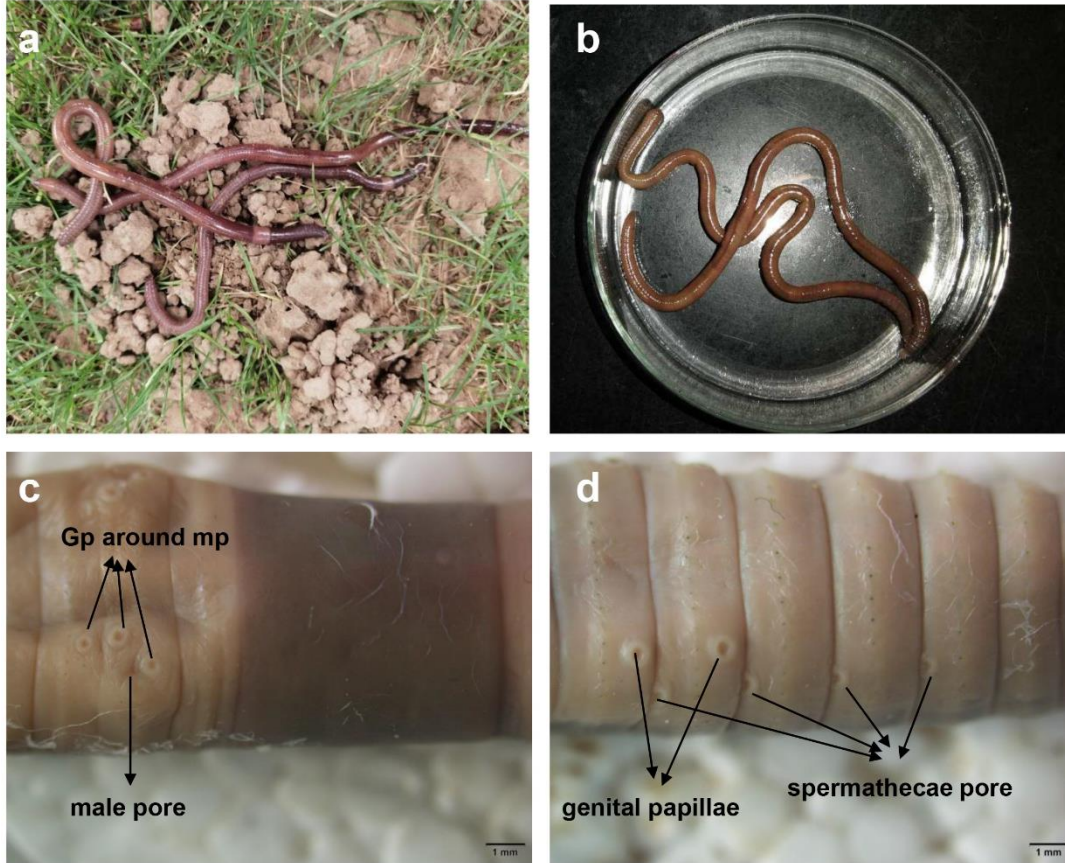


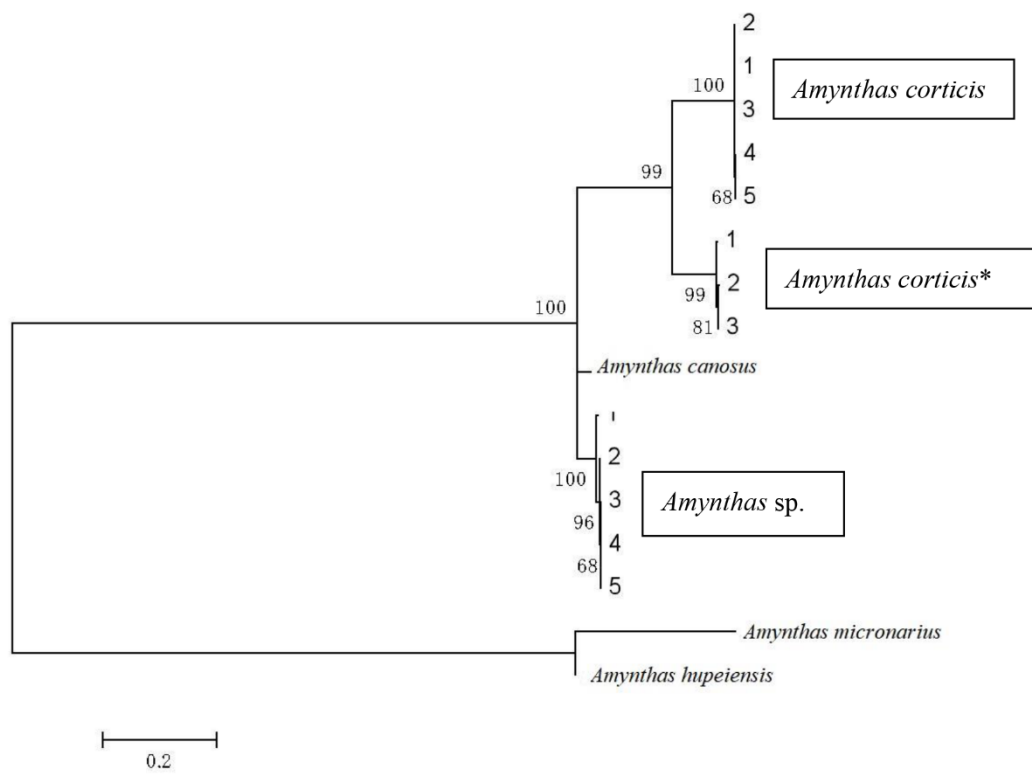
Supplementary figures

**Supplementary Fig. 1: Morphological and anatomical characters of earthworm
Amyntas corticis.**



a, *A. corticis* in the wild. **b**, *A. corticis* in the lab. **c**, Three small and circular genital papillae are present around each male pore of *A. corticis*. **d**, Four pairs of spermathecae locate in venter of intersegments (5/6, 6/7, 7/8 and 8/9) of *A. corticis*.

49 **Supplementary Fig. 2: COI barcode phylogeny for *Amyntas*.**



50 0.2

51 The gene tree was constructed based on COI barcode sequences of *A. corticis* as well

52 as other earthworms in the genus of *Amyntas*. The number labeled at tips of the gene

53 tree meant different earthworm individuals. COI barcode sequences were aligned via

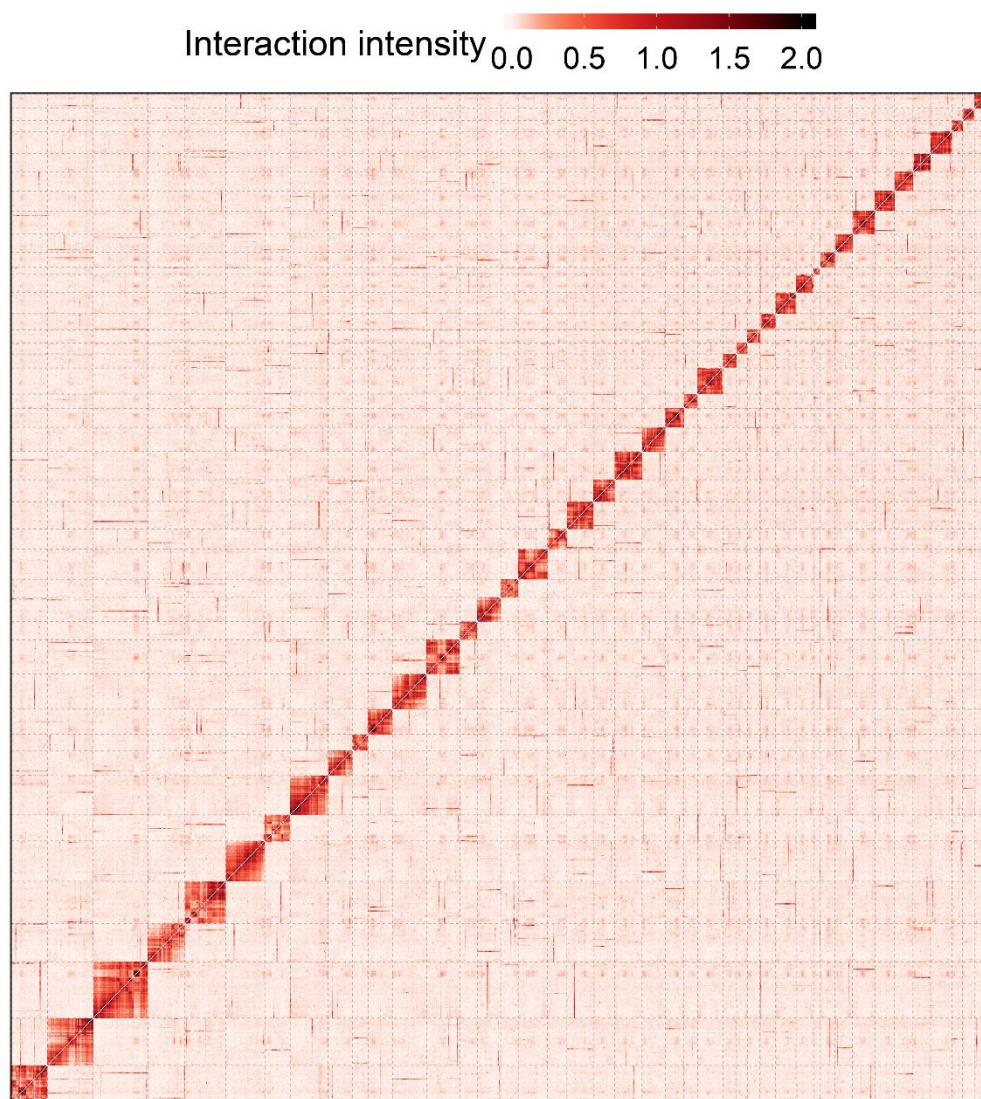
54 ClustalW. The gene tree was constructed via MEGA v.6.0 with maximum likelihood

55 (ML) algorithm. Branch support values were calculated through bootstrap testing with

56 1000 replications. * represents the COI barcode extracted from the *A. corticis* genome

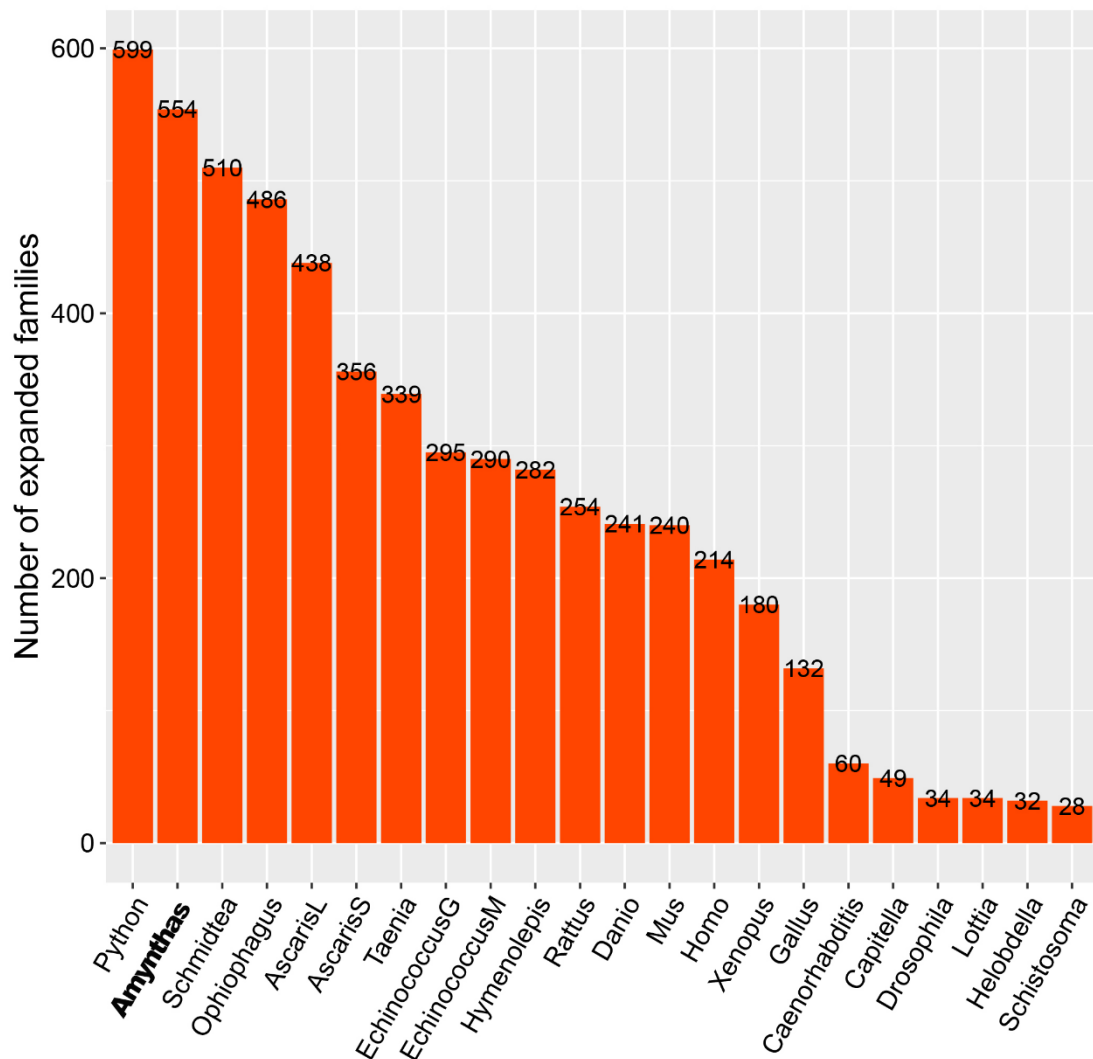
57 discussed in this paper, others represent COI from GenBank.

Supplementary Fig. 3: Interaction intensities of scaffolds.



After assembly of contigs using PacBio SMRT long reads and polishing using paired-end reads sequenced from short-insert libraries, the genome of *Amyntas corticis* was scaffolding using Hi-C sequencing data ([Supplementary Information S5](#)). The heatmap shows interaction intensities calculated based on the mapping of Hi-C sequencing data between all pairs of the longest 42 scaffolds, in which each row and each column represents one scaffold, respectively. The darker the color is, the stronger the interaction intensity is. Interaction intensities within the same scaffold are much stronger than ones between different scaffolds, indicating correct placing of contigs on scaffolds.

80 **Supplementary Fig. 4: The number distribution of expanded gene families.**



81
82 Gene families with significantly accelerated rates of gene duplication and loss indicated
83 by CAFE were selected to do this statistic. For each species, if the count of gene
84 members in a family is larger than that of the MRCA of all species, this family was
85 labeled as expanded for the species ~~(Supplementary Information S15)~~. *Amynthis*
86 *corticis* is represented by earthworm in this figure, and it has the second most expanded
87 families, which is a distinguished feature compared with other two annelids. In this
88 figure, Earthworm: *Amynthis corticis*, Helobdella: *Helobdella robusta*, Capitella:
89 *Capitella teleta*, EchinococcusG: *Echinococcus granulosus*, EchinococcusM:
90 *Echinococcus multilocularis*, Hymenolepis: *Hymenolepis microstoma*, Taenia: *Taenia*
91 *solium*, Schistosoma: *Schistosoma mansoni*, Schmidtea: *Schmidtea mediterranea*,
92 AscarisL: *Ascaris lumbricoides*, AscarisS: *Ascaris suum*, Caenorhabditis:

93 *Caenorhabditis elegans*, Lottia: *Lottia gigantea*, Drosophila: *Drosophila melanogaster*,
94 Danio: *Danio rerio*, Xenopus: *Xenopus tropicalis*, Ophiophagus: *Ophiophagus hannah*,
95 Python: *Python bivittatus*, Gallus: *Gallus gallus*, Mus: *Mus musculus*, Rattus: *Rattus*
96 *norvegicus* and Homo: *Homo sapiens*.

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

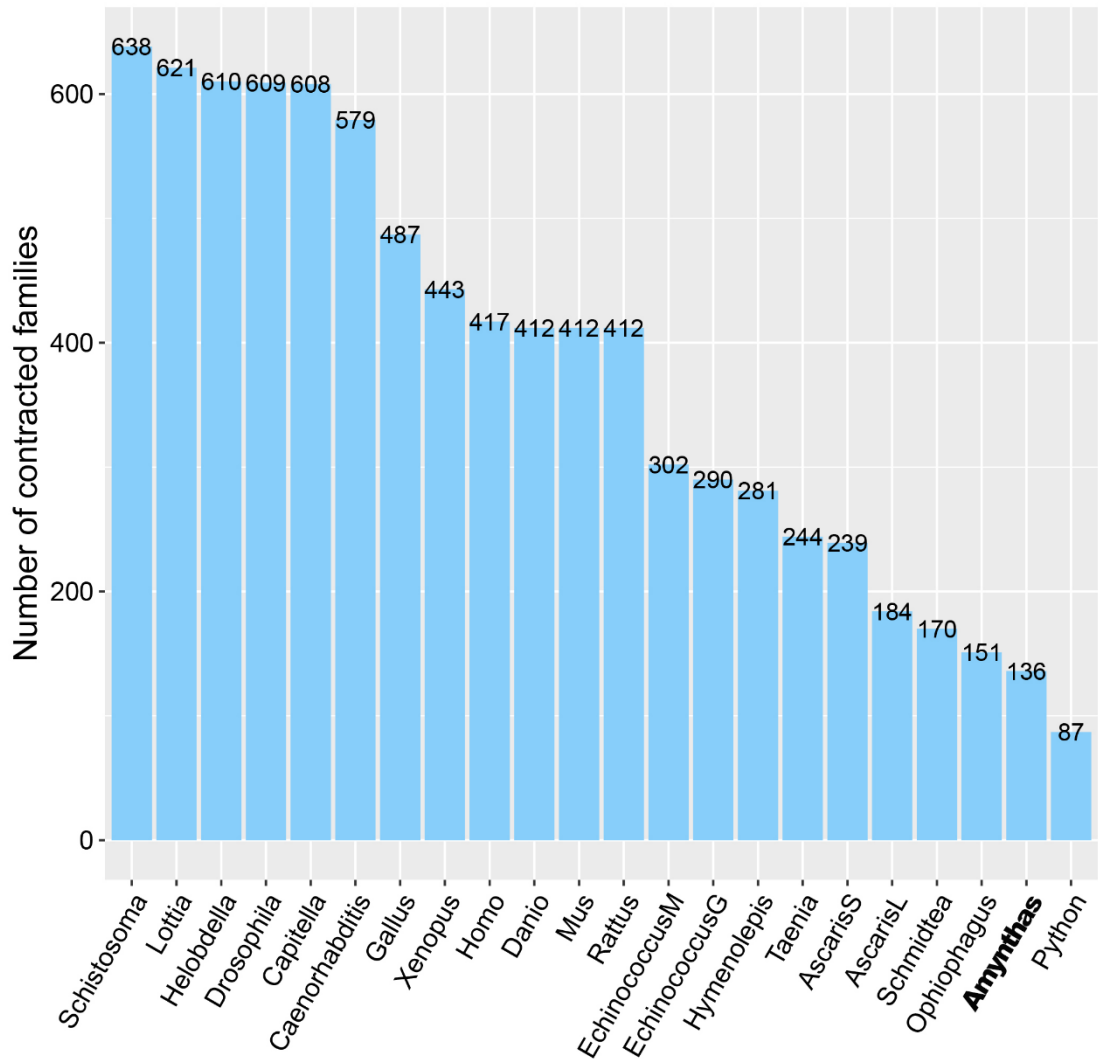
119

120

121

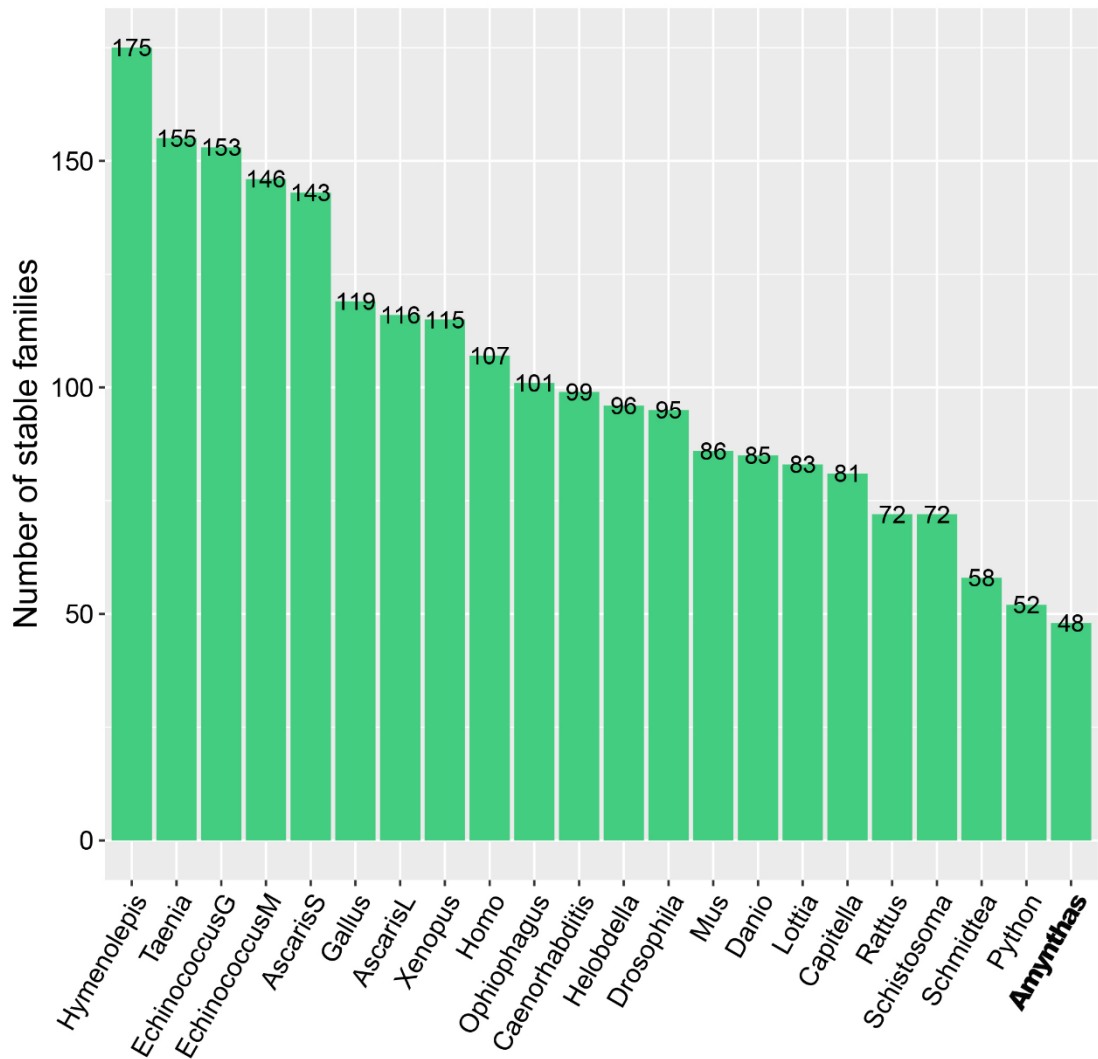
122

Supplementary Fig. 5: The number distribution of contracted gene families.



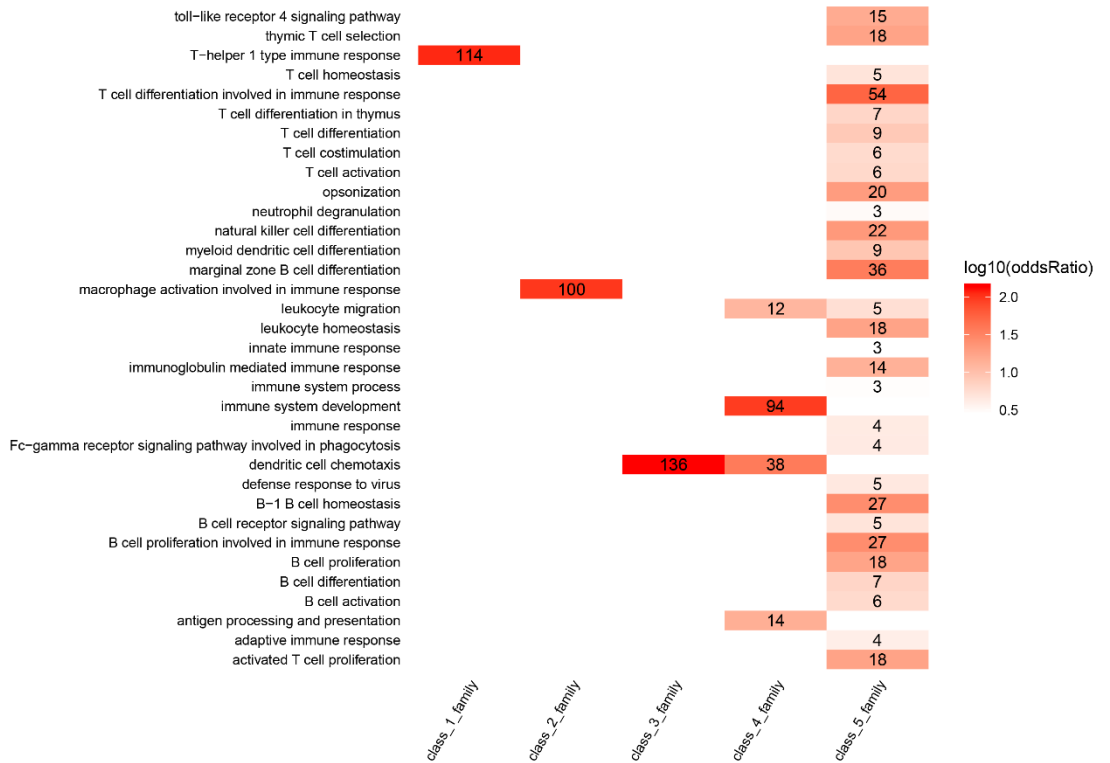
Gene families with significantly accelerated evolutionary rate indicated by CAFE were selected to do this statistics. For each species, if the count of gene members in a family is smaller than that of the MRCA of all species, this family was labeled as contracted for the species. *Amyntas corticis* is represented by earthworm in this figure, and it has the second fewest contracted families, which is a distinguished feature compared with other two annelids. Names of species in this figure are the same as Supplementary Fig. 4.

Supplementary Fig. 6: The number distribution of stable gene families.



Gene families with significantly accelerated evolutionary rate indicated by CAFE were selected to do this statistics. For each species, if the count of gene members in a family is equal to that of the MRCA of all species, this family was labeled as stable for the species. *Amynthis corticis* is represented by earthworm in this figure, and it has the fewest stable families, suggesting evolutionary changes of gene families in *A. corticis* have been dramatic. Names of species in this figure are the same as Supplementary Fig 4.

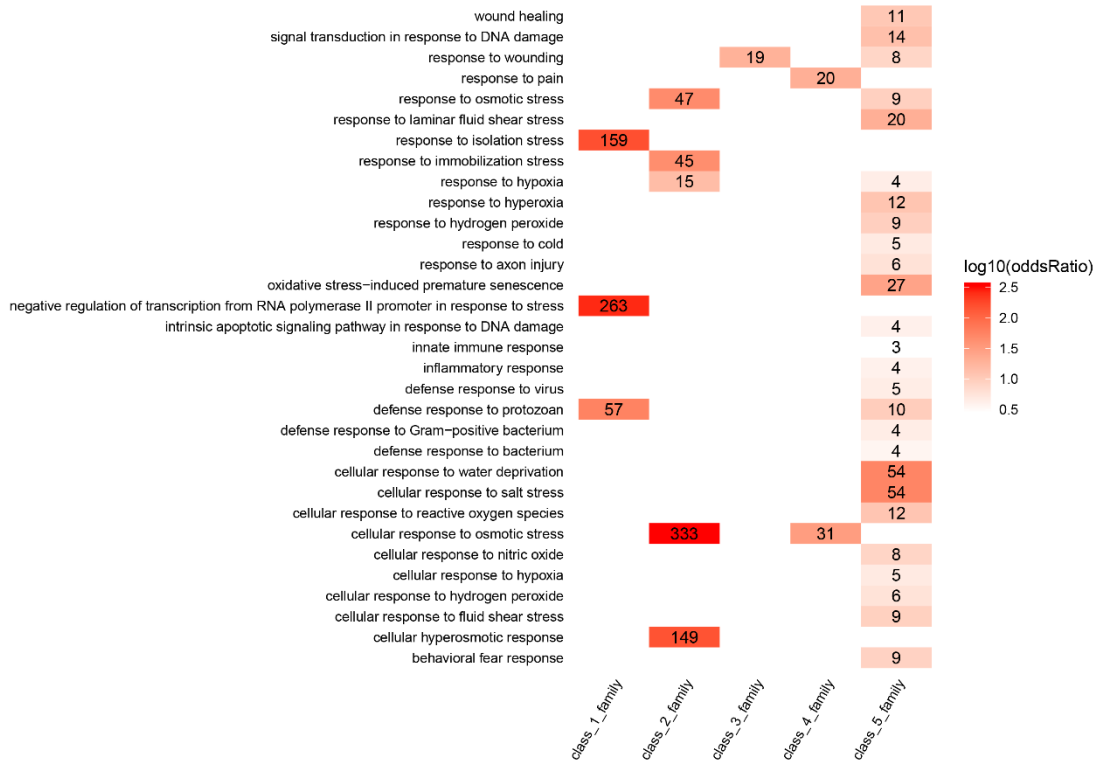
Supplementary Fig. 7: Odds ratio of significantly enriched GO terms attributed to GOSlim of immune system process.



GO enrichment analysis was conducted for five subsets of gene families with positive variable loading in the first component. Odds ratio of significantly enriched GO terms attributed to GOSlim of immune system process was colored and labeled for each subset of gene families, in which the redder the color is, the higher the odds ratio is.

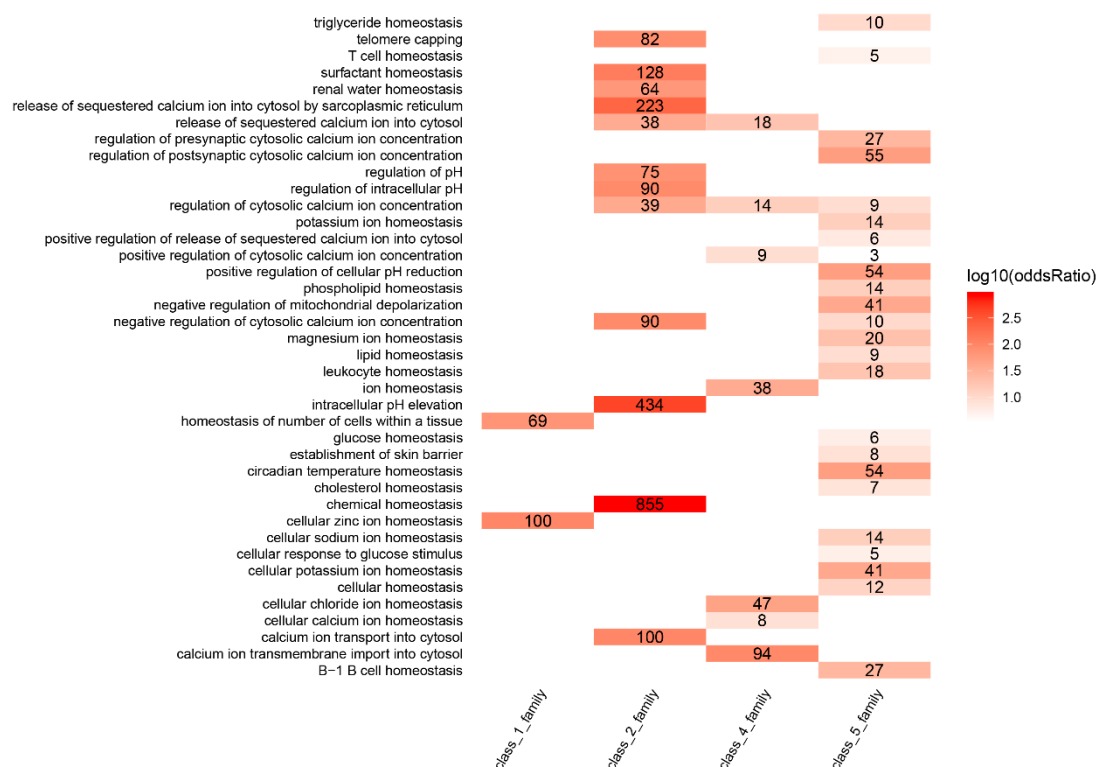
Supplementary Fig. 8: Odds ratio of significantly enriched GO terms attributed

to GOSlim of response to stress.



GO enrichment analysis was conducted for five subsets of gene families with positive variable loading in the first component. Odds ratio of significantly enriched GO terms attributed to GOSlim of response to stress was colored and labeled for each subset of gene families, in which the redder the color is, the higher the odds ratio is.

Supplementary Fig. 9: Odds ratio of significantly enriched GO terms attributed to GOSlim of homeostatic process.



186

187 GO enrichment analysis was conducted for five subsets of gene families with positive
188 variable loading in the first component. Odds ratio of significantly enriched GO terms
189 attributed to GOSlim of homeostatic process was colored and labeled for each subset
190 of gene families, in which the redder the color is, the higher the odds ratio is.

191

192

193

194

195

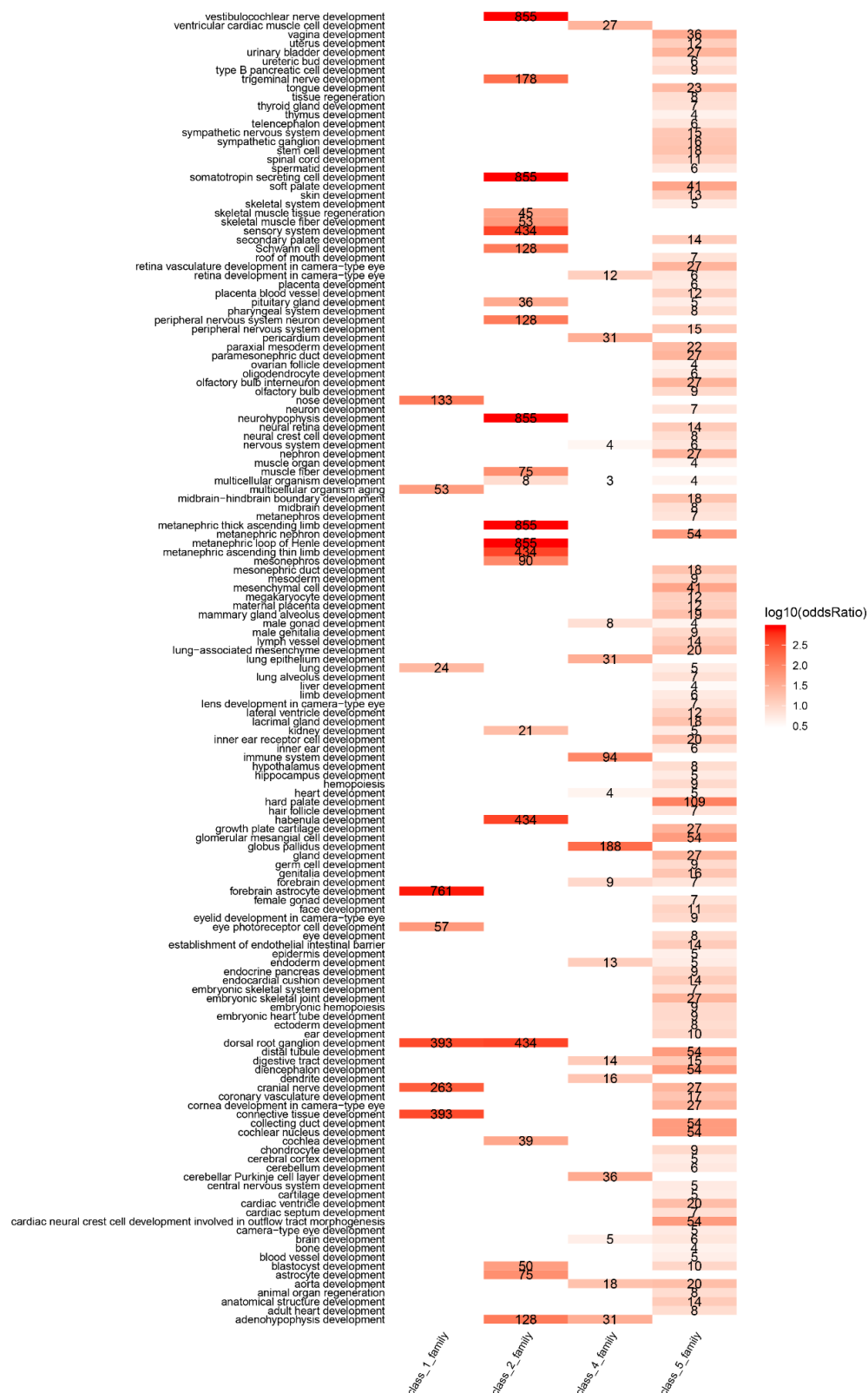
196

197

198

199

200 **Supplementary Fig. 10: Odds ratio of significantly enriched GO terms attributed**
201 **to GOSlim of anatomical structure development.**



202

203 GO enrichment analysis was conducted for five subsets of gene families with positive
 204 variable loading in the first component. Odds ratio of significantly enriched GO terms
 205 attributed to GOSlim of anatomical structure development was colored and labeled for
 206 each subset of gene families, in which the redder the color is, the higher the odds ratio

207 is.

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

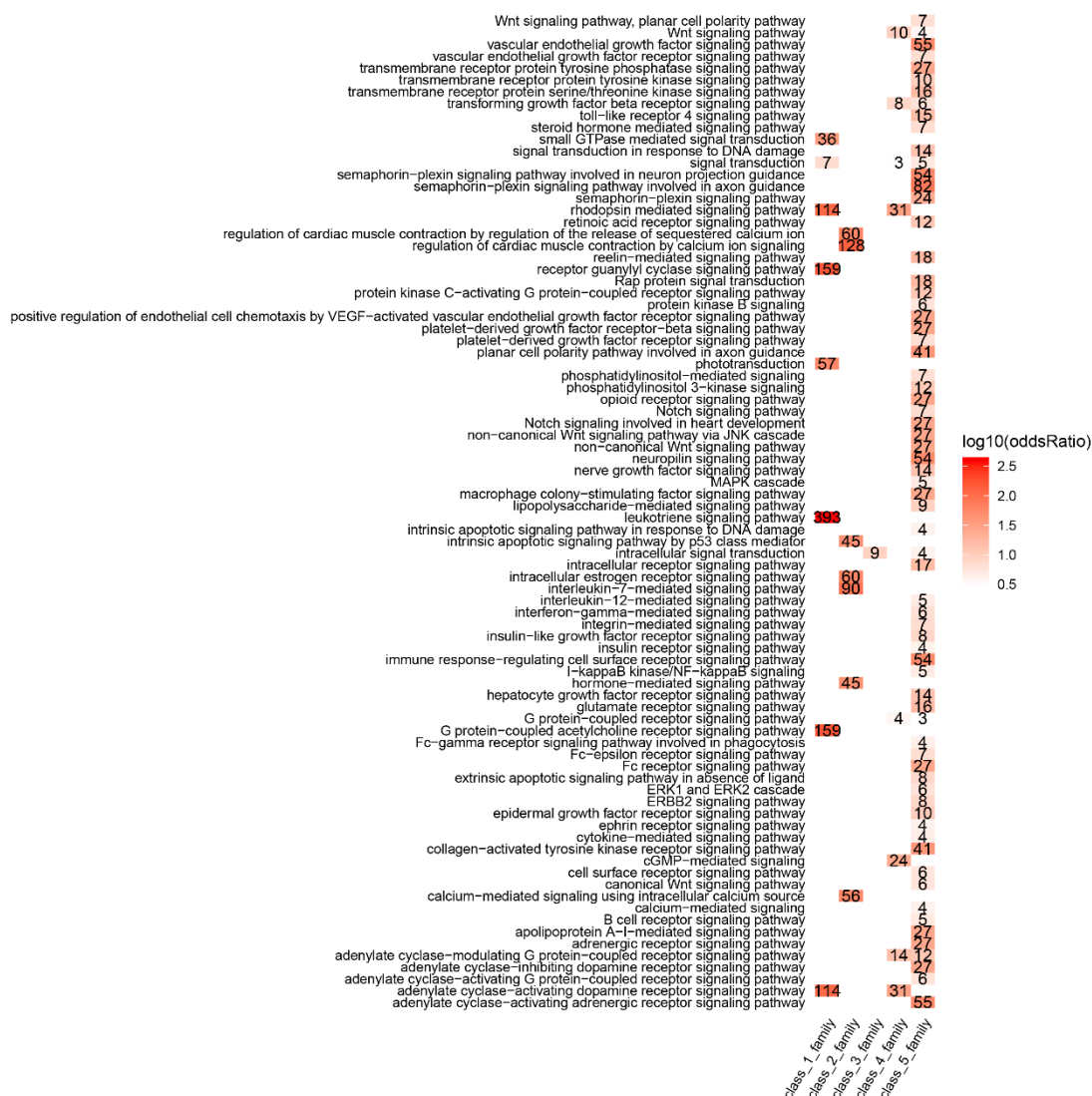
231

232

233

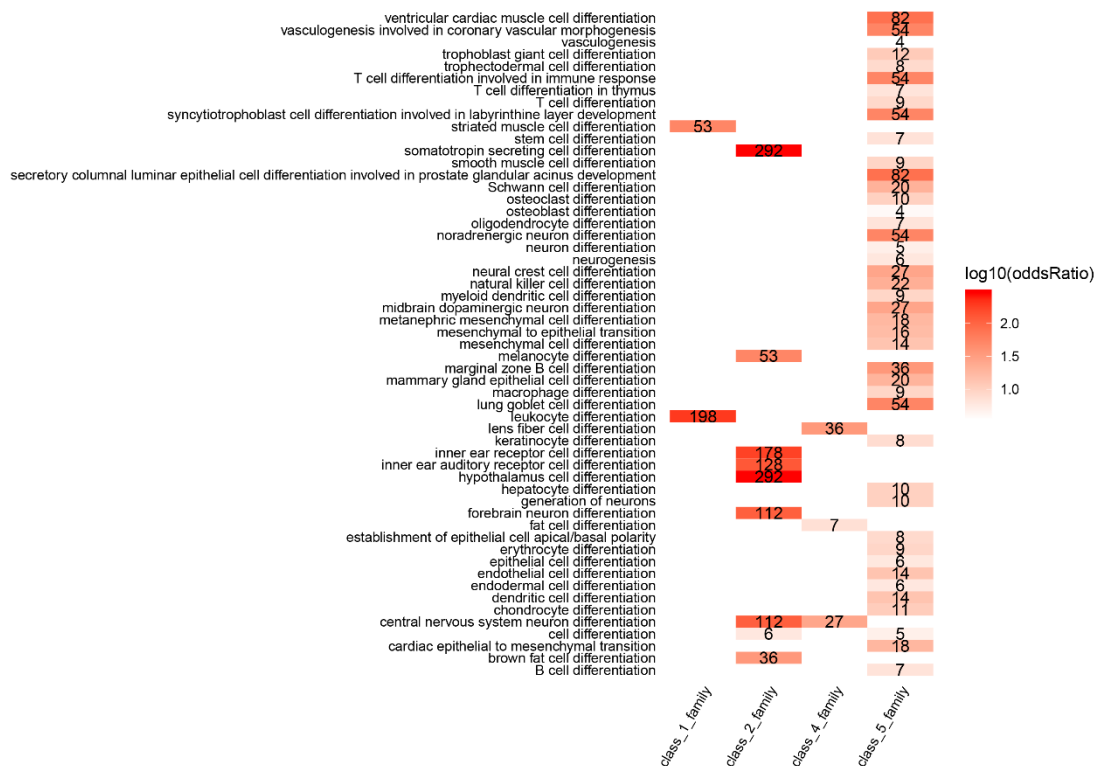
234

235 **Supplementary Fig. 11: Odds ratio of significantly enriched GO terms attributed**
236 **to GOSlim of signal transduction.**



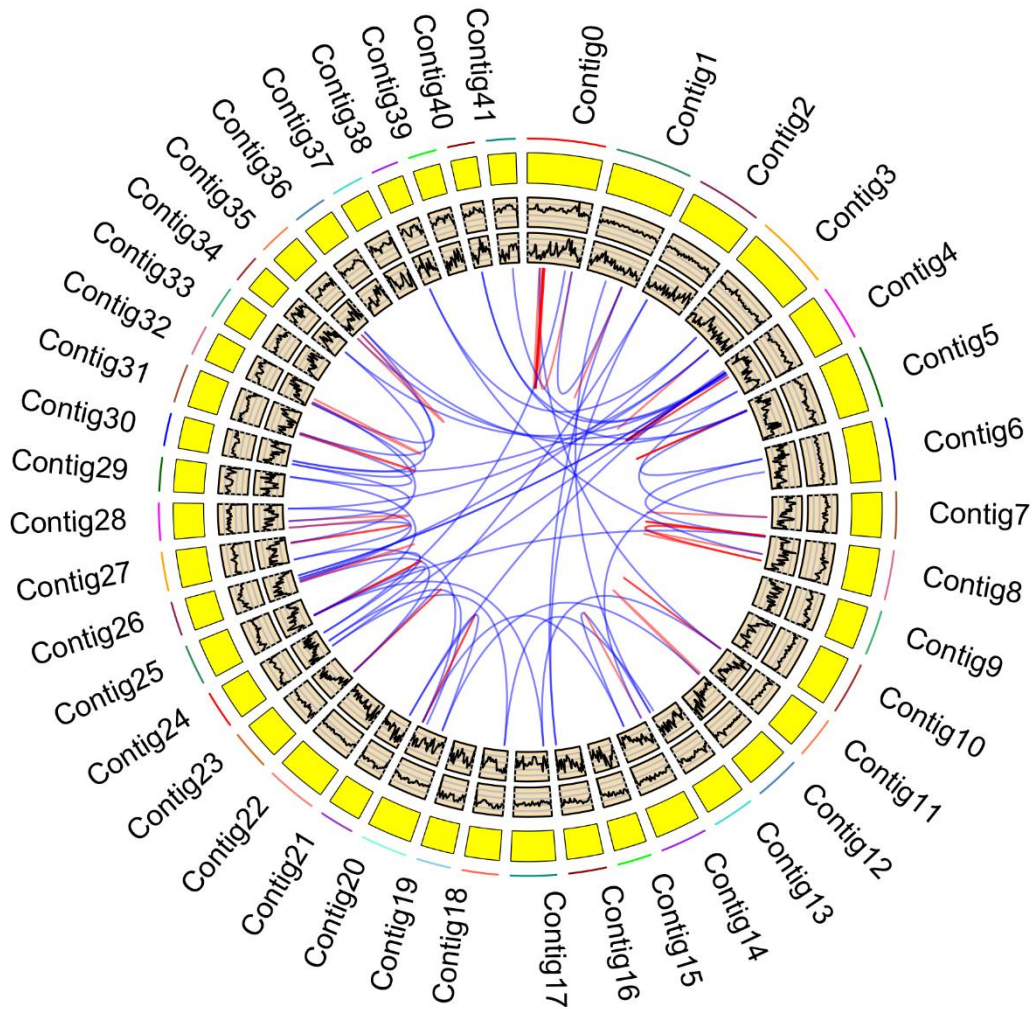
GO enrichment analysis was conducted for five subsets of gene families with positive variable loading in the first component. Odds ratio of significantly enriched GO terms attributed to GOSlim of signal transduction was colored and labeled for each subset of gene families, in which the redder the color is, the higher the odds ratio is.

Supplementary Fig. 12: Odds ratio of significantly enriched GO terms attributed to GOSlim of cell differentiation.



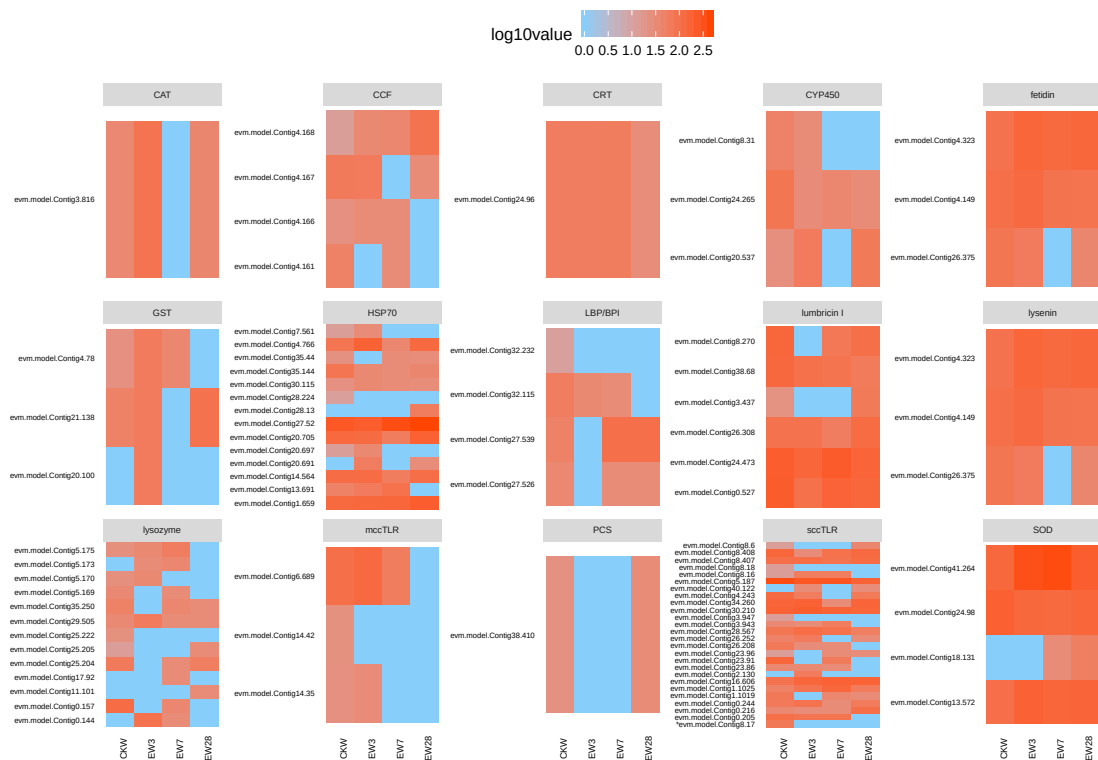
GO enrichment analysis was conducted for five subsets of gene families with positive variable loading in the first component. Odds ratio of significantly enriched GO terms attributed to GOSlim of cell differentiation was colored and labeled for each subset of gene families, in which the redder the color is, the higher the odds ratio is.

Supplementary Fig. 13: Genomic locations of orthologous hits of well determined earthworm defensive genes.



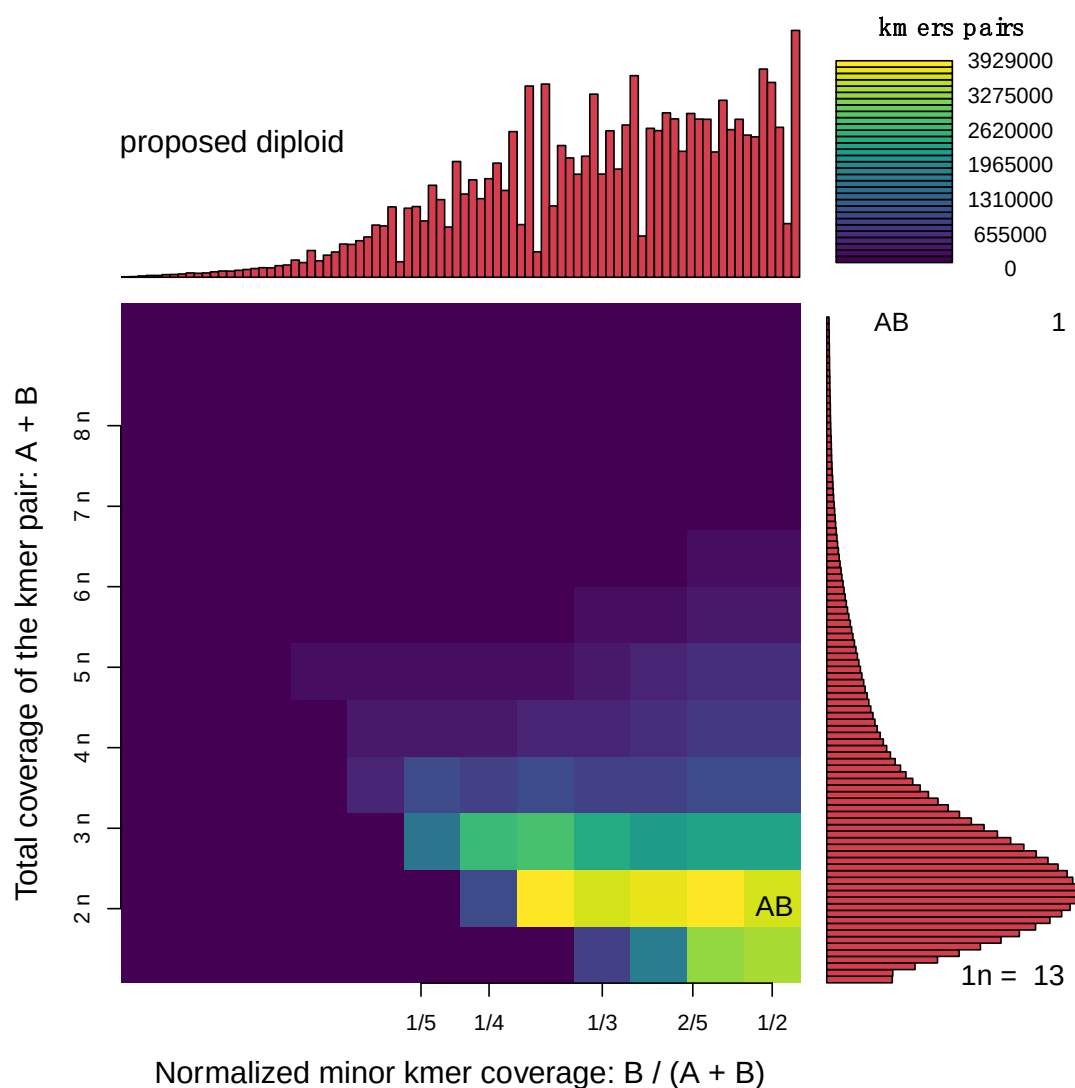
Genomics locations of 111 orthologous hits of 15 well determined earthworm defensive genes were plotted. The outermost ring represents the longest 42 scaffolds. Then, from outside to inside, two rings represent the density distribution of repeat sequences and SNPs. Density distributions of SNPs in genomic regions containing orthologous hits and other regions are significantly different. Lines in the center connect orthologous hits of the same well determined earthworm defensive gene, in which hits on the same chromosome are connected by red lines and hits on different chromosomes are connected by blue lines.

Supplementary Fig. 14: Expression of orthologous hits of well determined earthworm defensive genes.



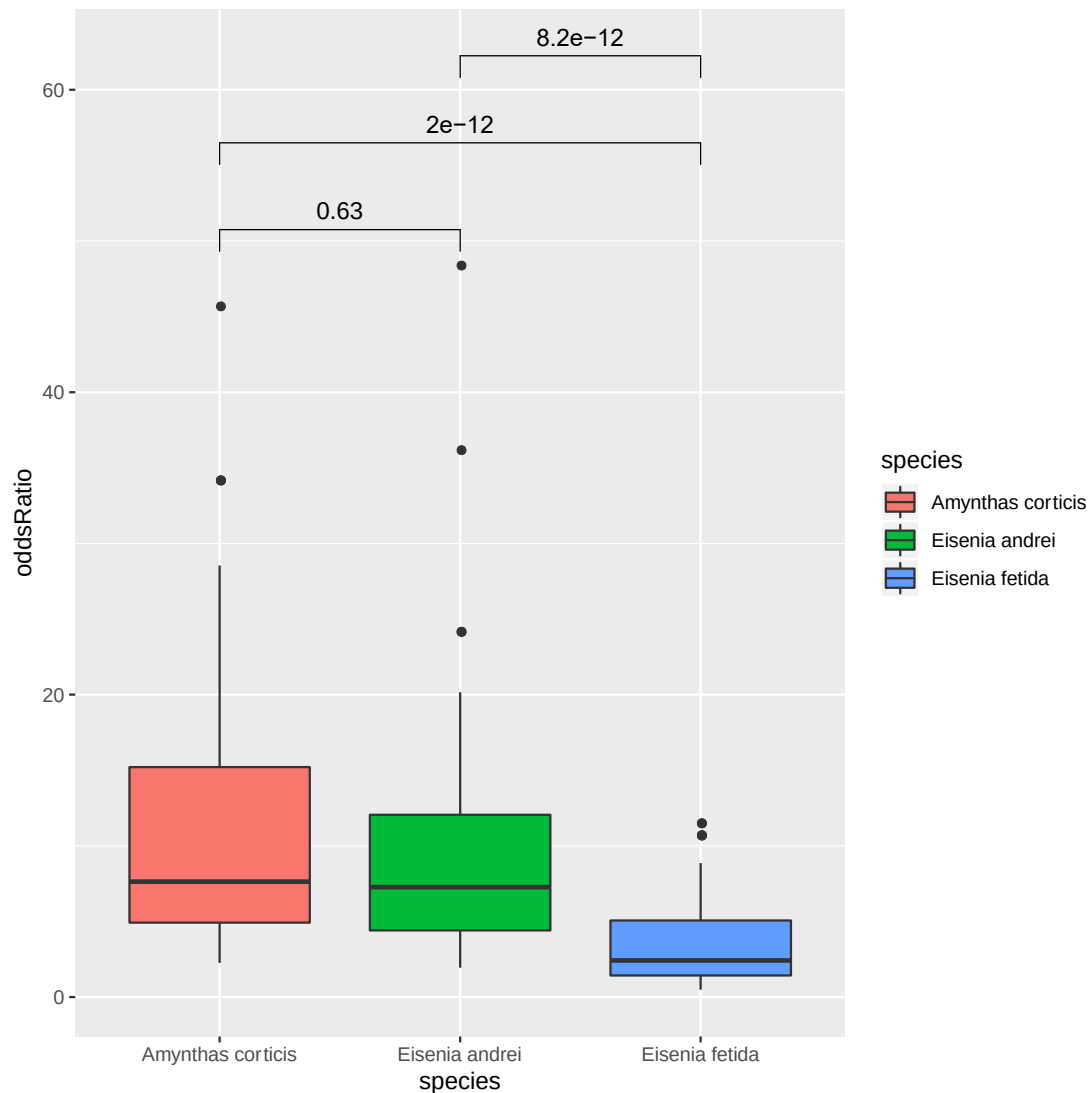
Expression of orthologous hits of well determined earthworm defensive genes were detected by protein mass spectrum of gut removed body along the time course, including before incubation (CKW) in infected soil with pathogenic *E. coli* O157:H7, on the third (EW3), seventh (EW7) and twenty-eighth (EW28) day after incubation. Each panel in this figure represents expression of orthologous hits of each well determined earthworm defensive gene, in which the name of defensive gene is labeled on the top and mean expression levels at different time points are colored from blue (low expression level) to red (high expression level). The name of orthologous hit with significantly varied expression identified in time series analysis was labeled with a star signal.

Supplementary Fig. 15: Proofs supporting that *Eisenia fetida* is diploid.



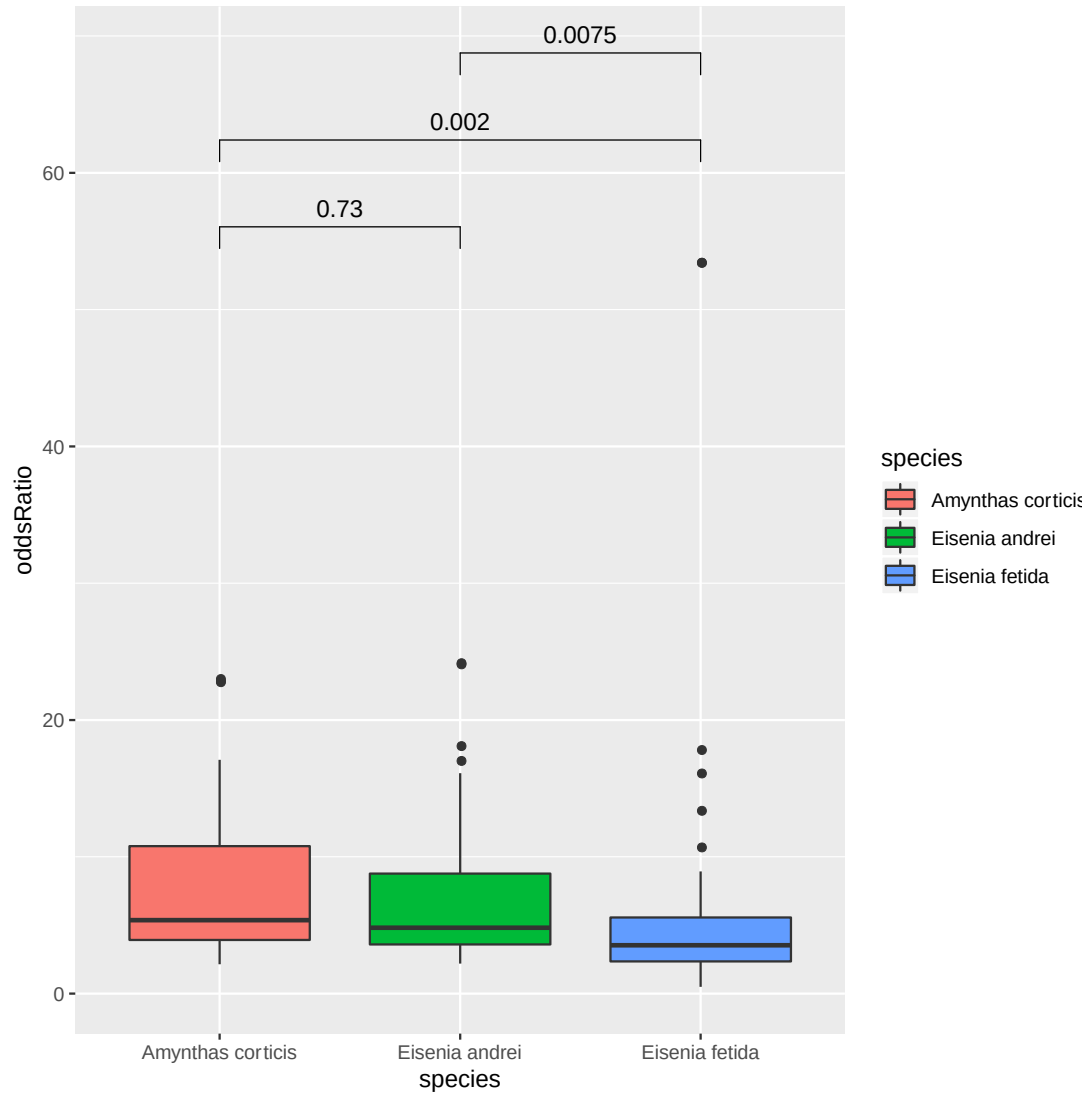
Heatmap for coverage pattern of heterozygous k-mer pairs, in which X axis indicates the normalized minor k-mer coverage, while Y axis indicates the total k-mer pairs coverage. The location on the heatmap colored by bright yellow represents that two are the most k-mer pairs supporting the normalized minor k-mer coverage of 1/2 and the total k-mer pairs coverage of 2n.

Supplementary Fig. 16: Odds ratio distribution of enriched functions related to immune system.



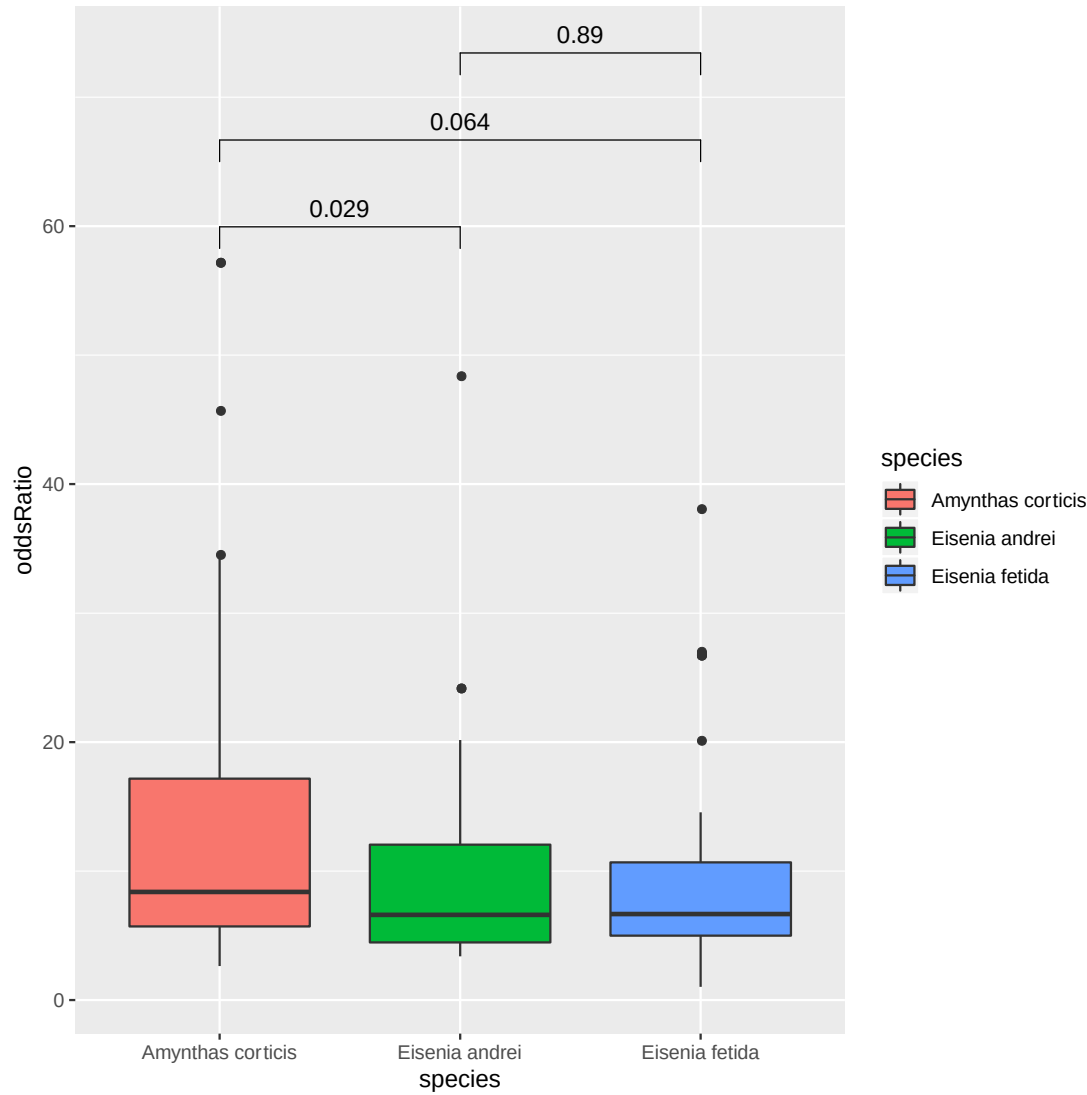
Odds ratio distribution of enriched functions related to immune system process among genomes of *A. corticis*, *Eisenia fetida* and *Eisenia andrei*, in which Wilcoxon rank sum test was adopted to do the comparing analysis.

Supplementary Fig. 17: Odds ratio distribution of enriched functions related to response to stress.



Odds ratio distribution of enriched functions related to response to stress among genomes of *A. corticis*, *Eisenia fetida* and *Eisenia andrei*, in which Wilcoxon rank sum test was adopted to do the comparing analysis.

Supplementary Fig. 18: Odds ratio distribution of enriched functions related to homeostatic process.



Odds ratio distribution of enriched functions related to homeostatic process among genomes of *A. corticis*, *Eisenia fetida* and *Eisenia andrei*, in which Wilcoxon rank sum test was adopted to do the comparing analysis.