Celal Bayar University Journal of Science
Volume 14, Issue 1, p 135-140

Genome-wide EST-SSR Marker Identification in Red Wiggler Worm *Eisenia fetida* (Savigny, 1826)

Vahap Eldem

Istanbul University, Faculty of Science, Department of Biology, 34134, Istanbul, Turkey,
Phone: +90 212 4555700 (15086), vahap.eldem@istanbul.edu.tr

Received: 5 February 2018
Accepted: 16 March 2018
DOI: 10.18466/cbayarfbe.390277

Abstract

Earthworms belong to the main groups of soil-dwelling invertebrates and are an important element for soil biota. The ecological importance of earthworms on terrestrial ecosystem is mainly attributed to their close synergistic interaction with soil biota. Among earthworms, *Eisenia fetida* has been considered as preferred for vermiculture and vermicomposting practices due to their reproductive potentials, short life cycle and wide temperature and density pressure tolerances. Although the genome and transcriptome data for *E. fetida* are available, the EST-SSRs composition and frequency of this species, which used for phylogenetic studies, genus-level taxonomy and population genetics, remain poorly defined. In current study, we mined publicly available transcriptome data and characterised genome-wide EST-SSR markers for this epigeic species. A total of 13,060 EST-SSRs were identified from 162,609 contigs. The most abundant EST-SSR types were found to be trimeric repeats (5,998, 46%) followed by dimeric (4,762, 36%), tetrameric (1,716, 13%), pentameric (478, 4%) and hexameric repeats (106, 1%). For SSR motifs, the most prevalent motifs were AC/GT (14.72%), followed by AT/AT (13.20%), ATC/ATG (13.03%), AAT/ATT (12.30%) and AG/CT (8.46%). Following validation of the species using COI sequences and to test the designed primer pairs, the eight primer pairs were designed, but only six of them were successfully amplified. Although EST-SSR information is scarce for annelids, the EST-SSR patterns of *E. fetida* seem to be similar to annelids. Overall, the EST-SSR markers help in taxonomic resolution of *Eisenia* genus from other earthworm genera and in studying the population structure and geographic distributions of *E. fetida*.

Keywords: EST-SSRs, *Eisenia fetida*, Annelids, Sequence repeats

1. Introduction

With more than 8000 species in around 800 genera, earthworms are considered as integral components of terrestrial ecosystems and constitute approximately >80 % of the invertebrate biomass [1]. Earthworms play an essential role in terrestrial ecosystem health and serving various functions such as organic matter decomposition, soil turnover, soil aeration and drainage [2]. Among these functions, the organic waste decomposition or "vermicomposting" is drawing interest due to its potential use in agricultural applications as organic fertilizers. Currently, vermicomposting can be defined as bio-oxidative and degradation processes that involve the conversion of organic substances (generally, large amount of organic debris) into environmentally friendly humus-like material by joint action of soil microbiome and earthworms [3]. In ecological aspect, earthworms are categorised into three types based on feeding habits; epigeic, endogeic and anecic species. Epigeic species are important for vermicomposting because they characteristically feed on surface organic matter while endogeic and anecic species preferentially feed on subsurface materials and burrows in the soil [4].

The earthworms *Eisenia fetida* (Savigny, 1826) are epigeic species (i.e., non-burrowing earthworm) that extensively used for vermiculture and vermicomposting practices. Although some other epigeic species, such as *Eudrtilus eugeniae*, *Perionyx excavatus* and *Perionyx sansibaricus*, were reported to have vermicomposting potential [5,6,7], *E.fetida* is the most demanded species because of their high reproductive potential (in terms of cocoon production), short life cycle and incubation periods of cocoons, wide temperature tolerance and density pressure [8]. However, the morphology of epigeic earthworms highly resemble to each other, thus, the morphological differences alone appear to be insufficient to correctly identify *Eisenia* taxa. Resolving this ambiguity at sequence level and developing reliable markers can facilitate *Eisenia* genus identification, which is highly demanded by commercial vermiculture industry.

Simple sequence repeats (SSRs), also known as microsatellites, tandemly repeated nucleotide sequences (1-6 bases in lenght) and widely distributed across eukaryotic genomes. Currently, SSRs are extensively used as molecular markers for analysing level of genetic diversity and population genetic structure, cultivar and species identification as well as hybridity determination due to their locus-specific and multi-allelic nature and high rate of transferability across species [9]. Unlike SSR, expressed sequence tag-simple sequence repeat markers (hereafter, EST-SSRs) are originated from coding (i.e., transcribed) regions of the genome and the recent studies showed that EST-SSRs are more conserved and may have a higher rate of transferability across close-species than genomic SSR. These markers are, therefore, well suited for application in cross-species phylogenetic studies and genus-level taxonomy [10].

In this study, we have mined *E. fetida* transcriptome and characterised genome-wide EST-SSR markers for this epigeic species, which is prime choice of earthworm for vermicomposting practices. The main objective of this study was to generate a valuable EST-SSRs resource for *E. fetida*, but due to a high rate of transferability across species in a genus, developed EST-SSRs serve as a powerful marker to allow a clear differentiation of *Eisenia* taxa from other morphologically similar earthworm genera.

2. Materials and Methods

2.1. RNA-Seq-based Transcriptome Data for *E. fetida*

Paired-end raw transcriptome sequences for *E. fetida* were obtained from from Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession number of PRJNA304461. The sequenced read length for forward and reverse sequences of paired-end data is 100 bp in length.

2.2. Data filtering and *De novo* Transcriptome Assembly

FASTQ-formatted all raw sequences were firstly subjected to quality filtering to eliminate low-quality reads using Trimmomatic software (version 0.36) [11] prior to *de novo* transcriptome assembly. The resulting high-quality sequences were combined and assembled into contigs (contiguous sequence) to generate a reference transcriptome using Trinity software (version 0.36) [12] with default parameters. To generate a non-redundant reference transcript set (or assembled contigs) for *E. fetida*, repetitive, identical or near-identical transcripts were removed by clustering sequences with 97 % similarity cut-off using CD-HIT-EST [13] software. Besides, the gVolante [14] server was applied to measure the completeness assessment of *E. fetida* transcriptome sequences and to test whether transcripts included core gene set of metazoa.

2.3. Functional SSR Marker Detection and Primer Design

Following transcriptome assembly, all transcript sequences were subject to identify potential microsatellite or SSR motifs using a perl script called MIcroSAtellite (http://pgrc.ipk-gatersleben.de/misa/). The parameters for identifying simple repeat motifs were as follows: (i) the SSR motifs are expected to contain di-, tri-, tetra-, penta-, and hexa-nucleotides with minimum repeat numbers of 6, 5, 5, 4 and 4, respectively, (ii) a compound SSR motif was defined if the number of bases between two adjacent SSR motifs was ≤100. The EST-SSR primers were designed at the flanking regions of SSR motifs using Primer 3 software (http://bioinfo.ut.ee/primer3/). The following criteria were taken into account to design appropriate primer pairs: (i) the sequence length for primer pairs ranged from 18 to 26 bases, (ii) melting temperature between primer pairs varied from 56 °C and 60°C (iii) maximum melting temperature discrepancy between the pairs was set as 4°C (iv) the amplicon sizes ranged from 100 to 280 bp in length.

2.4. High-quality Genomic DNA Isolation and PCR reaction

All reagents and consumables used for genomic DNA isolation and PCR reaction were sterilized before use. The *E. fetida* samples were obtained from a commercial local farm and the samples were washed twice in sterile ultrapure water to remove debris and soil, and then air-dried. The ten samples were transversely cut into rectangular muscle pieces and 50 mg muscle tissues were used for genomic DNA isolation. Genomic DNA were isolated using Qiagen DNEasy® Blood & Tissue kit (Cat. No. 60504) according to the manufacturers “mouse tail protocol” instructions. The quality and quantity of isolated DNA were measured using by the NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). Prior to EST-SSR primer testing, the partial mitochondrial COI (cytochrome c oxidase subunit 1) sequences were amplified from *E. fetida* samples using the primer pair of KkSF (5-AAGTGGGCTTCGACTAGAAC-3) and KkSR (5-AAGTGGGCTTCGACTAGAAC-3) to verify species identifications. The PCR reactions for COI sequences was performed in 25 µl reaction mixtures containing 2 µl total DNA (~100 ng/µl), 2.5 µl 10X Taq Buffer (ThermoFisher, Cat.No. EP0402), 2 µl $MgCl_2$ (25 mM), 2 µl dNTP (10 mM), 1 µl Taq DNA Polymerase (5 U/µl), 1 µl forward and 1 µl reverse primers (0.01 µM) and 13,5 µl sterile ultrapure water. The amplification protocol was carried out under the following conditions: 35 cycles of 45 s at 94 °C for denaturation, 45 s at 51 °C for annealing, and 1 min at 72 °C for extension. The amplification procedure were as follows initial denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 1 min, 56°C for 1 min and 72°C for 1 min, and a final elongation at 72°C for 10 min. Then, the amplicons were purified using High Pure PCR Product Purification Kit (Roche, Germany)

and sequenced in BGI (BGI, Hong Kong). All COI sequences were deposited in GenBank under accession numbers: MG737857-MG737875. Following species identification based on COI, the eight EST-SSR primer pairs were selected and screened in *E. fetida* samples to test whether successful amplification can be achieved. The PCR reaction was performed in a 50 µl reaction volume containing 4 µl genomic DNA (50 ng/µl), 5 µl 10X Taq Buffer (ThermoFisher, Cat.No. EP0402), 5 µl $MgCl_2$ (25 mM), 5 µl dNTP (10 mM), 2 µl Taq DNA Polymerase (5 U/µl), 2 µl forward and 2 µl reverse primers (0.01 µM) and 25 µl sterile ultrapure water. Cycling parameters for all primer pairs started with 5 min at 94 °C for initial denaturation, followed by 30 cycles of denaturation at 94 °C for 1 min, annealing at 56 °C for 1 min and extension at 72 °C for 1 min, and a final extension for 10 min at 72 °C. The amplicons were purified using High Pure PCR Product Purification Kit (Roche, Germany) and sequenced in BGI (BGI, Hong Kong).

3. Results and Discussion

In order to determine the genome-wide EST–SSR markers of *E. fetida*, the publicly available raw transcriptome data have been mined and characterized in detail. Above 36.5 million high-quality reads, each which are 100 bp in length, were *de novo* assembled after pre-processing and filtering of raw reads. The *de novo* transcriptome assembly result revealed that a total of 162,609 contigs were obtained with total length exceeding 104 megabases. The length of contig sequences varied from 201 to 8,327 bp with an average length of 641 bp and 41.32% GC content. To assess the completeness and accuracy transcriptome assembly, 978 single-copy ortholog genes evolutionarily conserved across metazoan animals were searched against these contig sequences using BUSCO tool [15] incorporated into gVolante server. The analysis results showed 755 (77.20%) of single-copy ortholog genes were completely detected in the assembly, but 91 (9.30%) single-copy ortholog genes were missing in the assembly. The presence of high proportion of conserved single-copy ortholog genes (>75%) demonstrated that the final assembly is largely complete and overall these contig sequences can be used for further EST-SSR analysis (Table 1).

Table 1. Basic metrics and summary statistics of contigs and EST-SSRs found in *E. fetida*.

Description	Numbers
Total number of contigs (transcripts)	162,609
Total length of contigs (bp)	104,365,354
Minimum sequence length (bp)	201
Maximum sequence length (bp)	8,327
Average sequence length (bp)	641.8
GC content (%)	41.32
Total number of core single-copy genes queried	978
Total number of complete core single-copy genes	755 (77.20%)
Total number of complete + partial core single-copy genes	887 (90.70%)
Total number of missing core single-copy genes	91 (9.30%)
Total number of contgis containing EST-SSRs	10,697
Total number of identified EST-SSRs	13,060
Total number of contig containing more than one EST-SSR	1,854
Dimeric repeats	4,762
Trimeric repeats	5,998
Tetrameric repeats	1,716
Pentameric repeats	478
Hexameric repeats	106

The EST-SSR contents of *E. fetida* indicated that the most abundant EST-SSR types were trimeric repeats (5,998, 46%) followed by dimeric (4,762, 36%), tetrameric (1,716, 13%), pentameric (478, 4%) and hexameric repeats (106, 1%) (Figure 1A). This SSR-patterns is also observed in an annelid worm, *Capitella teleta*, indicationg that the most prevalent SRR motifs are trinucleotide repeats [16]. Among these EST-SSRs, the most dominant motifs were found to be AC/GT (1922, 14.72%), followed by AT/AT (1724, 13.20%), ATC/ATG (1702, 13.03%), AAT/ATT (1606, 12.30%) and AG/CT (1105, 8.46%). These top five motifs accounted nearly for 61.72% of total motifs whereas the remaining fourteen motifs accounted for 32.64% (Figure 1B).

Prior to EST-SSR testing, the partial COI sequences of *E. fetida* samples were sequenced and deposited in NCBI database. A total of nineteen COI sequences with average 453 bp in length were obtained from *E. fetida* samples. No insertions, deletions or stop codons were observed in any of the sequence after translating all sequences to amino acids using the invertebrate mitochondrial genetic code. All these sequences were blasted against *E. fetida* reference COI-sequences (FJ214228.1) in GenBank and BlastN analysis showed average 98.5% identity to *E. fetida*. Furthermore, the partial COI sequences were queried against the Barcode of Life Database (http://www.boldsystems.org). According to taxon assignment result, it was found to be that matched sequences belong to *Eisenia* genus at above 99.10% identity. All results suggested that the *E. fetida* samples analysed here can directly used for SSR testing. Following species identification, to validate the stability of EST-SSR loci and determine alleles per locus, a total of 4,267 primer pairs were successfully designed based on the 10,697 EST-SSR containing contigs. Among these primer pairs, the eight primer pairs were selected and screened *E. fetida* individuals.

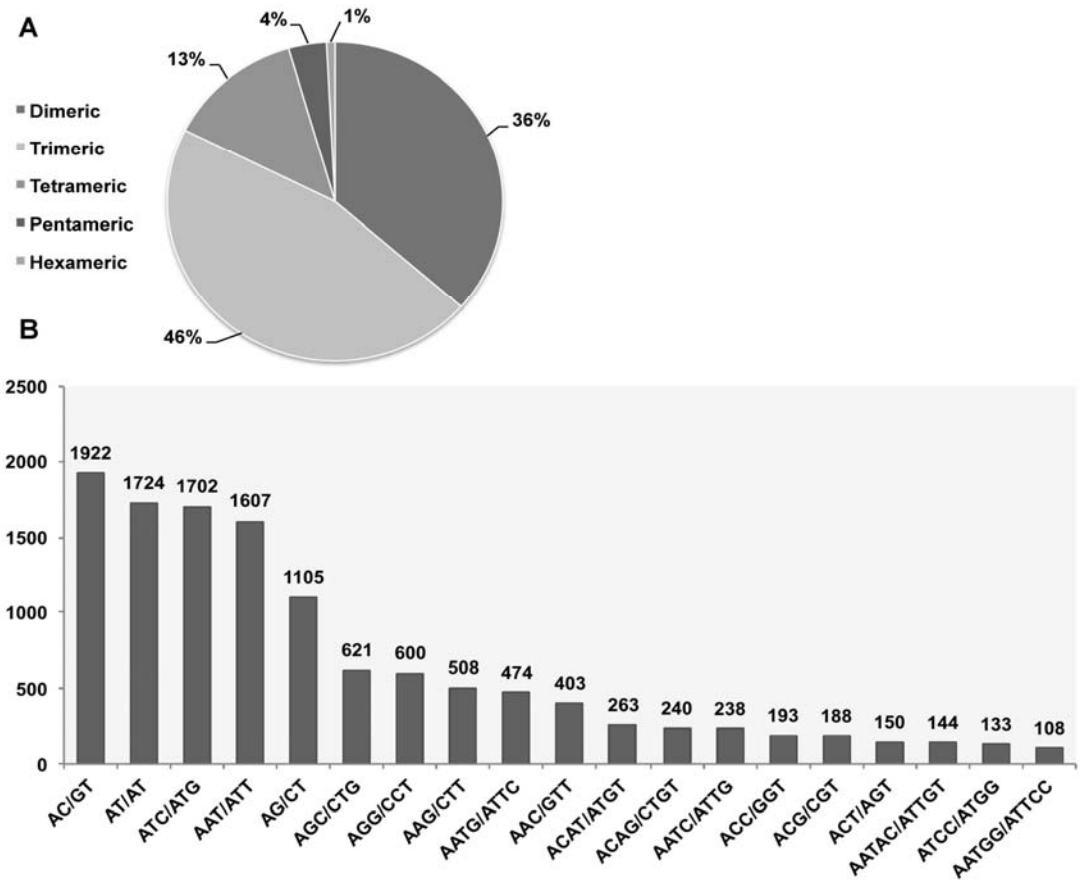

Figure 1. (A) The frequency and distribution of EST-SSR types. (B) Basic characteristics and number of top 19 dominant EST-SSR motifs in *E. fetida* (The motif number below 100 were neglected).

Table 2. Details on the six tested EST-SSR markers for *E. fetida* including primer sequences, amplicon sizes, annealing temperature and repeat motif.

Contig Name	Primer Sequence (Forward and Reverse)	Tm (°C)	Contig Length (bp)	Amplicon Length (bp)	Repeat Motif	Repeat Size	Repeat Start	Repeat End	Putative Functions
Efetida.157721	F: GAGAGCAAGACAACTCTGTG R: ACACTGCATAGCATGATCTG	57	1477	317	(AC)12	24	1153	1177	Heavy-metal-associated domain
Efetida.43077	F: TTGTGTTCGATGGCATTAGG R: AGGATGATGGAGATGAGCAA	57	1565	388	(TGC)8	24	899	923	Ion transport protein
Efetida.44776	F: CCAACTCAGACCAATTCCTG R: TGGCAAGATTAAGCTGTCTG	59	2101	343	(ATA)14	42	1000	1042	Ubiquitin carboxyl-terminal hydrolase
Efetida.92477	F: GTCGTTCACATAAGACAGCA R: GGAGATTGTGTGTCGAACAA	59	2117	295	(AGC)10	30	1609	1639	Protein kinase domain
Efetida.47040	F: TCGGAAGTGCCTCTTAATCT R: ATAGGCCTATGGTGACTACG	59	3285	280	(ATC)11	33	1098	1131	Cytochrome b5-like Heme/Steroid binding domain
Efetida.67884	F: TTGTCATCTGCGTCATCATT R: TGGCCTTGAAGCTAGTCAT	59	1732	333	(ATAC)13	52	1646	1698	WAP-type, Whey Acidic Protein

The selected primer pairs were used for amplification of contigs based on following criteria: (i) more than six tandem repeats and (ii) repeat motif lengths ≥ 20 bp. Out of tested primers, six primer pairs were successfully amplified, with an amplification efficiency rate of 75%, whereas remaining primer pairs (two) give no amplification product (Supplementary File). The negative results might be due to PCR conditions or primer design in spite of despite multiple effort to optimize PCR conditions with various Tm and $MgCl_2$ gradients. Besides, all primer pairs were blasted against the publicly available genomesequences of six annelide worms; *Amynthas corticis*, *Capitella teleta*, *Hydroides elegans*, *Helobdella robusta* and *E. fetida* to confirm specificity of the primer pairs. *In silico* BLASTn analysis showed that no *in silico* amplification products were obtained or very low similarity (base pair matching) between primers and these annelid genomes were observed. As for *E. fetida*, it showed excellent specificity as expected. The primer pairs, amplicon size and annealing temperature are represented in Table 2. The repeat-containing amplicon sequences were also compared with the whole genome shotgun sequence of *E. fetida* (GCA_900000155.1; SoapK31 Assembly) [17] and above 95% similarity was obtained. The overall results suggested that EST-SSR primers designed in the present study could be used for mining simple sequence repeats in *E. fetida*.

The EST-SSR identification, development and characterization in the earthworms (Oligochaeta) very scarce, mainly due to the limited number of transcriptomic and genomic studies. In an extensive database search across Sequence Read Archive (SRA https://www.ncbi.nlm.nih.gov/sra), genome-wide transcriptome studies are available for only thirty-four species and vast majority of transcriptome data (above 90%) belong to *E. fetida*. Among Oligochaeta species, the main reason why researchers have tended to focus on *E. fetida* is that it is considered as a model species for annelids and its close relationships with soil biota as well as vermicomposting potential. Therefore, characterising the EST-SSR composition of *E. fetida* will be a valuable resource for researchers studying *Eisenia* genus. Recently, the comprehensive taxonomical studies on Lumbricidae family were made by Mısırlıoğlu et al. [18,19] and according to this study, a total of ten *Eisenia* species including *E. fetida* were reported in Turkey. We suppose that the EST-SSR markers provided here can be species-specific markers for taxonomic identification of *E. fetida* from other *Eisenia* species. The EST-SSR data can also be used for genus (*Eisenia*) specific markers. Besides, the researchers in India have recently deciphered the completed mitochondrial genome of *E. fetida* yielded ~15.3 Kb, it can also be used for phylogenetic or species level identification [20]. However, EST-SSRs are the preferred method at genome-wide level because of their multi-allelic and co-dominant nature aw well as fairly high level of transferability across species/genera.

4. Conclusion

Overall, in this study we comprehensively mined and characterised the publicly available transcriptome data from *E. fetida* to identify EST-SSR markers at genome-scale. The EST-SSR markers not only can serve as a powerful marker to allow taxonomic resolution to *Eisenia* genus from other earthworm genera, but also these markers provide reliable information to study the population structure and geographic distributions of *E. fetida*.

Acknowledgement

I gratefully acknowledge the cooperation and technical assistance of GENOKS Inc. (http://www.genoks.com.tr) by providing Sanger sequence service for amplicons.

References

1. Edwards, C. A, The importance of earthworms as key representatives of the soil fauna. In: Edwards C.A.(ed) Earthworm Ecology, 2nd edn. CRC press, NY, Washington, 2004, pp 3-11.

2. Dominguez, J, Edwards, C.A, Subler, S, A comparison of vermicomposting and composting, *Biocycle*, 1997, 38(4), 57-59.

3. Blouin, M, Hodson, M.E, Delgado, E.A, Baker, G, Brussaard, L, Butt, K.R, Dai, J, Dendooven, L, Peres, G, Tondoh, J.E, Cluzeau, D, Brun, J.J, A review of earthworm impact on soil function and ecosystem services, *European Journal of Soil Science*, 2013, 64(2), 161-82.

4. Gómez-Brandón, M, Domínguez, J, Recycling of solid organic wastes through vermicomposting: microbial community changes throughout the process and use of vermicompost as a soil amendment, *Critical Reviews in Environmental Science and Technology*, 2014, 44(12), 1289-1312.

5. Bouché, M.B, Strategies lombriciennes. In: U, Lohm, T, Persson(ed) Soil Organism as Components of Ecosystems, 1nd edn, Stockholm, 1977, pp 122-132.

6. Deka, H, Deka, S, Baruah, C. K, Das, J, Hoque, S, Sarma, H, Sarma, N. S, Vermicomposting potentiality of Perionyx excavatus for recycling of waste biomass of Java citronella--an aromatic oil yielding plant, *Bioresource Technology*, 2011, 102(24), 11212-11217.

7. Ponmani, S, Udayasoorian, C, Jayabalakrishnan, R. M, Kumar, K. V, Vermicomposting of paper mill solid waste using epigeic earthworm Eudrilus eugeniae, *Journal of Environmental Biology*, 2014, 35(4), 617-622.

8. Suthar, S, Vermicomposting potential of Perionyx sansibaricus (Perrier) in different waste materials, *Bioresource Technology*, 2007, 98(6), 1231-1237.

9. Cheng, J, Zhao, Z, Li, B, Qin, C, Wu, Z, Trejo-Saavedra, D. L, Luo, X, Cui, J, Rivera-Bustamante, R. F, Li, S, Hu, K, A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in Capsicum, *Scientific Reports*, 2016, 6, 18919.

10. Eujayl, I, Sledge, M. K, Wang, L, May, G. D, Chekhovskiy, K, Zwonitzer, J. C, Mian, M. A, Medicago truncatula EST-SSRs reveal cross-species genetic markers for Medicago spp, *Theoretical and Applied Genetics*, 2004, 108(3), 414-422.

11. Bolger, A. M, Lohse, M, Usadel, B, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics*, 2014, 30(15), 2114-2120.

12. Haas, B. J, A, Papanicolaou, M, Yassour, M, Grabherr, P. D, Blood, J, Bowden, M. B, Couger, D, Eccles, B, Li, M, Lieber, M. D, MacManes, M. Ott, J, Orvis, N, Pochet, F, Strozzi, N, Weeks, R, Westerman, T, William, C. N, Dewey, R, Henschel, R. D, LeDuc, N, Friedman, A, Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, *Nature Protocols*, 2013, 8(8), 1494-512.

13. Huang, Y, Niu, B, Gao, Y, Fu, L, Li, W, CD-HIT Suite: a web server for clustering and comparing biological sequences, *Bioinformatics*, 2010, 26(5), 680-682.

14. Nishimura, O, Hara, Y, Kuraku, S, gVolante for standardizing completeness assessment of genome and transcriptome assemblies, *Bioinformatics,* 2017, 33(22), 3635-3637.

15. Simão, F. A, Waterhouse, R. M, Ioannidis, P, Kriventseva, E. V, Zdobnov, E. M, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics*, 2015, 31(19), 3210-3212.

16. Jiang, Q, Li, Q, Yu, H, Kong, L, Genome-wide analysis of simple sequence repeats in marine animals-a comparative approach, *Marine Biotechnology*, 2014, 16(5), 604-619.

17. Zwarycz, A. S, Nossa, C. W, Putnam, N. H, Ryan, J. F, Timing and Scope of Genomic Expansion within Annelida: Evidence from Homeoboxes in the Genome of the Earthworm Eisenia fetida, *Genome Biology and Evolution*, 2015, 8(1), 271-281.

18. Csuzdi, C, Zicsi, A, Misirlioğlu, M, An annotated checklist of the earthworm fauna of Turkey (Oligochaeta: Lumbricidae), *Zootaxa*, 2006, 1175, 1-29.

19. Mısırlıoğlu, M, Valchovski, H, Contributions to Earthworm (Clitellata; Annelida) Fauna of Turkish Thrace, *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2017, 21(3), 385-388.

20. Bhambri, A, Dhaunta, N, Patel, S. S, Hardikar, M, Srikakulam, N, Shridhar, S, Vellarikkal, S, Suryawanshi, H, Pandey, R, Jayarajan, R, Verma, A, Kumar, V, Bhatt, A, Gautam, P, Rai, M, Khan, J. A, Fromm, B, Peterson, K. J, Scaria, V, Sivasubbu, S, Pillai, B, Insights into regeneration from the genome, transcriptome and metagenome analysis of *Eisenia fetida*, *bioRxiv*,2017.

Supplementary File. The fasta files indicated below are amplified contig sequences including EST-SSR motifs. The yellow and blue highlighted sequences are forward and reverse sequences, respectively whereas underlined sequences show EST-SSR motif.

>Efetida.157721

GCTT**GAGAGCAAGACAACTCTGTG**TGAATTTTGCAT
ACGCTTAGATACCTTAAGGGGGCTATATAAATGTTG
ACAAGCTGAAATGTAAACCAGAAAGTTATGAACTCG
CAGAAATTGGACTTTTTTTGTTTGTGAGTTACATGAA
AATTACTCTAACTTAAAATGAAGAAATATATATGCAT
ACACATTC**ACACACACACACACACACACACACAC**AAA
CAAACACACACACATAGATATATACCCCATATAA
TCTTTTTCTGTGATTCAAGGTTACTTAGCTCTACAGG
ATGGAGAATAT**CAGATCATGCTATGCAGTGT**ACTG
CTTCTGAAA

>Efetida.43077

AGC**TTGTGTTCGATGGCATTAGG**AGTGGCAAATGA
TTGTGCTGGTTTGCAGTTGAAATGTTCATGCAAGGTA
CCAGTCCTGTTCTCTGGCAATGAATGCAGCGAAGGA
CTCGTTTTCAGGTGCACGTCATGCAGCGAAGGACTCG
TTTTCAGGTGCACGTCATGCAGTGAAGGACTCGTTTT
CAGGTGCACGTCATGCAGTGAAGGACTCGTTTTCAG
AGGCACGCCCTCTGTTGTCAACTTTGGAATGCTGAAG
CTATGACTAAATGTGGCCTTATTCCCTGAGTTTAGAA
AATTCATCTCTGTTCCTGTATGTGAAGGCAGATCT**TG**
CTGCTGCTGCTGCTGCTGCTGCTGAAGCGGCTGCT
GCTCCTC**TTGCTCATCTCCATCATCCT**TCT

>Efetida.44776

TC**CCAACTCAGACCAATTCCTG**ATGTTCTGGAGGTC
TCCATTCTATCTTGATCAAATATGTGATTCTTCAAGC
ATCAATTCTATAATTACAATATCACTA**ATAATAATAA**
TAATAATAATAATAATAATAATAATAATAAAA
GCAATAACATTCATAATAAATACATTTCCATTTTACT
AAAATAATAATAGTGTCATATAAATAACAGTCAGCA
ACTACGATACAGGAGGAGTTTCAGTGGACTCTAAGC
CACAAATATGCATCTGAATTACATTAACCACTTAGAA
ATCTATGAACAGCAAGAAACACAACAAACGACCA**CA**
GACAGCTTAATCTTGCCAAGTG

>Efetida.92477

GAAAAACCG**GTCGTTCACATAAGACAGCA**TATCAT
ATGATGATCTCTGC**AGCAGCAGCAGCAGCAGCAGC**
AGCAGCAGCAAACATGATTATGCAATAGACGAGGA
GTCATCAACAGAAGATGAGGATGTTAACGACTGGTC
TTTGCCGAGTGATGCAGGAGGCAGAGGCAATGTGCT
GCTGCAGTCCAGCAATGACGAGAGGTAAACGGAGGT
ACGGTTGAAACGCTCAGAGATGCATTCAGCTGTTAG
ACGCGCTTCTGGATCTTGGTCCCAGCATTCCTCTA**TT**
GTTCGACACACAATCTCCAAAT

>Efetida.47040

CTCGGG**TCGGAAGTGCCTCTTAATCT**CTGAGTAGC
ACTATTTAC**ATCATCATCATCATCATCATCATC**
ATCATCCACTATGGCTATTTACATATTCAATTACAAT
TACAAATTGTTACAATGGAGTGACAATATAACTGCT
AGTTTGTTCTGGCTTAGAGCTCGTGATAATAAATTCT
GTGACCTTATACTCTGTATAACCTCTAAACAAAGGCT
ATAGCTTTAACTAAACTCTGTAATTTCAGAATAATTT
CTGAATATACG**CGTAGTCACCATAGGCCTAT**TAA

>Efetida.67884

ATTCT**TTGTCATCTGCGTCATCATT**ATGACTGTGCT
CTACAGAAGACGTCGTCGACGTGAAAAAACCAACAT
CATCTACAAAAAGGCTCCGACTGCAGAAGATCCTAC
AGCTACATCGAAAATGTGAAGATAACTTGAAATCGT
GACCTTGTCAAACACTCAGTCAAGAAAACGCCACGA
AACGTCTTGTCTCTAGTACATCAAAAACTATATAAAT
TATGTATATTGTATATATATATGTATATATACATTCTT
ATAT**ATACATACATACATACATACATACATACATA**
CATACATACATACATACATACATAATACG**ATGACT**
AGCTTCAAGGCCACAT