Send Orders for Reprints to reprints@benthamscience.net

Current Genomics, 2022, *23*, 118-125

RESEARCH ARTICLE

Draft Genome Sequence of the Earthworm *Eudrilus eugeniae*

Arun Arumugaperumal[1], Dinesh Kumar Sudalaimani[1], Vaithilingaraja Arumugaswami[2] and Sudhakar Sivasubramaniam[1,*]

[1]*Department of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu- 627012, India;* [2]*Department of Molecular and Medical Pharmacology, University of California, Los Angeles, California, CA 90095, USA*

ARTICLE HISTORY

Received: December 01, 2021
Revised: January 30, 2022
Accepted: February 14, 2022

DOI:
10.2174/1389202923666220401095626

Abstract: ***Background*:** Earthworms are annelids. They play a major role in agriculture and soil fertility. Vermicompost is the best organic manure for plant crops. *Eudrilus eugeniae* is an earthworm well suited for efficient vermicompost production. The worm is also used to study the cell and molecular biology of regeneration, molecular toxicology, developmental biology, *etc*., because of its abilities like high growth rate, rapid reproduction, tolerability toward wide temperature range, and less cost of maintenance.

***Objective*:** The whole genome has been revealed only for *Eisenia andrei* and *Eisenia fetida.*

***Methods*:** In the present work, we sequenced the genome of *E. eugeniae* using the Illumina platform and generated 160,684,383 paired-end reads.

***Results*:** The reads were assembled into a draft genome of size 488 Mb with 743,870 contigs and successfully annotated 24,599 genes. Further, 208 stem cell-specific genes and 3,432 non-coding genes were identified.

***Conclusion*:** The sequence and annotation details were hosted in a web application available at https://sudhakar-sivasubramaniam-labs.shinyapps.io/eudrilus_genome/.

Keywords: Earthworm, genome, next-generation sequencing, *Eudrilus eugeniae*, genome resource, gene annotation.

1. INTRODUCTION

Earthworm belongs to the phylum Annelida. They are widely used in vermicomposting. It is called farmer's friend. The genome of the earthworm *E. andrei* has been completed [1] and also draft genome of *E. fetida* has been published [2]. The genome size of the *E. fetida* and *E. andrei* are 1.05 and 1.3 Gb, respectively. The earthworm, *Eudrilus eugeniae*, commonly known as the "African nightcrawler", belongs to the phylum Annelida. The interesting abilities like high growth rate, rapid reproduction, tolerability toward wide temperature range, and less cost of maintenance in the laboratory make it a suitable animal model [3]. The presence of fully developed organs and the ability to regenerate make it the most agreeable model for the study of regeneration. The regeneration ability comes from the presence of stem cells in the area bordered by the epithelial and circular muscle layer of the worm [4].

They mark their niche by giving out fluorescence signals. The cause of fluorescence is the presence of riboflavin and its derivatives. The source of riboflavin is the endosymbiotic bacterium *Bacillus endophyticus* [5]. The starvation resulting from food deprivation induces the sporulating *B. endophyticus* to produce more of this riboflavin in the worm's gut. The riboflavin helps the worm to regenerate faster. But the clitellum of *E. eugeniae* is indispensable for the regeneration of the worm, without which the process fails. The segments trailing the clitellum, during regeneration, express a protein known as TCTP at the tip of blastema, which supports regeneration, enhances wound healing, epithelial granular cell differentiation, and cell migration [6]. The transcriptome sequencing of regenerating worm confirmed the presence of TCTP in the anterior regeneration blastema [7]. These studies imply that *E. eugeniae* is an economic model for regeneration studies. The worm stages self-assemblage when introduced to the water together. The worms that have their heads cut off lose this social ability which they regain once the brain is formed [8]. This proves its potency as a model for studying mental disorders. The suspected anti-mitotic compounds were found to inhibit blastema formation in the regenerating *E. eugeniae* [9]. Thus the worm serves as a model to experiment with the anti-mitotic properties of chemical compounds. Certain chemical compounds are toxic to the biological system. *E. eugeniae* concentrates such chemicals to its posterior segments and sacrifices those tail segments through the process of autotomy. The worm may regenerate and survive depending on the toxic severity of the chemicals [10]. This ability helps to characterize the compounds. The earthworm *E. eugeniae* biomass has been utilized as feeding stock for fish

*Address correspondence to this author at the Department of Biotechnology, Manonmaniam Sundaranar University, Tirunelveli, Tamilnadu- 627012, India; Tel: +91 9940998936; E-mail: sudhakar@msuniv.ac.in

1875-5488/22 $65.00+.00

fingerlings [11]. In the present work, the DNA of the worm was extracted from the prostate gland and was subjected to sequencing by the Illumina platform. The cleaned reads were assembled and annotated. Totally 24,599 genes were annotated. A web resource was developed to browse the details of genes, mRNA, proteins, and non-coding RNA molecules.

2. MATERIALS AND METHODS

2.1. *De novo* Genome Sequencing

Ten years ago, two earthworms of *E. eugeniae* species were taken and reared to form a colony in our laboratory. The earthworm bed consisted of soil, cow dung, and leaf litters and an outlet drain was provided for the removal of excess water. The bed was watered every morning and covered with a moist jute bag to maintain the moisture. The bed material was changed every week. A matured earthworm from the colony weighing 1 g was selected for the experiment. The worm was surface sterilized with 10% ethanol six times. Then it was dissected aseptically under a laminar flow hood and the prostate gland was dissected separately. The total DNA from the prostate gland tissue was extracted using the protocol given by Mathur *et al.* [1]. The concentration of the DNA in the extracted sample was 1061 ng/µL in TE buffer. The sample was treated with RNase before library preparation. The sequencing library prepared was ≤ 800 bp insert normal library. The whole genome of the earthworm *E. eugeniae* was performed using the Illumina HiSeq 2500 system.

2.2. Genome Assembly

The reads were trimmed to remove the adapter sequences. The quality of the reads was checked using the FASTQC tool version 0.11.7 available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. The trimmed reads were assembled using CLC Genomics Workbench ver 11. The genome assembly statistics were generated using NGS QC Toolkit version 2.3.3 [2]. Another assembly was generated using Geneious version 9.0. The mitochondria genome was obtained by aligning the reads to the mitochondria genome of *Lumbricus terrestris* with the NCBI accession NC_001673 [3]. Then the consensus sequence was obtained.

2.3. Repeat Masking

Repeat Masker module available with Omicsbox software with default values was used for masking the repeats in the genome assembly. Dfam 3.0 database was used for repeat identification through rmblastn version 2.9.0+.

2.4. Genome Annotation

Augustus was used to predict the genes present in the assembled contigs. The gene models of humans were taken as templates. The predicted sequences were subjected to sequence similarity search using blastx with E-value 1E-5 against the nr database. Omicsbox version 1.2 was used for the functional annotation of the predicted genes. The promoter sequences were identified using FPROM software integrated with MolQuest version 2.4.5. The promoter sequences reported for humans were taken as the template for the identification of promoters in *E. eugeniae*.

2.5. Identification of Non-coding Genes

The non-coding RNA (ncRNA) genes present in the *E. eugeniae* genome were identified by searching against the Rfam 13.0 database (https://rfam.xfam.org/) using Infernal software version 1.1.2. The E value for the target hits was set to 1E-5 and the clan information was also retrieved. The command used in the Ubuntu shell was cmscan --tblout contTable.txt -o Contigsrenamed.txt --fmt 2 -E 0.00001 --rfam --clanin Rfam.clanin --oclan Rfam.cm Contigs-renamed.fa

2.6. Identification of Stem Cell-specific Genes

The CDS sequences were compared with a list of stem cell-specific gene sequences reported by Maguire *et al.* [4]. Blastx program available with NCBI Blast+ software integrated with Omicsbox ver 1.2 was used for the sequence similarity search with the E-value cut-off set at 1E-05, leaving all other options as default.

2.7. Creation of webApp

The webApp hosting the genome information of *E. eugeniae* was built in R version 4.0.3 [5] using the shiny package [6]. The assembly contigs, coding sequences, and protein sequences were placed in separate tabs. The sequence information and annotations were loaded in table format and were displayed using the DT package [7]. The assembly from geneious software and its annotation information was deposited in a web application, and the link was given as 'Assembly with gaps' in the former web application. The software code and data files used were deposited at https://doi.org/10.6084/m9.figshare.16385280

3. RESULTS AND DISCUSSION

The earthworm *Eudrilus eugeniae* has about 120 to 340 body segments. The worm is shown in Fig. (**1**). The worm has clitellar segments from 13 to 18. The anterior section to the clitellum contains critical organs such as the mouth, brain, heart, testis, ovary, and seminal vesicle, while the posterior region contains the worm's downstream segments till the anus. The species is an epigeic feeder and can be easily identified by the metallic sheen on its back, and they are extremely sensitive to sunlight.

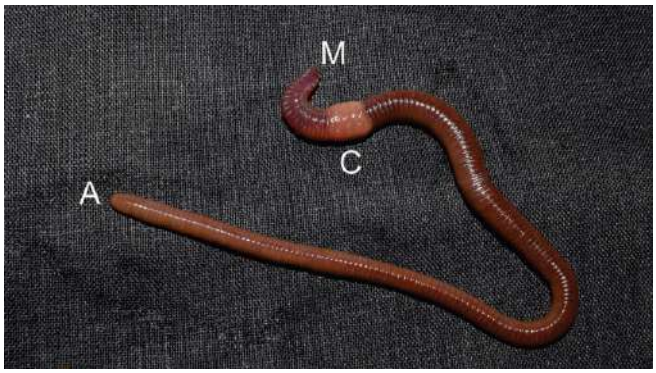

Fig. (1). The earthworm *Eudrilus eugeniae*. M - Mouth; C - Clitellum; A - Anus. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

The earthworm lives in the soil environment, which is rich in microbes. Although the body surface of the earthworm was washed with 10% ethanol, there is a high chance of the meager presence of microbes. This tiny amount of microbes is sufficient to contribute nucleotide, which may get amplified and appear as a contaminant. "Hence, carefully selecting the cleanest body tissue for genome extraction was needed". We examined twenty worms by histological studies and found that the prostate gland is free from the parasite. Also, the organs such as skin have a rich population of symbiotic microbes. We extracted the whole genome of *Verminephrobacter eiseniae* from the draft genome of *E. fetida* and published a paper [8]. Hence, we examined the hundreds of Gram-stained tissue sections of the prostate gland and checked them for the presence of any microbes. We did not find any bacterial cells inside and it is concluded that the prostate gland is better than the skin for extraction of DNA for sequencing. Hence, the prostate gland was selected for DNA extraction. The worm *E. eugeniae* has 18 pairs of chromosomes. The summary statistics of the genome assembly of the earthworm *Eudrilus eugeniae* are shown in Table **1**.

The adapter-trimmed reads were quality checked and the read quality was found to be in the range of 28 to 40 nucleotides. The sequencing reads were 160,684,383 in number. Geneious software assembly produced an assembly with 1,454,384 contigs and a size of 527,700,658. A total of 21,899 genes were predicted in the assembly. However,the assembly had many gaps, denoted by a sequence of N's, in the contigs. Hence, for improving the quality, another assembly was created using CLC genomics workbench software. The summary statistics of the genome assembly of the earthworm *Eudrilus eugeniae* are shown in Table **1**. The N50 value increased from 457 to 938 in the latter assembly and there were no gaps within the contig sequences. The assembly had 743,870 contigs and 488,099,286 bases. The sum of the lengths of the contigs was 488 Mbp, which is the draft genome size of the *E. eugeniae.* The draft genome sizes of *Eisenia fetida* and the complete genome of *Eisenia andrei* were 1.05 Gbp and 1.3 Gbp, respectively [9, 10]. The GC % of the genomes were 40.7 % and 41.1 %, respectively. The number of contigs was 1,659,527 and 3,633, respectively. The GC % of *E. eugeniae* was 43.2 %.

The contig sequence length was in the range of 200 to 33,504. The largest contigs are lengthier enough to contain full genes inside them. The N50 length was 938 bases. The quality of the assembly gets better with increased N50 length. The N50 length that approaches the size of a gene indicates that about half of the contigs are above this size and contain a full-sized gene. There was no gap in the assembled contigs, as indicated by the absence of Ns. The genome sequencing reads were submitted to NCBI SRA. BioProject ID and BioSample ID are PRJNA751411, SAMN20517504, respectively.

The earthworm mitochondria genome has 37 genes, out of which 21 were tRNAs. So far, 24 mitogenomes are available for earthworms. The size ranges from 14,998 bp to 15,188 bp [11]. The mitochondria genome was obtained by aligning the reads to the mitochondria genome of *Lumbricus terrestris* with the NCBI accession NC_001673 and size 14,998 bp [3]. Consensus sequence was then obtained. The size obtained was 15,027 bp (Fig. **2**). The mitochondria genome was deposited at https://doi.org/10.6084/m9.figshare.16385280.

The major portion of the eukaryotic genome contains repeat sequences. These repeat sequences make the assembly process challenging. The Dfam database contains information about the reported sequence models representing the repeats in known model organisms. The repeat models from humans were considered for finding the repeats in *E. eugeniae*. The same strategy was used in the assembly of the ge-

Table 1. Summary statistics of genome assembly of *Eudrilus eugeniae.*

Statistics	Geneious Assembly	CLC Genomics Workbench Assembly
Total reads	160,684,383	160,684,383
Total contigs	1,454,384	743,870
Total bases	527,700,658	488,099,286
Min sequence length	141	200
Max sequence length	14864	33,504
Average sequence length	362.83	656.16
Median sequence length	243.00	391.00
N25 length	876	1,876
N50 length	457	938
N75 length	247	438
GC%	42.30%	43.18%
Ns	2.59%	0%

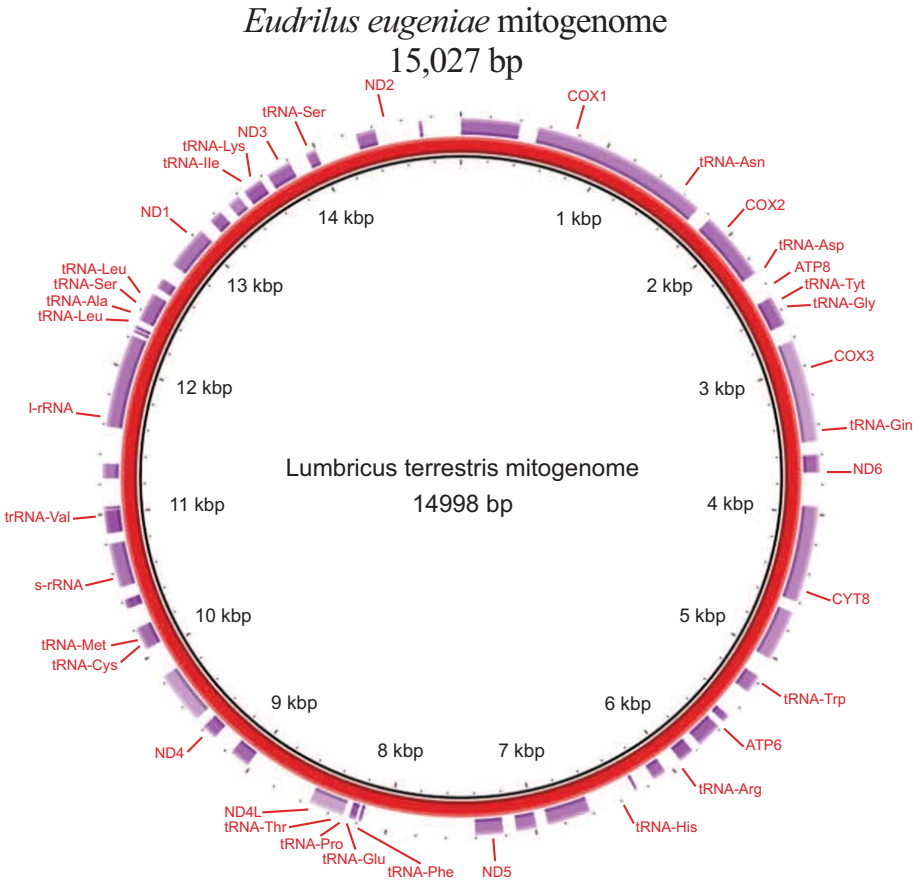

Fig. (2). The mitogenome of *Eudrilus eugeniae* aligned to *Lumbricus terrestris* mitogenome. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

nome of the earthworms *E. fetida,* and *E. andrei.* The repeats found more in the *E. eugeniae* genome are the simple repeats among other repeats like low complexity repeats, satellites, transposons, LTRs, LINEs, and SINEs (Table **2**). DNA transposons were the major repeats in the *Eisenia fetida* genome [12]. The number of satellites found in the *E. eugeniae* genome (58) is more than that in *Eisenia fetida* [11]. In contrast, the LTRs, LINEs, and SINEs identified were low in the *E. eugeniae* genome (Table **2**) and were less than that in *Eisenia fetida* (365,755, 204,621, and 86,934, respectively). LINE2 elements in *Eisenia andrei* were found to regulate the nearby-gene expression, thereby involving it in the regeneration-regulatory network [10]. The presence of tandem repeats hinders the construction of lengthier contigs. Also, they make the gene-finding process difficult by decreasing the ability of the gene finding software. Hence, to overcome this difficulty and improve the chances of gene finding, the repeat sequences were masked beforehand. A total of 28,675,132 bases were masked in the assembly, amounting to 5.87% of the bases in the genome (Table **2**).

There are many gene prediction software programs available. Augustus is one of the well-performing software for *de novo* gene prediction of eukaryotic genomes. It predicted 95,634 genes in the *E. eugeniae* genome. 37,632 CDS sequences had complete ORF. The number of unigenes predicted and annotated was 24,599, whereas the number of genes reported in *E. fetida and E. andrei* was 29,552 and 26,926, respectively [10, 12]. Even though the genome is a draft one, the number of genes predicted is comparable with that predicted in the complete genome of the other worms. The structure of a gene in contig 177418 is shown in Fig. (**3**). The gene had 13 exons, and it was identified as protein disulfide-isomerase A3-like. The predicted coding sequences were subjected to Blast search to find their identity based on similarity with homologous sequences. This identified a total of 10,510 genes with already reported homologous sequences. 62,872 sequences had at least one Blast hit. The gene ontology (GO) terms associated with these sequences were used for annotating the remaining sequences. The distribution of this data is shown in Fig. (**4A**). The GO terms obtained from this mapping cannot be blindly assigned to the sequences. There will be repetitive entries for the same gene and also the chance of false positives. The filtering algorithm available with Omicsbox software was used to clear these anomalies and assign only the correct annotations. Finally, 42,549 GO terms were considered, and 9,813 among them were successfully mapped (Fig. **4B**). The annotation details are given in Table **S1**. The distribution of a number of GO terms assigned to individual sequences is shown in Fig. (**4D**). This distribution is right-skewed, indicating that less number of sequences had more GO terms assigned to them. The E-value distribution for the Blast search is shown in Fig. (**4C**). A right-skewed distribution was observed, indicating that 92.2 % of sequences had a match with high confidence. This sequence similarity distri-

bution is shown in Fig. (**4E**). The number of hits increased with increasing alignment length, while there were no hits below the alignment length of 30. The search operation did not rely on matches with poor alignment length. The genes can also be annotated based on their protein domains. Interpro scan retrieved domain details for 65,774 sequences, out of which 29,715 sequences had their GOs retained after the filtering operation. The protein disulfide-isomerase A3-like gene identified in contig 177418 had a PDI family protein domain (Fig. **3**). The domain details of all the contigs were given in Table **S2**.

Table 2. Overview of repeat elements in the genome of *Eudrilus eugeniae.*

Elements	Numbers
Total sequence length (bp)	488,099,286
Total bases masked (bp)	28,675,132
Percentage of bases masked	5.87%
Simple repeats	333,507
Low complexity repeats	39,691
Satellites	58
DNA transposons	1,966
LTRs	1,152
LINEs	2,295
SINEs	1,812

The top 10 protein domains present were ankyrin repeat domain, 7TM domain, ABC transporter domain, zinc finger domain, protein kinase domain, Ig like domain, major facilitator superfamily domain, LysR substrate binding domain, and EF hand domain (Fig. **4F**). Ankyrin repeat containing domain was found to occur more than any other protein domain. This domain mediates protein-protein interactions and is found to play a major role in the progression of human cancer [13]. *E. eugeniae* has a potent regenerative ability and the presence of ankyrin repeat domain coincides with this ability. The progression of regeneration should also be connected with the ankyrin repeat domain containing proteins. The transmembrane seven helix is the next abundant protein domain which is present in a wide range of receptor proteins. This domain has already been reported to be widely present in the *Eisenia andrei* as an EGFR receptor [10]. ABC transporter domain is present in the cell membrane proteins and helps in the transport of materials across the membrane [14]. The earthworm requires the transport of metabolites across the membrane to different body layers during regeneration. The regeneration process will be heavily affected if the metabolites are not transported at a specific time point. The extensive presence of the ABC transporter domain is evidence of the existence of an effective transport mechanism that aids in the regeneration process. The zinc finger domain is generally abundant in eukaryotes with a prominent role in functions like development, differentiation, autophagy, and metabolism. They function as transcription factors and mediate the expression of genes required for specific functions [15]. In regeneration, specific genes need to be activated at specific time points to get successful regeneration. The protein kinase domain is known to add a phosphate group from ATP to proteins and activate them in their respective pathways. The regeneration process also includes complex pathways that need control through activation by proteins having protein kinase domain. *E. eugeniae* is also known as farmer's friend. It consumes organic matter as plant remains in the soil, digests, and leaves out the vermicast as manure for the soil. Vermicompost is a healthy manure for the agriculture sector. The worm produces manure on a massive scale. It is possible only with efficient metabolic processes. The protein kinase domain-containing proteins help in these extensive metabolic processes also. *E. eugeniae* has been used to treat tannery waste and the efficiency was attributed to earthworm gut enzymes [16]. Immunoglobulin-like domain is known to function in cell-cell recognition, act as cell surface receptors, muscle structure, and immune response [17]. *E. eugeniae* is well known for studies on innate immune responses. Cell type recognition is an important function for regenerating tissue, and the Ig-like domain is inevitable in the design. The major facilitator superfamily domain is responsible for the movement of small solutes across the cell membrane. *E. eugeniae* is a sensitive worm and it ejects out its coelomic fluid when intimidated. The fluid contains antibacterial peptides and also venom peptides that may aid in self-defense. The worm is known to harbor riboflavin producing bacteria in its intestinal tract. The bacteria produce riboflavin in large quantities as a supplement for the worm [18]. Riboflavin promotes the migration of stem cells from one region to another, thereby enhancing the regeneration process. It also has the necessity to transport the sequestered riboflavin from the gut to different body layers. The EF hand domain binds to Ca^{2+} ions and functions by sequestering the calcium ions or supplying the calcium ions to interacting proteins [19]. Calcium ions were established to function in the wound healing process [20]. They are also required for cell migration in the regeneration process [21]. The top ten protein domain families are shown in Fig. (**4G**). P-loop containing nucleoside triphosphate hydrolase (IPR027417) was the top domain family. They are involved in the removal of beta and gamma phosphate bonds from nucleotide triphosphates and release energy. This family of proteins is very much needed for an organism like *E. eugeniae,* which regenerates at a faster rate and require a lot of energy. The concentration of phosphorus was found to be higher in vermicast compared to that in soil [22].

The expression of genes at specific time points is under the control of promoters. The sequence detail of promoters in a genome is essential for designing molecular experiments to understand the expression of the genes. The promoter search in *E. eugeniae* genome identified a total of 208,363 promoter sequences at various positions. The predicted promoter positions are given in the supplementary file available at https://doi.org/10.6084/m9.figshare.16385280. The promoters can be widely classified as TATA box-containing promoters and TATA-less promoters [23]. Of the 208,363 identified promoters, 124,286 were TATA-containing promoters and 84,077 genes did not have the TATA box. The protein disulfide-isomerase A3-like gene present in contig 177418 had a TATA promoter at position 6547 and an enhancer sequence at 6738 (Fig. **3**).

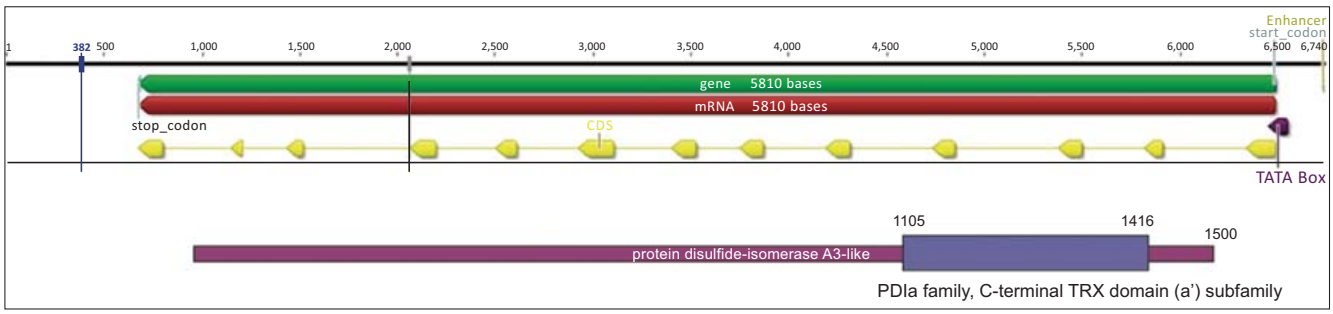

Fig. (3). The structure of protein disulfide-isomerase A3-like gene present in Contig_177418. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

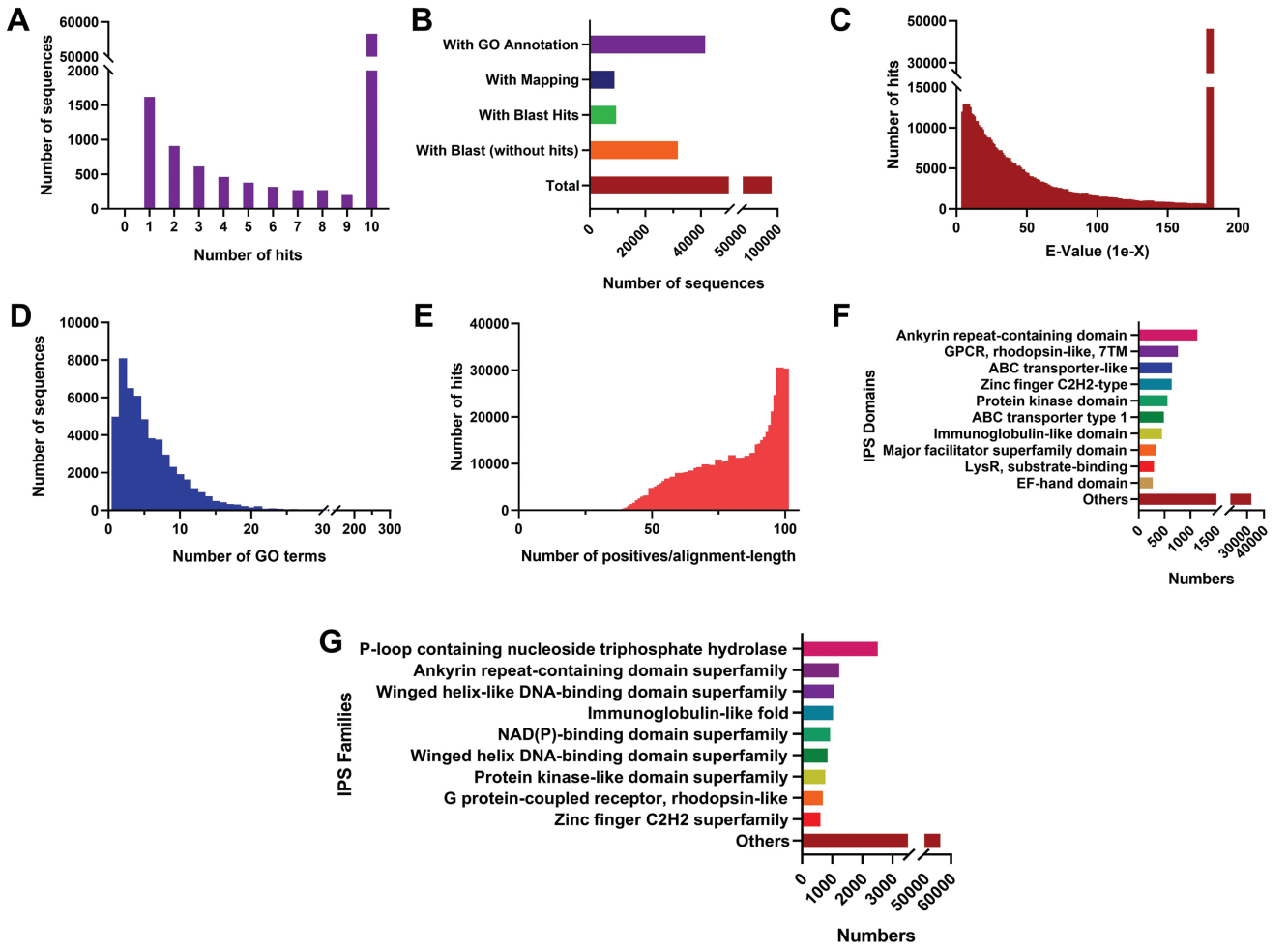

Fig. (4). The genome annotation details for *Eudrilus eugeniae*. (**A**) Distribution of Blast hits, (**B**) Distribution of Blast annotation results, (**C**) Distribution of Blast e-value, (**D**) Distribution of gene ontology mapping, (**E**) Distribution of sequence similarity, (**F**) Distribution of Interpro domains, (**G**) Distribution of Interpro families. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

The individual gene products function together in pathways to achieve a physiological function. The pathways enacted by the genes of *E. eugeniae* were mapped and the KEGG pathway annotation details are given in Table **S3**. The pathways include important processes like the synthesis of secondary metabolites, degradation of toxic compounds, energy metabolism, cellular component synthesis like fatty acids, and antimicrobial peptide synthesis. The earthworm was known to contain riboflavin derivatives during regeneration time [18]. The presence of 6 genes in the riboflavin metabolism and transport pathways is evidence and accounts for the observation. The earthworm is known to take organic material as food. Here the presence of 12 genes in the starch and sucrose metabolism pathway gives a molecular map of the process. It has been shown that *E. eugeniae* efficiently decomposed the distillation waste of plant origin [24]. A total of 11 genes were observed in the nitrogen metabolism pathway. Experiments on vermicomposting carried out with high nitrogenous feed also suggested that *E. eugeniae* was able to assimilate the feed efficiently [25, 26]. There

is a need for a huge amount of nucleotide synthesis during regeneration. The presence of 28 genes in purine metabolism and 19 genes in pyrimidine metabolism indicates the importance of the pathways. It has been reported that the symbiotic bacteria of *Eisenia fetida* was able to synthesize vitamin B2, thiamine, biotin, propionate, pyruvate, folate, and pantothenate and may provide to the worm [8]. In *E. eugeniae,* pathways for the metabolism of all these metabolites except propionate were observed. Also, an endosymbiont for *E. eugeniae* has been reported [27]. The significance of the presence of pathways for secondary metabolites' metabolism was exemplified by the studies on improved reproduction success of *E. eugeniae* in vermibeds supplemented with organic plant materials [28, 29]. The worm was reported to be used in the bioremediation of petroleum hydrocarbons [30]. Accordingly, genes were found to be involved in the toluene degradation pathway.

The stem cells in *E. eugeniae* were identified and characterized by Johnson *et al.* in 2012 [18]. In the genome of *E. eugeniae,* a total of 2,757 genes were identified as stem cell-specific genes. The database contained 250 stem cell-specific genes [4]. Out of the 250 genes, 208 were found in the genes of *E. eugeniae*. The list is provided in Table **S4**. A total of 3,432 non-coding genes were identified in the genome and are listed in Table **S5**.

The genome sequence and the gene annotation details were hosted in a web application at https://sudhakar-sivasubramaniam-labs.shinyapps.io/eudrilus_genome/. The contig sequences, coding sequence and annotation, protein sequence and annotation, non-coding RNA annotation, and other resources are displayed in individual tabs. The data are provided in a table format, with separate columns for annotation and sequence, which can be sorted. The users can search for any gene using the gene name or the exact partial sequence. The challenge existing in hosting large scale genome sequences is their large size. The display and sorting operations consume a lot of memory. Hence, the contig sequences are separated into 15 parts of 50,000 sequences each. The users can choose between the parts to display. A link in the side panel named 'Assembly with gaps' takes to a web application hosting the genome assembly produced by geneious software which can be separately accessed at https://sudhakar-sivasubramaniam-labs.shinyapps.io/eudrilus_genome2/. The other resources include the transcriptome sequences and annotation of regenerating *E. eugeniae* reported by Paul *et al.* [31]. The transcriptome resource can separately be accessed at https://sudhakar-sivasubramaniam-labs.shinyapps.io/eudrilus_transcriptome/. The gene sequences marked with 'Unknown' in the description field have scope in novel gene discovery. The web application will be helpful in the easy browsing of the data for researchers working with earthworms. Recent works show that computer simulations can be used to complement experimental methods for genes. Large biological systems demand the combined usage of CPU and GPU computing powers to perform molecular simulations [32]. Homology modeling was done on the sequences of the SGLT2 gene to find inhibitors for use as anti-diabetic drugs [33]. Similarly, the gene sequences of the earthworm might be helpful in molecular biology research with earthworms.

CONCLUSION

The genomic resources of the earthworm *E. eugeniae* will be helpful to extend the research in the area of regeneration, molecular toxicity, developmental biology, *etc*. The accumulating evidence suggest that *E. eugeniae* has many molecular mechanisms involved in tissue regeneration. Hence, there are numerous biological processes in this worm that are yet to be discovered. The unannotated novel genes can be characterized by the sequence information available. Individual genes can be studied easily as the sequence is available for designing PCR primers. The genomic resource reported here will aid the process.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

Not applicable.

FUNDING

This study was supported by the Department of Biotechnology, India, for Senior Research Fellowship (DBT/2015/MSU/447).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

AA is thankful to Department of Biotechnology, India for Senior Research Fellowship (DBT/2015/MSU/447).

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Mathur, A.; Bhat, R.; Verma, S.; Prakash, A.; Prasad, G.; Dua, V. Standardization of method for genomic DNA extraction in *Eudrilus eugeniae*. *Indian Pharmacists.,* **2010**, *9*, 49-51.

[2] Patel, R.K.; Jain, M. NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS One,* **2012**, *7*(2), e30619. http://dx.doi.org/10.1371/journal.pone.0030619 PMID: 22312429

[3] Boore, J.L.; Brown, W.M. Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. *Genetics,* **1995**, *141*(1), 305-319. http://dx.doi.org/10.1093/genetics/141.1.305 PMID: 8536978

[4] Maguire, C.T.; Demarest, B.L.; Hill, J.T.; Palmer, J.D.; Brothman, A.R.; Yost, H.J.; Condic, M.L. Genome-wide analysis reveals the unique stem cell identity of human amniocytes. *PLoS One,* **2013**, *8*(1), e53372.

http://dx.doi.org/10.1371/journal.pone.0053372 PMID: 23326421

[5] Team, R.C.R. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, **2013**.

[6] Chang, W.; Cheng, J.; Allaire, J.; Sievert, C.; Schloerke, B.; Xie, Y. Shiny: Web application framework for R. R package version 160: **2021**.

[7] Xie, Y.; Cheng, J.; Tan, X. DT A Wrapper of the JavaScript Library 'DataTables'. R package version 018. **2021**.

[8] Arumugaperumal, A.; Paul, S.; Lathakumari, S.; Balasubramani, R.; Sivasubramaniam, S. The draft genome of a new *Vermiephrobacter eiseniae* strain: A nephridial symbiont of earthworms. *Ann. Microbiol.,* **2020**, *70*(1), 1-18. http://dx.doi.org/10.1186/s13213-020-01549-w

[9] Zwarycz, A.S.; Nossa, C.W.; Putnam, N.H. Timing and scope of genomic expansion within Annelida: Evidence from homeoboxes in the genome of the earthworm. *Eisenia fetida*. **2016**, *8*(1), 271-81.

[10] Shao, Y.; Wang, X.-B.; Zhang, J.-J.; Li, M.-L.; Wu, S.-S.; Ma, X.-Y. Genome and single-cell RNA-sequencing of the earthworm *Eisenia andrei* identifies cellular mechanisms underlying regeneration. **2020**, *11*(1), 1-15. http://dx.doi.org/10.1038/s41467-020-16454-8

[11] Zhang, L.; Sechi, P.; Yuan, M.; Jiang, J.; Dong, Y.; Qiu, J. Fifteen new earthworm mitogenomes shed new light on phylogeny within the Pheretima complex. *Sci. Rep.,* **2016**, *6*(1), 20096. http://dx.doi.org/10.1038/srep20096 PMID: 26833286

[12] Paul, S.; Arumugaperumal, A.; Rathy, R.; Ponesakki, V.; Arunachalam, P.; Sivasubramaniam, S. Data on genome annotation and analysis of earthworm. *Eisenia fetida. Data Brief,* **2018**, *20*, 525-534. http://dx.doi.org/10.1016/j.dib.2018.08.067 PMID: 30191166

[13] Kariri, Y.A.; Joseph, C.; Kurozumi, S.; Toss, M.S.; Alsaleem, M.; Raafat, S. Prognostic significance of KN motif and ankyrin repeat domains 1 (KANK1) in invasive breast cancer. *Breast Cancer Res. Treat.,* **2020**, *179*(2), 349-57.

[14] Hollenstein, K.; Dawson, R.J. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.,* **2007**, *17*(4), 412-8.

[15] Laity, J.H.; Lee, B.M.; Wright, P.E.J. Zinc finger proteins: New insights into structural and functional diversity. *Curr. Opin. Struct. Biol.,* **2001**, *11*(1), 39-46.

[16] Ravindran, B.; Contreras-Ramos, S.; Sekaran, G. Changes in earthworm gut associated enzymes and microbial diversity on the treatment of fermented tannery waste using epigeic earthworm *Eudrilus eugeniae*. *Ecol. Eng.,* **2015**, *74*, 394-401. http://dx.doi.org/10.1016/j.ecoleng.2014.10.014

[17] Davis, S.J.; Ikemizu, S.; Wild, M.K.; van der Merwe, P.A. CD2 and the nature of protein interactions mediating cell-cell recognition. *Immunol. Rev.,* **1998**, *163*(1), 217-36.

[18] Johnson Retnaraj Samuel, S.C.; Elaiya Raja, S.; Beryl Vedha, Y.; Edith Arul Jane, A.; Amutha, K.; Dinesh, S.M.; Jackson Durairaj, S.C.; Kalidas, R.M.; Tharmaraj, V.; Pitchumani, K.; Sudhakar, S. Autofluorescence in BrdU-positive cells and augmentation of regeneration kinetics by riboflavin. *Stem Cells Dev.,* **2012**, *21*(11), 2071-2083. http://dx.doi.org/10.1089/scd.2011.0485 PMID: 22150027

[19] Nelson, M.R.; Chazin, W.J.J.B. Structures of EF-hand Ca^{2+}-binding proteins: Diversity in the organization, packing and response to Ca^{2+} binding. *Biometals,* **1998**, *11*(4), 297-318.

[20] Lansdown ABG. Calcium: A potential central regulator in wound healing in the skin. *Wound. Rep. Regen*., **2002**, *10*(5), 271-85.

[21] Wei, C.; Wang, X.; Zheng, M.; Cheng, H.J. Calcium gradients underlying cell migration. *Curr. Opin. Cell Biol.,* **2012**, *24*(2), 254, 61.

[22] Tiwari, S.; Tiwari, B.; Mishra, R. Microbial populations, enzyme activities and nitrogen-phosphorus-potassium enrichment in earthworm casts and in the surrounding soil of a pineapple plantation. *Biol. Fertil. Soils,* **1989**, *8*(2), 178-182. http://dx.doi.org/10.1007/BF00257763

[23] Yella, V.R.; Bansal, M.J. DNA structural features of eukaryotic TATA-containing and TATA-less promoters. *FEBS Open Biol.,* **2017**, *7*(3), 324-334.

[24] Deka, H.; Deka, S.; Baruah, C.K.; Das, J.; Hoque, S.; Sarma, N.S. Vermicomposting of distillation waste of citronella plant (*Cymbopogon winterianus* Jowitt.) employing *Eudrilus eugeniae*. *Bioresour. Technol.,* **2011**, *102*(13), 6944-6950. http://dx.doi.org/10.1016/j.biortech.2011.04.027 PMID: 21550233

[25] Balachandar, R.; Baskaran, L.; Yuvaraj, A.; Thangaraj, R.; Subbaiya, R.; Ravindran, B.; Chang, S.W.; Karmegam, N. Enriched pressmud vermicompost production with green manure plants using. *Eudrilus eugeniae*. *Bioresour. Technol.,* **2020**, *299*, 122578. http://dx.doi.org/10.1016/j.biortech.2019.122578 PMID: 31865155

[26] Coulibaly, S.S.; Bi, I.A.Z. Influence of animal wastes on growth and reproduction of the African earthworm species *Eudrilus eugeniae* (Oligochaeta). *Eur. J. Soil Biol.,* **2010**, *46*(3-4), 225-229. http://dx.doi.org/10.1016/j.ejsobi.2010.03.004

[27] Subramanian, E.R.; Sudalaimani, D.K.; Selvan Christyraj, J.R.S.; Ramamoorthy, K.; Gopi Daisy, N.; Selvan Christyraj, J.D.; Renganathan, K.; Krishnan, S.; Sivasubramaniam, S. Studies on organogenesis during regeneration in the earthworm, *Eudrilus eugeniae*, in support of symbiotic association with Bacillus endophyticus. *Turk. J. Biol.,* **2017**, *41*(1), 113-126. http://dx.doi.org/10.3906/biy-1604-44

[28] Sonia, V.; Felix, S.; Antony, C. Comparative study of growth and reproduction of earthworm *Eudrilus eugeniae* in different organic substrate. *Int. J. Appl. Sci.,* **2016**, *4*(1), 7. http://dx.doi.org/10.21013/jas.v4.n1.p7

[29] Kabi, F.; Kayima, D.; Kigozi, A.; Mpingirika, E.Z.; Kayiwa, R.; Okello, D. Effect of different organic substrates on reproductive biology, growth rate and offtake of the African night crawler earthworm (*Eudrilus eugeniae*). *Org. Agric.,* **2020**, *10*(3), 395-407. http://dx.doi.org/10.1007/s13165-020-00284-5

[30] Ekperusi, O.A.; Aigbodion, I.F. Bioremediation of heavy metals and petroleum hydrocarbons in diesel contaminated soil with the earthworm: *Eudrilus eugeniae*. *Springerplus,* **2015**, *4*(1), 540. http://dx.doi.org/10.1186/s40064-015-1328-5 PMID: 26413446

[31] Paul, S.; Balakrishnan, S.; Arumugaperumal, A.; Lathakumari, S.; Syamala, S.S.; Arumugaswami, V.; Sivasubramaniam, S. The transcriptome of anterior regeneration in earthworm *Eudrilus eugeniae*. *Mol. Biol. Rep.,* **2021**, *48*(1), 259-283. http://dx.doi.org/10.1007/s11033-020-06044-8 PMID: 33306150

[32] Allec, S.I.; Sun, Y.; Sun, J.; Chang, C.A.; Wong, B.M. Chang C-eA, Wong BM. Heterogeneous CPU+ GPU-enabled simulations for DFTB molecular dynamics of large chemical and biological systems. *J. Chem. Theory Comput.,* **2019**, *15*(5), 2807-2815. http://dx.doi.org/10.1021/acs.jctc.8b01239 PMID: 30916958

[33] Bhattacharya, S.; Asati, V.; Mishra, M.; Das, R.; Kashaw, V.; Kashaw, S.K. Integrated computational approach on sodium-glucose co-transporter 2 (SGLT2) Inhibitors for the development of novel antidiabetic agents. *J. Mol. Struct.,* **2021**, *1227*, 129511. http://dx.doi.org/10.1016/j.molstruc.2020.129511