



Annotation of nerve cord transcriptome in earthworm *Eisenia fetida*



Vasanthakumar Ponesakki¹, Sayan Paul¹, Dinesh Kumar Sudalai Mani, Veeraragavan Rajendiran, Paulkumar Kanniah, Sudhakar Sivasubramaniam*

Department of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu 627012, India

ARTICLE INFO

Keywords:
Transcriptome
Nerve cord
Eisenia fetida
Annotation

ABSTRACT

In annelid worms, the nerve cord serves as a crucial organ to control the sensory and behavioral physiology. The inadequate genome resource of earthworms has prioritized the comprehensive analysis of their transcriptome dataset to monitor the genes express in the nerve cord and predict their role in the neurotransmission and sensory perception of the species. The present study focuses on identifying the potential transcripts and predicting their functional features by annotating the transcriptome dataset of nerve cord tissues prepared by Gong et al., 2010 from the earthworm *Eisenia fetida*. Totally 9762 transcripts were successfully annotated against the NCBI nr database using the BLASTX algorithm and among them 7680 transcripts were assigned to a total of 44,354 GO terms. The conserve domain analysis indicated the over representation of P-loop NTPase domain and calcium binding EF-hand domain. The COG functional annotation classified 5860 transcript sequences into 25 functional categories. Further, 4502 contig sequences were found to map with 124 KEGG pathways. The annotated contig dataset exhibited 22 crucial neuropeptides having considerable matches to the marine annelid *Platynereis dumerilii*, suggesting their possible role in neurotransmission and neuromodulation. In addition, 108 human stem cell marker homologs were identified including the crucial epigenetic regulators, transcriptional repressors and cell cycle regulators, which may contribute to the neuronal and segmental regeneration. The complete functional annotation of this nerve cord transcriptome can be further utilized to interpret genetic and molecular mechanisms associated with neuronal development, nervous system regeneration and nerve cord function.

1. Introduction

The cellular components of nervous system play regulatory role in conducting sensory signal transduction, synaptic transmission, muscle contraction and neuronal communication [1]. Among the invertebrate species, the annelid earthworms are occupying pivotal position within soil faunal biomass due to their immense contribution in maintaining the soil quality & fertility and their ecotoxicological importance in assessing the soil contaminants and toxicity [2,3]. In addition, the earthworms are also renowned for their astonishing phenomenon of regenerating the amputated organs, hence, used as a key model for stem cell and regeneration research [4]. From the past few decades, the nervous system of the earthworms has been extensively studied to unveil their importance in maintaining the sensory and behavioral physiology of the species. The nervous system of the worm is composed of bilobed cerebral ganglion connected to the circumpharyngeal connectives followed by the sub-esophageal ganglion and the nerve cord extended throughout the length of the worm [5,6]. The complex

regeneration process of the bilobed cerebral ganglion commences from the amputated nerve cord and the worm regenerates the functional brain in eight days [7]. Acknowledging the correlation between nerve cord and escape reflex of earthworm, Drewes, 1984 indicated the activation of the giant nerve fibers in the ventral nerve cord is critically required to elicit the escape reflexes in the species under threatening external stimuli [8]. Subsequent studies have confirmed the transplantation of the ventral nerve cord in heterotrophic and homotropic manner has the capability of restoring the sensory and motor activities to reconstruct the crucial behavioral reflex pathways associated with the escape mechanism of the worm [9].

The earthworm *Eisenia fetida*, popularly known as the red wiggler worm, has always remained a species of interest to the environmental and regeneration biologists because of their easy availability, effective usage in controlling the waste management, extensive regeneration capacity and rapid growth and reproductive ability [10–13]. The morphological analysis has demonstrated the presence of both peptidergic and amine secreting neurons in the ventral nerve cord of the

* Corresponding author.

E-mail address: sudhakar@msuniv.ac.in (S. Sivasubramaniam).

¹ Authors having equal contribution.

worm [14]. The immunocytochemical analysis revealed that the organization of serotonergic system in earthworm *E. fetida* is correlated with the embryogenesis of the species and during the post embryonic development/hatching period the number of serotonin immunoreactive cells and total serotonin content increase in the ganglia of the ventral nerve cord [6]. The stomogastric nervous system of earthworm *E. fetida* showed the distribution of monoamine neurotransmitters like serotonin, dopamine and octopamine, amino acid neurotransmitter like GABA and neuropeptides like proctolin, FMRFamide and ETP having the excitatory and inhibitory effects on the muscle cells of alimentary canal [15]. Herbert et al., 2009 identified six novel neuropeptides in the ventral nerve cord ganglia of worm having significant sequence homology with the insect neuropeptides like pyrokinins (PKs) and periviscerokinins (PVKs) [16]. Varhalmi et al., (2008) demonstrated the expression of neurotrophic factor PACAP (pituitary adenylate cyclase-activating polypeptide) in both normal and regenerated ventral nerve cord ganglia of *E. fetida* suggesting the peptide may act as a crucial neuro-modulator and assist during the caudal regeneration of the worm [17].

Transcriptome analysis is widely used powerful tool that provides a speedy and meticulous gene expression data in a given tissue sample. During specific developmental and physiological events, it is an effective way for studying the gene expression, gene regulation, characterizing gene functions, and discovering of novel genes and biomarkers [18]. The recent advancements in next generation sequencing techniques accompanied with relevant bioinformatics tools have enabled to explore the non-model species at the transcriptome level [19]. Especially in the field of neurological research the transcriptome sequencing has emerged as a valuable technique to explore the differentially expressed genes controlling the cellular and molecular architecture of the invertebrate nervous system. Among molluscs species, the de novo transcriptome analysis of the central nervous system (CNS) of *Lymnaea stagnalis* generated 116,355 contigs of which only 18% sequences had significant match to the known proteins in BLAST search including some newly identified monoamine synthesis enzymes like tyrosine hydroxylase and dopa decarboxylase [20]. Among the arthropods the transcriptome based EST database prepared from the CNS of the desert locust *Schistocerca gregaria* displayed 4000 functionally annotated transcripts with a considerable proportion of sequences associated with the neuronal signaling and signal transduction to regulate the physiological process of the species [21]. The deep sequencing transcriptome study of the tissue samples obtained from the nervous system of *Cancer borealis* (Jonah crab) and *Homarus americanus* (American lobster) identified 9489 and 11,061 protein coding transcripts along with the genes belong to ion channels, amine and GABA receptors and neurotransmitter receptors to regulate the neuronal properties and behavioral physiology of the crustacean species [22].

The inadequate information at genome level for most of the annelid worms have prioritized the transcriptome analysis and annotation as the key to understanding the crucial genes and transcripts associated with several biological events like metabolism, regeneration and development. Few significant research works based on the transcriptome and expressed sequence tag analysis of various earthworm species have been performed in the recent past. These researches have revealed some of the key findings like expression of globin-related genes and fibrinolytic enzymes in the midgut sample of earthworm *Eisenia andrei* [18], differential expression of putative genes under inorganic, organic and agrochemical xenobiotic exposure in earthworm *Lumbricus rubellus* [23], genes regulating the anterior regeneration in *Perionyx excavatus* [24] and identification of genes responsible for immune response in *Eisenia andrei* [25]. Considering the ecological and environmental aspects of the earthworm *E. fetida* several transcriptome analysis were undertaken to extract maximum information regarding their genetic and molecular resources. Pirooznia et al., 2007 performed the cloning and sequencing of the ESTs (expressed sequence tags) of the worm to monitor the genes associated with environmental toxicity [26]. Gong

et al., 2010 constructed the transcriptome based EST library for designing and validating the oligo-probes associated with the given ESTs [27]. More recently the comprehensive de novo transcriptome analysis and annotation study have captured the differential expression of toxin genes by giving priority to the ecotoxicological impact of the species [28]. Despite of certain elaborative efforts the nervous system and the nerve cord specific genes of the earthworms have not been explored to a great extent at the transcriptome level. Among the few important studies, the transcriptome sequencing of earthworm *E. fetida* upon the exposure to the neurotoxic chemical Perfluorooctanesulfonic acid (PFOS) exhibited altered expression in neuronal development and calcium homeostasis related genes and resulted in the neurodegeneration of the species [29]. Subsequently, the Perfluorooctanoic acid (PFOA) exposure in the worm led to the differential regulation of different neuronal development specific genes associated with synaptogenesis, synaptic transmission and cellular morphogenesis [30]. However, the data is still quite insufficient and can be improved further with comprehensive transcriptome analysis and annotation. Notably, the transcriptome data prepared by Gong et al. [27] were obtained from the 454 sequencing of the nerve cord tissue comprising both neuronal and peripheral cells of earthworm *E. fetida*, but the complete functional annotation details of the assembled contigs were remained unaccomplished in their study. The present study focuses on the functional annotation of the assembled transcriptome dataset of earthworm *E. fetida*, generated by Gong et al., 2010 to identify the nerve cord specific genes and predict their functional features in regulating different cellular and physiological processes like signal transduction, reproduction and sensory perception. Simultaneously with extensive regeneration capacity and well known metabolic activity of the earthworm *E. fetida* the annotation overview based on biological pathway assessment and gene ontology analysis can be informative to identify and categorize the crucial stem cell markers, developmental genes and metabolic markers reside in the nerve cord the species.

2. Materials and methods

2.1. Acquiring de novo assembled contig sequences of *Eisenia fetida*

We have selected the Newbler and Seqman assembled contig sequences obtained from the 454 transcriptome sequencing of the nerve cord cDNA sample of earthworm *E. fetida* (Gong et al., 2010; Table S1 and S2). The statistical information regarding the length, median value, N50 value and total number of bases of the given contigs was generated by using the NGS QC Toolkit [31].

2.2. Annotation of the transcriptome data

The functional annotation study of the Newbler and Seqman assembled contigs of earthworm *E. fetida* was performed through 1. Sequence based alignment and 2. Domain based alignment. The sequence based similarity searching was carried out by aligning the sequences against NCBI non-redundant protein (nr) database, and Swiss Prot database using the BLASTx algorithm with *E*-value threshold of 1E-5. Depending upon the significant BLAST hits obtained from the nr database the Gene Ontology (GO) (www.geneontology.org) annotation study was performed by using BLAST2GO functional annotation software [32,33]. The GO annotation step was accomplished with two sub steps: 1. retrieving the GO terms associated with the BLAST hits (Mapping) and 2. assigning the GO terms and Enzyme Commission number (EC number) to the query sequences based on certain parameters like Annotation Cut-Off: 55; E value hit filter: 1.0E-6; GO Weight: 5; HSP-Hit Coverage Cut-Off: 0 and Hit Filter: 500.

The domain details and motif information of the query contig sequences were obtained by scanning the sequences against InterPro protein signatures databases using the InterProScan annotation embedded in BLAST2GO [34]. The resulting GO terms from InterProScan

annotation were further transferred and merged with the already existing GO terms acquired from GO annotation. To predict and classify the functions in a more reliable manner the assembled contigs were annotated based on their orthologous groups (COGs). The screening and annotation of the corresponding orthologous groups for the given contigs of earthworm *E. fetida* were performed by using the EggNog database (Evolutionary genealogy of genes) implanted within the BLAST2GO annotation software [35]. The Newbler and Seqman assembled contigs were aligned to the EggNog database using the *E*-value threshold of $1E-5$ and minimum similarity percentage of 50%.

The KEGG database acts as a major knowledgebase to provide significant details regarding the biological pathways and molecular networks associated with the given genes or transcripts (36). The KEGG pathway annotation of the assembled contigs was achieved by mapping the sequences assigned with EC numbers against the online KEGG pathway database using the BLAST2GO program.

The enrichment of the gene ontology terms associated with the nerve cord transcriptome of earthworm *E. fetida* in comparison to their draft genome [37] was analyzed by using the Fisher's exact test [38] implemented in the BLAST2GO version 4.1. The annotated draft genome of *E. fetida* was used as the reference set and the nerve cord specific transcripts were used as the test set for the analysis. The *P*-values were corrected by Benjamini and Hochberg FDR correction method and the $FDR < 0.05$ was considered as statistically significant. Further the enrichments of the KEGG pathways were confirmed by KOBAS 3.0 web tool [39].

2.3. Identification of neuropeptides and stem cell markers

In invertebrate organisms the neuropeptides are largely localized in the central and peripheral nervous system and act as essential neurotransmitters to regulate the cellular and physiological activities [40]. The Newbler and Seqman assembled transcripts of earthworm *E. fetida* annotated against the nr database were thoroughly screened to monitor the neuropeptide genes present in the nerve cord tissue of the species. Simultaneously the potential stem cell specific genes participate in the cellular differentiation and segmental regeneration processes of the worm were identified by comparing the assembled contigs with the previously reported 250 stem cell marker genes obtained from human amniocytes [41] using the local BLASTX with *E*-value cutoff of $1E-5$.

2.4. Identification of calcium binding proteins and phylogenetic analysis

The nr annotated contigs of earthworm *E. fetida* were further screened to identify the calcium binding proteins containing the EF hand domain. The identified contigs were translated into their corresponding protein sequences by using the ExPasy translate tool. The translated protein sequences of earthworm *E. fetida* were aligned to their orthologs in leech *Helobdella robusta* and the polychaete *Capitella teleta* using the Clustal W multiple sequence alignment tool [42] with default settings. The phylogenetic tree was constructed using the UPGMA method [43] with 100 bootstrap replicates.

3. Results and discussion

3.1. Quality assessment and length distribution of contigs

To create the complete annotation framework of the transcriptome dataset obtained from the nerve cord cDNA sample of earthworm *E. fetida* we have selected total 94,716 contigs generated from the de novo assembly of 562,327 filtered reads using Newbler and Seqman assemblers (Gong et al., 2010; Tables S1 and S2) [27]. Both the Newbler and Seqman are efficient assemblers for the 454 transcriptome dataset. The Newbler assembly is comparatively faster and produces less number of contigs with low redundancy and better alignment scores whereas the Seqman assembly usually generates higher amount of small and

redundant contigs along with many significant novel transcripts [44,45]. The quality assessment statistics (Table 1) obtained from the NGS QC Toolkit reported that the N50 and N75 values for the Newbler assembled contigs were 146 bp and 109 bp and Seqman assembled contigs were 170 bp and 119 bp respectively. The GC% for both the Newbler and Seqman assembled contigs were observed as 39.89. The distribution of contigs length is a considerable feature, which can be connected with retrieving maximum number of BLAST matches. Hence, we have opted for analyzing the contig length distribution and correlated the information with number of sequences matched in the BLAST search against the non-redundant (nr) database. Among the total 94,716 assembled contig sequences a large proportion of contigs (87,183, 92.04%) showed sequence length ≤ 300 bp, whereas only 5000 (5.27%), 1504 (1.58%), 586 (0.61%) and 443 (0.46%) contigs had sequence lengths ranging from 301 to 500 bp, 501–700 bp, 701–900 bp and ≥ 901 bp respectively.

3.2. Annotation of *E. Fetida* contigs against public databases

In transcriptomic study, the annotation is a crucial step to gain maximum information regarding the sequence identity and features of the resulting transcripts. The assembled contig sequences of earthworm *E. fetida* were annotated through similarity searching against the NCBI non-redundant protein (nr) database and Swiss Prot database using the BLASTX algorithm with cut-off *E*-value $< 1E^{-5}$. The annotation quality based on BLAST search was further improved by comparing the contigs against the EST database using the BLASTN algorithm with *E*-value threshold $1E-5$. A total of 17,189 (18.14%) contigs were annotated from BLAST search against these three databases. The overall sequence alignment data represented 9762 (10.30%), 8992 (9.49%) and 9560 (10.09%) *E. fetida* transcripts with significant BLAST matches against the nr, SwissProt and EST databases respectively. A three way Venn diagram showed the overlap of the assembled contigs annotated against the nr, SwissProt and EST databases (Fig. 1A). The Venn diagram represented that 2689 contigs had homologous sequences in all three databases and 395, 276 and 5075 contig sequences showed the annotation overlap between nr and EST, EST and SwissProt and nr and SwissProt databases respectively. Among the contigs with significant BLAST hits total 6199 (36.06%) sequences showed exclusive BLAST matches with the dbEST as they were not annotated to the nr and Swiss Prot databases. The Sequence based similarity analysis depicted that most of the assembled contigs did not match to the known proteins available in these databases probably due to their short sequence length or the deficiency of the genome information for earthworm *E. fetida*. After correlating the contigs length with the retrieved BLAST hits we have observed that 51.57% of the contigs with sequence length ≥ 900 bp showed BLAST matches against nr database, whereas the matching efficiency gradually decreased to 10.43% for the contigs with sequence length between 100 and 300 bp (Fig. 1B). The retrieval of low number of BLAST hits for the contigs with shorter sequence length is mainly due to the lack of conserved domains representing the functional properties of the annotated proteins [46]. The overall data distribution details of the annotated contigs obtained from BLAST2GO analysis was outlined in Fig. 1C. The complete list of the Newbler and Seqman assembled transcripts annotated against the nr database was documented in Table S1.

The *E*-value distribution and sequence similarity distribution are considerable statistical parameters to evaluate the quality of BLAST analysis [47]. The *E*-value distribution of the top BLAST hits obtained from nr database revealed that among the contigs with BLAST hits 9040 (92.60%) sequences showed significant homology with an *E*-value ranging from $1E-5$ to $1E-45$ (Fig. 2A). The similarity distribution data indicated that 53.69% of the assembled contigs with BLAST hits showed sequence similarity $\geq 80\%$ to the known protein sequences in nr database (Fig. 2B). The analysis of the top hit species distribution data for the transcriptome dataset of earthworm *E. fetida* identified the leech,

Table 1
Summary Statistics of the Newbler and Seqman assembled contigs.

Statistics	Newbler contigs	Seqman contigs
Total sequences	31,114	63,602
Total bases	4,828,470	10,463,348
Min sequence length	72	40
Max sequence length	1395	2167
Average sequence length	155.19	164.51
Median sequence length	115.00	128.00
N25 length	277	303
N50 length	146	170
N75 length	109	119
GC%	39.89%	39.89%

Helobdella robusta as the top hit species with maximum number of BLAST hits (20%). Among the other species a commendable number of BLASTX hits were observed to match with the nr protein sequences of polychaete worm, *Capitella teleta* (16%), brachiopod species, *Lingula anatina* (6%), oyster, *Crassostrea gigas* (4%) and owl limpet, *Lottia gigantea* (3%) respectively (Fig. 2C). *Helobdella robusta* and *Capitella teleta* are the most extensively studied annelid species with complete draft genome information giving detailed insights regarding their sensory perception, signal transduction, embryonic development and metabolic activities [48]. In addition, the genome datasets of these two annelids have represented the expression of several neuro hormones and neuropeptides precursors providing a significant clue regarding the

neuropeptide evolution in the early lineage [49]. The top hit species distribution statistics obtained from the BLASTX annotation clearly indicated that the nerve cord specific transcripts of earthworm *E. fetida* are evolutionarily more closely related to the other annelid organisms and the distribution of the species depends on the completeness of their genome information available in the databases. Interestingly we have found that only 141 (1%) BLAST hits for the identified transcripts matched with the known protein sequences of earthworm *E. fetida* available in the nr databases probably due to the lack of annotated draft genome and insufficient proteome information of the species in the database.

3.3. Conserved domain annotation of *E. Fetida* transcripts

The annotation based on evolutionary conserve domains provides detailed insight regarding the domain architecture and functional sites conserved across the protein families [50]. To further understand the functional properties of the given transcripts of earthworm *E. fetida* the conserved domains are identified against the InterPro database using the BLAST2GO software [34]. The top 30 InterPro domains/families for the Newbler and Seqman assembled contigs were represented in Fig. 3. The domain details obtained from InterProScan analysis denoted “P-loop containing nucleoside triphosphate hydrolase” (IPR027417) as the most conserved domain with 222 contigs, followed by the “EF-hand domain pair” (IPR011992) (160 contigs), “EF-hand domain” (IPR002048) (132 contigs) and “NAD(P)-binding domain” (IPR016040)

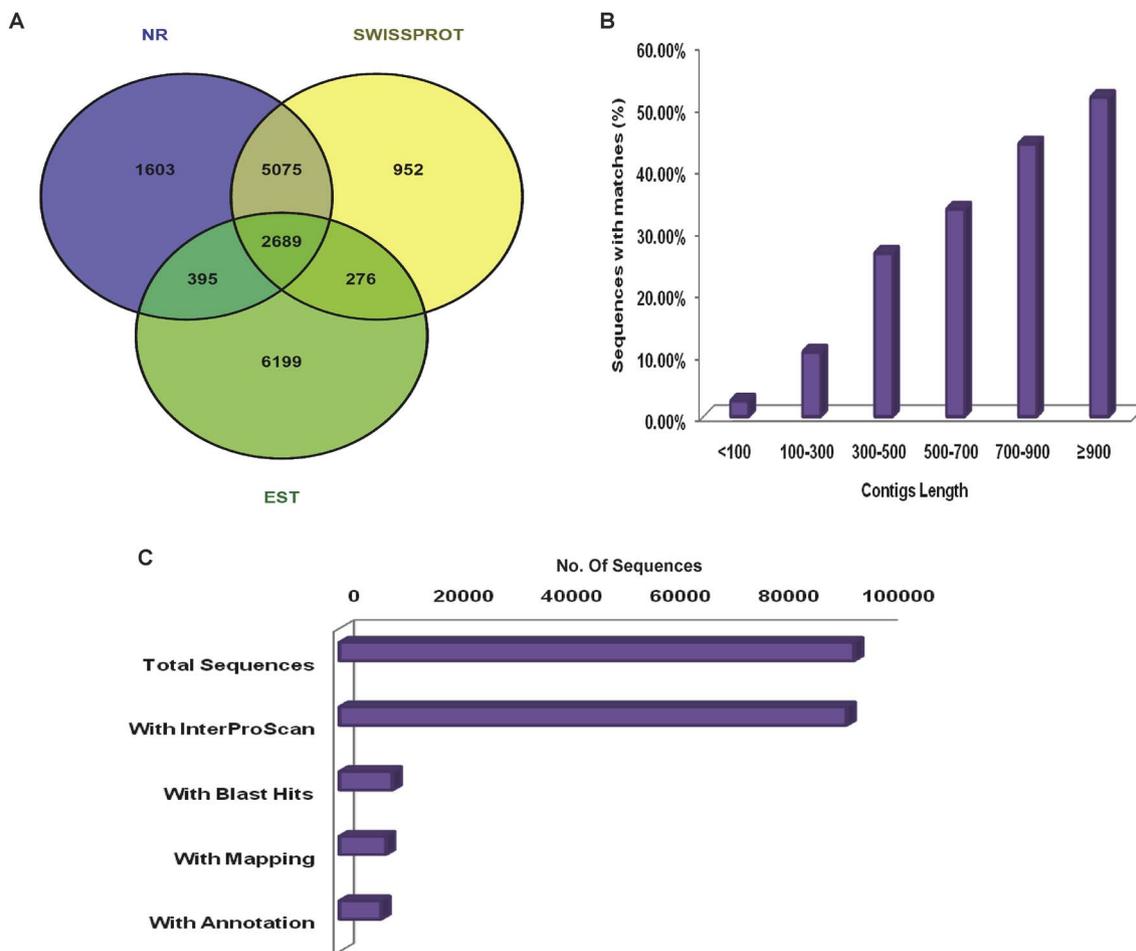


Fig. 1. (A) A three way Venn diagram denoting the unique and overlapped transcripts annotated against the public databases nr, SwissProt and EST (BLASTX algorithm; E -value threshold $1E^{-5}$). (B) The correlation between contigs length and percentage of sequences with BLAST matches. The longer assembled contigs have more number of sequences with BLAST matches against NCBI nr database (cut-off E -value $< 1E^{-5}$). (C) The data distribution details obtained from BLAST2GO annotation depicts the InterProScan annotation, BLAST hits, mapping and annotation summary of the assembled contigs.

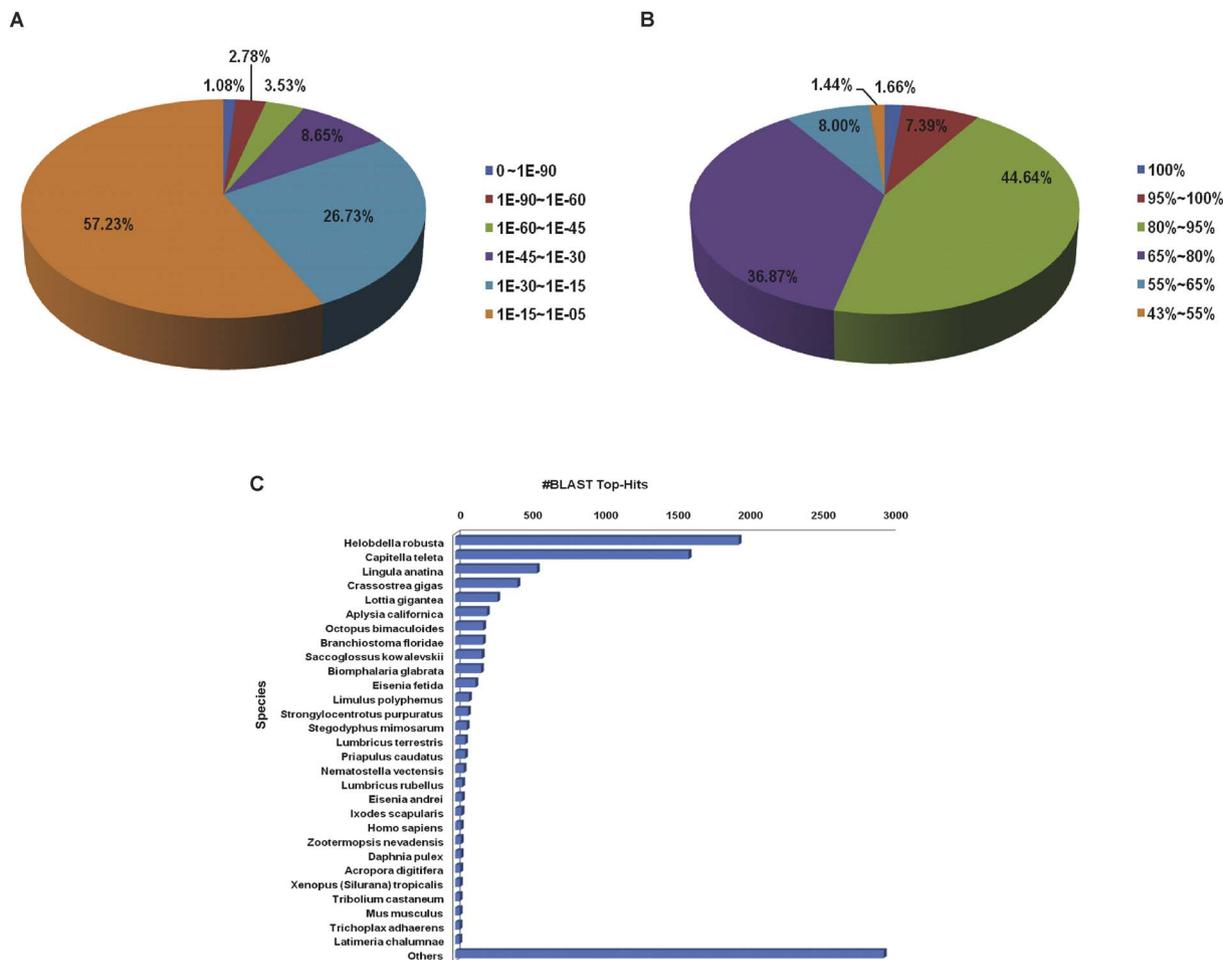


Fig. 2. (A) E-value distribution of the nr BLAST hits for each sequence of contigs with a cut-off E-value of 1E-5. (B) Sequence similarity distribution (in percentage) of top BLAST hits for each contigs sequences. (C) Top hit species distribution summary of the *E. fetida* contigs annotated against nr database with an E-value cut-off of 1E-5.

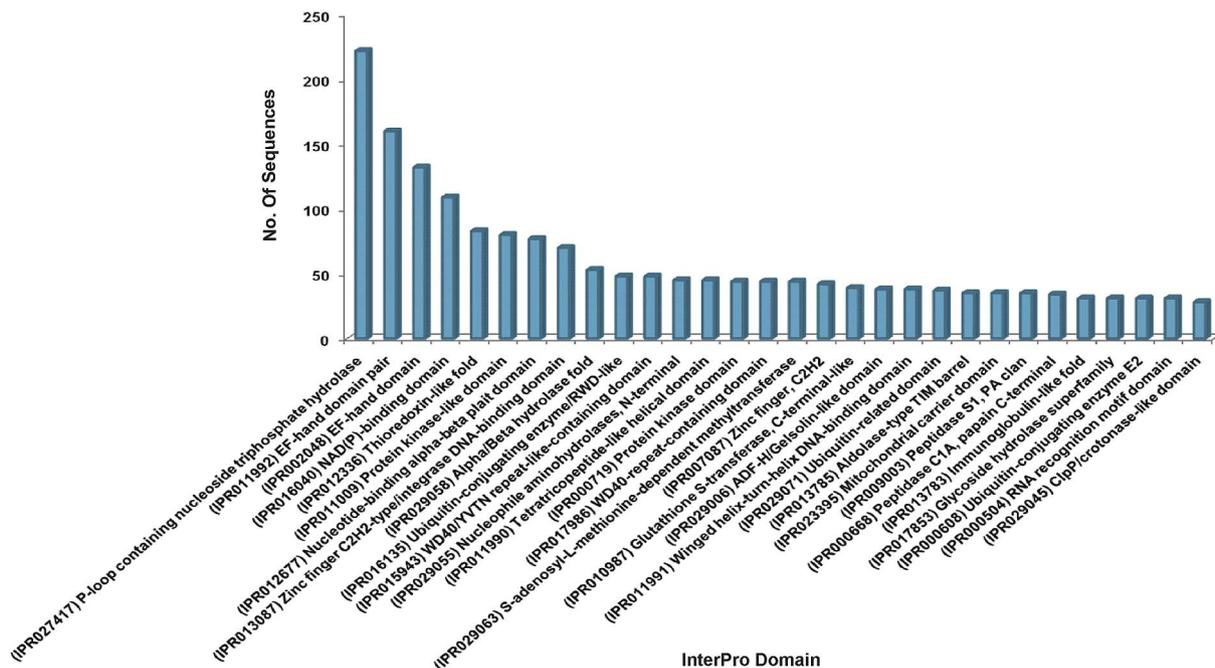


Fig. 3. Histogram of top 30 InterPro domains distribution obtained from InterProScan annotation of the *E. fetida* nerve cord transcriptome dataset using the BLAST2GO software.

(109 contigs). Detailed analysis of the identified domain characteristics in the transcript sequences of earthworm *E. fetida* depicted that most of the abundant domains are associated with several critical biological functions like nucleotide binding, signal transduction and transcriptional regulation to carry out the genetic and environmental information processing. More specifically, the P-loop NTPase domain plays a crucial role in energy production through NTP hydrolysis and acts as a substrate for nucleotide binding [51]. Simultaneously the domain also participates in signal transduction, stress response and bacterial transcriptional regulation [52]. The abundance of this domain in the transcriptome dataset of earthworm *E. fetida* indicated that most of the transcripts expressed in the nerve cord of the worm are associated with signal transduction, protein transport and localization, signal recognition, chromosome partitioning and membrane transport [53]. Besides the InterPro analysis data depicted the presence of several ATP synthase subunits and molecular motor proteins with P-loop NTPase domain in the transcriptome dataset. The ATP and the related nucleotides are important neurotransmitters of the central and peripheral nervous system to conduct the purinergic neurotransmission, smooth muscle contraction and sensory perception to external stimuli [54,55]. Among the molecular motor proteins myosin is a conserved motor protein that regulates the muscle contraction and motility in most of the living organisms through ATP hydrolysis [56,57]. Since the backbone of our transcriptome dataset is the nerve cord sample of earthworm *E. fetida*, the physiological energy resource generated from the nucleotide hydrolysis may assist in the neurotransmission, sensory signal transduction and muscle contraction of the species. The second most representative domain in our study, the EF-hand domain pair mainly involves in signal transduction, transcriptional regulation and muscle contraction through calcium ion binding [58]. Simultaneously the EF-hand domain acts as the hallmark for the calcium binding proteins participating in neuronal stimulation and neural circuit development through their calcium buffering activities. Especially in invertebrate like *Drosophila* the calcium buffering assist in the axonal growth during the neuronal development [59,60]. Subsequent studies have reported the EF-hand domain containing calcium binding proteins play regulatory role in the maturation of nervous system and formation of chemosensory neurons in nematodes like *Caenorhabditis elegans* to facilitate the serotonin synthesis [61]. Hence the expression of calcium binding proteins containing the EF-hand domain in the transcriptome dataset may assist in the neuronal development and locomotory activity of the worm to maintain the behavioral physiology of the species.

3.4. Gene ontology annotation

The gene ontology (GO) is an international functional classification system aims to annotate and classify the functional features of the genes and gene products in species independent manner [32]. BLAST2GO software was used to functionally categorize these annotated contigs into three different ontologies known as Biological Process, Molecular Function and Cellular Component. In our gene ontology study, we made use of the BLAST2GO mapping step to retrieve GO terms associated with the contigs having BLAST hits from the nr database. The GO data illustrated that out of 9762 contigs with nr BLAST hits 8819 (90.34%) contig sequences were mapped with the associated GO terms and among them 7680 (78.67%) contigs were successfully annotated to total 44,354 GO terms. The functional categorization study of the obtained GO terms represented that among these 44,354 GO terms assigned to the annotated contigs 19,786 (44.60%) GO terms belong to the Biological Process, 12,937 (29.16%) GO terms were categorized into Molecular Function category and 11,631 (26.22%) GO terms belong to the Cellular Component. The major functional subcategories specifying the predicted functional features of the annotated contigs were obtained from GO level class 2 and representing total 55 functional groups (20 Biological Process, 15 Molecular Function and 20 Cellular Component) in the gene ontology annotation study (Fig. 4).

The overall GO distribution analysis of the annotated contig dataset of earthworm *E. fetida* denoted that within the Biological Process the dominating subcategories were “cellular Process” (58.45%), “metabolic process” (56.17%) and “single-organism process” (40.06%); within the Cellular Component the most represented GO terms were “cell” (56.88%), “cell part” (56.30%) and “organelle” (44.80%) and within Molecular Function most of the annotated contigs are associated with the subcategories like “binding” (53.34%) and “catalytic activity” (45.96%). Interestingly the transcriptome annotation data obtained from sterile cultured earthworm *Eisenia andrei* also identified these GO terms as dominant functional subcategories for the reported transcripts (25). Apart from these over represented subcategories a substantial percentage of the annotated contigs were assigned to the significant GO terms like “biological regulation” (GO:0065007), “response to stimulus” (GO:0050896), “structural molecule activity” (GO:0005198) and “macromolecular complex” (GO:0032991).

The environmental and cellular signals play pivotal role in regulating the physiological process of the living organisms. In earthworm, the sensory perception to the external stimuli is a notable physiological feature carried out mainly by the sensory nerve fibers residing in the ventral nerve cord ganglion [62]. The external stimuli coming from the tactile, photo and chemical stimulus creates electrical impulses in the form of the action potential to excite the giant nerve fibers and assist in the sensory and escape reflex of the worm [63]. Simultaneously several neuropeptides and monoamine neurotransmitters serve as important chemical messengers to transmit the neuronal signal and conduct synaptic modulation [16]. Hence the nerve cord of earthworm *E. fetida* is expected to contain a large number of contigs associated with signal transduction, cell signaling and sensory perception of the species. In the annotated transcriptome dataset we have observed 323 contigs were assigned to the function “signal transduction” (GO:0007165), 31 contigs were assigned to the function “cell-cell signaling” (GO:0007267) and 30 contigs were assigned to different sensory perceptions (GO:0007601, GO:0050908, GO:0050916, GO:0019233, GO:0001580, GO:0007600, GO:0007605, GO:0050909) (Table S2). A total of 410 transcripts in the transcriptome dataset were found to associate with different signaling pathways like cytokine-mediated signaling pathway (GO:0019221), G-protein coupled receptor signaling pathway (GO:0007186), fibroblast growth factor receptor signaling pathway (GO:0008543), epidermal growth factor receptor signaling pathway (GO:0007173), Notch signaling pathway (GO:0007219), Wnt signaling pathway (GO:0016055) and smoothened signaling pathway (GO:0007224). The G protein-coupled receptors (GPCRs) are conserved seven transmembrane domain receptors that mediate sensory perception in response to the environmental cues by altering the neural circuits [64]. Several odor molecules, pheromones, light sensitive components, neurohormones and neurotransmitters serve as the potential ligands for binding and activating the GPCRs to conduct a wide range of cellular and physiological processes like photoreception, olfaction, chemoreception, synaptic transmission, neuromodulation and cellular communication [65,66]. Besides the notch pathway acts as a key signaling mechanism to determine the cell fate during development. In invertebrate like *Drosophila* the notch pathway assists in the patterning of neurogenesis and lateral inhibition of neural fate to promote the embryo and larval development [67]. Interestingly in annelid *Platynereis* the notch signaling is associated with the development of chaetal sac and formation of bristles to support the locomotory activity [68].

Previous studies confirmed that ventral nerve cord has pronounced effect on the reproductive behavior of several invertebrate species. In insect tobacco budworm the ventral nerve cord regulates the emission of sex pheromone and enhances the egg production and deposition [69]. The egg laying behavior in nematode *C. elegans* is modulated by different environmental stimuli activating the neural circuits, motor neurons and skeletal muscles. The ventral nerve cord plays major role in regulating the hermaphrodite-specific motor-neurons to conduct the synaptic transmission for the contraction of egg laying muscle in *C.*

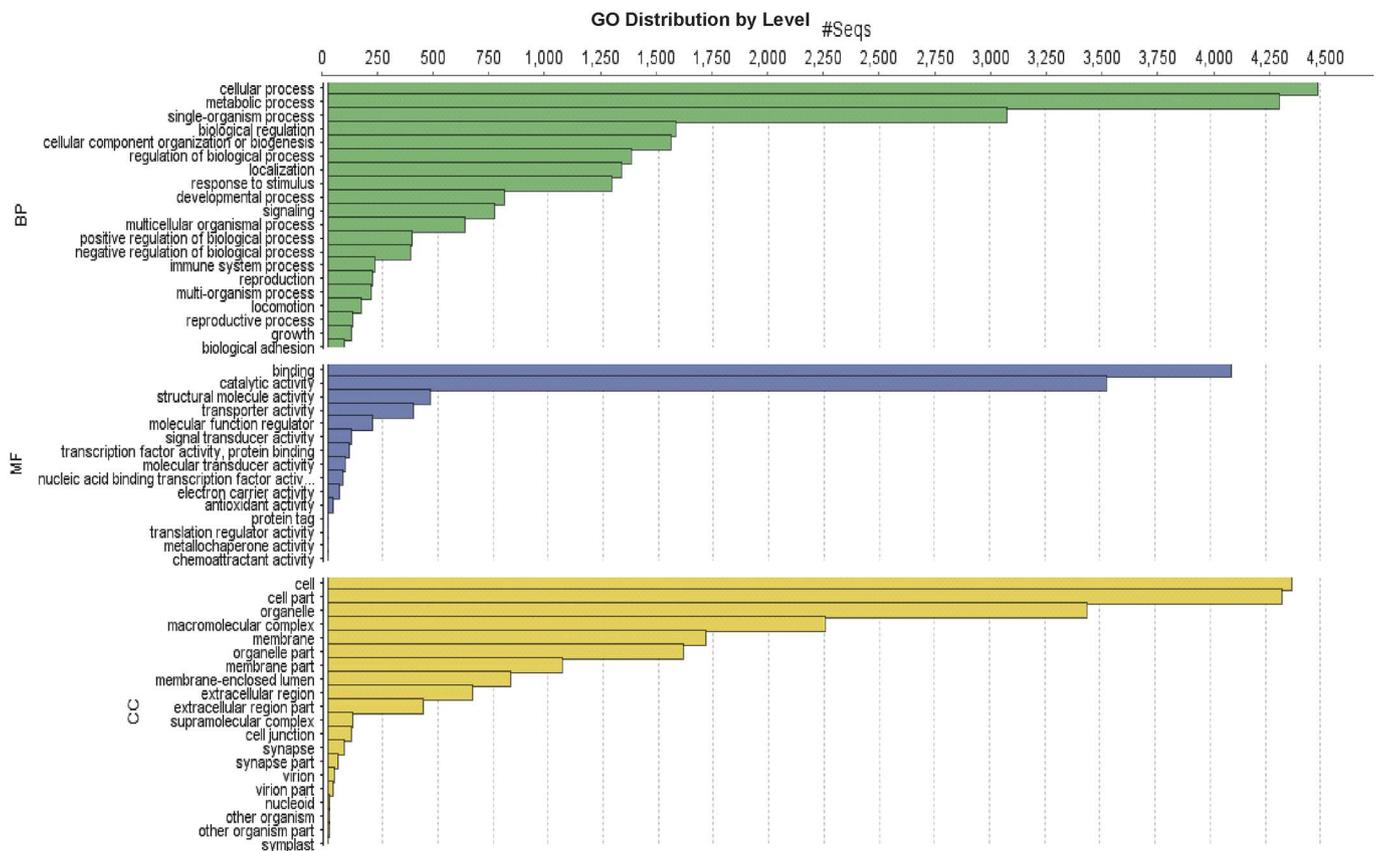


Fig. 4. Histogram representing the gene ontology distribution of the assembled *E. fetida* contigs. The functionally annotated contigs were classified into three main GO categories: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC).

elegans [70]. The transcriptome analysis have pinpointed the regulatory role of pheromones (Attractin, Temptin and Seductin), sexual differentiation markers (DMRT3, SPATA2 and SOX3), germ line determination factors (vasa, PL10) and fertilization factors (Fertilin and Acrosin) in reproduction and sexual maturity of the Mediterranean earthworms [71]. The functional annotation details of earthworm *E. fetida* denoted that total 89 contigs in the dataset were assigned to the gene ontology term reproduction (GO:0000003) and 61 contigs were associated with the function embryo development (GO:0009790, GO:0009795). Based on the sequence homology and functional importance of the annotated contigs we have shortlisted some crucial genes (Table 2) regulating the functions like anatomical structure formation involved in morphogenesis (GO:0048646), developmental maturation (GO:0021700), gonad development (GO:0008406), hatching (GO:0035188), nematode larval development (GO:0002119) and hermaphrodite genitalia development (GO:0040035). Among the identified genes Ras-related Rab-2, dihydrolipoyl dehydrogenase, DEAD box, Centrin-2 and Calmodulin showed complete (100%) sequence match with the nr database. Calmodulin is a well-known calcium binding protein which can alter the Smad signaling pathway to trigger the ventral mesoderm formation in *Xenopus* species [72]. In addition calmodulin plays decisive role in modulating the calcium function to allow the compaction essential for blastocyst formation [73]. Among the other embryonic regulators the 14-3-3 zeta protein affects the RAS/MAP kinase pathway to trigger the eye and embryo differentiation in *Drosophila* [74].

3.5. Enzyme based classification

The enzyme commission number is a hierarchical classification system that classifies the enzymes depending upon their chemical reactions [75]. The complete analysis of the assigned EC numbers to the annotated contigs indicated that among the 7680 contigs with

functionally annotated gene ontology terms, 2638 contig sequences were assigned to 535 EC numbers. The enzyme code distribution analysis for the annotated contig datasets were represented in Fig. 5A. The overall data denoted that among the enzyme classes the enzymes associated with the hydrolase activity were the most represented (1265, 47.96%) followed by the enzymes associated with transferase (629, 23.84%) and oxidoreductase (513, 19.44%) activities. The over-representation of these enzyme classes in our study clearly substantiates that most of the annotated contigs of earthworm *E. fetida* are involved in crucial metabolic processes, signal transduction and immune response process of the species. Our functional interpretation can be strongly supported by the previous studies explicating the roles of oxidoreductase and hydrolases in regulating the metabolism [76] and signal transduction process [77].

3.6. COG functional classification

The Clusters of Orthologous Groups (COGs) database is a useful resource which allows the prediction and classification of the possible functions of new species depending upon the orthologous relationship of the gene products [78]. The COGs functional annotation was performed by using the EggNog database integrated in BLAST2GO annotation pipeline [35]. The overall data from EggNog analysis illustrated that total 5860 transcript sequences of earthworm *E. fetida* (60.02% of the contigs with BLAST hit) were assigned to 25 COG functional groups (Fig. 5B). Among these 25 functional categories the cluster for “Post-translational modification, protein turnover, chaperones” (1991, 33.98%) comprises the largest functional group. The heat shock proteins (HSPs) are notable molecular chaperons with anti-apoptotic activity to provide protection to the cells under different stress conditions like altered temperature, pH and oxygen deprivation [79]. These proteins express abundantly in the glial and neuronal cells and involved in

Table 2
List of transcripts associated with embryonic development and reproduction.

SeqName	Description	E-value	Sim mean
Newbler_Contig00646	Ras-related Rab-2	5.46E-13	100%
Newbler_Contig08599	cAMP-dependent kinase catalytic subunit beta isoform × 3	3.47E-21	97%
Newbler_Contig11354	Dihydropolypol dehydrogenase	6.86E-15	100%
Newbler_Contig12349	DEAD box ATP-dependent RNA helicase	3.51E-13	100%
Newbler_Contig17280	Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16	1.47E-14	98.05%
Newbler_Contig19026	Guanine nucleotide-binding subunit beta-2-like 1	2.42E-20	99.85%
Newbler_Contig19125	Ras-related Rac1	3.65E-15	99.85%
Newbler_Contig20803	Transmembrane emp24 domain-containing 2	1.25E-100	92%
Newbler_Contig21587	Ras-related Rap-1b isoform × 2	1.86E-74	94.65%
Newbler_Contig23046	DNA excision repair ERCC-1 isoform × 1	5.65E-15	97%
Newbler_Contig24204	Structural maintenance of chromosomes 1A	3.58E-21	97.30%
Newbler_Contig25563	Centrin-2 isoform × 2	1.02E-35	100%
Newbler_Contig28743	Calmodulin	1.00E-51	100%
Newbler_Contig28993	Ubiquitin-conjugating enzyme E2 2	6.26E-13	99.80%
Newbler_Contig30617	BRICK1	1.58E-20	92.75%
Seqman_Contig_4254	TALDO_DROME ame: Full = Probable transaldolase	1.49E-35	80.40%
Seqman_Contig_16470	Elongation factor 1- oocyte form	1.10E-23	98.60%
Seqman_Contig_23100	frataxin mitochondrial	6.56E-14	87.10%
Seqman_Contig_31323	14-3-3 zeta	1.46E-26	97.70%
Seqman_Contig_40681	Casein kinase I isoform alpha	1.23E-20	95%
Seqman_Contig_50528	Angiopoietin-related protein 2	1.45E-19	97%
Seqman_Contig_60730	GTP-binding 128sup	2.73E-25	87.70%
Seqman_Contig_60731	14-3-3 GF14 omicron	3.95E-12	99.70%

regulating the inflammatory responses, cell homeostasis and cell survival. In addition the heat shock proteins show altered expression in the central and peripheral nervous system to support the axonal regeneration and remyelination upon nerve injuries like cerebral ischemia [80]. The transcriptome dataset of earthworm *E. fetida* denoted the presence of total 65 heat shock protein transcripts. Among the other functional groups the clusters for “Translation, ribosomal structure and biogenesis” (1233, 21.04%), “Function unknown” (1180, 20.13%), “Intracellular trafficking, secretion, and vesicular transport” (909, 15.51%), “Signal transduction mechanisms” (822, 14.02%), “General Function Prediction Only” (710, 12.11%) and “Energy production and conversion” (700, 11.94%) were the highly represented functional categories (Fig. 5B). Simultaneously the COG functional annotation data delineated that the clusters for “Cell wall/membrane/envelope biogenesis”, “Defense mechanisms”, “Extracellular structures”, “Cell motility” and “Nuclear structure” were the under-represented categories which constituted the smallest functional groups in our study. The detailed interpretation of the COG analysis deciphers that most of the contigs of earthworm *E. fetida* transcriptome dataset are associated with information storage and processing. In metabolism category the clusters for “Carbohydrate transport and metabolism”, “Lipid transport and metabolism” and “Amino acid transport and metabolism” were the well represented functional groups. In addition a considerable number of contigs were assigned to poorly characterized functional categories probably due to phylogenetically distant relationship of *E. fetida* with other species available in COG database. In eukaryotic system the ribosome biogenesis is an essential metabolic activity connected with pivotal cellular processes like cell proliferation and differentiation [81]. Especially in invertebrate species like *C. elegans* the ribosomal biogenesis regulates gonadogenesis by assisting in germline proliferation and pattern formation [82]. The transcriptome

annotation details of earthworm *E. fetida* have represented several important translational regulatory factors like 60S ribosomal L5, 60S ribosomal L3, 40S ribosomal S13, 40S ribosomal S5 and 40S ribosomal S20, which may serve as the regulatory factors to assist in the biogenesis and structural constitution of ribosome in the worm species.

3.7. KEGG pathway annotation

The KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway is an informative knowledgebase that displays the biological pathways connected to the gene or gene products obtained from molecular interaction network analysis [36]. These biological pathways provide significant functional information by graphically representing the crucial cellular and metabolic process of the species. In order to identify the biological pathways associated with the transcriptome dataset generated from the nerve cord tissue sample of earthworm *E. fetida*, the assembled contigs were annotated with corresponding EC numbers against the KEGG Pathway database using the BLAST2GO annotation tool. Total 4502 contig sequences were mapped to 124 KEGG pathways. Fig. 6 denoted the top 20 KEGG pathways highly represented by the annotated contig sequences of earthworm *E. fetida*. Among these pathways Purine Metabolism (617 contigs), Thiamine metabolism (500 contigs) and Biosynthesis of antibiotics (313 contigs) were the most dominant KEGG pathways observed in the transcriptome dataset. In addition the other highly represented pathways were Glycolysis/Gluconeogenesis, Oxidative phosphorylation and Aminobenzoate degradation. Notably, purines are small organic molecules involve in a wide range of neuronal functions like neurotransmission, neuromodulation, controlling neurite outgrowth and glial cell proliferation. The purine metabolism is firmly connected to the neuronal differentiation and neural development and any deficiency in the purine biosynthesis and recycling can lead to the degeneration of motor control, behavior and cognition [83]. Besides the purinergic signaling in central and peripheral nervous systems are essential to regulate diverse physiological and cellular processes like mechanosensory transduction, neurohormones secretion, cell proliferation, differentiation and neuronal protection [84]. Further meticulous analysis of the mapped KEGG pathways indicated that majority of the annotated contig sequences were attributed to the metabolism pathway (4229 contigs), whereas only few contigs were mapped to the pathways like organismal systems (162 contigs), genetic information processing (83 contigs) and Environmental Information Processing (27 contigs) (Table S3).

3.8. Functional enrichment analysis

The functional enrichment analysis is a powerful analytical technique which provides valuable information regarding the over and under representation of the annotated functions and biological pathways associated with a group of genes compare to their background dataset [85]. The enrichment analysis of the nerve cord specific GO terms and KEGG pathways in comparison to the entire genome of earthworm *E. fetida* [37] was performed using the Fisher's exact test with FDR corrected *P* value < 0.05. The top 10 over represented GO terms from the three different ontologies were shown in Fig. S1. The GO enrichment data denoted that “ATP binding” (GO:0005524), “extracellular exosome” (GO:0070062), “structural constituent of ribosome” (GO:0003735) and “GTP binding” (GO:0005525) were the most enriched GO terms for the nerve cord transcripts. The enrichment of the transcripts associated with molecular functions like ATP and GTP binding activities in the nerve cord transcriptome dataset is extremely significant as they might be the potential source for the energy and phosphate to support the crucial physiological functions like neuronal sensory signal transduction and muscle contraction and relaxation of the worm [86]. The details of the functionally upregulated GO terms and the biological pathways were represented in Table S4. The pathway enrichment study suggested that among the 124 KEGG pathways

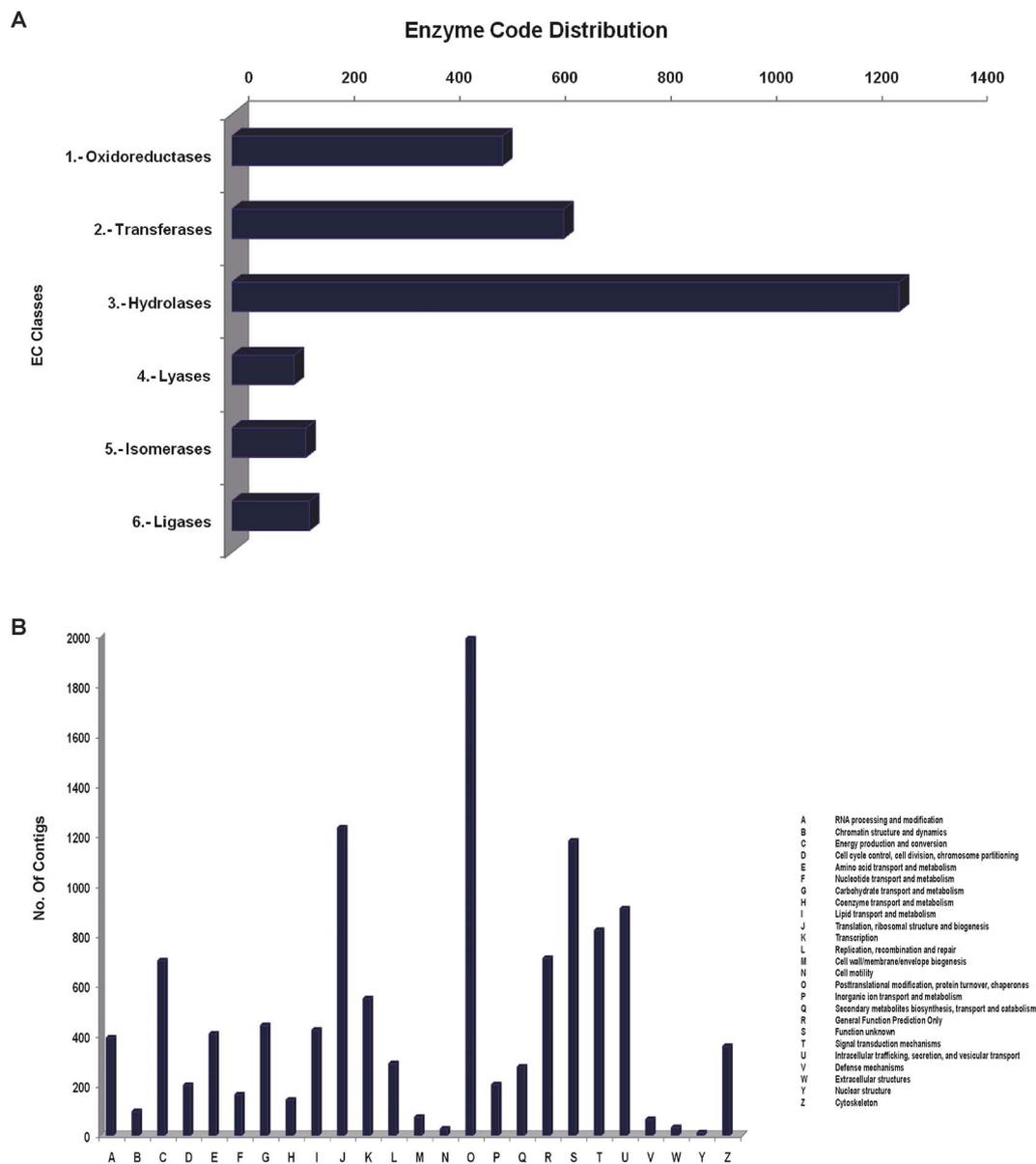


Fig. 5. (A) Distribution of the enzyme classes in the annotated contig sequences of earthworm *E. fetida*. (B) Histogram presenting the Clusters of Orthologous Groups (COG) classification of the assembled contig sequences.

mapped to the nerve cord transcripts, 23 pathways were functionally enriched and the transcripts associated with nucleotide metabolism, lipid metabolism and metabolism of cofactors and vitamins were abundant in the transcriptome dataset.

3.9. Identification of neuropeptide specific genes

Neuropeptides are peptidergic neurotransmitters capable of transmitting the cellular signals to facilitate the neuronal communication [87]. The neuropeptides are widely distributed in the central and peripheral nervous system and regulate a wide variety of cellular, physiological and behavioral processes like metabolism, energy expenditure, reproduction, circadian rhythmicity, pain sensation, memory and learning ability [88]. In invertebrate species the neuropeptides are mainly confined in the neurosecretory cells, inter neurons and motor neurons and act as a synaptic modulator to affect the neural circuits and neuromuscular junction in the central nervous system [89]. Especially in nematodes like *C. elegans* the neuropeptides control the muscle contraction and convey the convulsive behavior through

excitatory and inhibitory imbalance of the cholinergic and GABAergic motor neurons of the species [90]. In the annelid organisms like earthworm, leeches and other polychaetes > 40 neuropeptides were identified so far, showing high sequence similarity to the reported neuropeptides of arthropods, molluscs and vertebrates [91]. Among the earthworm species *E. fetida* has been thoroughly investigated to identify the putative neuropeptides and interpret their function in the physiological regulation of the species. The neuropeptides are mainly observed in the ventral nerve cord ganglia and somatogastric nervous system of the worm and known to regulate the physiological and reproductive activities like muscle contraction, gut motility, osmoregulation and egg laying behaviors associated with changes in body shapes, rotatory movements and mucosal secretion [15,16,92,93]. In addition the neuropeptides also participate in the embryogenesis and caudal regeneration process of the species [16,94]. The nerve cord specific transcriptome sequences of earthworm *E. fetida* annotated against the nr database have marked the presence of total 37 contigs homologous to 22 neuropeptides (Table 3). Most of the neuropeptide homologs in the nr BLAST alignment were retrieved from the marine

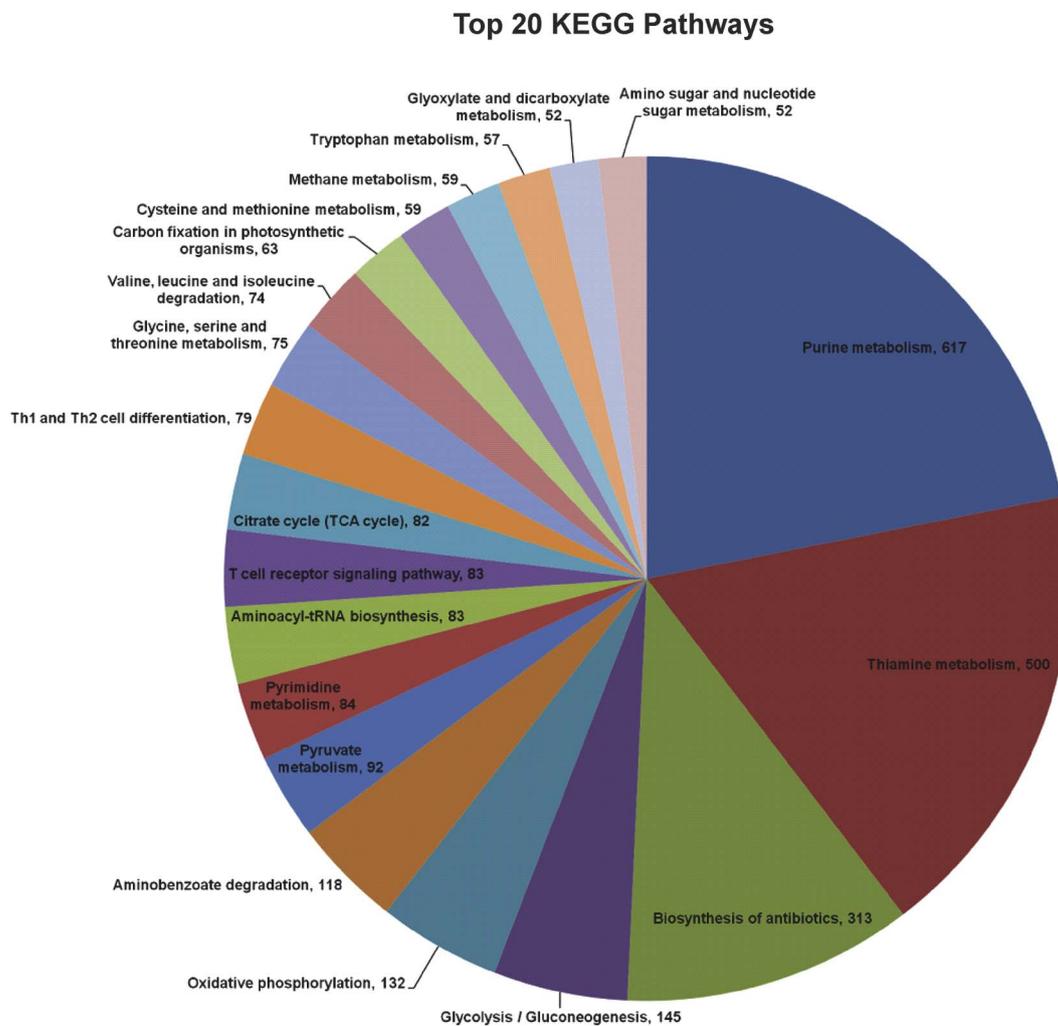


Fig. 6. Pie chart denoting the distribution of top 20 KEGG pathways associated with the annotated contigs of earthworm *Eisenia fetida*.

annelid *Platynereis dumerilii*. Among the identified neuropeptides XPRLamide neuropeptide precursor, whitin-1 neuropeptide precursor SPTR, QERAS neuropeptide precursor, WI neuropeptide and allatostatin-C neuropeptide precursor showed high sequence homology (80–100%) in the BLAST search. XPRLamide is an insect pyrokinin related neuropeptide that plays regulatory role in insect development through diapause termination and assist in sexual communication through pheromone biosynthesis [95]. Whitin gene is a proctolin neuropeptide homolog commonly observed in the annelid *Platynereis dumerilii* and gastropod *Haliotis asinina* [96,97]. Proctolin functions as the myotropic neuromuscular transmitter to regulate the visceral and skeletal muscle contraction through neuromodulatory activity [98]. Allatostatin-C neuropeptide is widely distributed throughout the peripheral and ventral nerve cord ganglia of arthropod species *Calanus finmarchicus* and modulates their feeding and locomotory activity for physiological adaptation [99]. In lobster *Homarus americanus* Allatostatin-C has modulatory role on the cardiac ganglion to trigger the rhythmic heart contraction [100]. In addition the transcriptome data also denoted the expression of neuropeptide Y and FMRF-amide neuropeptides in the nerve cord tissue. Both these neuropeptides along with proctolin are known to be expressed in the somatogastric nervous system of *E. fetida* to regulate the muscle cells in the alimentary canal of the worm [15].

3.10. Identification of potential stem cell markers

Annelids are renowned for their phenomenal ability to regenerate the amputated body parts [4]. Among the annelid species the earthworm *E. fetida* have been immensely exploited to monitor the different aspects of annelid regeneration. The regeneration potential and rate of survival of *E. fetida* are positively correlated with the number of segments retained in worm post amputation [13]. Confirming the certain influence of nerve cord in regeneration process, Morgan, 1902 reported the presence of ventral nerve cord at the amputated region is essentially required for the head regeneration in earthworm *Alloobophora foetida* [101]. Subsequent studies have indicated that the regeneration of ventral nerve cord in worm upon amputation is correlated with the conduction of impulses for regulating the coordinated movement and locomotory activity [102]. Recently, Nino et al., 2016 demonstrated that during the anterior regeneration in the brain amputated earthworm *Eudrilus eugeniae* the nerve cord takes control of the neurological activities and regulates the segmental nerves to assist in worm's survival [7]. The process of regeneration in annelids is a combinatorial outcome of epimorphosis and morphallaxis and the formation of blastema during epimorphosis is supported by the neoblast stem cells residing particularly in the epidermal cell layers [103–105]. Especially the post amputation and wound healing process enables the synthesis and activation of several growth factors and signaling molecules, which assist in signal transduction by triggering the stem cell proliferation to facilitate the process of tissue regeneration and organogenesis [106]. The tissue

Table 3
List of Neuropeptides with their gene ontology details.

Sl. No.	SeqName	Description	Length	E-value	Sim mean	GO names
1	Newbler_Contig0720505	CLCCY neuropeptide precursor [<i>Platynereis dumerilii</i>]	235	4.58E-07	70	P:neuropeptide signaling pathway; C:extracellular region
2	Newbler_Contig08652	XPRLamide neuropeptide precursor [<i>Eisemia fetida</i>]	183	1.03E-32	100	P:signal transduction
3	Newbler_Contig11852	XPRLamide neuropeptide precursor [<i>Eisemia fetida</i>]	146	1.03E-21	100	P:signal transduction
4	Newbler_Contig13284	FMRF-amide neuropeptides-like [<i>Biomphalaria glabrata</i>]	237	7.94E-16	67.5	P:neuropeptide signaling pathway
5	Newbler_Contig18293	whitinin-1 neuropeptide precursor SPTR [<i>Platynereis dumerilii</i>]	254	1.54E-16	71.67	C:cellular_component
6	Newbler_Contig24715	FMRFamide precursor [<i>Lumbricillus variegatus</i>]	485	2.48E-06	66	P:neuropeptide signaling pathway
7	Newbler_Contig25668	neuropeptide F [<i>Loxia gigantea</i>]	309	4.55E-08	69.33	P:neuropeptide signaling pathway; C:membrane; C:integral component of membrane; C:extracellular region; F:hormone activity
8	Newbler_Contig26118	SYPGVILRF-amide 3 [<i>Aplysia californica</i>]	305	4.87E-17	61.56	C:extracellular space; P:neuropeptide signaling pathway; P:embryo development ending in birth or egg hatching; F:neuropeptide hormone activity; C:extracellular region; F:neuropeptide receptor binding; F:molecular_function; F:neuropeptide Y receptor binding
9	Newbler_Contig26513	whitinin-1 neuropeptide precursor SPTR [<i>Platynereis dumerilii</i>]	577	7.10E-20	77.83	P:signal transduction; C:cellular_component
10	Newbler_Contig29900	WI neuropeptide partial [<i>Platynereis dumerilii</i>]	351	1.41E-15	77	P:neuropeptide signaling pathway
11	Newbler_Contig30157	prohormone-3 neuropeptide precursor [<i>Platynereis dumerilii</i>]	481	1.83E-43	60.5	P:neuropeptide signaling pathway
12	Seqman_Contig_11518	neuropeptide Y [<i>Lymnaea stagnalis</i>]	224	4.04E-14	73.25	P:neuropeptide signaling pathway; C:extracellular region; F:hormone activity
13	Seqman_Contig_11560	MIP allatostatin B neuropeptide precursor [<i>Platynereis dumerilii</i>]	816	1.01E-17	52.5	P:neuropeptide signaling pathway
14	Seqman_Contig_18656	pedal peptide neuropeptide precursor 2 MLDpeptide [<i>Platynereis dumerilii</i>]	231	1.10E-10	61	P:neuropeptide signaling pathway
15	Seqman_Contig_1992	WI neuropeptide partial [<i>Platynereis dumerilii</i>]	253	2.76E-12	84	P:neuropeptide signaling pathway
16	Seqman_Contig_20437	Mesotoxin-neurophysin partial [<i>Corvus brachyrhynchos</i>]	280	7.46E-10	57.75	P:neuropeptide signaling pathway; F:neurohypophyseal hormone activity; C:extracellular region
17	Seqman_Contig_2485	whitinin-1 neuropeptide precursor SPTR [<i>Platynereis dumerilii</i>]	695	6.60E-28	72.2	P:neuropeptide signaling pathway
18	Seqman_Contig_27117	allatostatin-C neuropeptide precursor [<i>Platynereis dumerilii</i>]	682	1.59E-15	67	P:neuropeptide signaling pathway
19	Seqman_Contig_3225	FVRLamide neuropeptide precursor [<i>Platynereis dumerilii</i>]	296	2.35E-07	57.75	P:neuropeptide signaling pathway
20	Seqman_Contig_36821	oxytocin-neurophysin 1 [<i>Xenopus (Silurana) tropicalis</i>]	153	2.43E-08	68.27	P:neuropeptide signaling pathway; F:neurohypophyseal hormone activity; C:extracellular region
21	Seqman_Contig_39560	allatostatin-C neuropeptide precursor [<i>Platynereis dumerilii</i>]	604	1.80E-06	80	P:neuropeptide signaling pathway
22	Seqman_Contig_40507	XPRLamide neuropeptide precursor [<i>Eisemia fetida</i>]	214	7.63E-39	80	P:neuropeptide signaling pathway
23	Seqman_Contig_40845	LYamide FDSIG neuropeptide precursor [<i>Platynereis dumerilii</i>]	136	2.23E-09	79.5	P:neuropeptide signaling pathway
24	Seqman_Contig_5359	CLCCY neuropeptide precursor [<i>Platynereis dumerilii</i>]	314	2.46E-07	64.5	P:neuropeptide signaling pathway; F:ion channel inhibitor activity; C:extracellular region; P:pathogenesis
25	Seqman_Contig_55026	whitinin-1 neuropeptide precursor SPTR [<i>Platynereis dumerilii</i>]	160	9.93E-12	70.75	C:membrane
26	Seqman_Contig_57817	FMRF-amide neuropeptides-like [<i>Biomphalaria glabrata</i>]	258	1.16E-17	62.63	P:neuropeptide signaling pathway
27	Seqman_Contig_58801	whitinin-1 neuropeptide precursor SPTR [<i>Platynereis dumerilii</i>]	605	2.32E-12	95.5	P:neuropeptide signaling pathway; C:integral component of membrane
28	Seqman_Contig_59269	QERAS neuropeptide precursor [<i>Platynereis dumerilii</i>]	235	5.04E-08	94	P:neuropeptide signaling pathway
29	Seqman_Contig_59283	neuropeptide F [<i>Loxia gigantea</i>]	591	6.28E-14	70.4	P:neuropeptide signaling pathway; C:extracellular region; F:hormone activity
30	Seqman_Contig_59977	prohormone-3 neuropeptide precursor [<i>Platynereis dumerilii</i>]	473	1.87E-12	49	P:neuropeptide signaling pathway
31	Seqman_Contig_60850	CLCCY neuropeptide precursor [<i>Platynereis dumerilii</i>]	340	3.14E-09	72	P:neuropeptide signaling pathway; C:extracellular region
32	Seqman_Contig_61369	Neuroendocrine 7B2 [<i>Zootermopsis nevadensis</i>]	344	1.19E-34	63.15	P:neuropeptide signaling pathway; C:secretory granule
33	Seqman_Contig_61694	prohormone-3 neuropeptide precursor [<i>Platynereis dumerilii</i>]	619	1.78E-28	52.5	P:neuropeptide signaling pathway
34	Seqman_Contig_62842	XPRLamide neuropeptide precursor [<i>Eisemia fetida</i>]	233	2.59E-26	95	P:neuropeptide signaling pathway

(continued on next page)

Table 3 (continued)

Sl. No.	SeqName	Description	Length	E-value	Sim mean	GO names
35	Seqman_Contig_7798	neuropeptide precursor [<i>Platyneris dumerilii</i>]	830	1.08E-10	54	P:neuropeptide signaling pathway
36	Seqman_Contig_7835	MIP allatostatin B neuropeptide precursor [<i>Platyneris dumerilii</i>]	191	6.26E-09	68.5	P:neuropeptide signaling pathway
37	Seqman_Contig_8949	MIP allatostatin B neuropeptide precursor [<i>Platyneris dumerilii</i>]	1026	4.21E-10	53.5	P:neuropeptide signaling pathway

homogenate of earthworm *E. fetida* showed the activation of EGF and FGF to carry out the signal transduction during regeneration [107]. In addition the crucial pluripotent stem cell markers like oct4, c-myc, sox2 and nanog plays regulatory role in assisting the regeneration process of the species [108]. In order to screen the genes associated with the stem cell differentiation we have compared the entire transcriptome dataset of the worm with 250 essential stem cell markers present in the human amniocytes [41]. The corresponding protein sequences of these stem cell markers were obtained from the UniProt database and locally blasted with the assembled contig dataset of earthworm *E. fetida* using E-value threshold 1E-5. Total 487 contigs homologous to 108 stem cell markers (Table S5) were identified from the BLAST Search. In the transcriptome dataset of earthworm *E. fetida* we have marked the presence of some crucial epigenetic regulators like *Wdr5*, *Kat5*, *Rbbp4* and *Cbx*; Transcriptional repressors like *Cnot1*, *Tle1* and *Yy1* and cell cycle specific genes like *Smc1a*, *Cdk2* and *Pcna* having high sequence homology with human in BLAST search. The BLAST data retrieved against the stem cell markers in the transcriptome dataset were further confirmed by matching them with the nr BLAST hits. Among the trithorax group protein *Wdr5* acts as a transcriptional activator for the self-renewal of the embryonic stem cells in mammalian system [109]. Simultaneously in invertebrate species like planaria the *Wdr5* forms complex with the *Smed* gene and knocking down of the *Smed-wdr5* complex has inhibitory effect on regeneration by reducing the blastema formation of the species [110]. The close evolutionary relationship between planaria and earthworm and the mechanistic similarity in their regeneration and organogenesis have prioritized the role of the gene in assisting the blastema formation of the worm.

3.11. Phylogenetic analysis of calcium binding proteins

Calcium acts as a crucial messenger molecule to facilitate the process of synaptic transmission, neuronal growth and axonal regeneration [59]. The domain based annotation of *E. fetida* nerve cord transcriptome has demonstrated the abundance of calcium binding proteins containing the EF hand domain. We have selected total 91 *E. fetida* contigs representing the calcium binding proteins and performed BLAST search to identify their orthologs in leech *H. robusta* and the polychaete *C. teleta*. The BLAST analysis has shown significant orthology to 74 *H. robusta* proteins and 84 *C. teleta* proteins respectively. The protein sequences were aligned together using Clustal W tool and a phylogenetic tree was constructed using the UPGMA method with bootstrap replicate of 100 (Fig. S2). The phylogenetic tree represented total 168 internal nodes and 170 leaf nodes. A few of the nodes lack the proper bootstrap support probably due to the construction of the phylogenetic tree from large set of gene families [111]. Among the EF hand domain containing transcripts the calbindin family proteins (Seqman_Contig_17763, Newbler_Contig27185, Seqman_Contig_9416, Newbler_Contig27868 and Seqman_Contig_24001) and the striated muscle regulatory proteins (Seqman_Contig_61120 and Newbler_Contig23758) were evolutionary closely related to their orthologs in the phylogenetic tree. Among the members of calbindin family, the multifunctional protein calretinin is widely distributed in the central and peripheral nervous system and it performs heterogeneous functions ranging from regulating the calcium homeostasis to trigger the nerve impulses, maintaining the neuronal physiology, ensuring neuronal protection by reducing the neuronal cell death and supporting neurogenesis [112,113]. The protein is evolutionary conserved and having high sequence identity with its invertebrate ortholog calbindin 32 [113]. The close evolutionary relationship of the calretinin and calbindin 32 proteins in our phylogenetic analysis suggested that the protein is functionally conserved in the annelids and it may play essential role in regulating their neurophysiology and neuronal development. Simultaneously our evolutionary data also depicted the event of tandem gene duplication for the sarcoplasmic calcium-binding proteins (Seqman_Contig_3135, Newbler_Contig21647, Newbler_Contig26378,

Seqman_Contig_60438, Seqman_Contig_61888, Seqman_Contig_60208, Newbler_Contig24773, Newbler_Contig07358 and Seqman_Contig_6452) in the transcriptome dataset of earthworm *E. fetida*. Among the other proteins the 20 kDa calcium-binding (Antigen SM20) protein (Newbler_Contig24366) was clustered together with the troponin C protein (Seqman_Contig_8254) as a monophyletic clade and showed significant evolutionary relatedness with their orthologs. In contrast, the neuronal calcium sensors (Newbler_Contig25072 and Seqman_Contig_10943) and neurocalcin (Seqman_Contig_31579) proteins were phylogenetically distantly related with their orthologs in *H. robusta* and *C. teleta*.

4. Conclusion

In the scarcity of proper genome information for the annelid earthworms the transcriptome sequencing has emerged as a perfect substitute to explore them at cellular, molecular and genetic level. The cerebral ganglion and nerve cord of earthworms are crucial organs for regulating their movement, rhythmicity, regeneration, social interaction and physiological behavior. However, the nerve cord specific transcripts and their function features were not explored in great details. The present study emphasizes the role of earthworm nerve cord by providing a complete annotation summary of the nerve cord transcripts and their functions. In our study 1. we have represented the complete functional information regarding 8819 nerve cord transcripts, 2. rationalized the involvement of earthworm nerve cord in maintaining the sensory and behavioral physiology of the species based on their transcriptomic perspective and 3. identified 22 neuropeptide transcripts and 108 stem cell markers reside in the nerve cord tissue of earthworm *Eisenia fetida*. The complete transcriptome annotation summary of the nerve cord specific transcripts of earthworm can be further utilized as a useful resource to interpret the genetic and biochemical pathways associated with the neurobiological and physiological process of the species.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gdata.2017.10.002>.

List of Abbreviations

GABA	Gamma-Amino butyric acid
ETP	Eisenia tetradecapeptide
EST	Express Sequence Tag
GPCR	G protein Couple receptors
NGS	Next Generation Sequencing
GO	Gene Ontology
COGs	Clusters of Orthologous Groups
EggNog	Evolutionary genealogy of genes
KEGG	Kyoto Encyclopedia of Genes and Genomes
Wdr5	WD repeat-containing protein 5
Kat5	Histone acetyltransferase KAT5
RBBP4	Histone-binding protein RBBP4
Cbx	Chromobox family
Cnot1	CCR4-NOT transcription complex subunit 1
Tle1	Transducin-like enhancer protein 1
Yy1	Transcriptional repressor protein YY1
Smc1a	Structural maintenance of chromosomes 1A
Cdk2	Cyclin-dependent kinase 2
Pcna	Proliferating cell nuclear antigen

Conflict of interests

The authors declare no potential conflicts of interest.

Transparency document

The <http://dx.doi.org/10.1016/j.gdata.2017.10.002> associated

with this article can be found, in the online version.

Acknowledgements

This work was financially supported by the Bioinformatics Infrastructure Facility (BIF), Department of Biotechnology (DBT) New Delhi, India [BT/B1/04/055/2001].

References

- [1] T. Matheson, Invertebrate nervous systems, eLS (2002).
- [2] C.A. Edwards, P.J. Bohlen, Biology and Ecology of Earthworms, Springer Science & Business Media, 1996.
- [3] A.J. Reinecke, S.A. Reinecke, Earthworms as test organisms in ecotoxicological assessment of toxicant impacts on ecosystems, Earthworm Ecol. (2004) 299–320.
- [4] G. Gates, Regeneration in an earthworm, *Eisenia foetida* (Savigny) 1826. I. Anterior regeneration, Biol. Bull. 96 (1949) 129–139.
- [5] W.N. Hess, Nervous system of the earthworm, *Lumbricus terrestris* L., J. Morphol. 40 (1925) 235–259.
- [6] A. Koza, M. Wilhelm, L. Hiripi, K. Elekes, M. Csoknya, Embryogenesis of the serotonergic system in the earthworm *Eisenia fetida* (Annelida, Oligochaeta): Immunohistochemical and biochemical studies, J. Comp. Neurol. 497 (2006) 451–467.
- [7] N.G. Daisy, E.R. Subramanian, J.D.S. Christyraj, D.K.S. Mani, J.R.S.S. Christyraj, K. Ramamoorthy, V. Arumugaswami, S. Sivasubramanian, Studies on regeneration of central nervous system and social ability of the earthworm *Eudrilus eugeniae*, Invertebr. Neurosci. 16 (2016) 1–13.
- [8] C.D. Drewes, Escape reflexes in earthworms and other annelids, Neural mechanisms of startle behavior, Springer, 1984, pp. 43–91.
- [9] E. Vining, C. Drewes, Restoration of sensory and motor function in earthworm escape reflex pathways following ventral nerve cord transplantation, J. Neurobiol. 16 (1985) 301–315.
- [10] E.F. Neuhauser, R. Hartenstein, D.L. Kaplan, Growth of the earthworm *Eisenia foetida* in relation to population density and food rationing, Oikos (1980) 93–98.
- [11] J. Siddique, A.A. Khan, I. Hussain, S. Akhter, Growth and reproduction of earthworm (*Eisenia fetida*) in different organic media, Pak. J. Zool. 37 (2005) 211.
- [12] R. Hartenstein, E.F. Neuhauser, D.L. Kaplan, Reproductive potential of the earthworm *Eisenia foetida*, Oecologia 43 (1979) 329–340.
- [13] N. Xiao, F. Ge, C.A. Edwards, The regeneration capacity of an earthworm, *Eisenia fetida*, in relation to the site of amputation along the body, Acta Ecol. Sin. 31 (2011) 197–204.
- [14] H. Vries-Schoumacker, Fluorescence and ultrastructural localization of aminergic neurons in the nerve cord of *Eisenia foetida* (Annelida—Oligochaeta), Cell Tissue Res. 185 (1977) 351–360.
- [15] J. Barna, M. Csoknya, Z. Lázár, L. Bartho, J. Hamori, K. Elekes, Distribution and action of some putative neurotransmitters in the stomatogastric nervous system of the earthworm, *Eisenia fetida* (Oligochaeta, Annelida), J. Neurocytol. 30 (2001) 313–325.
- [16] Z. Herbert, E. Pollák, A. Zougman, A. Boros, N. Kapan, L. Molnár, Identification of novel neuropeptides in the ventral nerve cord ganglia and their targets in an annelid worm, *Eisenia fetida*, J. Comp. Neurol. 514 (2009) 415–432.
- [17] E. Varhalmi, I. Somogyi, G. Kiszler, J. Nemeth, D. Reglodi, A. Lubics, P. Kiss, A. Tamas, E. Pollak, L. Molnar, Expression of PACAP-like compounds during the caudal regeneration of the earthworm *Eisenia fetida*, J. Mol. Neurosci. 36 (2008) 166–174.
- [18] M.S. Lee, S.J. Cho, E.S. Tak, J.A. Lee, H.J. Cho, B.J. Park, C. Shin, D.K. Kim, S.C. Park, Transcriptome analysis in the midgut of the earthworm (*Eisenia andrei*) using expressed sequence tags, Biochem. Biophys. Res. Commun. 328 (2005) 1196–1204.
- [19] R. Ekblom, J. Galindo, Applications of next generation sequencing in molecular ecology of non-model organisms, Heredity 107 (2011) 1–15.
- [20] H. Sadamoto, H. Takahashi, T. Okada, H. Kenmoku, M. Toyota, Y. Asakawa, De novo sequencing and transcriptome analysis of the central nervous system of mollusc *Lymnaea stagnalis* by deep RNA sequencing, PLoS One 7 (2012) e42546.
- [21] L. Badisco, J. Huybrechts, G. Simonet, H. Verlinden, E. Marchal, R. Huybrechts, L. Schoofs, A. De Loof, J.V. Broeck, Transcriptome analysis of the desert locust central nervous system: production and annotation of a *Schistocerca gregaria* EST database, PLoS One 6 (2011) e17274.
- [22] A.J. Northcutt, K.M. Lett, V.B. Garcia, C.M. Diester, B.J. Lane, E. Marder, D.J. Schulz, Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation, BMC Genomics 17 (2016) 868.
- [23] J. Owen, B.A. Hedley, C. Svendsen, J. Wren, M.J. Jonker, P.K. Hankard, L.J. Lister, S.R. Stürzenbaum, A.J. Morgan, D.J. Spurgeon, Transcriptome profiling of developmental and xenobiotic responses in a keystone soil animal, the oligochaete annelid *Lumbricus rubellus*, BMC Genomics 9 (2008) 266.
- [24] S.-J. Cho, M.S. Lee, E.S. Tak, E. Lee, K.S. Koh, C.H. Ahn, S.C. Park, Gene expression profile in the anterior regeneration of the earthworm using expressed sequence tags, Biosci. Biotechnol. Biochem. 73 (2009) 29–34.
- [25] Y. Mikami, A. Fukushima, T. Kuwada-Kusunose, T. Sakurai, T. Kitano, Y. Komiyama, T. Iwase, K. Komiyama, Whole transcriptome analysis using next-generation sequencing of sterile-cultured *Eisenia andrei* for immune system research, PLoS One 10 (2015) e0118587.

- [26] M. Pirooznia, P. Gong, X. Guan, L.S. Inouye, K. Yang, E.J. Perkins, Y. Deng, Cloning, analysis and functional annotation of expressed sequence tags from the earthworm *Eisenia fetida*, *BMC Bioinf.* 8 (2007) S7.
- [27] P. Gong, M. Pirooznia, X. Guan, E.J. Perkins, Design, validation and annotation of transcriptome-wide oligonucleotide probes for the oligochaete annelid *Eisenia fetida*, *PLoS One* 5 (2010) e14266.
- [28] M. Thunders, J. Cavanagh, Y. Li, De novo transcriptome assembly, functional annotation and differential gene expression analysis of juvenile and adult *E. fetida*, a model oligochaete used in ecotoxicological studies, *Biol. Res.* 50 (2017) 7.
- [29] S. Mayilswami, K. Krishnan, M. Megharaj, R. Naidu, Chronic PFOS exposure alters the expression of neuronal development-related human homologues in *Eisenia fetida*, *Ecotoxicol. Environ. Saf.* 110 (2014) 288–297.
- [30] S. Mayilswami, K. Krishnan, M. Megharaj, R. Naidu, Gene expression profile changes in *Eisenia fetida* chronically exposed to PFOA, *Ecotoxicology* 25 (2016) 759–769.
- [31] R.K. Patel, M. Jain, NGS QC toolkit: a toolkit for quality control of next generation sequencing data, *PLoS One* 7 (2012) e30619.
- [32] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, Gene ontology: tool for the unification of biology, *Nat. Genet.* 25 (2000) 25–29.
- [33] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, BLAST2GO: a universal tool for annotation, visualization and analysis in functional genomics research, *Bioinformatics* 21 (2005) 3674–3676.
- [34] N.J. Mulder, R. Apweiler, The InterPro database and tools for protein domain analysis, *Curr. Protoc. Bioinformatics* 21 (2008) 2.7.1–2.7.18 (2.7. 1–2.7. 18).
- [35] J. Huerta-Cepas, D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M.C. Walter, T. Rattei, D.R. Mende, S. Sunagawa, M. Kuhn, eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences, *Nucleic Acids Res.* (2015) (gkv1248).
- [36] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.* 28 (2000) 27–30.
- [37] A.S. Zwarycz, C.W. Nossa, N.H. Putnam, J.F. Ryan, Timing and scope of genomic expansion within Annelida: evidence from homeoboxes in the genome of the earthworm *Eisenia fetida*, *Genome Biol. Evol.* 8 (2015) 271–281.
- [38] K. Glass, M. Girvan, Annotation enrichment analysis: an alternative method for evaluating the functional properties of gene sets, *Sci Rep* 4 (2014).
- [39] C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.-Y. Li, L. Wei, KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases, *Nucleic Acids Res.* 39 (2011) W316–W322.
- [40] S. Al-Said, H. Al-Yousuf, Neuropeptides and immunocytochemistry in annelids, *Int. Rev. Cytol.* 133 (1992) 231–308.
- [41] C.T. Maguire, B.L. Demarest, J.T. Hill, J.D. Palmer, A.R. Brothman, H.J. Yost, M.L. Condic, Genome-wide analysis reveals the unique stem cell identity of human amniocytes, *PLoS One* 8 (2013) e53372.
- [42] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.* 22 (1994) 4673–4680.
- [43] Y. Loewenstein, E. Portugaly, M. Fromer, M. Linal, Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space, *Bioinformatics* 24 (2008) 141–149.
- [44] S. Kumar, M.L. Blaxter, Comparing de novo assemblers for 454 transcriptome data, *BMC Genomics* 11 (2010) 571.
- [45] J.R. Miller, S. Koren, G. Sutton, Assembly algorithms for next-generation sequencing data, *Genomics* 95 (2010) 315–327.
- [46] R. Che, Y. Sun, R. Wang, T. Xu, Transcriptomic analysis of endangered Chinese salamander: identification of immune, sex and reproduction-related genes and genetic markers, *PLoS One* 9 (2014) e87940.
- [47] A. Sayadi, E. Immonen, H. Bayram, G. Arnqvist, The de novo transcriptome and its functional annotation in the seed beetle *Callosobruchus maculatus*, *PLoS One* 11 (2016) e0158565.
- [48] O. Simakov, F. Marletaz, S.-J. Cho, E. Edsinger-Gonzales, P. Havlak, U. Hellsten, D.-H. Kuo, T. Larsson, J. Lv, D. Arendt, Insights into bilaterian evolution from three spiralian genomes, *Nature* 493 (2013) 526–531.
- [49] J.A. Veenstra, Neuropeptide evolution: neurohormones and neuropeptides predicted from the genomes of *Capitella teleta* and *Helobdella robusta*, *Gen. Comp. Endocrinol.* 171 (2011) 160–175.
- [50] M.K. Derbyshire, C.J. Lanczycki, S.H. Bryant, A. Marchler-Bauer, Annotation of functional sites with the Conserved Domain Database, *Database* 2012 (2012) bar058.
- [51] D.D. Leipe, E.V. Koonin, L. Aravind, STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer, *J. Mol. Biol.* 343 (2004) 1–28.
- [52] D.D. Leipe, Y.I. Wolf, E.V. Koonin, L. Aravind, Classification and evolution of P-loop GTPases and related ATPases, *J. Mol. Biol.* 317 (2002) 41–72.
- [53] E. Pathak, N. Atri, R. Mishra, Role of highly central residues of P-loop and its flanking region in preserving the archetypal conformation of Walker A motif of diverse P-loop NTPases, *Bioinformation* 9 (2013) 23–28.
- [54] B. Sperlágh, S.E. Vizi, Neuronal synthesis, storage and release of ATP, P.Pp, *Semin. Neurosci.* (1996) 175–186 (Elsevier).
- [55] P.P. Bertrand, ATP and sensory transduction in the enteric nervous system, *Neuroscientist* 9 (2003) 243–260.
- [56] H. Lodish, S.L. Zipursky, Molecular cell biology, *Biochem. Mol. Biol. Educ.* 29 (2001) 126–133.
- [57] J.M. Berg, J.L. Tymoczko, L. Stryer, *Biochemistry*, 5th edition, WH Freeman, 2002 P.Pp. (Available online through the NCBI Bookshelf: link).
- [58] A. Lewit-Bentley, S. Réty, EF-hand calcium-binding proteins, *Curr. Opin. Struct. Biol.* 10 (2000) 637–643.
- [59] C.R. Tessier, Considering calcium-binding proteins in invertebrates: multi-functional proteins that shape neuronal growth, *Neural Regen. Res.* 11 (2016) 208.
- [60] K.R. Hagel, J. Beriont, C.R. Tessier, *Drosophila* Cbp53E regulates axon growth at the neuromuscular junction, *PLoS One* 10 (2015) e0132636.
- [61] A.O. Estevez, R.H. Cowie, K.L. Gardner, M. Estevez, Both insulin and calcium channel signaling are required for developmental regulation of serotonin synthesis in the chemosensory ADF neurons of *Caenorhabditis elegans*, *Dev. Biol.* 298 (2006) 32–44.
- [62] M. Roberts, The giant fibre reflex of the earthworm, *Lumbricus Terrestris* L., *J. Exp. Biol.* 39 (1962) 229–237.
- [63] J.C. Eccles, R. Granit, J. Young, Impulsis in the Giant Nerve Fibres of Earthworms, (1932).
- [64] B. Martin, R.L. de Maturana, R. Brenneman, T. Walent, M.P. Mattson, S. Maudsley, Class II G protein-coupled receptors and their ligands in neuronal function and protection, *NeuroMolecular Med.* 7 (2005) 3–36.
- [65] M. Spehr, S.D. Munger, Olfactory receptors: G protein-coupled receptors and beyond, *J. Neurochem.* 109 (2009) 1570–1583.
- [66] D. Julius, J. Nathans, Signaling by sensory receptors, *Cold Spring Harb. Perspect. Biol.* 4 (2012) a005991.
- [67] E. Cau, P. Blader, Notch activity in the nervous system: to switch or not switch? *Neural Dev.* 4 (2009) 36.
- [68] E. Gazave, Q.I. Lemaître, G. Balavoine, The Notch pathway in the annelid *Platynereis*: insights into chaetogenesis and neurogenesis processes, *Open Biol.* 7 (2017) 160242.
- [69] P.E. Teal, N.T. Davis, J.A. Meredith, T.A. Christensen, J.G. Hildebrand, Role of the ventral nerve cord and terminal abdominal ganglion in the regulation of sex pheromone production in the tobacco budworm (Lepidoptera: Noctuidae), *Ann. Entomol. Soc. Am.* 92 (1999) 891–901.
- [70] W.R. Schafer, Genetics of egg-laying in worms, *Annu. Rev. Genet.* 40 (2006) 487–509.
- [71] M. Novo, A. Riesgo, A. Fernández-Guerra, G. Giribet, Pheromone evolution, reproductive genes, and comparative transcriptomics in Mediterranean earthworms (Annelida, Oligochaeta, Hormogastridae), *Mol. Biol. Evol.* (2013) (mst074).
- [72] A. Scherer, J.M. Graff, Calmodulin differentially modulates Smad1 and Smad2 signaling, *J. Biol. Chem.* 275 (2000) 41430–41438.
- [73] P. Pakrasi, S. Dey, Role of calmodulin in blastocyst formation in the mouse, *J. Reprod. Fertil.* 71 (1984) 513–517.
- [74] T.T. Su, D.H. Parry, B. Donahoe, C.-T. Chien, P.H. O'farrell, A. Purdy, Cell cycle roles for two 14-3-3 proteins during *Drosophila* development, *J. Cell Sci.* 114 (2001) 3445–3454.
- [75] O.F. Webb, T.J. Phelps, P.R. Bienkowski, P.M. Digrazia, D.C. White, G.S. Saylor, *Enzyme Nomenclature*, (1992).
- [76] M. Guppy, P. Hochachka, Role of dehydrogenase competition in metabolic regulation. The case of lactate and alpha-glycerophosphate dehydrogenases, *J. Biol. Chem.* 253 (1978) 8465–8469.
- [77] J.M. Berg, J.L. Tymoczko, L. Stryer, *Biochemistry*, P.Pp. W. H. Freeman and Company, New York, 2002.
- [78] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, The COG database: an updated version includes eukaryotes, *BMC Bioinf.* 4 (2003) 41.
- [79] T.B. Franklin, A. Krueger-Naug, D. Clarke, A.-P. Arrigo, R. Currie, The role of heat shock proteins Hsp70 and Hsp27 in cellular protection of the central nervous system, *Int. J. Hyperth.* 21 (2005) 379–392.
- [80] S.S. Ousman, A. Frederick, E.-M.F. Lim, Chaperone proteins in the central nervous system and peripheral nervous system after nerve injury, *Front. Neurosci.* 11 (2017).
- [81] E. Thomson, S. Ferreira-Cerca, E. Hurt, Eukaryotic ribosome biogenesis at a glance, P.Pp The Company of Biologists Ltd, 2013.
- [82] R. Voutev, D.J. Killian, J.H. Ahn, E.J.A. Hubbard, Alterations in ribosome biogenesis cause specific defects in *C. elegans* hermaphrodite gonadogenesis, *Dev. Biol.* 298 (2006) 45–58.
- [83] M. Göttle, H. Burhenne, D. Sutcliffe, H. Jinnah, Purine metabolism during neuronal differentiation: the relevance of purine synthesis and recycling, *J. Neurochem.* 127 (2013) 805–818.
- [84] G. Burnstock, A. Verkhatsky, Purinergic signaling, *Wiley Interdiscip. Rev.: Membr. Transp. Signal.* 1 (2012) 116–125.
- [85] J. Wang, *Functional Enrichment Analysis*, Encyclopedia of Systems Biology, Springer, 2013, p. 772.
- [86] K. Schmidt-Nielsen, *Animal Physiology: Adaptation and Environment*, Cambridge University Press, 1997.
- [87] N. Merlin, T. Joseph, S.S. Dharan, S. Mathan, Neuropeptides—a review, *Asian J. Pharm. Res.* 4 (2014) 198–200.
- [88] L.D. Fricke, Neuropeptides and other bioactive peptides: from discovery to function, P.Pp. Colloquium Series on Neuropeptides Morgan & Claypool Life Sciences, 2012, pp. 1–122.
- [89] D.R. Nässel, P.H. Taghert, Invertebrate neuropeptides. eLS, *Encyclopedia of life sciences & 2005*, John Wiley & Sons, Ltd., 27 Jan 2006, pp. 1–11, <http://dx.doi.org/10.1038/ngp.els.0004080> www.els.net.
- [90] T.M. Stawicki, S. Takayanagi-Kiyaya, K. Zhou, Y. Jin, Neuropeptides function in a homeostatic manner to modulate excitation-inhibition imbalance in *C. elegans*, *PLoS Genet.* 9 (2013) e1003472.
- [91] M. Salzet, Molecular aspect of annelid neuroendocrine system, (2007) (arXiv preprint arXiv:0708.2594).

- [92] Y. Fujino, T. Nagahama, T. Oumi, K. Ukena, F. Morishita, Y. Furukawa, O. Matsushima, M. Ando, H. Takahama, H. Satake, Possible functions of oxytocin/vasopressin-superfamily peptides in annelids with special reference to reproduction and osmoregulation, *J. Exp. Zool. A Ecol. Genet. Physiol.* 284 (1999) 401–406.
- [93] J.L. Quintanar, K. Gutiérrez-García, L. Castillo-Hernández, The neuropeptide Gonadotropin-releasing hormone modifies the spontaneous muscular contraction in the earthworm: *Eisenia foetida*, *Invertebr. Neurosci.* 11 (2011) 113–116.
- [94] A. Boros, D. Reglodi, Z. Herbert, G. Kiszler, J. Nemeth, A. Lubics, P. Kiss, A. Tamas, S. Shioda, K. Matsuda, Changes in the expression of PACAP-like compounds during the embryonic development of the earthworm *Eisenia fetida*, *J. Mol. Neurosci.* 36 (2008) 157–165.
- [95] J.-S. Sun, Q.-R. Zhang, T.-Y. Zhang, Z.-L. Zhu, H.-M. Zhang, M.-K. Teng, L.-W. Niu, W.-H. Xu, Developmental expression of FXPRLamide neuropeptides in peptidergic neurosecretory cells of diapause-and nondiapause-destined individuals of the cotton bollworm, *Helicoverpa armigera*, *Gen. Comp. Endocrinol.* 141 (2005) 48–57.
- [96] M. Conzelmann, E.A. Williams, K. Krug, M. Franz-Wachtel, B. Macek, G. Jékely, The neuropeptide complement of the marine annelid *Platynereis dumerilii*, *BMC Genomics* 14 (2013) 906.
- [97] P.S. York, S.F. Cummins, S.M. Degnan, B.J. Woodcroft, B.M. Degnan, Marked changes in neuropeptide expression accompany broadcast spawnings in the gastropod *Haliotis asinina*, *Front. Zool.* 9 (2012) 9.
- [98] D. Konopińska, G. Rosiński, Proctolin, an insect neuropeptide, *J. Pept. Sci.* 5 (1999) 533–546.
- [99] C.H. Wilson, A.E. Christie, Distribution of C-type allatostatin (C-AST)-like immunoreactivity in the central nervous system of the copepod *Calanus finmarchicus*, *Gen. Comp. Endocrinol.* 167 (2010) 252–260.
- [100] M. Stanhope, T.J. Lameyer, D.N. Shea, M. Chi, M.G. Pascual, D.J. Schulz, A.E. Christie, P.S. Dickinson, Mechanisms underlying differential responses to the neuropeptide Allatostatin-C (AST-C) in the cardiac ganglion of the lobster, *Homarus americanus*, *FASEB J.* 30 (2016) (760.761-760.761).
- [101] T. Morgan, Experimental studies of the internal factors of regeneration in the earthworm, *Archiv für Entwicklungsmechanik der Organismen* 14 (1902) 562–591.
- [102] A.R. Hall, Regeneration in the annelid nerve cord, *J. Comp. Neurol.* 33 (1921) 163–191.
- [103] V.G. Martinez, J. Manson, M.J. Zoran, Effects of nerve injury and segmental regeneration on the cellular correlates of neural morphallaxis, *J. Exp. Zool. B Mol. Dev. Evol.* 310 (2008) 520–533.
- [104] V.G. Martinez-Acosta, M.J. Zoran, Evolutionary aspects of annelid regeneration, *eLS* (2015).
- [105] E.M. Tanaka, P.W. Reddien, The cellular basis for animal regeneration, *Dev. Cell* 21 (2011) 172–185.
- [106] N. Odintsova, Stem cells of marine invertebrates: regulation of proliferation and differentiation processes *in vitro*, *Tsitologiya* 51 (2008) 367–372.
- [107] M. Grdiša, M. Popović, T. Hrženjak, Stimulation of growth factor synthesis in skin wounds using tissue extract (G-90) from the earthworm *Eisenia foetida*, *Cell Biochem. Funct.* 22 (2004) 373–378.
- [108] P. Zheng, Q. Shao, X. Diao, Z. Li, Q. Han, Expression of stem cell pluripotency factors during regeneration in the earthworm *Eisenia foetida*, *Gene* 575 (2016) 58–65.
- [109] Y.-S. Ang, S.-Y. Tsai, D.-F. Lee, J. Monk, J. Su, K. Ratnakumar, J. Ding, Y. Ge, H. Darr, B. Chang, Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network, *Cell* 145 (2011) 183–197.
- [110] A. Hubert, J.M. Henderson, K.G. Ross, M.W. Cowles, J. Torres, R.M. Zayas, Epigenetic regulation of planarian stem cells by the SET1/MLL family of histone methyltransferases, *Epigenetics* 8 (2013) 79–91.
- [111] M. Uliano-Silva, J.A. Americo, R. Brindeiro, F. Dondero, F. Prosdocimi, M. de Freitas Rebelo, Gene discovery through transcriptome sequencing for the invasive mussel *Limnoperna fortunei*, *PLoS One* 9 (2014) e102973.
- [112] M.B. Levanti, G. Montalbano, R. Laurà, E. Ciriaco, T. Cobo, O. García-Suarez, A. Germanà, J. Vega, Calretinin in the peripheral nervous system of the adult zebrafish, *J. Anat.* 212 (2008) 67–71.
- [113] B. Schwaller, Calretinin: from a “simple” Ca²⁺ buffer to a multifunctional protein implicated in many biological processes, *Front. Neuroanat.* 8 (2014).