Molecular mechanisms behind global distribution of earthworm revealed by the genome

Xing Wang^{1,2,†*}, Yi Zhang^{3,†}, Yufeng Zhang^{4,†}, Mingming Kang^{5,†}, Yuanbo Li^{1,2},
Samuel W James⁶, Yang Yang⁷, Yanmeng Bi⁸, Hao Jiang¹, Xin Zhou⁹, Yi Zhao^{1*},
Zhenjun Sun^{1,2*}

¹College of Resources and Environmental Sciences, China Agricultural University,
Beijing, China.

²Beijing Key Laboratory of Biodiversity and Organic Farming, Beijing, China.

³School of Environmental Science & Engineering, Southern University of Science
and Technology, Shenzhen, Guangdong, China.

⁴College of Biology, Langfang Normal University, Langfang, Hebei, China.

⁵College of Biological Sciences, China Agricultural University, Beijing, China.

⁶Maharishi University of Management, Fairfield, Iowa, USA.

⁷Beijing Gencode Diagnostics Laboratory, Beijing, China.

⁸Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences
and Peking Union Medical College, Beijing, China.

⁹College of Plant Protection, China Agricultural University, Beijing, China.

[†]These authors contributed equally: Xing Wang, Yi Zhang, Yufeng Zhang, Mingming
Kang.

* Correspondence to: College of Resources and Environmental Sciences, China
Agricultural University, Beijing 100193, China

E-mail address: swwangxing@cau.edu.cn (X. Wang); zhaoyis@163.com (Y. Zhao);
sun108@cau.edu.cn (Z.J. Sun).

Abstract

Earthworms (Annelida: Crassicitellata), are widely distributed around the world due to their great adaptability. However, lack of a high-quality genome sequence prevents gaining the many insights into physiology, phylogeny, and genome evolution that could come from a good earthworm genome. Herein, we report a complete genome assembly of the earthworm *Amyntas corticis* of about 1.2 Gb, based on a strategy combining third-generation long-read sequencing and Hi-C mapping. A total of 29,256 protein-coding genes are annotated in this genome. Analysis of resequencing data indicates that this earthworm is a triploid species. Furthermore, gene family evolution analysis shows that comprehensive expansion of gene families in the earthworm genome has produced more defensive functions compared with other species in Annelida. Quantitative proteomic iTRAQ analysis shows 97 immune related proteins and 16S rDNA sequences shows 88 microbes with significantly response to pathogenic *Escherichia coli* O157:H7. Our genome assembly provides abundant and valuable resources for the earthworm research community, serving as a first step toward uncovering the mysteries of this species, may explain its powerful defensive functions adapt to complex environment and invasion from molecular level.

Introduction

The earthworm belongs to oligochaete annelid distributed all over the world. Recently, a research focus on diversity of earthworm at the global scale showed that this underground species has more local richness and abundance in temperate zone, but more dissimilarity across different locations in tropical zone¹. More interestingly, it also showed that climate is the key factor affecting distribution of earthworm, in spite of the soil property and habitat cover through modeling based on integrated big data¹. Soil properties and habitat covers represent diverse environments, which means different threats and dangers to species, especially for underground ones. Thus, it could be proposed that there must be unique and robust molecular mechanisms in individual earthworm helping it face to and live in adverse and complex environments.

Earthworm has been recognized as the species with profound ecological and economic impacts on soil^{2,3} and considered as the “soil engineer”. On the other hand, earthworm is also an invasive species with not ignorable affections on soil ecosystem⁴⁻⁶, including local diversity of native species⁷. Behind the invasive characteristic, a close association between polyploidy and parthenogenetic reproduction has been identified in earthworm species of *Lumbricidae* family⁸⁻¹¹. Polyploidy of earthworm provides more genomic materials for evolving novel phenotypes, while parthenogenetic reproduction ensures the fixation of them, which is benefit for its rapid adaptation to a new environment.

Amyntas corticis, an earthworm species of *Megascolecidae* family, which is a wild species with global distribution. Particularly, *A. corticis* is also a cosmopolitan invasive species originating from Asian. There was a study suggested that the invasive ability of *A. corticis* is due to its greater mobility and polyploidy or parthenogenetic reproduction¹². To reveal the whole picture of molecular mechanisms behind the strongly adaptive ability of *A. corticis*, we sequenced its genome and generated a complete assembly with a length of 1.29 Gb in total, including 42 chromosome-level scaffolds with N50 length of 31 Mb (Table 1 and Supplementary Fig. 1). In addition to the genome, we also detected protein mass spectrum for the body of earthworm and sequenced 16S rDNA for the intestinal tract of earthworm under the treatment of a pathogenic *Escherichia coli* stress *O157:H7*¹³, which mimics the severely adverse environmental factor, to profile how earthworm genome functions when facing to stresses.

Results

A. corticis is triploid

Based on the data resequenced from two individual earthworms of *A. corticis*, we called genome wide SNPs and constructed the coverage distribution of heterozygous k-mers. Following that, we evaluated the ploidy of the genome of *A. corticis*. First of all, the frequency distribution of SNPs at biallelic loci showed that there were two peaks near

the frequency of 1/3 and 2/3 appeared (Fig. 1d), suggesting most of biallelic loci have three copies in the genome of *A. corticis*. In addition, the result of statistical analysis based on the frequency distribution of genome wide SNPs showed the feature of triploidy possesses the lowest delta log-likelihood (Fig. 1c), suggesting the genome of *A. corticis* is triploid with the largest probability. Furthermore, the coverage distribution related to 88% of heterozygous k-mers concentrated at 3n for the total coverage of k-mer pairs and 1/3 for the normalized coverage of minor k-mers (Fig. 1e), implying the triploid feature of the genome of *A. corticis*. Lastly, karyotype analysis of four earthworms showed that chromosome number of each individual was larger than 120 (Fig. 1a-b). In combination with that we have assembled 42 chromosome-level scaffolds, the large number of chromosomes also implied the triploid feature of *A. corticis* genome. Thus, multiple independent proofs supported the genome of *A. corticis* is triploid.

Abundant gene contents related to defensive

Compared with other two annelids, *Capitella teleta* and *Helobdella robusta*¹⁴, the genome size of earthworm *A. corticis* is significantly larger than them with 3.87 and 5.49 folds, respectively. This phenomenon suggests that there should be comprehensive expansion of gene families happened in earthworm genome. To make it clear, we constructed orthologous groups for genes in earthworm and other invertebrate lineages, including annelids, mollusc, flatworms, roundworms and vertebrates based on TreeFam database^{15,16}.

To reveal the evolutionary relationship of earthworm and other species, we construct the phylogenetic tree covering all of species mentioned above based on single-copy orthologous groups. As is shown in Figure 2a, the constructed tree revealed evolutionary relationship of all species consistent with the common sense. However, when focusing on orthologous groups with significantly evolutionary rate, the Spearman's rank correlation coefficient matrix of group size of different species shows earthworm was away from its phylogenetic tree neighbor of annelids and mollusc, but

close to flatworms, roundworms and two snakes in vertebrates (Fig. 2b). These contradictory phenomena imply that environmental or other factors had affected the evolutionary path of earthworm away from other annelids, resulting in a unique annelid genome.

Detailed analysis of orthologous groups with significantly rapid evolutionary rate showed that more than 70% of them in earthworm are expanded compared with the most recent common ancestor (MRCA) of annelids, which is consistent with its large genome size. Furthermore, the principle component analysis (PCA) using group size of different species also showed that earthworm is away from other annelids along the first principle component (Fig. 2c). What is more interesting, earthworm diverged away from other annelids with a stepwise manner. At each step, species of flatworms, roundworms or vertebrates appeared. For convenience, we defined five steps along the divergence, which are represented by *Echinococcus multilocularis*, *Ascaris lumbricoides*, *Ophiophagus Hannah*, *Schmidtea mediterranea* and itself. Then we extracted orthologous groups with the size in earthworm, which is larger than it in annelids and mollusc while smaller than or equal to it in representative species of each step, as expanded gene families of earthworm at this step. The result of functional enrichment analysis^{17,18} showed that there were the most enriched expanded functions at the last step. Besides fundamental functions such as anatomical structure development, signal transduction, cell differentiation and transport, many functions related to immune system process, response to stress and homeostatic process expanded (Fig. 2d). It could be suggested that the evolutionary process at the last step ensures abundant gene contents related to defensive in the genome of *A. corticis* and thus equips earthworm with large number of molecular arms and effective strategies to live in hostile environments. Furthermore, we reconstructed the evolutionary path of functions related to immune system process, response to stress and homeostatic process through plotting odds ratio at each step for significantly enriched ones (Supplementary Fig. 2-4).

Defensive to pathogenic *E. coli* O157:H7

We fed earthworm with pathogenic *E. coli* O157:H7¹³, which mimics the severely adverse environmental factor, and then detected protein mass spectrum for the body of earthworm and sequenced 16S rDNA for the intestinal tract of earthworm. To profile the dynamic pattern of expression and intestinal microorganism constitution, we selected samples at several time points after treatment of pathogenic *E. coli* O157:H7 to detect protein mass spectrum while sequence 16S rDNA. Following normalization of expression level and 16S rDNA abundance, we conducted analysis of variance (ANOVA) for both of them.

In terms of protein mass spectrum, we found 23 genes expressed with significant variation among samples (Fig. 3). First of all, none of these 23 genes expressed in control samples, indicating their specific roles in response to stresses. Furthermore, these genes could be classified as two categories according to whether expressed on the immediate 3rd day after feed of pathogenic *E. coli* O157:H7. Genes expressed on the 7th day or both of the 7th and 28th days but not in the 3rd day were classified as the first category, while genes expressed on the 3rd day were classified as the second category. There are 6 genes belonging to the first category, representing the pattern of later regulation to resist pathogenic bacteria. Two of these genes contained abhydrolase domain or F-BAR domain, while others were novel ones appeared in earthworm without any reliable functional counterparts in otherwise species. The left 17 genes belong to the second category, representing a pattern that immediate activation followed by subtle regulation. Genes classified as the second category include heat shock 70kDa proteins, M1 metallopeptidases, and SRCAP complex, which expressed not only on the 3rd day but on either the 7th or 28th day.

For 16S rDNA, we found the abundance of 9 kinds of microorganisms varied significantly among samples (Fig. 4). The dynamic changing patterns of abundance of these 9 kinds of microorganisms could be classified as 2 categories, including down-up trend and novel appearance on the 28th day. The pattern of down-up trend revealed the

reconstruction process of native microorganisms in the intestinal track of earthworm, which could be considered as an indicator representing defensive effects to the pathogenic bacteria. In particular, soil microorganisms belong to class *Acidobacteria*-6¹⁹ or order *Ellin6067*²⁰ reached the highest abundance on the 28th day compared with others, revealing the active uptake of earthworm to resist biotic stresses. Microorganisms with novel appearance on the 28th day, including drilosphere microorganisms belong to family *AKIW874*²¹, also revealed the active uptake of earthworm helping to survive under stresses.

Profile of defensive genes in earthworm genome

Great adaptation and strong vitality of earthworm have always attracted attentions of biologists and ecologists to make it clear. Thus, several well determined defensive genes were cloned and studied in detail²²⁻³⁷. However, due to the absence of a complete genome, it is failed to construct an overview on defensive systems of earthworm. Here, we systematically profiled copies of defensive genes with reliable evidence in earthworm from the assembled complete genome. Predicted genes were considered as orthologs of well determined defensive genes if there were reciprocally best hits between them, which were listed in Supplementary Table 1. Then, we analyzed genomic location and expression pattern for orthologous group members of defensive genes (Fig. 5).

Discussion

Profiling of molecule mechanisms behind defensive of earthworm help to better understand the global distribution of this species and the origin of sophisticated immunity in vertebrates. In addition, it can also facilitate the usage of earthworm as an animal model in comparative immunology studies and ecotoxicology. Except for its value in research communities, earthworm is useful in monitoring environmental pollutions, and furthermore, is considered as an abundant source of biological active compounds with potential usage in industrial or medical fields.

212 **Figures and Tables**

Figure 1 | The triploidy feature of *A. corticis*. **a**, Imaging of chromosomes in an individual *A. corticis* resulted from karyotype analysis. **b**, Chromosome numbers of four *A. corticis* individuals determined via karyotype analysis. **c**, Delta log-likelihood values under hypotheses with different ploidy features. **d**, Distribution of frequencies related to biallelic loci. **e**, Coverage pattern of heterozygous k-mer pairs, in which X axis indicates normalized minor k-mer coverage, while Y axis indicates the total coverage of heterozygous k-mer pairs.

225

226

227 **Figure 2 | Species tree and evolving pattern showed by orthologous groups with**
228 **significantly evolutionary rate. a**, Species tree constructed based on conserved single-
229 copy orthologous groups. The scale indicates 95 million years. **b**, Heatmap and
230 dendrogram of Spearman's rank correlation coefficient matrix between each pair of
231 species related to gene counts in orthologous groups with significantly evolutionary
232 rate. The distance between each pair of species was defined as the difference value
233 between 1 and Spearman's rank correlation coefficient. **c**, Clustering of species revealed
234 by principal component analysis (PCA) based on gene counts in orthologous groups
235 with significantly evolutionary rate. **d**, Count of enriched functions at each evolutionary
236 step, in which each line represents a category of GO slims. AscarisL or Alu: *Ascaris*
237 *lumbricoides*, AscarisS or Asu: *Ascaris suum*, Caenorhabditis or Cel: *Caenorhabditis*
238 *elegans*, Capitella or Cte: *Capitella teleta*, Danio or Dre: *Danio rerio*, Drosophila or
239 Dme: *Drosophila melanogaster*, Earthworm or Aco: *Amyntas corticis*,
240 EchinococcusG or Egr: *Echinococcus granulosus*, EchinococcusM or Emu:

241 *Echinococcus multilocularis*, Gallus or Gga: *Gallus gallus*, Helobdella or Hro:
 242 *Helobdella robusta*, Homo or Hsa: *Homo sapiens*, Hymenolepis or Hmi: *Hymenolepis*
 243 *microstoma*, Lottia or Lgi: *Lottia gigantea*, Mus or Mmu: *Mus musculus*, Ophiophagus
 244 or Oha: *Ophiophagus hannah*, Python or Pbi: *Python bivittatus*, Rattus or Rno: *Rattus*
 245 *norvegicus*, Schistosoma or Sma: *Schistosoma mansoni*, Schmidtea or Sme: *Schmidtea*
 246 *mediterranea*, Taenia or Tso: *Taenia solium*, Xenopus or Xtr: *Xenopus tropicalis*.

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

Figure 3 | Expression heatmap of earthworm genes with significantly differential expression along days after feeding of pathogenic bacteria. In which, normalized expression count produced by protein mass spectrum and gene annotation retrieved from HGNC database were labeled.

291

f:Polyangiaceae	6	0	92
f:Cryomorphaceae	6	0	98
f:Pelobacteraceae	19	11	152
c:Acidobacteria-6	78	100	738
o:Ellin6067	133	82	902
g:Chloronema	0	0	41
f:AKIW874	0	0	59
g:Rubellimicrobium	0	0	110
o:Gemmatimonadales	0	0	35
	ckw	ew3	ew28

292

293 **Figure 4 | Abundance of microorganisms in intestinal tract of earthworm with**
294 **significant difference along days after feeding of pathogenic bacteria.** In which,
295 normalized abundance and species origin of microorganisms were labeled.

296

297

298

299

300

301

302

303

304

305

306

307

308

309

Figure 5 | Expression heatmap of orthologous group members of well determined defensive genes in earthworm along days after feeding of pathogenic bacteria. In which, orthologous group members were clustered based on normalized expression count produced by protein mass spectrum.

Table 1 Summary of features related to genome assembly

Assembly feature	Contig	Scaffold
Total length (bp)	1,190,923,171	1,285,148,884
Number of sequences	16,882	1,042
Length of the longest sequence (bp)	1,225,102	55,651,430
N50 length (bp)	117,165	31,950,757
N90 length (bp)	33,690	21,414,444
GC content	40.34%	40.34%
N content	0	0.26%
Number of gaps	0	12,771
Length of gaps (bp)	0	3,283,278

Supplementary Materials

Supplementary Figure 1 | Genome assembly and annotation. **a**, Heatmap of interaction intensity of intra-chromosome and inter-chromosome contacts detected by Hi-C for 42 chromosome-level scaffolds. **b**, The genome-wide distribution of annotated functional elements. The outermost circle represents 42 assembled chromosome-level scaffolds. Then, from outside to inside, the four circles represent the distribution of repeat sequences, non-coding genes, coding genes and SNPs, respectively. Connecting lines in the center represent duplicated genomic regions. The gray, blue and red colors of the connecting lines represent alignment scores generated by MCSanX that are lower than 500, between 500 and 1,000, and greater than 1,000, respectively.

Supplementary Figure 2 | Odds ratios of significantly enriched functions related to the GO slim of immune system process along evolutionary steps.

Supplementary Figure 3 | Odds ratios of significantly enriched functions related to the GO slim of response to stress along evolutionary steps.

Supplementary Figure 4 | Odds ratios of significantly enriched functions related to the GO slim of homeostatic process along evolutionary steps.

**Supplementary Table 1 Information of orthologous group members of well
determined defensive genes in earthworm.**

349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378

References

- 1 Phillips, H. R. P. *et al.* Global distribution of earthworm diversity. *Science* **366**, 480-485, doi:10.1126/science.aax4851 (2019).
- 2 Darwin, C. *The formation of vegetable mould through the action of worms.* (Cambridge university press, 1881).
- 3 Vila, M. B., C.; Pysek, P. How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Frontiers in Ecology and the Environment* **8**, 135-144 (2010).
- 4 Hendrix, P. F., Callaham, M.A., Drake, J.M., et al. Pandora's box contained bait: the global problem of introduced earthworms. *Annual Review of Ecology Evolution and Systematics* **39**, 593-613 (2008).
- 5 Blouin M, H. M., Delgado EA, et al. A review of earthworm impact on soil function and ecosystem services. *Eur J Soil Sci* **64**, 161-182 (2013).
- 6 Qiu J.X, T. M. G. Effects of non-native Asian earthworm invasion on temperate forest and prairie soils in the Midwestern US. *Biological Invasions* **19**, 73-88 (2016).
- 7 Pejchar, L., Mooney, H.A. Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution* **24**, 497-504 (2009).
- 8 Viktorov, A. G. Diversity of polyploid races in the family Lumbricidae. *Soil Biology and Biochemistry* **29**, 217-221 (1997).
- 9 Terhivuo, J. S., A. Dispersal and clonal diversity of North-European parthenogenetic earthworms. *Biological Invasions* **8**, 1205-1218 (2006).
- 10 Garbar, A. V., Vlasenko, R.P. Karyotypes of three species of the genus Aporectodea (Oligochaeta: Lumbricidae) from the Ukraine. *Comp. Cyt.* **1**, 59-62 (2007).
- 11 Bakhtadze N.G., B. G. I., Kvavadze E.Sh. The chromosome numbers of Georgian earthworms (Oligochaeta: Lumbricidae). *Comparative Cytogenetics* **2**, 79-83 (2008).
- 12 M. Novo, L. C., A. Maceda-Veiga, et al. Multiple introductions and environmental factors affecting the establishment of invasive species on a volcanic island. *Soil Biology and Biochemistry* **85**, 89-100 (2015).
- 13 Lim, J. Y., Yoon, J. & Hovde, C. J. A brief overview of Escherichia coli O157:H7 and its plasmid O157. *J Microbiol Biotechnol* **20**, 5-14 (2010).
- 14 Simakov, O. *et al.* Insights into bilaterian evolution from three spiralian genomes. *Nature* **493**, 526-531, doi:10.1038/nature11696 (2013).
- 15 Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M. & Bateman, A. TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic Acids Res* **42**, D922-925, doi:10.1093/nar/gkt1055 (2014).
- 16 Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* **34**, D572-580, doi:10.1093/nar/gkj118 (2006).
- 17 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).
- 18 The Gene Ontology, C. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338, doi:10.1093/nar/gky1055 (2019).

- 423 19 Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A. & Kuramae,
424 E. E. The Ecology of Acidobacteria: Moving beyond Genes and Genomes.
425 *Front Microbiol* **7**, 744, doi:10.3389/fmicb.2016.00744 (2016).
- 426 20 Lezcano, M. A., Velazquez, D., Quesada, A. & El-Shehawy, R. Diversity and
427 temporal shifts of the bacterial community associated with a toxic
428 cyanobacterial bloom: An interplay between microcystin producers and
429 degraders. *Water Res* **125**, 52-61, doi:10.1016/j.watres.2017.08.025 (2017).
- 430 21 Schlatter, D. C. *et al.* Mining the Drilosphere: Bacterial Communities and
431 Denitrifier Abundance in a No-Till Wheat Cropping System. *Front Microbiol*
432 **10**, 1339, doi:10.3389/fmicb.2019.01339 (2019).
- 433 22 Bilej, M. *et al.* Distinct carbohydrate recognition domains of an invertebrate
434 defense molecule recognize Gram-negative and Gram-positive bacteria. *J Biol*
435 *Chem* **276**, 45840-45847, doi:10.1074/jbc.M107220200 (2001).
- 436 23 Cho, J. H., Park, C. B., Yoon, Y. G. & Kim, S. C. Lumbricin I, a novel proline-
437 rich antimicrobial peptide from the earthworm: purification, cDNA cloning and
438 molecular characterization. *Biochim Biophys Acta* **1408**, 67-76,
439 doi:10.1016/s0925-4439(98)00058-1 (1998).
- 440 24 Fisker, K. V., Holmstrup, M. & Sorensen, J. G. Variation in metallothionein gene
441 expression is associated with adaptation to copper in the earthworm
442 *Dendrobaena octaedra*. *Comp Biochem Physiol C Toxicol Pharmacol* **157**, 220-
443 226, doi:10.1016/j.cbpc.2012.11.007 (2013).
- 444 25 Joskova, R., Sileroova, M., Prochazkova, P. & Bilej, M. Identification and
445 cloning of an invertebrate-type lysozyme from *Eisenia andrei*. *Dev Comp*
446 *Immunol* **33**, 932-938, doi:10.1016/j.dci.2009.03.002 (2009).
- 447 26 Lassegues, M., Milochau, A., Doignon, F., Du Pasquier, L. & Valembois, P.
448 Sequence and expression of an *Eisenia fetida*-derived cDNA clone that encodes
449 the 40-kDa fetidin antibacterial protein. *Eur J Biochem* **246**, 756-762,
450 doi:10.1111/j.1432-1033.1997.00756.x (1997).
- 451 27 Li, Y., Zhao, C., Lu, X., Ai, X. & Qiu, J. Identification of a cytochrome P450
452 gene in the earthworm *Eisenia fetida* and its mRNA expression under
453 enrofloxacin stress. *Ecotoxicol Environ Saf* **150**, 70-75,
454 doi:10.1016/j.ecoenv.2017.12.020 (2018).
- 455 28 Prochazkova, P. *et al.* Developmental and Immune Role of a Novel Multiple
456 Cysteine Cluster TLR From *Eisenia andrei* Earthworms. *Front Immunol* **10**,
457 1277, doi:10.3389/fimmu.2019.01277 (2019).
- 458 29 Ricketts, H. J., Morgan, A. J., Spurgeon, D. J. & Kille, P. Measurement of
459 annetocin gene expression: a new reproductive biomarker in earthworm
460 ecotoxicology. *Ecotoxicol Environ Saf* **57**, 4-10,
461 doi:10.1016/j.ecoenv.2003.08.008 (2004).
- 462 30 Rorat, A., Vandenbulcke, F., Galuszka, A., Klimek, B. & Plytycz, B. Protective
463 role of metallothionein during regeneration in *Eisenia andrei* exposed to
464 cadmium. *Comp Biochem Physiol C Toxicol Pharmacol* **203**, 39-50,
465 doi:10.1016/j.cbpc.2017.10.002 (2017).
- 466 31 Roubalova, R. *et al.* The effect of dibenzo-p-dioxin- and dibenzofuran-

- contaminated soil on the earthworm *Eisenia andrei*. *Environ Pollut* **193**, 22-28, doi:10.1016/j.envpol.2014.05.026 (2014).
- 32 Silerova, M. *et al.* Characterization, molecular cloning and localization of calreticulin in *Eisenia fetida* earthworms. *Gene* **397**, 169-177, doi:10.1016/j.gene.2007.04.035 (2007).
- 33 Skanta, F., Prochazkova, P., Roubalova, R., Dvorak, J. & Bilej, M. LBP/BPI homologue in *Eisenia andrei* earthworms. *Dev Comp Immunol* **54**, 1-6, doi:10.1016/j.dci.2015.08.008 (2016).
- 34 Skanta, F., Roubalova, R., Dvorak, J., Prochazkova, P. & Bilej, M. Molecular cloning and expression of TLR in the *Eisenia andrei* earthworm. *Dev Comp Immunol* **41**, 694-702, doi:10.1016/j.dci.2013.08.009 (2013).
- 35 Wang, J. *et al.* Transcriptional responses of earthworm (*Eisenia fetida*) exposed to naphthenic acids in soil. *Environ Pollut* **204**, 264-270, doi:10.1016/j.envpol.2015.05.006 (2015).
- 36 Wang, K. Y., Tull, L., Cooper, E., Wang, N. & Liu, D. Recombinant protein production of earthworm lumbricin for potential antithrombotic application. *Evid Based Complement Alternat Med* **2013**, 783971, doi:10.1155/2013/783971 (2013).
- 37 Zheng, S. *et al.* Annetocin and TCTP expressions in the earthworm *Eisenia fetida* exposed to PAHs in artificial soil. *Ecotoxicol Environ Saf* **71**, 566-573, doi:10.1016/j.ecoenv.2007.10.025 (2008).

Acknowledgments

We thank NextOmics (Wuhan, China) for generating the PacBio and Hi-C data. We also thank Yan Sun, Qingxiao Li, and Qiong Yu for comments on the manuscript. We also thank Yanxia Hu, Xichun Zhang, Guozhen Xu, Yanqin Liu, Chong Wang, Xiaoping Diao, Yuhong Gao, Shuaizhang Li, Yanrui Luo, Xuelian Liu, Lan Yao, Feifan Guo, and so on, who are all previous members of Sun's earthworm research laboratory. **Funding:** This work was supported by grants from the National Natural Science Foundation of China (No. 31172091 and No. 31801190) and the Fundamental Research Funds for the Central Universities (No. 2018QC155 and 2018ZH003). **Author Contributions:** X.W., Y.Zhao and Z.J.S designed and supervised research. X.W., Y.Zhang, Y.B.L. collected materials for sequencing and generated transcriptome data, proteomics, 16S rDNA and karyotype data. M.M.K. performed the genome assembly, genome annotation and heterozygosity estimation. Y.Zhao performed the synteny analysis, species tree, transcriptome, proteomics, 16S rDNA analysis. X.W. performed karyotype analysis. X.W., Y.Zhao, Y.Zhang, Y.F. Zhang, M.M.K, and Z.J.S. wrote the paper with assistance

504 provided by co-authors. **Competing interests:** The authors declare no competing
505 interests.
506

a

b