**, *102*, 101776. [[CrossRef](#)]
6. Li, X.-Y.; Yu, R.-C.; Geng, H.-X.; Li, Y.-F. Increasing dominance of dinoflagellate red tides in the coastal waters of Yellow Sea, China. *Mar. Pollut. Bull.* **2021**, *168*, 112439. [[CrossRef](#)] [[PubMed](#)]

7. Hoagland, P.; Anderson, D.M.; Kaoru, Y.; White, A.W. The economic effects of harmful algal blooms in the United States: Estimates, assessment issues, and information needs. *Estuaries* **2002**, *25*, 819–837. [[CrossRef](#)]
8. Van Dolah, F.M. Marine algal toxins: Origins, health effects, and their increased occurrence. *Environ. Health Perspect.* **2000**, *108*, 133–141. [[CrossRef](#)]
9. Reguera, B.; Riobó, P.; Rodríguez, F.; Díaz, P.A.; Pizarro, G.; Paz, B.; Franco, J.M.; Blanco, J. Dinophysins toxins: Causative organisms, distribution and fate in shellfish. *Marine Drugs* **2014**, *12*, 394–461. [[CrossRef](#)]
10. Vale, P.; Botelho, M.J.; Rodrigues, S.M.; Gomes, S.S.; Sampayo, M.A.d.M. Two decades of marine biotoxin monitoring in bivalves from Portugal (1986–2006): A review of exposure assessment. *Harmful Algae* **2008**, *7*, 11–25.
11. Bauder, A.G.; Cembella, A.D.; Bricelj, V.M.; Quilliam, M.A. Uptake and fate of diarrhetic shellfish poisoning toxins from the dinoflagellate *Prorocentrum lima* in the bay scallop *Argopecten irradians*. *Marine Ecol. Prog. Ser.* **2001**, *213*, 39–52. [[CrossRef](#)]
12. Shumway, S.E. Phycotoxin-related shellfish poisoning: Bivalve molluscs are not the only vectors. *Rev. Fish. Sci.* **1995**, *3*, 1–31. [[CrossRef](#)]
13. Campos, A.; Freitas, M.; de Almeida, A.M.; Martins, J.C.; Domínguez-Pérez, D.; Osório, H.; Vasconcelos, V.; Reis Costa, P. OMICs approaches in diarrhetic shellfish toxins research. *Toxins* **2020**, *12*, 493. [[CrossRef](#)] [[PubMed](#)]
14. García, C.; Schonstedt, V.; Santelices, J.P.; Lagos, N. High amount of dinophysistoxin-3 in *Mytilus chilensis* collected in Seno de Reloncaví, Chile, during massive human intoxication associated with outbreak of *Vibrio parahaemolyticus*. *J. Toxicol. Sci.* **2006**, *31*, 305. [[PubMed](#)]
15. Yasumoto, T.; Murata, M.; Oshima, Y.; Sano, M.; Matsumoto, G.; Clardy, J. Diarrhetic shellfish toxins. *Tetrahedron* **1985**, *41*, 1019–1025. [[CrossRef](#)]
16. Young, N.; Sharpe, R.A.; Barciela, R.; Nichols, G.; Davidson, K.; Berdalet, E.; Fleming, L.E. Marine harmful algal blooms and human health: A systematic scoping review. *Harmful Algae* **2020**, *98*, 101901. [[CrossRef](#)] [[PubMed](#)]
17. Sukanuma, M.; Fujiki, H.; Okabe, S.; Nishiwaki, S.; Brautigam, D.; Ingebritsen, T.S.; Rosner, M.R. Structurally different members of the okadaic acid class selectively inhibit protein serine/threonine but not tyrosine phosphatase activity. *Toxicon* **1992**, *30*, 873–878. [[CrossRef](#)]
18. Blanco, J.; Mariño, C.; Martín, H.; Acosta, C.P. Anatomical distribution of diarrhetic shellfish poisoning (DSP) toxins in the mussel *Mytilus galloprovincialis*. *Toxicon* **2007**, *50*, 1011–1018.
19. McCarron, P.; Kilcoyne, J.; Hess, P. Effects of cooking and heat treatment on concentration and tissue distribution of okadaic acid and dinophysistoxin-2 in mussels (*Mytilus edulis*). *Toxicon* **2008**, *51*, 1081–1089.
20. Prego-Faraldo, M.; Vieira, L.; Eirin-Lopez, J.; Méndez, J.; Guilhermino, L. Transcriptional and biochemical analysis of antioxidant enzymes in the mussel *Mytilus galloprovincialis* during experimental exposures to the toxic dinoflagellate *Prorocentrum lima*. *Mar. Environ. Res.* **2017**, *129*, 304–315.
21. Wei, X.-M.; Lu, M.-Y.; Duan, G.-F.; Li, H.-Y.; Liu, J.-S.; Yang, W.-D. Responses of CYP450 in the mussel *Perna viridis* after short-term exposure to the DSP toxins-producing dinoflagellate *Prorocentrum lima*. *Ecotoxicol. Environ. Saf.* **2019**, *176*, 178–185. [[CrossRef](#)] [[PubMed](#)]
22. Zou, Y.; Wei, X.-M.; Weng, H.-W.; Li, H.-Y.; Liu, J.-S.; Yang, W.-D. Expression profile of eight glutathione S-transferase genes in *Crassostrea ariakensis* after exposure to DSP toxins producing dinoflagellate *Prorocentrum lima*. *Toxicon* **2015**, *105*, 45–55. [[CrossRef](#)] [[PubMed](#)]
23. Huang, L.; Liu, S.-L.; Zheng, J.-W.; Li, H.-Y.; Liu, J.-S.; Yang, W.-D. P-glycoprotein and its inducible expression in three bivalve species after exposure to *Prorocentrum lima*. *Aquat. Toxicol.* **2015**, *169*, 123–132. [[CrossRef](#)] [[PubMed](#)]
24. Huang, L.; Wang, J.; Chen, W.-C.; Li, H.-Y.; Liu, J.-S.; Jiang, T.; Yang, W.-D. P-glycoprotein expression in *Perna viridis* after exposure to *Prorocentrum lima*, a dinoflagellate producing DSP toxins. *Fish Shellfish Immunol.* **2014**, *39*, 254–262. [[CrossRef](#)]
25. Martínez-Escariuza, R.; Lozano, V.; Pérez-Parallé, M.L.; Blanco, J.; Sánchez, J.L.; Pazos, A.J. Expression Analyses of Genes Related to Multixenobiotic Resistance in *Mytilus galloprovincialis* after Exposure to Okadaic Acid-Producing *Dinophysis acuminata*. *Toxins* **2021**, *13*, 614.
26. Martins, J.C.; Domínguez-Pérez, D.; Azevedo, C.; Braga, A.C.; Costa, P.R.; Osório, H.; Vasconcelos, V.; Campos, A. Molecular Responses of Mussel *Mytilus galloprovincialis* Associated to Accumulation and Depuration of Marine Biotoxins Okadaic Acid and Dinophysistoxin-1 Revealed by Shotgun Proteomics. *Front. Mar. Sci.* **2020**, *7*, 589822. [[CrossRef](#)]
27. Miao, J.; Chi, L.; Pan, L.; Song, Y. Generally detected genes in comparative transcriptomics in bivalves: Toward the identification of molecular markers of cellular stress response. *Environ. Toxicol. Pharmacol.* **2015**, *39*, 475–481. [[CrossRef](#)] [[PubMed](#)]
28. Suarez-Ulloa, V.; Fernandez-Tajes, J.; Aguiar-Pulido, V.; Prego-Faraldo, M.V.; Florez-Barros, F.; Sexto-Iglesias, A.; Mendez, J.; Eirin-Lopez, J.M. Unbiased high-throughput characterization of mussel transcriptomic responses to sublethal concentrations of the biotoxin okadaic acid. *PeerJ* **2015**, *3*, e1429. [[CrossRef](#)]
29. Murgarella, M.; Puiu, D.; Novoa, B.; Figueras, A.; Posada, D.; Canchaya, C. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PLoS ONE* **2016**, *11*, e0151561.
30. Zhang, G.; Fang, X.; Guo, X.; Li, L.; Luo, R.; Xu, F.; Yang, P.; Zhang, L.-L.; Wang, X.; Qi, H.; et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **2012**, *490*, 49–54. [[CrossRef](#)]
31. Sellami, B.; Khazri, A.; Louati, H.; Dellali, M.; Driss, M.R.; Aïssa, P.; Mahmoudi, E.; Hamouda, B.; Coelho, A.V.; Sheehan, D. Effects of anthracene on filtration rates, antioxidant defense system, and redox proteomics in the Mediterranean clam *Ruditapes decussatus* (Mollusca: Bivalvia). *Environ. Sci. Pollut. Res.* **2015**, *22*, 10956–10968. [[CrossRef](#)] [[PubMed](#)]

32. Domínguez-Pérez, D.; Lippolis, J.; Dennis, M.; Miller, B.; Tiley, K.; Vasconcelos, V.; de Almeida, A.M.; Campos, A. The Queen Conch (*Lobatus gigas*) proteome: A valuable tool for biological studies in marine gastropods. *Protein J.* **2019**, *38*, 628–639. [[CrossRef](#)] [[PubMed](#)]
33. Vale, P. Differential dynamics of dinophysistoxins and pectenotoxins between blue mussel and common cockle: A phenomenon originating from the complex toxin profile of *Dinophysis acuta*. *Toxicon* **2004**, *44*, 123–134. [[CrossRef](#)]
34. Rossignoli, A.E.; Fernández, D.; Regueiro, J.; Mariño, C.; Blanco, J. Esterification of okadaic acid in the mussel *Mytilus galloprovincialis*. *Toxicon* **2011**, *57*, 712–720. [[CrossRef](#)] [[PubMed](#)]
35. Vale, P.; Sampayo, M.A.d.M. First confirmation of human diarrhoeic poisonings by okadaic acid esters after ingestion of razor clams (*Solen marginatus*) and green crabs (*Carcinus maenas*) in Aveiro lagoon, Portugal and detection of okadaic acid esters in phytoplankton. *Toxicon* **2002**, *40*, 989–996. [[CrossRef](#)]
36. Suzuki, T.; Mitsuya, T. Comparison of dinophysistoxin-1 and esterified dinophysistoxin-1 (dinophysistoxin-3) contents in the scallop *Patinopecten yessoensis* and the mussel *Mytilus galloprovincialis*. *Toxicon* **2001**, *39*, 905–908. [[CrossRef](#)]
37. Braga, A.C.; Lage, S.; Pacheco, M.; Rydberg, S.; Costa, P.R. Native (*Ruditapes decussatus*) and non-indigenous (*R. philippinarum*) shellfish species living in sympatry: Comparison of regulated and non-regulated biotoxins accumulation. *Mar. Environ. Res.* **2017**, *129*, 147–155.
38. Vale, P.; Sampayo, M.A.d.M. Seasonality of diarrhetic shellfish poisoning at a coastal lagoon in Portugal: Rainfall patterns and folk wisdom. *Toxicon* **2003**, *41*, 187–197. [[CrossRef](#)]
39. Oyaneder-Terrazas, J.; Polanco, C.; Figueroa, D.; Barriga, A.; García, C. In vitro biotransformation of OA-group and PTX-group toxins in visceral and non-visceral tissues of *Mytilus chilensis* and *Ameghinomya antiqua*. *Food Addit. Contam. Part A* **2020**, *37*, 1216–1228.
40. The UniProt Consortium. The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **2021**, *49*, D480–D489.
41. BioBam Bioinformatics. OmicsBox-Bioinformatics Made Easy (Version 1.4.11). Available online: <https://www.biobam.com/omicsbox> (accessed on 3 March 2019).
42. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [[CrossRef](#)]
43. Mitchell, A.L.; Attwood, T.K.; Babbitt, P.C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S.D.; Chang, H.-Y.; El-Gebali, S.; Fraser, M.I. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2019**, *47*, D351–D360. [[CrossRef](#)] [[PubMed](#)]
44. Prego-Faraldo, M.V.; Martínez, L.; Méndez, J. RNA-Seq analysis for assessing the early response to DSP toxins in *Mytilus galloprovincialis* digestive gland and gill. *Toxins* **2018**, *10*, 417.
45. Nielsen, P.; Krock, B.; Hansen, P.J.; Vismann, B. Effects of the DSP-toxic dinoflagellate *Dinophysis acuta* on clearance and respiration rate of the blue mussel, *Mytilus edulis*. *PLoS ONE* **2020**, *15*, e0230176. [[CrossRef](#)]
46. Huang, L.; Zou, Y.; Weng, H.-w.; Li, H.-Y.; Liu, J.-S.; Yang, W.-D. Proteomic profile in *Perna viridis* after exposed to *Prorocentrum lima*, a dinoflagellate producing DSP toxins. *Environ. Pollut.* **2015**, *196*, 350–357. [[CrossRef](#)] [[PubMed](#)]
47. Bork, P.; Holm, L.; Sander, C. The immunoglobulin fold: Structural classification, sequence patterns and common core. *J. Mol. Biol.* **1994**, *242*, 309–320. [[CrossRef](#)]
48. Halaby, D.; Poupon, A.; Mornon, J.-P. The immunoglobulin fold family: Sequence analysis and 3D structure comparisons. *Protein Eng.* **1999**, *12*, 563–571. [[CrossRef](#)]
49. Potapov, V.; Sobolev, V.; Edelman, M.; Kister, A.; Gelfand, I. Protein—Protein recognition: Juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. *J. Mol. Biol.* **2004**, *342*, 665–679.
50. Teichmann, S.A.; Chothia, C. Immunoglobulin superfamily proteins in *Caenorhabditis elegans*. *J. Mol. Biol.* **2000**, *296*, 1367–1383. [[CrossRef](#)]
51. Bodian, D.L.; Jones, E.Y.; Harlos, K.; Stuart, D.I.; Davis, S.J. Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5 Å resolution. *Structure* **1994**, *2*, 755–766. [[CrossRef](#)]
52. Smith, D.K.; Xue, H. Sequence profiles of immunoglobulin and immunoglobulin-like domains. *J. Mol. Biol.* **1997**, *274*, 530–545.
53. Sun, S.-C.; Lindstrom, I.; Boman, H.G.; Faye, I.; Schmidt, O. Hemolin: An insect-immune protein belonging to the immunoglobulin superfamily. *Science* **1990**, *250*, 1729–1732. [[CrossRef](#)]
54. Ladendorff, N.E.; Kanost, M.R. Bacteria-induced protein P4 (hemolin) from *Manduca sexta*: A member of the immunoglobulin superfamily which can inhibit hemocyte aggregation. *Arch. Insect Biochem. Physiol.* **1991**, *18*, 285–300. [[CrossRef](#)] [[PubMed](#)]
55. Schmidt, O.; Faye, I.; Lindström-Dinnetz, I.; Sun, S.-C. Specific immune recognition of insect hemolin. *Dev. Comp. Immunol.* **1993**, *17*, 195–200. [[CrossRef](#)]
56. Aathmanathan, V.S.; Jothi, N.; Prajapati, V.K.; Krishnan, M. Investigation of immunogenic properties of Hemolin from silkworm, *Bombyx mori* as carrier protein: An immunoinformatic approach. *Sci. Rep.* **2018**, *8*, 6957. [[CrossRef](#)]
57. Bettencourt, R.; Lanz-Mendoza, H.; Lindquist, K.R.; Faye, I. Cell adhesion properties of hemolin, an insect immune protein in the Ig superfamily. *Eur. J. Biochem.* **1997**, *250*, 630–637. [[CrossRef](#)] [[PubMed](#)]
58. Watthanasurorot, A.; Jiravanichpaisal, P.; Liu, H.; Söderhäll, I.; Söderhäll, K. Bacteria-induced Dscam isoforms of the crustacean, *Pacifastacus leniusculus*. *PLoS Pathog.* **2011**, *7*, e1002062. [[CrossRef](#)]

59. Hoek, R.M.; Smit, A.B.; Frings, H.; Vink, J.M.; de Jong-Brink, M.; Geraerts, W.P. A new Ig-superfamily member, molluscan defence molecule (MDM) from *Lymnaea stagnalis*, is down-regulated during parasitosis. *Eur. J. Immunol.* **1996**, *26*, 939–944. [CrossRef] [PubMed]
60. Kim, B.-M.; Kang, S.; Ahn, D.-H.; Kim, J.-H.; Ahn, I.; Lee, C.-W.; Cho, J.-L.; Min, G.-S.; Park, H. First insights into the subterranean crustacean *Bathynellacea transcriptome*: Transcriptionally reduced opsin repertoire and evidence of conserved homeostasis regulatory mechanisms. *PLoS ONE* **2017**, *12*, e0170424. [CrossRef]
61. Magalhães, L.; Correia, S.; de Montaudouin, X.; Freitas, R. Spatio-temporal variation of trematode parasites community in *Cerastoderma edule* cockles from Ria de Aveiro (Portugal). *Environ. Res.* **2018**, *164*, 114–123. [CrossRef]
62. De Sousa Magalhães, L.V. Inventory, Dynamics and Impact of the Trematodes Parasites in Bivalves with High Economic Importance. Ph.D. Thesis, University of Bordeaux, Bordeaux, France, 2018.
63. Dou, M.; Jiao, Y.-h.; Zheng, J.-w.; Zhang, G.; Li, H.-y.; Liu, J.-s.; Yang, W.-d. *De novo* transcriptome analysis of the mussel *Perna viridis* after exposure to the toxic dinoflagellate *Prorocentrum lima*. *Ecotoxicol. Environ. Saf.* **2020**, *192*, 110265.
64. Gerdol, M.; Venier, P.; Pallavicini, A. The genome of the Pacific oyster *Crassostrea gigas* brings new insights on the massive expansion of the C1q gene family in Bivalvia. *Dev. Comp. Immunol.* **2015**, *49*, 59–71. [CrossRef]
65. Kishore, U.; Gaboriaud, C.; Waters, P.; Shrive, A.K.; Greenhough, T.J.; Reid, K.B.; Sim, R.B.; Arlaud, G.J. C1q and tumor necrosis factor superfamily: Modularity and versatility. *Trends Immunol.* **2004**, *25*, 551–561. [CrossRef]
66. Gestal, C.; Pallavicini, A.; Venier, P.; Novoa, B.; Figueras, A. MgC1q, a novel C1q-domain-containing protein involved in the immune response of *Mytilus galloprovincialis*. *Dev. Comp. Immunol.* **2010**, *34*, 926–934. [CrossRef]
67. He, Z.-b.; Duan, G.-f.; Liang, C.-y.; Li, H.-y.; Liu, J.-s.; Yang, W.-d. Up-regulation of Nrf2-dependent antioxidant defenses in *Perna viridis* after exposed to *Prorocentrum lima*. *Fish Shellfish Immunol.* **2019**, *90*, 173–179. [CrossRef]
68. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [PubMed]
69. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [CrossRef] [PubMed]
70. Lockwood, B.L.; Sanders, J.G.; Somero, G.N. Transcriptomic responses to heat stress in invasive and native blue mussels (genus *Mytilus*): Molecular correlates of invasive success. *J. Exp. Biol.* **2010**, *213*, 3548–3558. [CrossRef] [PubMed]
71. Lockwood, B.L.; Somero, G.N. Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). *Mol. Ecol.* **2011**, *20*, 517–529. [CrossRef]
72. European Union Reference Laboratory for Marine Biotoxins (EU-RL-MB). *EU-Harmonised Standard Operating Procedure for Determination of Lipophilic Marine Biotoxins in Molluscs by LC-MS/MS*; Spanish Ministry of Consumer Affairs: Madrid, Spain, 2015.
73. Andrews, S. FastQC: A quality control tool for high throughput sequence data [Online]. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 13 January 2021).
74. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
75. Martis, M.M.; Tarbiat, B.; Tydén, E.; Jansson, D.S.; Höglund, J. RNA-Seq de novo assembly and differential transcriptome analysis of the nematode *Ascaridia galli* in relation to in vivo exposure to flubendazole. *PLoS ONE* **2017**, *12*, e0185182. [CrossRef] [PubMed]
76. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from rRNA-seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]
77. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness. In *Gene Prediction; Humana*: New York, NY, USA, 2019; pp. 227–245.
78. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; Simão, F.A.; Zdobnov, E.M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **2019**, *47*, D807–D811. [CrossRef] [PubMed]
79. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef]
80. Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
81. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]
82. Haas, B.; Papanicolaou, A. TransDecoder 5.5.0 [Online]. 2015. Available online: <https://github.com/TransDecoder/TransDecoder/wiki> (accessed on 3 February 2021).
83. Punta, M.; Coghill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J. The Pfam protein families database. *Nucleic Acids Res.* **2012**, *40*, D290–D301. [CrossRef]
84. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37.
85. Langmead, B.; Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Meth. Nat. Res.* **2012**, *9*, 357–359.

-
86. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 1–16. [[CrossRef](#)]
 87. Tarazona, S.; García-Alcalde, F.; Dopazo, J.; Ferrer, A.; Conesa, A. Differential expression in RNA-seq: A matter of depth. *Genome Res.* **2011**, *21*, 2213–2223. [[CrossRef](#)] [[PubMed](#)]
 88. Tarazona, S.; Furió-Tarí, P.; Turrà, D.; Pietro, A.D.; Nueda, M.J.; Ferrer, A.; Conesa, A. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **2015**, *43*, e140. [[CrossRef](#)] [[PubMed](#)]
 89. Roninson, M.; Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **2010**, *11*, R25. [[CrossRef](#)] [[PubMed](#)]
 90. Kalvari, I.; Nawrocki, E.P.; Ontiveros-Palacios, N.; Argasinska, J.; Lamkiewicz, K.; Marz, M.; Griffiths-Jones, S.; Toffano-Nioche, C.; Gautheret, D.; Weinberg, Z. Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **2021**, *49*, D192–D200.