First line of text
Authors
Chang Liu, Xuan Wang, Jiu-Bo Tian, Pei-Yuan Ma, Fan-Xiu Meng, Qi Zhang, Bao-Feng Yu, Rui Guo, Zhi-Zhen Liu, Hai-Long Wang, Jun Xie, Niu-Liang Cheng, Jian-Hua Wang, Bo Niu, and Gongqin Sun
Construction of a cDNA library and preliminary analysis of the expressed sequence tags of the earthworm Eisenia fetida (Savigny, 1826)

CHANG LIU1*, XUAN WANG1*, JIU-BO TIAN1, PEI-YUAN MA1, FAN-XIU MENG1, QI ZHANG1, BAO-FENG YU1, RUI GUO1, ZHI-ZHEN LIU1, HAI-LONG WANG1, JUN XIE1, NIU-LIANG CHENG1, JIAN-HUA WANG2, BO NIU1,2 and GONG-QIN SUN1,3

1Department of Biochemistry and Molecular Biology, Shanxi Medical University, Taiyuan, Shanxi 030001; 2Department of Biotechnology, Beijing Municipal Key Laboratory of Child Development and Nutriomics, Capital Institute of Pediatrics, Beijing 100020, P.R. China; 3Department of Cell and Molecular Biology, University of Rhode Island, Kingston, RI 02881, USA

Received June 24, 2018; Accepted January 11, 2019

DOI: 10.3892/mmr.2019.9938

Abstract. Earthworms are useful indicator organisms of soil health and Eisenia fetida have been extensively used as test organisms in ecotoxicological studies. In order to gain insight into the gene expression profiles associated with physiological functions of earthworms, a full-length enriched cDNA library of the Eisenia fetida genome was successfully constructed using Switching Mechanism at 5’End of RNA Template technology. Construction of a cDNA library and analysis of Expressed Sequence Tags (ESTs) are efficient approaches for collecting genomic information and identifying genes important for a given biological process. Furthermore, analysis of the expression abundance of ESTs was performed with the aim of providing genetic and transcriptomic information on the development and regenerative process of earthworms. Phrep and Crossmatch were used to process EST data and a total of 1,140 high-quality EST sequences were determined by sequencing random cDNA clones from the library. Clustering analysis of sequences revealed a total of 593 unique sequences including 225 contiguous and 368 singleton sequences. Basic Local Alignment Search Tool analysis against the Kyoto Encyclopedia of Genes and Genomes database resulted in 593 significant hits (P-value <1x10^-8), of which 168 were annotated through Gene Ontology analysis. The STRING database was used to determine relationships among the 168 ESTs, identifying associated genes involved in protein-protein interactions and gene expression regulation. Based on nucleic acid and protein sequence homology, the mutual relationships between 287 genes could be obtained, which identified a portion of the ESTs as known genes. The present study reports on the construction of a high-quality cDNA library representative of adult earthworms, on a preliminary analysis of ESTs and on a putative functional analysis of ESTs. The present study is expected to enhance our understanding of the molecular basis underlying the biological development of earthworms.

Introduction

Earthworms are terrestrial annelids in the oligochaeta subclass, with a generally preferred habitat of damp and loose soil. They include ~3,000 species worldwide, with 229 species in China (1,2). In a wide variety of soil types, earthworms serve vital roles in converting large pieces of organic matter into rich humus to enrich soil fertility. The earthworms are the highest evolutionary species capable of regenerating an anterior portion containing the central nervous system, heart and clitellum (3). The anterior regeneration is a unique developmental process that requires cell proliferation, re-differentiation and sophisticated cell-cell communication. This process can serve as a useful model for investigating normal development and differentiation (4).

Over the past several years, cDNA library construction and analysis have become established as indispensable methods for functional genome analysis since they provide detailed information about the genomic mechanisms underlying the diverse processes of an organism (5). However, conventionally
generated cDNA libraries contain a high percentage of 5'-truncated clones, limiting the utility of such libraries. The Switching Mechanism at 5'End of RNA Template (SMART) technique (6) amplifies and enriches the full-length mRNA, and thus generates cDNA libraries with a significantly improved ratio of full-length to partial cDNA sequences. In the present study, the SMART technique was adopted to construct a high quality library of full-length cDNAs representative of adult earthworms, namely of the earthworm Eisenia fetida (Savigny, 1826).

Unlike other model organisms, none of the oligochaete genomes have been sequenced to the best of our knowledge, and genomic research on earthworms lags behind that of other model species such as Mus musculus. In the absence of the full genome sequences, expressed sequence tags (ESTs) aid the rapid detection of expressed genes via sequence analysis, and are a significant resource for comparative and functional genomic studies (2). In addition, among the biological techniques for transcriptome analysis, the determination of ESTs is considered the simplest method for profiling the transcriptome, which is also particularly useful in the development of cDNA microarrays for systematic identification of differentially expressed genes (7). Analysis of ESTs is an effective method for rapidly analyzing gene expression, characterizing gene functions and discovering new genes that are important for specific developmental and physiological processes (8). The present study established 593 ESTs, representing 168 genes and 425 unknown tags, providing a gene expression profile of earthworm development. This collection of ESTs may provide a valuable basis for future research on the physiology of earthworms.

Materials and methods

Isolation of total RNA and mRNA. Eisenia fetida earthworms were purchased from Beijing Shuangqiao Farm (Beijing, China). Fully developed adult Eisenia fetida earthworms weighing 0.3–0.6 g (live weight) were selected for all experiments. All earthworm tissues were harvested and total RNA was isolated using TRIzol reagent (Invitrogen; Thermo Fisher Scientific, Inc., Waltham, MA, USA). The integrity of total RNA was analyzed by electrophoresis using 1% agarose gels. Isolation of poly(A) mRNA from total RNA was carried out using an MN-NucleoTrap® mRNA kit (Machery-Nagel GmbH & Co. KG, Düren, Germany) according to the manufacturer’s protocol. Oligo(dT) beads suspension was applied to total RNA and incubated at 68˚C for 5 min before eluting mRNA. Then ethidium bromide (EB) staining was applied and 1% agarose gel was used to visualize the result. The isolated mRNA was further vacuum concentrated using Concentrator plus™ (Eppendorf, Hamburg, Germany). The quantity and integrity of isolated mRNA were determined using a nanodrop spectrophotometer and agarose gel electrophoresis, respectively.

cDNA library construction. A total of ~8,048.4 ng mRNA was used for single-stranded cDNA synthesis. The purified mRNA was used as a template, Oligo(dT)18 with XhoI cleavage site was used as the primer, and first strand cDNA was transcribed at 42°C using SuperScript™ II RnaseH-Reverse (Thermo Fisher Scientific, Inc.). Then the mRNA was digested using RNaseH, and the resultant mRNA fragments were used as further primers. The first cDNA chain was used as a template for double-stranded cDNA synthesis, using DNA Polymerase I (Takara Biotechnology Co., Ltd., Beijing, China). The ends of the double-stranded cDNA were ligated by T4 DNA polymerase and the ligation products were purified by phenol/chloroform/isoamyl alcohol to remove excess impurities such as protein. Subsequently, the double-stranded DNA fragments were ligated into EcoRI Adaptor using T4 DNA ligase at 4°C overnight. Then the double-stranded DNA fragments were phosphorylated with T4 Polynucleotide Kinase and digested with XhoI. Following XhoI digestion of the double-stranded cDNA, producing XhoI sticky ends, a QIAEXII Gel Extraction kit (Beijing BioDev-Tech, Beijing, China) was used to recycle 0.5-4 Kb fragments. The recycled cDNA was preserved at -20°C. Then the cDNA was ligated into the pBluescript II SK+(+) XR vector (Promega Corporation, Madison, WI, USA) in a 3:1 molar ratio with T4 DNA ligase at 4°C overnight. To reduce the redundancy and avoid the under-representation of different transcript species, cDNA fragments with different fractionated sizes were balanced and subjected to library construction (9,10). Prior to transformation, mixing of all ligated products with microporous membranes was performed to remove salt ions. Subsequently the products were transformed into 5x10^7/ml DH10B competent cells (Thermo Fisher Scientific, Inc.), plated on agar plates (10 cm diameter) by pipetting the cells onto the middle of the plate and spreading, and monoclonal colonies were selected for PCR amplification. The inserted sequences in the plasmids were amplified by PCR using T3 primers (5’-ATTACCCCTACTAAAGGGA-3’) and T7 primers (5’-TATACGACTCACTATTAGGG-3’). The total volume of PCR reaction mixture was 20 µl, containing 1 µl template, 10 µl 2X Taq MasterMix (CWBio, Beijing, China), 1 µl T3 primers (10 pmol), 1 µl T7 primers (10 pmol) and 7 µl ddH₂O (CWBio). Cycling conditions were: 94°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 55°C for 40 sec and 72°C for 60 sec, followed by 72°C for 5 min.

Bioinformatic analysis. cDNA clones were selected randomly from the cDNA library and the vector sequences were trimmed from the raw sequence data using Vecscreen tool (www.ncbi.nlm.nih.gov/tools/vecscreen/) from the National Center for Biotechnology Information. The sequence of each EST was also edited, mainly to remove ambiguous bases and poor-quality sequences (nucleotide sequences <100 bp). All edited sequences were assembled into groups using SeqMan software version 8.0 (DNASTAR, Madison, WI, USA). The processed cDNA sequences were used to perform a BLAST search in the GenBank database to compare all available ESTs and genes to date (11). The Basic Local Alignment Search Tool (BLAST; blast.ncbi.nlm.nih.gov/BLAST.cgi) results with P-values <1x10^-8 were generally regarded as a significant match (12,13). A large-scale Unigene assembly of the ESTs was initiated to identify and functionally annotate as many unique transcripts as possible. BLAST analysis against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, and protein and nucleic acid databases was conducted for examination of biological functions. The ESTs homologous to known proteins were further annotated for Gene Ontology.
(GO; www.geneontology.org) terms and the GO analysis was carried out using WEB-based GEne SeT AnaLysis Toolkit (WebGestalt; www.webgestalt.org/option.php) (14,15).

Results

Construction of cDNA library. Obtaining an adequate quantity of high quality mRNA initially is the key to yielding a sufficient quantity of first-strand full-length cDNA by reverse transcription. In the present study, total RNA was extracted from the tissues of earthworms. As shown in Fig. 1, 28s and 18s bands were clearly visible in the electrophoresis gel of total RNA, indicating that the total RNA was obtained. The optical density (OD)\textsubscript{260}/OD\textsubscript{280} ratio for the total RNA was 2.04, well within the range of 1.8-2.1, indicating that the isolated total RNA was suitable for cDNA library construction.

Once the double-stranded cDNA was synthesized as described in the Materials and methods, the present study determined the size distribution of the products. Diffuse strips between 0.5-4.0 kb could be detected by 1% agarose gel electrophoresis, which indicated that double-stranded cDNAs were successfully synthesized. A cDNA library of 4.12x10\textsuperscript{5} clones was obtained and half of the bacteria were cultured for amplification, which produced a total of 1.4x10\textsuperscript{11} clones. Several colonies were selected, and the inserted sequences in the plasmids were amplified by PCR using T3 and T7 primers. The PCR products were detected using 1% agarose gel electrophoresis as clear bands. No nonspecific bands were identified and the recombinant rate was 97% (Fig. 2).

To investigate the quality of the full-length cDNA library, the lengths of the cDNA inserts were assessed. Sequence outputs were manually edited to remove vector and ambiguous sequences. Then, the sequence data of the cDNA clones obtained by random partial sequencing were searched in the NCBI GenBank using BLAST to identify similarities with sequences in the nucleic acid databases. An evaluation of cDNA insert size and its distribution revealed a low level of insert size bias in the final cDNA library. The majority of the cDNA inserts were larger than 500 bp.

EST analysis. Instead of the amplified library, the primary cDNA library was used to generate ESTs to reduce the redundancy of cDNA clones. Following the removal of the redundant sequences and low-quality sequences (<100 bp), 1,148 effective sequences (>100 bp) from the total cDNA sequences were obtained. As shown in Fig. 3, 53 ESTs were 100-300 bp, 261 ESTs were 300-500 bp, 828 ESTs were 500-700 bp and 6 ESTs were larger than 700 bp. Taken together, 1,148 ESTs were larger than 100 bp. Among them, the shortest sequence was 100 bp, the longest was 718 bp and the average length was ~452 bp. Following sequencing, a homology BLAST search and assembling of the data, 368 singletons and 225 contigs were obtained out of the 1,140 high-quality ESTs, as shown in Table I. Additionally, a total of 593 individual ESTs were analyzed and 168 ESTs annotated in GenBank with nematode homology (Table I).

GO annotation and bioinformatic analysis. The cDNA functions were classified using the GO database into the three

| Table I. Summary of ESTs obtained from the cDNA library of earthworms. |
|---------------------------------------------------------------|
| ESTs                         | Number |
|-------------------------------|--------|
| Total number of ESTs          | 1,256  |
| Total length of ESTs (bp)     | 568,140|
| Average length of ESTs (bp)   | 452.34 |
| Unique genes                  | 593    |
| Contigs                       | 225    |
| Singletons                    | 368    |
| Annotation                    | 168    |

ESTs, expressed sequence tags.

Figure 1. Result of the electrophoresis of total RNA. A total of 1.056 µg RNA was obtained from the tissues of earthworms. The optical density values were A\textsubscript{260}=0.880 and A\textsubscript{280}=0.430; A\textsubscript{260}/A\textsubscript{280}=2.04.

Figure 2. Amplified inserts of cDNA clones from the constructed cDNA library. M, DNA marker; Lanes 1-32, insert cDNA (>500 bp) clones from phage plaques.

Figure 3. Different groups of ESTs, separated by length. ESTs, expressed sequence tags.
The main categories of molecular functions, cell components and biological processes.

GO annotation of genes associated with molecular functions indicated that among the 168 ESTs, 46% (77/168 ESTs) were associated with growth and metabolic pathways, with the distribution of the 77 ESTs shown in Fig. 4A. Out of the 77 ESTs, 21% (16 ESTs) were associated with ‘proteolytic enzymes’, 16% (12 ESTs) with ‘protein ligases’, 14% (11 ESTs) with ‘oxido-reductases’, 13% (10 ESTs) with ‘energy release’, 10% (8 ESTs) with ‘signal transduction and cell communication’, 5% (4 ESTs) with ‘transport’ and only 3% (2 ESTs) with ‘post-translational modification’, ‘protein turnover’ and ‘chaperones’ (Fig. 4B).

Cellular components associated with the cDNAs included ‘myosin’, the ‘citrate lyase compound’, the ‘mitochondrial inner membrane translocase compound’, ‘microtubules’, the ‘mitochondrial inner membrane’ and ‘nucleosomes’. The proportions of cellular components are presented in Fig. 5. It can be observed that the proportion of ‘myosin’ among cellular components was the largest. Myosin, an actin-dependent molecular motor, is involved in a number of important functions in earthworms. In particular, the myosin network can drive movement and support different moving speeds of earthworms. This specific feature is closely related to the free moving ability of earthworms (16).

Regarding biological processes, known genes were determined as those presenting significant matches to protein sequences with known functions in non-redundant nucleotide databases. According to these biological functions, the biological processes component was divided into different functions including ‘larval development’ (46%), ‘changes of cell morphogenesis’ (6%), ‘the process of cytokinesis’ (12%), ‘post-translational protein modification’ (15%), ‘stress response’ (5%), ‘cell redox homeostasis’ (2%), ‘protein polymerization’ (2%) and ‘protein catabolism’ (12%), as shown in Fig. 6. It was concluded from the above data that promoting growth was considered to be an important biological function of genes associated with biological processes.

**KEGG pathway annotation.** KEGG is a collection of online databases describing pathways associated with biochemical, genomic and enzymatic processes. Furthermore, it provides annotations of biochemical pathways for the species in which the genome has been sequenced (17). In this analysis, proteins are not viewed as individual gene products but are organized into pathways and networks according to their biological function(s). In the present study, from the data in Table II, 15 of the 168 ESTs were revealed to be involved in metabolism. Notably, 4 of the 15 ESTs (27%) were involved in the ubiquitin-mediated proteolysis pathway, which was the most
represented. The glutathione metabolism (3 ESTs) and arachidonic acid metabolism (2 ESTs) pathways were the second and third most represented pathways, respectively. Additionally, chondroitin sulfate biosynthesis, heparan sulfate biosynthesis, riboflavin metabolism, selenoamino acid metabolism, \(\gamma\)-hexachlorocyclohexane degradation, and the fructose and mannose metabolism pathways were also represented.

**Mutual relationship between 287 genes.** Relationships among the 168 ESTs were analyzed using the STRING database, and a functional association network was determined with 287 nodes, as shown in Table III. Each node corresponds to a gene and each (weighted) edge represents the evidence of a functional association between the gene pair. Predicted potential regulators are presented in Fig. 7. It can be seen from the STRING results that the most associated nodes (blue) included the 19 genes: Ribosomal protein L (rpl)-1, ribosomal protein S (rps)-0, rpl-4, rpl-5, rps-13, rps-2, acidic ribosomal protein (rha)-1, translocon-associated protein-4, transcription factor BTF3 homolog, rpl-7A, iff-1, rps-17, elongation factor 1\(\alpha\)3, polyadenylate-binding protein-1, rps-28, translationally-controlled tumor protein homolog-1, rpl-18,
Table III. Relationships between the 287 associated genes, including several important parameters.

| node1     | node2     | node1_string_id | node2_string_id | node1_external_id | node2_external_id | coexpression | combined_score |
|-----------|-----------|-----------------|-----------------|-------------------|-------------------|--------------|---------------|
| rpl-5     | atp-2     | 502,561         | 496,479         | F54C9.5.1         | C34E10.6.3        | 0.866        | 0.871         |
| rpl-18    | snr-2     | 510,406         | 508,903         | Y45F10D.12.2      | W08E3.1           | 0.426        | 0.468         |
| trap-4    | rps-2     | 511,330         | 497,379         | Y56E3A.21.2       | C49H3.11.1        | 0.243        | 0.919         |
| sod-2     | daf-21    | 498,735         | 497,248         | F10D11.1.1        | C47E8.5.1         | 0            | 0.572         |
| rla-2     | rps-4     | 511,588         | 510,280         | Y62E10A.1.1       | Y43B11AR.4.2      | 0.981        | 0.998         |
| rpl-1     | daf-1     | 511,889         | 493,869         | Y71F9AL.13a.1     | B0464.1.1         | 0.492        | 0.523         |
| prdx-6    | sod-2     | 509,858         | 498,735         | Y38C1AA.11        | F10D11.1.1        | 0.233        | 0.726         |
| act-4     | rpo-5     | 505,013         | 493,798         | M03F4.2a          | B0393.1.1         | 0.426        | 0.425         |
| daf-21    | rps-13    | 497,248         | 495,308         | C47E8.5.1         | C16A3.9.1         | 0.402        | 0.474         |
| rpl-1     | daf-1     | 511,889         | 493,528         | Y71F9AL.13a.1     | B0041.4.1         | 0.8          | 0.999         |
| cyc-1     | atp-2     | 497,719         | 496,479         | C54G4.8.1         | C34E10.6.3        | 0.849        | 0.873         |
| rla-1     | act-4     | 509,817         | 505,013         | Y37E3.7.2         | M03F4.2a          | 0.538        | 0.538         |
| rpl-7A    | rps-17    | 509,604         | 506,860         | Y24D9A.4a         | T08B2.10.1        | 0.872        | 0.923         |
| rps-28    | rps-5     | 510,228         | 502,561         | Y41D4B.5.2        | F54C9.5.1         | 0.77         | 0.887         |
| rpl-18    | cyc-1     | 510,406         | 497,719         | Y45F10D.12.2      | C49H3.11.1        | 0.387        | 0.411         |
| rpl-1     | rps-0     | 509,457         | 493,798         | Y17G7B.7.2        | B0393.1.1         | 0.477        | 0.49          |
| daf-21    | daf-1     | 497,248         | 493,869         | C47E8.5.1         | B0464.1.1         | 0.281        | 0.543         |
| lsm-3     | rla-2     | 511,591         | 511,588         | Y62E10A.12.2      | Y62E10A.1.1       | 0            | 0.408         |
| rpl-1     | rps-4     | 511,889         | 510,280         | Y71F9AL.13a.1     | Y43B11AR.4.2      | 0.819        | 0.989         |
| tpi-1     | rps-0     | 509,457         | 493,798         | Y17G7B.7.2        | B0393.1.1         | 0.477        | 0.49          |
| rpl-1     | daf-1     | 511,889         | 497,248         | Y71F9AL.13a.1     | C47E8.5.1         | 0.251        | 0.415         |
| rpl-1     | daf-1     | 511,889         | 506,602         | Y62E10A.10.1      | T05G5.10          | 0.436        | 0.521         |
| rpl-1     | daf-1     | 511,889         | 509,187         | Y71F9AL.13a.1     | Y37E3.7.2         | 0.794        | 0.997         |
| tpi-1     | rha-4     | 509,457         | 506,234         | Y17G7B.7.2        | T01H3.1.1         | 0.483        | 0.483         |
| vha-4     | atp-2     | 506,234         | 496,479         | T01H3.1.1         | C34E10.6.3        | 0.304        | 0.724         |
| rps-2     | daf-1     | 511,889         | 497,248         | Y71F9AL.13a.1     | C47E8.5.1         | 0.251        | 0.415         |
| rps-2     | daf-1     | 511,889         | 505,013         | Y62E10A.11.1      | T05G5.10          | 0.429        | 0.469         |
| rpl-1     | rps-13    | 509,457         | 506,234         | Y17G7B.7.2        | T01H3.1.1         | 0.483        | 0.483         |
| rps-2     | daf-1     | 511,889         | 497,248         | Y71F9AL.13a.1     | C47E8.5.1         | 0.251        | 0.415         |
| rpl-1     | daf-1     | 511,889         | 506,234         | Y17G7B.7.2        | T01H3.1.1         | 0.483        | 0.483         |
| rps-2     | daf-1     | 511,889         | 505,013         | Y62E10A.11.1      | T05G5.10          | 0.429        | 0.469         |
| rpl-1     | rps-13    | 509,457         | 497,248         | Y17G7B.7.2        | T01H3.1.1         | 0.483        | 0.483         |
| rps-2     | daf-1     | 511,889         | 497,248         | Y71F9AL.13a.1     | C47E8.5.1         | 0.251        | 0.415         |
| rpl-1     | rps-13    | 509,457         | 497,248         | Y17G7B.7.2        | T01H3.1.1         | 0.483        | 0.483         |
Table III. Continued.

| node1  | node2  | node1_string_id | node2_string_id | node1_external_id | node2_external_id | coexpression | combined_score |
|--------|--------|-----------------|-----------------|-------------------|-------------------|--------------|----------------|
| crt-1  | rpl-5  | 509,852         | 502,561         | Y38A10A.5.1       | F54C9.5.1         | 0.452        | 0.465          |
| rps-28 | tpi-1  | 510,228         | 509,457         | Y41D4B.5.2        | Y17G7B.7.2        | 0.278        | 0.451          |
| rps-4  | iff-1  | 510,280         | 506,602         | Y43B11AR.4.2      | T05G5.10          | 0.64         | 0.888          |
| icd-1  | rps-13 | 497,806         | 495,308         | C56C10.8.1        | C16A3.9.1         | 0.856        | 0.856          |
| trap-4 | rla-1  | 511,330         | 509,817         | Y56A3A.21.2       | Y37E3.7.2         | 0.176        | 0.923          |
| nmy-1  | unc-54 | 502,225         | 498,853         | F52B10.1          | F11C3.3.1         | 0.102        | 0.565          |
| trap-4 | rps-13 | 511,330         | 495,308         | Y56A3A.21.2       | C16A3.9.1         | 0.148        | 0.942          |
| iff-1  | rps-13 | 506,602         | 495,308         | T05G5.10          | C16A3.9.1         | 0.614        | 0.614          |
| maoc-1 | ftn-2  | 498,218         | 497,902         | E04F6.3           | D1037.3.4         | 0.997        | 0.996          |
| rpl-18 | eft-3  | 510,406         | 505,324         | Y45F10D.12.2      | R03G5.1a.2        | 0.428        | 0.938          |
| exos-2 | cpf-1  | 512,052         | 500,247         | Y73B6BL.3         | F28C6.3           | 0.403        | 0.403          |
| rla-2  | rla-1  | 511,588         | 509,817         | Y62E10A.1.1       | Y37E3.7.2         | 0.999        | 0.999          |
| rla-2  | rps-17 | 509,182         | 506,860         | Y106G6H.2a.4      | T08B2.10.1        | 0.217        | 0.936          |
| tpi-1  | rps-17 | 509,457         | 506,860         | Y17G7B.7.2        | T08B2.10.1        | 0.341        | 0.412          |
| act-4  | rpl-4  | 510,182         | 509,723         | Y38A10A.5.1       | C56C10.8.1        | 0.432        | 0.523          |
| rpl-7A | rps-2  | 509,852         | 497,379         | Y38A10A.5.1       | Y56A3A.21.2       | 0.402        | 0.462          |
| trap-4 | rps-0  | 511,330         | 493,798         | Y24D9A.4a         | Y56A3A.21.2       | 0.977        | 0.536          |
| rla-2  | tct-1  | 509,604         | 499,979         | Y24D9A.4a         | Y56A3A.21.2       | 0.793        | 0.794          |
| exos-2 | tsm-3  | 512,052         | 511,581         | Y73B6BL.3         | Y62E10A.12.2      | 0.164        | 0.449          |
| rpl-18 | atp-2  | 510,406         | 496,479         | Y45F10D.12.2      | Y45F10D.12.2      | 0.476        | 0.659          |
| rps-2  | drs-1  | 497,379         | 493,869         | F57B9.6a.3        | F54C9.5.1         | 0.409        | 0.409          |
| rpl-5  | rps-2  | 502,561         | 497,379         | Y38A10A.5.1       | Y56A3A.21.2       | 0.174        | 0.911          |
| rpl-18 | rpl-5  | 510,406         | 497,379         | Y45F10D.12.2      | Y45F10D.12.2      | 0.476        | 0.659          |
| rps-2  | atp-2  | 509,852         | 493,798         | Y38A10A.5.1       | Y56A3A.21.2       | 0.273        | 0.949          |
| snr-2  | rps-0  | 508,903         | 493,798         | W08E3.1           | B0393.1.1         | 0.35         | 0.414          |
| unc-54 | daf-21 | 498,853         | 497,248         | F11C3.3.1         | C47E8.5.1         | 0.322        | 0.674          |
| rpl-18 | rps-0  | 510,406         | 493,798         | Y45F10D.12.2      | B0393.1.1         | 0.999        | 0.999          |
| rpl-7A | icd-1  | 509,604         | 497,379         | Y45F10D.12.2      | B0393.1.1         | 0.999        | 0.999          |
| rps-4  | rps-2  | 509,457         | 493,798         | Y17G7B.7.2        | C47E8.5.1         | 0.458        | 0.458          |
| rps-0  | rps-0  | 508,903         | 497,379         | F08E3.1           | B0393.1.1         | 0.35         | 0.414          |
| unc-54 | daf-21 | 498,853         | 497,248         | F11C3.3.1         | C47E8.5.1         | 0.322        | 0.674          |
| rpl-18 | rps-0  | 510,406         | 493,798         | Y45F10D.12.2      | B0393.1.1         | 0.999        | 0.999          |
| rpl-7A | icd-1  | 509,604         | 497,379         | Y45F10D.12.2      | B0393.1.1         | 0.999        | 0.999          |
| rps-0  | rps-0  | 508,903         | 497,379         | F08E3.1           | B0393.1.1         | 0.35         | 0.414          |
| unc-54 | daf-21 | 498,853         | 497,248         | F11C3.3.1         | C47E8.5.1         | 0.322        | 0.674          |
| node1   | node2   | node1_string_id | node2_string_id | node1_external_id | node2_external_id | coexpression | combined_score |
|---------|---------|-----------------|-----------------|-------------------|-------------------|--------------|----------------|
| rla-1   | tct-1   | 509,817         | 499,979         | Y37E3.7.2         | F25H2.11.2        | 0.998        | 0.999          |
| rps-17  | iff-1   | 506,860         | 506,602         | T08B2.10.1        | T05G5.10          | 0.569        | 0.678          |
| rps-28  | rla-1   | 510,228         | 509,817         | Y41D4B.5.2        | Y37E3.7.2         | 0.785        | 0.963          |
| try-1   | cyp-31A2| 513,205         | 503,484         | ZK546.15          | H02I12.8          | 0.543        | 0.543          |
| rpl-1   | rpl-5   | 511,889         | 502,561         | Y71F9AL.13a.1     | F54C9.5.1         | 0.799        | 0.995          |
| rps-28  | rps-0   | 510,228         | 493,798         | Y41D4B.5.2        | B0393.1.1         | 0.76         | 0.966          |
| act-4   | unc-54  | 505,013         | 498,853         | M03F4.2a          | F11C3.3.1         | 0.168        | 0.474          |
| rpl-5   | tct-1   | 502,561         | 499,979         | Y41C9.5.1         | F25H2.11.2        | 0.996        | 0.996          |
| rla-1   | rps-13  | 509,817         | 493,308         | Y37E3.7.2         | C16A3.9.1         | 0.999        | 0.999          |
| rpl-1   | rps-17  | 511,889         | 506,802         | Y71F9AL.13a.1     | T08B2.10.1        | 0.799        | 0.993          |
| rpl-1   | rla-2   | 511,889         | 511,588         | Y71F9AL.13a.1     | Y62E10A.1.1       | 0.772        | 0.996          |
| cey-1   | cpi-1   | 500,586         | 500,247         | M03F4.2a          | F11C3.3.1         | 0.168        | 0.474          |
| rpl-5   | rpl-7A  | 510,228         | 509,604         | Y41C9.5.1         | C16A3.9.1         | 0.999        | 0.999          |
| rpl-1   | rps-13  | 510,280         | 495,308         | Y43B11AR.4.2      | C16A3.9.1         | 0.999        | 0.999          |
| rpl-1   | rps-2   | 510,280         | 493,798         | Y41D4B.5.2        | C16A3.9.1         | 0.999        | 0.999          |
| rpl-1   | rps-0   | 506,860         | 493,798         | Y41D4B.5.2        | C16A3.9.1         | 0.999        | 0.999          |
| sqv-8   | ret-1   | 512,932         | 508,785         | ZK1307.5          | W06A7.3f          | 0            | 0.579          |
| rps-28  | rpl-7A  | 510,228         | 509,604         | Y41D4B.5.2        | Y24D9A.4a         | 0.752        | 0.838          |
| iced-1  | rps-0   | 497,806         | 493,308         | C56C10.8.1        | B0393.1.1         | 0.875        | 0.88           |
| rpl-1   | rps-13  | 509,182         | 495,308         | Y106G6H.2a.4      | C16A3.9.1         | 0.327        | 0.928          |
| sqv-8   | ret-1   | 506,256         | 496,567         | T02C12.1          | C35D10.16         | 0            | 0.421          |
| rpl-1   | rps-13  | 510,280         | 495,308         | Y43B11AR.4.2      | C16A3.9.1         | 0.999        | 0.999          |
| rpl-1   | rps-2   | 510,280         | 493,798         | Y41D4B.5.2        | C16A3.9.1         | 0.999        | 0.999          |
| rpl-5   | rps-13  | 502,561         | 495,308         | F54C9.5.1         | C16A3.9.1         | 0.999        | 0.999          |
| hum-5   | arx-6   | 506,256         | 496,567         | T02C12.1          | C35D10.16         | 0            | 0.421          |
| rpl-1   | rps-13  | 510,280         | 495,308         | Y43B11AR.4.2      | C16A3.9.1         | 0.999        | 0.999          |
| rpl-1   | rps-2   | 510,280         | 493,798         | Y41D4B.5.2        | C16A3.9.1         | 0.999        | 0.999          |
| rpl-5   | rps-13  | 502,561         | 495,308         | F54C9.5.1         | C16A3.9.1         | 0.999        | 0.999          |
| node1_ string_id | node2_ string_id | node1_ external_id | node2_ external_id | combined_score | coexpression | combined_score | coexpression |
|-----------------|-----------------|--------------------|--------------------|----------------|--------------|----------------|--------------|
| Y43B11AR.4.2    | C49H3.11.1      | 0.999              | 0.999              |
| Y37E3.7.2       | Y24D9A.4a       | 0.927              | 0.912              |
| Y41D4B.5.2      | T08B2.10.1      | 0.805              | 0.989              |
| Y45F10D.12.2    | F25H2.11.2      | 0.963              | 0.966              |
| Y62E10A.1.1     | Y41D4B.5.2      | 0.799              | 0.879              |
| R03G5.1a.2      | B0041.4.1       | 0.713              | 0.97               |
| Y17G7B.7.2      | B0041.4.1       | 0.489              | 0.948              |
| Y17G7B.7.2      | K10B3.7.2       | 0.608              | 0.988              |
| F11C3.3.1       | C38C3.5b.1      | 0                   | 0.587              |
| Y37E3.7.2       | F33A8.3.2       | 0                   | 0.403              |
| Y24D9A.4a       | F11C3.3.1       | 0                   | 0.54               |
| Y41D4B.5.2      | Y106G6H.2a.4    | 0.344              | 0.955              |
| Y43B11AR.4.2    | C56C10.8.1      | 0.809              | 0.81               |
| Y24D9A.4a       | F54C9.5.1       | 0.934              | 0.994              |
| R03G5.1a.2      | B0041.4.1       | 0.187              | 0.916              |
| Y45F10D.12.2    | B0041.4.1       | 0.999              | 0.999              |
| Y62E10A.1.1     | Y106G6H.2a.4    | 0.38               | 0.548              |
| Y43B11AR.4.2    | F25H2.11.2      | 0.941              | 0.946              |
| Y41D4B.5.2      | Y41D4B.5.2      | 0.76               | 0.971              |
| Y62E10A.1.1     | Y41D4B.5.2      | 0.967              | 0.999              |
| Y43B11AR.4.2    | F54C9.5.1       | 0.999              | 0.999              |
| Y106G6H.2a.4    | Y106G6H.2a.4    | 0.29               | 0.929              |
| Y62E10A.1.1     | C49H3.11.1      | 0.991              | 0.999              |
| Y71F9AL.13a.1   | Y106G6H.2a.4    | 0.237              | 0.918              |
| Y41D4B.5.2      | C49H3.11.1      | 0.732              | 0.981              |
| Y43B11AR.4.2    | B0041.4.1       | 0.999              | 0.999              |
| Y62E10A.1.1     | T08B2.10.1      | 0.956              | 0.999              |
| Y43B11AR.4.2    | T08B2.10.1      | 0.956              | 0.996              |
| Y41D4B.5.2      | Y106G6H.2a.4    | 0.796              | 0.96               |
| F57B9.6a.3      | B0393.1.1       | 0.414              | 0.433              |
| F11B7.4d        | F08B6.4a        | 0.413              | 0.421              |
| Y37E3.7.2       | B0041.4.1       | 0.997              | 0.999              |
| Y62E10A.1.1     | C49H3.11.1      | 0.998              | 0.999              |
| Y43B11AR.4.2    | Y106G6H.2a.4    | 0.796              | 0.96               |
| Y03E6.7.2       | F28C6.3         | 0                   | 0.408              |
| Y56B6.4a        | F33A8.3.2       | 0                   | 0.55               |
| T02C12.1        | F52B10.1        | 0                   | 0.544              |
| Y62E10A.1.1     | Y17G7B.7.2      | 0.444              | 0.444              |
| Y62E10A.1.1     | Y56A3A.21.2     | 0.076              | 0.916              |
| Y41D4B.5.2      | B0041.4.1       | 0.756              | 0.847              |
| B0464.1.1       | B0393.1.1       | 0.464              | 0.494              |
| F33A8.3.2       | C54G4.8.1       | 0.382              | 0.54               |
| Y71F9AL.13a.1   | Y106G6H.2a.4    | 0.777              | 0.787              |
| Y41D4B.5.2      | C56C10.8.1      | 0.799              | 0.999              |
| Y17G7B.7.2      | B0393.1.1       | 0.099              | 0.999              |
| Y43B11AR.4.2    | Y106G6H.2a.4    | 0.444              | 0.444              |
| Y62E10A.1.1     | Y56A3A.21.2     | 0.076              | 0.916              |
| Y41D4B.5.2      | B0041.4.1       | 0.756              | 0.847              |
| B0464.1.1       | B0393.1.1       | 0.464              | 0.494              |
| F33A8.3.2       | C54G4.8.1       | 0.382              | 0.54               |
| Y71F9AL.13a.1   | Y106G6H.2a.4    | 0.777              | 0.787              |
| Y41D4B.5.2      | B0393.1.1       | 0.099              | 0.999              |
| Y17G7B.7.2      | Y17G7B.7.2      | 0.442              | 0.444              |
| Y43B11AR.4.2    | B0041.4.1       | 0.59               | 0.634              |
| Y62E10A.1.1     | Y56A3A.21.2     | 0.308              | 0.951              |
Table III. Continued.

| node1  | node2  | node1_string_id  | node2_string_id  | node1_external_id  | node2_external_id  | coexpression | combined_score |
|--------|--------|------------------|------------------|--------------------|--------------------|--------------|----------------|
| rpl-1  | rps-28 | 511,889          | 510,228          | Y71F9AL.13a.1      | Y41D4B.5.2        | 0.793        | 0.939          |
| rps-28 | tct-1  | 510,228          | 499,979          | Y41D4B.5.2        | F25H2.11.2        | 0.683        | 0.683          |
| eft-3  | rpl-5  | 505,324          | 502,561          | R03G5.1a.2        | F54C9.5.1         | 0.701        | 0.974          |
| sod-2  | cyc-1  | 498,735          | 497,719          | F10D11.1.1        | C54G4.8.1         | 0.456        | 0.604          |
| pdi-3  | phy-2  | 503,539          | 500,798          | H06O01.1.3        | F35G2.4.1         | 0            | 0.4            |
| rpl-18 | daf-21 | 510,406          | 497,248          | Y45F10D.12.2      | C47E8.5.1         | 0.387        | 0.401          |
| rpl-1  | tct-1  | 511,889          | 499,979          | Y71F9AL.13a.1     | Y2F152.11.2       | 0.751        | 0.757          |
| rpl-1  | iff-1  | 511,889          | 506,602          | Y71F9AL.13a.1     | T05G5.10          | 0.612        | 0.624          |
| rpl-1  | tpi-1  | 511,889          | 509,457          | Y71F9AL.13a.1     | Y17G7B.2.11.2     | 0.751        | 0.757          |
| rpl-1  | rps-0  | 493,798          | 493,528          | B0393.1.1         | B0041.4.1         | 0.999        | 0.999          |
| eft-3  | rps-13 | 505,324          | 495,308          | R03G5.1a.2        | C16A3.9.1         | 0.362        | 0.956          |
| act-4  | unc-60 | 505,013          | 496,717          | M03F4.2a          | C38C3.5b.1        | 0            | 0.863          |
| rps-4  | eft-3  | 505,324          | 509,182          | Y71F9AL.13a.1     | Y2F152.11.2       | 0.751        | 0.757          |
| rps-4  | rla-1  | 511,588          | 493,798          | Y62E10A.11.1      | T08B2.10.1        | 0.997        | 0.999          |
| rps-17 | rps-2  | 506,860          | 497,379          | Y71F9AL.13a.1     | Y2F152.11.2       | 0.751        | 0.757          |
| cyc-1  | eat-6  | 497,719          | 493,780          | C54G4.8.1         | B0365.3.2         | 0.295        | 0.425          |
| trap-4 | rpl-18 | 511,330          | 510,406          | Y56A3A.21.2       | Y45F10D.12.2      | 0.37         | 0.956          |
| rpl-18 | rps-0  | 509,457          | 493,528          | Y71F9AL.13a.1     | Y2F152.11.2       | 0.751        | 0.757          |
| rps-17 | rps-13 | 505,324          | 495,308          | R03G5.1a.2        | C16A3.9.1         | 0.362        | 0.956          |
| rps-17 | rps-0  | 509,182          | 493,798          | Y106G6H.2a.4      | Y106G6H.2a.4      | 0.425        | 0.453          |
| rpl-1  | trap-4 | 511,889          | 511,330          | Y71F9AL.13a.1     | Y2F152.11.2       | 0.751        | 0.757          |
| rpl-1  | tct-1  | 511,889          | 496,717          | M03F4.2a          | C38C3.5b.1        | 0            | 0.863          |
| rpl-1  | rps-28 | 511,889          | 510,228          | Y71F9AL.13a.1     | Y2F152.11.2       | 0.751        | 0.757          |
| rps-28 | tct-1  | 510,228          | 499,979          | Y41D4B.5.2        | F25H2.11.2        | 0.683        | 0.683          |
| eft-3  | rps-13 | 505,324          | 495,308          | R03G5.1a.2        | C16A3.9.1         | 0.362        | 0.956          |
| act-4  | unc-60 | 505,013          | 496,717          | M03F4.2a          | C38C3.5b.1        | 0            | 0.863          |
| rps-17 | rps-28 | 509,182          | 493,798          | B0393.1.1         | B0041.4.1         | 0.999        | 0.999          |
| eft-3  | rps-13 | 505,324          | 495,308          | R03G5.1a.2        | C16A3.9.1         | 0.362        | 0.956          |
| act-4  | unc-60 | 505,013          | 496,717          | M03F4.2a          | C38C3.5b.1        | 0            | 0.863          |
rps-4 and adenosine triphosphate synthase subunit β mitochondrial precursor-2. Of note, the genes rpl-4, rps-13, rpl-5, rpl-1, rps-0, rla-1 and rpl-7A served critical roles in overall gene expression networks.

Discussion

The present study reported on the construction of a high-quality cDNA library from tissues of the earthworm Eisenia fetida (Savigny, 1826), following preliminary analysis of ESTs, putative functional analysis of the ESTs and the gene expression pattern associated with the physiological functions of this organism. cDNA libraries are widely used to identify genes and splice variants and are considered to be a physical resource for the construction of full-length clones (18,19). In the present study, a cDNA library was utilized to provide a molecular resource for the analysis of genes involved in the specific biology of earthworms in terms of their development, survival, pathogenicity and virulence. There are two main factors to consider when assessing the quality of a cDNA library: Representation and cDNA lengths. According to Clarke-Carbon’s formula (20), a cDNA library should contain at least 1.7x10^5 independent clones to ensure that 99% of low-abundance mRNAs will be represented in the library (21). Furthermore, the average length of the inserted cDNAs should be no less than 1.0 kb to ensure the integrity of cDNAs, indicating that in the present study the fragment sizes were effective for ensuring full-length cDNAs in the cDNA library. Since selection bias could favor the smaller cDNAs, the present study used fewer PCR cycles to minimize such bias as previously suggested (10). In addition, up to 25 PCR amplification cycles were used to generate an adequate amount of cDNA for cloning.

Table III. Continued.

| node1 | node2 | node1_string_id | node2_string_id | node1_external_id | node2_external_id | coexpression | combined_score |
|-------|-------|----------------|----------------|-------------------|-------------------|--------------|----------------|
| rps-2 | atp-2 | 497,379        | 496,479        | C49H3.11.1        | C34E10.6.3        | 0.943        | 0.947          |
| unc-54| unc-87| 498,853        | 498,495        | F11C3.3.1         | F08B6.4a         | 0.566        | 0.776          |
| tct-1 | rps-2 | 499,979        | 497,379        | F25H2.11.2        | C49H3.11.1        | 0.999        | 0.999          |
| trap-4| rps-4 | 511,330        | 510,280        | Y56A3A.21.2       | Y43B11AR.4.2      | 0.245        | 0.921          |
| rps-28| iff-1 | 510,228        | 506,602        | Y41D4B.5.2        | T05G5.10          | 0.432        | 0.635          |
| rla-1 | daf-21| 509,817        | 497,248        | Y37E3.7.2         | C47E8.5.1         | 0.394        | 0.543          |
| rps-4 | crt-1 | 510,280        | 509,852        | Y43B11AR.4.2      | Y38A10A.5.1       | 0.491        | 0.5            |
| rpl-18| crt-1 | 510,406        | 509,852        | Y45F10D.12.2      | Y38A10A.5.1       | 0.595        | 0.597          |
| rps-2 | rpl-0 | 497,379        | 493,798        | C49H3.11.1        | B0393.1.1         | 0.999        | 0.999          |
| lev-11| unc-87| 509,147        | 498,495        | Y105E8B.1d        | F08B6.4a         | 0.272        | 0.454          |
| trap-4| rps-17| 511,330        | 506,860        | Y56A3A.21.2       | T08B2.10.1        | 0.308        | 0.929          |
| tct-1 | daf-21| 499,979        | 497,248        | F25H2.11.2        | C47E8.5.1         | 0.648        | 0.713          |
| rpl-1 | ral-4 | 509,852        | 493,528        | Y38A10A.5.1       | Y43B11AR.4.2      | 0.999        | 0.999          |
| rps-4 | rpl-0 | 509,406        | 506,802        | Y45F10D.12.2      | Y38A10A.5.1       | 0.595        | 0.597          |
| rps-2 | rpl-0 | 505,324        | 493,798        | R03G5.1a2         | B0393.1.1         | 0.999        | 0.999          |
| rpl-1 | rpl-1 | 506,602        | 497,806        | Y11F9AL.13a1      | R03G5.1a2         | 0.348        | 0.948          |
| rps-2 | rps-4 | 509,604        | 494,528        | Y38A10A.5.1       | B0401.1.1         | 0.472        | 0.604          |
| rpl-18| rpl-7A| 509,852        | 493,528        | Y38A10A.5.1       | Y43B11AR.4.2      | 0.999        | 0.999          |
| eft-3 | rps-4 | 505,324        | 497,902        | R03G5.1a2         | D1037.3.4         | 0.999        | 0.999          |
| rps-4 | daf-2 | 509,324        | 497,902        | Y38A10A.5.1       | B0401.1.1         | 0.995        | 0.999          |
| rpl-18| eft-3 | 507,806        | 497,379        | Y43B11AR.4.2      | B0393.1.1         | 0.999        | 0.999          |
| act-2 | act-4 | 506,426        | 505,013        | T04C12.5          | M03F4.2a          | 0.999        | 0.999          |
| rpl-18| rpl-7A| 509,852        | 493,528        | Y38A10A.5.1       | Y43B11AR.4.2      | 0.999        | 0.999          |
| act-2 | act-4 | 506,426        | 505,013        | T04C12.5          | M03F4.2a          | 0.999        | 0.999          |
The generation of ESTs is an effective and unique approach in molecular studies as it allows for the analysis and measurement of gene expression, as well as simultaneous discovery of new genes. As each EST represents a copy of the functional part of a genome, the study of ESTs is believed to be a more effective way to discover functional genes. Furthermore, analysis of the expression of a large number of genes combined with the knowledge of their functions enables insight into the overall situation in terms of biological processes, for the current purposes in earthworms. In the present study, ~91% of the ESTs generated were sequences with known or putative functions, while the remainder represented unknown proteins or sequences with no similarities to those in the databases. Although close to 600 ESTs were reported, this is actually far from what could be considered as a 'complete' transcriptome (which usually includes between 15,000-20,000 to >100,000 ESTs). Therefore, the present characterization of this seemingly partial transcriptome may far from reflect the full transcriptomic profile of tissues in earthworms.

A comparison of the classification of ESTs with a *C. elegans* cDNA library based on their putative functions was conducted. Based on identification of clusters via GO analysis, 168 ESTs were matched to *C. elegans* genes by BLASTx. It is well known that earthworms serve significant roles in organic matter decomposition and mineral cycling, and thus are considered to be important contributors to soil fertility and humification processes. In the present study, hydrolytic enzyme activity, conjugating protein ligase activity, oxidation reduction and energy release activity of metabolic enzymes accounted for a large proportion of the molecular functions component. These molecular functions are considered to be a key part of the physiological functions of earthworms, which allow earthworms to survive in different soil environments.

Cell components, as part of GO annotation, are mainly categorized based on subcellular location (including cell cytoplasm, mitochondria, lysosomes, nucleus, microtubules, plasma membrane and myosin), which is highly important for the study of protein functions. The results of subcellular localization analysis can provide significant clues to aid the understanding of protein functions. In the present study, several cell components were determined through GO analysis of the 168 ESTs (annotated genes). It was evident that these cell components had strong associations with the regulation of gene expression during the biological development of earthworms, enabling the regeneration of the anterior portion, alterations in movement ability and tissue differentiation.

Due to the temporal specificity of gene expression and interactions with other gene products, the specific pathway undertaken, sequence of gene expression and expression pattern may ultimately change the effects of multiple pathways in...
earthworms (2). Therefore, the cDNA library and transcription profile of genes representative of fully developed adult earthworms may differ markedly to those of juvenile earthworms. In the present study, gene expression profiles representative of adult earthworm development were generated; however, gene expression profiling of juvenile earthworms was not performed, nor were analyses of the expected differential gene expression between juveniles and adults. In the present study, 168 individual ESTs of earthworms were analyzed by KEGG pathway annotation, which identified 9 corresponding categories. Among them, glutathione metabolism is involved in antioxidant defense systems in *Eisenia Andrei*, and the associated enzymes are mainly identified in cytosolic fractions (26). Chondroitin sulfate and heparan sulfate biosynthesis are involved in biosynthesis pathways, which have important effects on growth and regeneration. In addition, heparan sulfate is also considered to be a type of anticoagulant, which may also be a function of arachidonic acid (27). Therefore, understanding the molecular function of earthworms may provide some basis for the treatment of thrombotic disorders.

Information on functional annotation and relevant biological interactions associated with a particular gene is available from many online resources. The gene network comprises a collection of genes that cooperate with each other to control the main biological processes. The STRING database suggested a functional context for earthworm lumbrokinase with unknown specificity. A previous study revealed that lumbrokinase and dilo administration can efficiently reduce the incidence of cardiac disease among nonsmokers exposed to second-hand smoke (28). In this regard, the discovery of genes and protein interactions in earthworms has provided a basis for further investigation into human diseases.

**Acknowledgements**

Not applicable.

**Funding**

The present study was supported by the National Natural Science Foundation of China (grant nos. 30472251, 30901821 and 81370312), the Natural Science Foundation of Shanxi Province, China (grant no. 2010021035-2), National Science Foundation of Shanxi Province, China (grant no. 2015011113), International Scientific and Technological Cooperative Foundation of Shanxi Province, China (grant no. 201703D421023), the Fund for Shanxi ‘1331 Project’ Key Subjects Construction and the Fund for Shanxi Key Subjects Construction.

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Authors’ contributions**

BFY, BN and JX conceived and designed the experiments and bioinformatics analysis. CL and FXM performed bioinformatics analysis, and wrote the manuscript. XW, PYM, QZ and JBT performed the experiments. RG, ZZL, HLW and NLC contributed to designing the present study and revising the manuscript. JHW and GQS analyzed and interpreted the data.

**Ethics approval and consent to participate**

Not applicable.

**Patient consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**References**

1. Torres-Leguizamón M, Mathieu J, Decaens T and Dupont L: Genetic structure of earthworm populations at a regional scale: Inferences from mitochondrial and microsatellite molecular markers in *Aporrectodea icterica* (Savigny 1826). PLoS One 9: e101597, 2014.
2. Pirooznia M, Gong P, Guan X, Inouye LS, Yang K, Perkins EJ and Deng Y: Cloning, analysis and functional annotation of expressed sequence tags from the *Earthworm Eisenia fetida*. BMC Bioinformatics 8 (Suppl 7): S7, 2007.
3. Cho SJ, Koh KS, Lee E and Park SC: Differential expression of three galabin genes during earthworm head regeneration. Biosci Biotechnol Biochem 73: 2609-2614, 2014.
4. Zheng P, Shao Q, Diao X, Li Z and Han Q: Expression of stem cell pluripotency factors during regeneration in the earthworm *Eisenia fetida*. Gene 575: 58-65, 2016.
5. Kim TH, Kim NS, Lim D, Lee KT, Oh JH, Park HS, Jang GW, Kim HY, Jeon M, Choi BH, et al: Generation and analysis of large-scale expressed sequence tags (ESTs) from a full-length enriched cDNA library of porcine backfat tissue. BMC Genomics 7: 36, 2006.
6. Wellenreuther R, Schupp I, Pouستka A and Wiemann S: German cDNA Consortium: SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. BMC Genomics 5: 36, 2004.
7. Dai Y, Su W, Yang C, Song B, Li Y and Fu Y: Development of novel polymorphic EST-SSR markers in *bailinggou* (*Pleurotus tuliensis*) for crossbreeding. Genes (Basel) 8: pii: E235, 2017.
8. Liu C, Liu D, Guo Y, Lu T, Li X, Zhang M, Ma J, Ma Y and Guan W: Construction of a full-length enriched cDNA library and preliminary analysis of expressed sequence tags from *Bengal Tiger* *Panthera tigris* tigris. Int J Mol Sci 14: 11072-11083, 2013.
9. Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watabiki A, Shibata K, Konno H, Muramatsu M and Hayashizaki Y: Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis. Genomics 77: 79-90, 2001.
10. Ling P, Wang M, Chen X and Campbell KG: Construction and characterization of a full-length cDNA library for the wheat stripe rust pathogen (*Puccinia striiformis* f. sp. *tritici*). BMC Genomics 8: 145, 2007.
11. Ye J, McGinnis S and Madden TL: BLAST: Improvements for better sequence analysis. Nucleic Acids Res 34 (Web Server Issue): W6-W9, 2006.
12. NCBI Resource Coordinators: Database resources of the national center for biotechnology information. Nucleic Acids Res 41: (Database Issue): D8-D20, 2013.
13. Thompson JD, Higgins DG and Gibson TJ: 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: 4673-4680, 1994.
14. Wang J, Duncan D, Shi Z and Zhang B: WEB-based GEne AnaLysis Toolkit (WebGestalt): Update 2013. Nucleic Acids Res 41 (Web Server Issue): W77-W83, 2013.
15. Zhang B, Kirov S and Snoddy J: WebGestalt: An integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33 (Web Server Issue): W741-W748, 2005.

16. Ravaux J, Hassanin A, Deutsch J, Gaill F and Markmann-Mulisch U: Sequence analysis of the myosin regulatory light chain gene of the vestimentiferan Riftia pachyptila. Gene 263: 141-149, 2001.

17. Kanehisa M, Goto S, Kawashima S and Nakaya A: The KEGG databases at GenomeNet. Nucleic Acids Res 30: 42-46, 2002.

18. Wiemann S, Mehrle A, Bechtle S, Wellenreuther R, Pepperkok R and Pousta A; German cDNA Consortium: cDNAs for functional genomics and proteomics: The German Consortium. Comptes Rendus Biologies 326: 1003-1009, 2003.

19. Blair MW, Fernandez AC, Ishitani M, Moreta D, Seki M, Aylings S and Shinozaki K: Construction and EST sequencing of full-length, drought stress cDNA libraries for common beans (Phaseolus vulgaris L.). BMC Plant Biol 11: 171, 2011.

20. Clarke L and Carbon J: A colony bank containing synthetic Col El hybrid plasmids representative of the entire E. coli genome. Cell 9: 91-99, 1976.

21. Li YP, Xia RX, Wang H, Li XS, Liu YQ, Wei ZJ, Lu C and Xiang ZH: Construction of a full-length cDNA Library from Chinese oak silkworm pupa and identification of a KK-42-binding protein gene in relation to pupa-diapause termination. Int J Biol Sci 5: 451-457, 2009.

22. Hata F, Tossel-Klopp G, Clouscard-Martino K, Mulsant P and Gasser F: Expressed sequence tags for genes: A review. Genet Selec Evol 30: 521-524, 1998.

23. Thanh T, Chi VT, Abdullah MP, Omar H, Noroozi M, Ky H and Napis S: Construction of cDNA library and preliminary analysis of expressed sequence tags from green microalga Ankistrodesmus convolutus Corda. Mol Biol Rep 38: 177-182, 2011.

24. Duo L, Yin L, Zhang C and Zhao S: Ecotoxicological responses of the earthworm *Eisenia fetida* to EDTA addition under turf-grass growing conditions. Chemosphere 220: 56-60, 2018.

25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29, 2000.

26. Homa J, Stalmach M, Wilczek G and Kolaczkowska E: Effective activation of antioxidant system by immune-relevant factors reversely correlates with apoptosis of Eisenia andrei coelomocytes. J Comp Physiol B 186: 417-430, 2016.

27. Li P, Sheng J, Liu Y, Li J, Liu J and Wang F: Heparosan-derived heparan sulfate/heparin-like compounds: One kind of potential therapeutic agents. Med Res Rev 33: 665-692, 2013.

28. Lai CH, Han CK, Shiubu MA, Pai PY, Ho TJ, Day CH, Tsai FJ, Tsai CH, Yao CH and Huang CY: Lumbrokinase from earthworm extract ameliorates second-hand smoke-induced cardiac fibrosis. Environ Toxicol 30: 1216-1225, 2015.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.