Data on the identification of microsatellite markers in Eisenia fetida and Eisenia andrei.

Marta Jaskulak, Franck Vandenbulcke, Agnieszka Rorat, Maxime Pauwels, K. Zorena, P. Grzmil, B. Plytycz

Data in Brief 45 (2022) 108612

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data on the identification of microsatellite markers in *Eisenia fetida* and *Eisenia andrei*

Marta Jaskulak[a,b,*], Franck Vandenbulcke[a], Agnieszka Rorat[a], Maxime Pauwels[a,c], Kararzyna Zorena[b], Paweł Grzmil[d], Barbara Płytycz[e]

[a] *Univ. Lille, IMT Lille Douai, Univ. Artois, Yncrea Hauts-de-France, ULR4515 - LGCgE, Laboratoire de Génie Civil et géo-Environnement, Lille F-59000, France*
[b] *Department of Immunobiology and Environment Microbiology, Faculty of Health Sciences with Institute of Maritime and Tropical Medicine, Medical University of Gdansk, Gdansk 80-210, Poland*
[c] *Univ. Lille, CNRS, UMR 8198 - EEP - Laboratoire Evolution Ecologie Paléontologie, Lille F-59000, France*
[d] *Laboratory of Genetics and Evolutionism, Institute of Zoology and Biomedical Research, Jagiellonian University, Krakow, Poland*
[e] *Department of Evolutionary Immunology, Institute of Zoology and Biomedical Research, Jagiellonian University, Krakow, Poland*

ARTICLE INFO

Article history:
Received 19 February 2022
Revised 11 September 2022
Accepted 12 September 2022
Available online 18 September 2022

Keywords:
SSR markers
Microsatellites
Molecular markers
Simple sequence repeats
Earthworms

ABSTRACT

Eisenia fetida and *Eisenia andrei* are closely related earthworm species that play a crucial part in soil and influence its structure and organic matter cycling. Due to their essential environmental role, they are widely used as model organisms in a vast spectrum of research areas. In this work, we partially sequenced genomes of *E. fetida* and *E. andrei*, using Illumina technology (Nano 2 × 250 v2 - MiSeq) and *de novo* assembly strategy. A total of 3785 and 4258 microsatellite or Simple Sequence Repeat (SSR) markers were identified within *E. fetida* and *E. andrei* genomic DNA, respectively. The microsatellite markers will facilitate the analyses of genetic diversity and population genetics studies for the two selected earthworm species and their interspecific hybrids.

* Corresponding author.
E-mail address: marta.jaskulak@gumed.edu.pl (M. Jaskulak).

https://doi.org/10.1016/j.dib.2022.108612

Specifications Table

Subject	Omics: Genomics
Specific subject area	Microsatellite markers data
Type of data	Tables - data from Next-Generation Sequencing, SSR markers, microsatellite markers, and designed primers
How data were acquired	Partial genome sequencing was performed by Illumina Nano 2 × 250 v2 – MiSeq (Illumina, Inc. USA). The sequences were assembled with the PrinSeq software (PRINSEQ software v0.20.4 (http://prinseq.sourceforge.net/). The final analysis and the design of the primers were carried out with QDD v3 software (http://www.imbe.fr/~emeglecz/qdd)
Data format	Raw Analysed Filtered
Parameters for data collection	Supravitally amputated tail tips of adult earthworms, DNA extraction and *de novo* sequencing
Description of data collection	Genomic DNA was extracted from amputated tail tips of two earthworm species with Eur_x Universal DNA/RNA/Protein extraction kit (Eur_x, Poland) after homogenisation in liquid nitrogen. Nano 2 × 250 v2 - MiSeq. The sequences were assembled with the PrinSeq software (PRINSEQ software v0.20.4 (http://prinseq.sourceforge.net/). The final analysis and the design of the primers were carried out with QDD v3 software (http://www.imbe.fr/~emeglecz/qdd).
Data source location	Adult composting *E. andrei* and *E. fetida* earthworms deriving from laboratory stocks at the University in Lille (France), cultured for a decade in the laboratory of the Institute of Zoology and Biomedical Research of the Jagiellonian University (Krakow, Poland), and in parallel for last four years in the laboratories of Rzeszow University (Poland).
Data accessibility	The nucleotide sequences of raw reads, the designed primer pairs and the products of their amplification are presented in S1 and S2 tables. The primer sets are also deposited in Mendeley Data: https://data.mendeley.com/datasets/hs67kkzkr5/1 The raw sequences are deposited in NCBI under the following accession numbers: for *Eisenia andrei* – BioProject PRJNA843202, for *Eisenia fetida* – BioProject PRJNA842580, links to both Bioprojects are provided below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA843202 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA842580
Related research article	Jaskulak, M., Rorat, A., Vandenbulcke, F., Pauwels, M., Grzmil, P., Plytycz, B. (2022). Polymorphic microsatellite markers demonstrate hybridization and interspecific gene flow between lumbricid earthworm species, Eisenia andrei and E. fetida. PLoS ONE 17(2): e0262493. https://doi.org/10.1371/journal.pone.0262493

Value of the Data

- Newly developed microsatellite markers of *E. fetida* and *E. andrei* can be useful for molecular studies, population genetics, the construction of linkage maps, QTL mapping and for studies on hybridization of these two earthworm species.
- The information on the sequenced and identified SSR motifs and markers will be useful for the assessment of genetic diversity in those species of earthworms.
- Data presented in this manuscript can be used as the training data in SSR and population structure studies.

1. Data Description

Eisenia Andrei and *E. fetida* lumbricid earthworms are important model species used in a vast range of research studies including in comparative immunology, ecotoxicology, biomedicine and environmental studies, thus their proper identification is a crucial task for many scientific purposes [1–3].

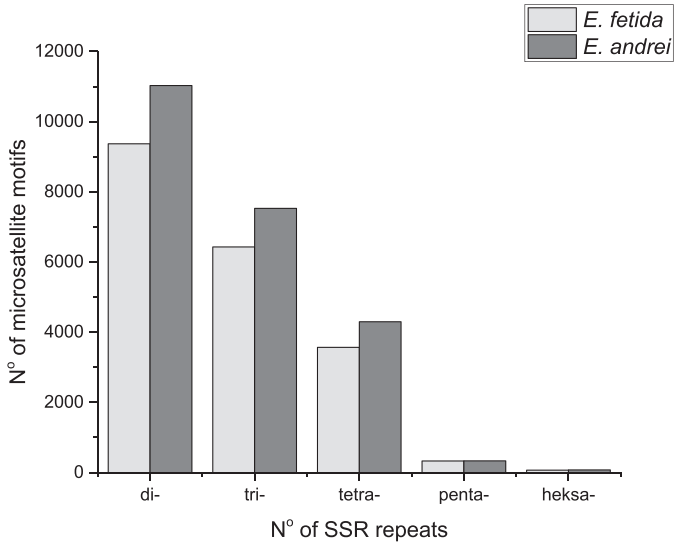

Fig. 1. Frequency distribution of SSR loci by motif length on the assembled genomic sequences of *E. fetida* and *E. andrei*. The graph is based on a total of 19 758 and 23 255 SSR markers detected in non-redundant genomic DNA of *E. fetida* and *E. andrei*, respectively. Di, tri, tetra, penta, and hexa, refer to dinucleotides, trinucleotides, tetranucleotides, pentanucleotides, and hexanucleotides, respectively.

Table 1
Data of number of reads obtained by Illumina Nano 2 × 250 v2 - MiSeq.

Id	Number of sequences	Total number of base pairs	Min length(bp)	Max length(bp)	Mean length(bp)	%GC	%N
E. fetida R1	1 709 140	425 014 601	35	250	248,67	41,27	4,36E-04
E. fetida R2	1 709 140	425 103 615	35	250	248,72	41,29	1,89E-04
E. andrei R1	2 094 624	519 503 343	35	250	248,02	41,41	8,62E-04
E. andrei R2	2 094 624	519 661 143	35	250	248,09	41,47	2,09E-04
Ef assembled	918 733	308 601 541	153	440	335,90	40,55	
Ea assembled	1 140 654	377 768 818	156	440	331,19	40,61	

Raw sequencing data for *Eisenia fetida* and *Eisenia andrei* was produced by *de novo* sequencing using Illumina Nano 2 × 250 v2 – MiSeq (Illumina, Inc. USA). The obtained data were quality trimmed, filtered, and assembled. The raw sequences were quality checked and assembled with the FastQC v0.11.9 software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) [4]. The trimming and filtering of raw data was performed using fastqp (https://github.com/OpenGene/fastp). All reads containing more than 5% of unknown nucleotides, and low-quality reads were discarded. Short reads (<35 bp) were removed from the filtered data [5]. Repetitive microsatellite sequences raging from dinucleotides to hexanucleotides were identified for *E. fetida* and *E. andrei*. Only SSRs with an overall repeat motif size from 2 to 7 bp and a length larger than 12 bp were considered adequate (Figure 1). For *E. fetida* 1 709 140 million 2 × 250 bp reads were generated whereas for *E. andrei* 2 094 624 million 2 × 250 bp reads (Table 1, Table 2). The raw sequences are deposited in NCBI under the following accession numbers: for *Eisenia andrei* – BioProject PRJNA843202, for *Eisenia fetida* – BioProject PRJNA842580.

Specific primer pairs were designed from flanking sequences of di to hexanucleotides of *E. fetida* and *E. andrei* (supplementary data, tables 1 and 2). The final analysis and the design of the primers were carried out with QDD v3 software and Primer 3 software (https://www.imbe.fr/~emeglecz/qdd), (https://primer3.ut.ee/). Overall, 19 758 pairs of primers were designed for *E. fetida* and 23 255 for *E. andrei*. After that, 3 777 primer pairs for *E. fetida* and 4 258 primer pairs

Table 2
Summary of the frequency of SSRs from *E. fetida* and *E. andrei* with different numbers of tandem repeats. Di, tri, tetra, refer to dinucleotides, trinucleotides, and tetranucleotides, respectively.

	Largest SSRs	
Motif length	*E. fetida*	*E. andrei*
Di-	(AG)26	(AG)32
Tri-	(AAG)44	(AAT)40
Tetra-	(AGAT)26	(AATG)26

were validated with QDD v3 software ((https://www.imbe.fr/~emeglecz/qdd)). Primer sequence, product size and the sequence of each product are available in supplementary data (S1, S2) [6]. Moreover, the primers products sequences are presented in the supplementary data and the supplementary material is available on Mendeley Data portal [6].

2. Experimental Design, Materials and Methods

2.1. Earthworm material and DNA extraction

To create two separate pools of DNA for sequencing, twelve samples of *E. fetida* earthworms and 12 samples of *E. andrei* earthworms were used. Out of 12 specimens of each species, 4 derived from a previously genotyped lineage from a collection of the Institute of Zoology and Biomedical Research of the Jagiellonian University (Krakow, Poland), and 8 derived from two different lineages (2×4) from a previously genotyped collection of University of Lille (Lille, France). The genomic DNA was extracted from supravitally amputated tail tips via using approximately 50 mg of tissue per sample. Tissue homogenization was performed via freezing the tissue in liquid nitrogen and homogenization in a mortar. Afterward, the genomic DNA was extracted with a Universal DNA/RNA/Protein kit (Eur_x, Poland), following the manufacturer's instructions. DNA quality and quantity were then checked on the SPECTROstar Nano spectrometer (BMG LABTECH, Germany). Prior to sequencing, the quality and quantity of extracted DNA were rechecked using 2100 Bioanalyzer (Agilent Technologies, USA).

2.2. Next-generation sequencing

The Illumina paired-end libraries for both species were prepared with the MiSeq Reagent Kits v2 (Illumina, San Diego, CA, USA). The paired-end libraries were sequenced using Illumina HiSeq 2500 Sequencer (Macrogen Inc., Seoul, Korea). The quality of raw genomic sequences following sequencing was assessed using FastQC software version 0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc). Afterward, data was trimmed and filtered with PRINSEQ software v0.20.4 (https://prinseq.sourceforge.net/) [5].

2.3. In silico identification of putative SSRs and primer design

The contig sequences in FASTAq files were screened with a repeat motif size range of 2–6 bp and a length of >12 bp, using MIcroSAtellite software which allowed to identify the potential SSR markers (microsatellite markers) in both earthworm species. The program allowed for direct primer design using PRIMER 3 software (https://primer3.ut.ee/) [7] by searching for microsatellite repeats and primer annealing sites in the flanking regions (S1 and S2 Tables) [8,9].

Ethics Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have influenced the work reported in this article.

Data Availability

Eisenia fetida and Eisenia andrei microsatellites (sequences and primers) (Original data) (Mendeley Data).

CRediT Author Statement

Marta Jaskulak: Investigation, Data curation, Visualization, Writing – original draft; **Franck Vandenbulcke:** Conceptualization, Resources, Writing – review & editing, Supervision; **Agnieszka Rorat:** Conceptualization, Writing – review & editing; **Maxime Pauwels:** Methodology, Resources; **Kararzyna Zorena:** Writing – review & editing; **Paweł Grzmil:** Resources, Writing – review & editing; **Barbara Płytycz:** Conceptualization, Writing – original draft, Writing – review & editing, Resources, Formal analysis, Funding acquisition, Investigation, Supervision.

Acknowledgments

Research was financially supported the National Centre of Science in Poland (2016/23/B/NZ8/00748) and by the Jagiellonian University (K/ZDS/005405).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2022.108612.

References

[1] B. Plytycz, J. Bigaj, A. Rysiewska, A. Osikowski, S. Hofman, A. Podolak, P. Grzmil, Impairment of reproductive capabilities in three subsequent generations of asymmetric hybrids between *Eisenia andrei* and *E. fetida* from French, Hungarian and Polish laboratory colonies, Plos One 15 (7) (2020), doi:10.1371/journal.pone.0235789.
[2] A. Podolak, J. Kostecka, S. Hofman, A. Osikowski, J. Bigaj, B. Plytycz, Annual reproductive performance of *Eisenia andrei* and *E. fetida* 2 in intra- and inter-specific pairs and lack of reproduction of isolated virgin earthworms, Folia Biol. 68 (1) (2020) 1–6, doi:10.3409/fb_68-1.01.
[3] M. Jaskulak, A. Rorat, L. Kurianska-Piatek, S. Hofman, J. Bigaj, F. Vandenbulcke, B. Plytycz, Species-specific Cd-detoxification mechanisms in lumbricid earthworms Eisenia andrei, Eisenia fetida and their hybrids, Ecotoxicol. Environ. Saf. 208 (2021) 111425, doi:10.1016/j.ecoenv.2020.111425.
[4] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) 884-i890.
[5] R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets, Bioinformatics 27 (6) (2011) 863–864.
[6] E. Rozen, H. Skaletsky, Primer 3 on the WWW for general users and for biologist programmers, Methods Mol. Biol. 132 (2000) 365–386.
[7] Marta Jaskulak, Eisenia fetida and Eisenia andrei microsatellites (sequences and primers)", Mendeley Data (2022) V1, doi:10.17632/hs67kkzkr5.1.

[8] B. Plytycz, J. Bigaj, A. Osikowski, S. Hofman, A. Falniowski, T. Panz, P. Grzmil, F. Vandenbulcke, The existence of fertile hybrids of closely related model earthworm species, Eisenia andrei and E. fetida, PLOS ONE 13 (1) (2018) e0191711, doi:10.1371/journal.pone.0191711.

[9] M. Jaskulak, A. Rorat, F. Vandenbulcke, M. Pauwels, P. Grzmil, B. Plytycz, Polymorphic microsatellite markers demonstrate hybridization and interspecific gene flow between lumbricid earthworm species, *Eisenia andrei and E. fetida*, PLoS ONE 17 (2) (2022) e0262493, doi:10.1371/journal.pone.0262493.