Reinvestigating the phylogeny of Myriapoda with more extensive taxon sampling and novel genetic perspective

Jiajia Wang [1] , **Yu Bai** [1] , **Haifeng Zhao** [2] , **Ruinan Mu** [3] , **Yan Dong** [Corresp. 1]

[1] College of Biology and Food Engineering, Chuzhou University, Chuzhou, Anhui, China

[2] Key Laboratory of Space Utilization, Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing, Beijing, China

[3] University of Chinese Academy of Sciences, Beijing, Beijing, China

Corresponding Author: Yan Dong
Email address: dongyan_bio@126.com

Background. There have been extensive debates on the interrelationships among the four major classes of Myriapoda—Chilopoda, Symphyla, Diplopoda, and Pauropoda. The core controversy is the position of Pauropoda; that is, whether it should be grouped with Symphyla or Diplopoda as a sister group. Two recent phylogenomic studies separately investigated transcriptomic data from 14 and 29 Myriapoda species covering all four groups along with outgroups, and proposed two different topologies of phylogenetic relationships. **Methods.** Building on these studies, we extended the taxon sampling by investigating 39 myriapods and integrating the previously available data with three new transcriptomic datasets generated in this study. Our analyses present the phylogenetic relationships among the four major classes of Myriapoda with a more abundant taxon sampling and provide a new perspective to investigate the above-mentioned question, where visual genes' identification were conducted. We compared the appearance pattern of genes, grouping them according to their classes and the visual pathways involved. Positive selection was detected for all identified visual genes between every pair of 39 myriapods, and 14 genes showed positive selection among 27 pairs. **Results.** From the results of phylogenomic analyses, we propose that Symphyla is a sister group of Pauropoda. This stance has also received strong support from tree inference and topology tests.

1 **Reinvestigating the phylogeny of Myriapoda with more extensive taxon**

2 **sampling and novel genetic perspective**

3 Wang Jiajia[1], Bai Yu[1], Zhao Haifeng[2], Mu Ruinan[3], Dong Yan[1*]

4 [1]College of Biology and Food Engineering, Chuzhou University, Chuzhou ,Anhui , China;

5 [2]Key Laboratory of Space Utilization, Technology and Engineering Center for Space Utilization, Chinese

6 Academy of Sciences, Beijing, China;

7 [3]University of Chinese Academy of Sciences, Beijing, China;

9 *Correspondence to: Dong Yan[1], College of Biology and Food Engineering, Chuzhou

10 University, Chuzhou 239000, China. Email: dongyan_bio@126.com.

12 **Abstract**

13 **Background.** There have been extensive debates on the interrelationships among the four major

14 classes of Myriapoda—Chilopoda, Symphyla, Diplopoda, and Pauropoda. The core controversy is

15 the position of Pauropoda; that is, whether it should be grouped with Symphyla or Diplopoda as a

16 sister group. Two recent phylogenomic studies separately investigated transcriptomic data from

17 14 and 29 Myriapoda species covering all four groups along with outgroups, and proposed two

18 different topologies of phylogenetic relationships.

19 **Methods.** Building on these studies, we extended the taxon sampling by investigating 39

20 myriapods and integrating the previously available data with three new transcriptomic datasets

21 generated in this study. Our analyses present the phylogenetic relationships among the four major

22 classes of Myriapoda with a more abundant taxon sampling and provide a new perspective to

23 investigate the above-mentioned question, where visual genes' identification were conducted. We

24 compared the appearance pattern of genes, grouping them according to their classes and the visual

25 pathways involved. Positive selection was detected for all identified visual genes between every

26   pair of 39 myriapods, and 14 genes showed positive selection among 27 pairs.

27   **Results.** From the results of phylogenomic analyses, we propose that Symphyla is a sister group

28   of Pauropoda. This stance has also received strong support from tree inference and topology tests.

29   Key words: Myriapoda, phylogenetic relationships, transcriptomic, positive selection, visual genes

30   Introduction

31   Myriapoda is a diverse group of terrestrial arthropods with more than 16,000 extant

32   species(Moore, 2006) including millipedes and centipedes, which are familiar with our daily life.

33   The presence of numerous legs (range from 6 to 750), which has given the myriapods their name,

34   is obviously a symplesiomorphy(Marek & Bond, 2006). Myriapoda are widely distributed on all

35   continents except Antarctica, and their diversity is concentrated in tropical and temperate regions,

36   where you can find evidence of their habitat in soil, tree barks and trunks, fields and pastures,

37   deserts, caverns, and coastal areas(Santos-Silva et al., 2019). There is an extensive debate on the

38   sister group to monophyletic Myriapoda. Pancrustacea and Chelicerata are the two candidates with

39   most support(Giribet & Edgecombe, 2019). There are four major classes of Myriapoda: Chilopoda

40   (also known as centipedes, CHI), Diplopoda (also known as millipedes, DIP), Pauropoda (PAU),

41   and Symphyla (SYM). Although the described extant species of all four classes are abundant,

42   especially in CHI and DIP, the phylogenomic data are scarce(Szucsich et al., 2020). To date, only

43   two phylogenomic studies have collected genome-wide or transcriptome-wide sequencing data

44   covering all four classes for phylogeny investigation. Results from both studies supported

45   monophyletic Myriapoda and the monophyly of each major class (DIP, CHI, PAU, and

46   SYM)(Fernández, Edgecombe & Giribet, 2018; Szucsich et al., 2020),(Bäcker, Fanenbruck &

47   Wägele, 2008). However, the interrelationships among the four classes are still controversial.

48   Previous molecular analyses proposed the PAU+SYM grouping (named Edafopoda), which

49   strongly contradicted the sister-group relationship DIP+PAU (named Dignatha). And the

50   hypothesis Edafopoda was supported by morphology and development(Regier et al., 2010; Regier

51   & Zwick, 2011; Dong et al., 2012; Zwick, Regier & Zwickl, 2012; Miyazawa et al., 2014; Szucsich

52 et al., 2020).

53   With the emergence of Next-Generation Sequencing (NGS) technology, some clarity has

54 been gained in recent years. Fernández et al. sequenced 12 myriapods, which greatly enriched the

55 available data for phylogenomic analyses(Fernández, Edgecombe & Giribet, 2018). Their results

56 strongly support Dignatha topology with grouping PAU+DIP. A strong dependence on the choice

57 of outgroups was emphasised in their study. In 2020, Szucsich et al. generated 22 Myriapod RNA-

58 Sequencing data by analysing 59 species(Szucsich et al., 2020). In addition to tree inference and

59 outgroup selection impact testing, they conducted two topology tests: approximate unbiased (AU)

60 tests and four-cluster likelihood-mapping (FcLM). Their results were consistent with Edafopoda

61 topology, thereby grouping PAU+SYM(Szucsich et al., 2020). It is worth noting that both studies,

62 although suggesting diverging topologies of the interrelationships among Myriapoda, placed

63 Myriapoda as a sister group to Pancrustacea.

64   These two seminal works on the interrelationships of Myriapoda have laid a good foundation

65 in phylogenomic data for further study. In our study, 60 species were investigated, including 39

66 Myriapoda members and 21 outgroups. We integrated data from the two aforementioned studies

67 with three newly sequenced transcriptome data (one chilopod: *Scolopendra* sp.; and two diplopods:

68 *Epanerchodus* sp., *Skleroprotopus* sp.). We compiled two concatenated supermatrices covering all

69 four major classes of Myriapoda and three clades of outgroups, one including 20 gene partitions

70 and the other, 369. We performed phylogenetic tree inference using Maximum-Likelihood

71 method. The resulting trees had the same topology as Edafopoda (PAU+SYM) and a sister group

72 of DIP+CHI, which was consistent with previous research results proposed by Szucsich et

73 al(Szucsich et al., 2020). As shown in the previous study proposed by Szucsich et al, Pancrustacea

74 is the closest relative to Myriapoda. Furthermore, topology tests including an AU test, weighted

75 Kishino-Hasegawa (KH) test, and weighted Shimodaira-Hasegawa (SH) test were conducted on

76 six topology hypotheses derived from the two most controversial phylogenetic relationships

77 (Edafopoda and Dignatha). The results showed that almost all hypotheses derived from Dignatha

78 were rejected with high probability. The topologies of PAU+SYM and DIP+CHI, which were

79 determined from our best Maximum-Likelihood (ML) tree, survived all the tests. We attempted to

80 find additional evidence to support PAU+SYM, and found that almost all species of PAU and

81 SYM were small-sized, blind and soil-dwelling, which may have a significant impact on visual

82 capabilities. For the evolution of vision-related genes, we performed Light Interaction Toolkit

83 (LIT)gene identification on each of the 39 Myriapoda species and conducted positive selection

84 analyses on the identified LIT genes. The distribution of LIT gene identification shared a very

85 similar pattern among the four major classes, however, positive selection evidence was narrowed

86 in CHI&DIP, CHI&PAU, CHI&SYM, DIP&PAU, and DIP&SYM.

87

Material and methods

Taxon sampling

90 Building upon previous works by Fernández et al. and Szucsich et al.(Fernández, Edgecombe &

91 Giribet, 2018; Szucsich et al., 2020), where more than 36 species representing the four major

92 groups of myriapods were included in taxon sampling, we sequenced three additional species (one

93 chilopod: *Scolopendra* sp.; and two diplopods: *Epanerchodus* sp., *Skleroprotopus* sp.) in this

94 study. Our sampling was designed to maximise the representation of myriapod groups. Information

95 on sampling localities and accession numbers in the Sequence Read Archive (SRA) database for

96 each transcriptome is shown in Table 1, including four genomes from http://metazoa.ensembl.org.

97 Twenty-one outgroups were also included: eight chelicerates (*Liphistius malayanus*, *Centruroides*

98 *vittatus*, *Damon diadema*, *Archegozetes longisetosus*, *Araneus diadematus*, *Egaenus convexus*,

99 *Euscorpius sicanus*, and *Nymphon gracile*), two onychophorans *(Peripatopsis capensis* and

100 *Peripatoides novaezealandiae*), and 11 pancrustaceans (*Daphnia pulex*, *Folsomia candida*,

101 *Drosophila melanogaster*, *Eubranchipus grubii*, *Triops cancriformis*, *Nebalia bipes*, *Anaspides*

102 *tasmaniae*, *Hemidiaptomus amblyodon*, *Tisbe furcata*, *Vargula hilgendorfii*, and *Xibalbanus*

103 *tulumensis*).

104

105 **RNA extraction and sequencing**

106 Following the manufacturer's instructions, total RNA was extracted using a commercial RNA

107 extraction kit (TAKARA). Samples were treated with Ambion turbo DNA-free DNase to remove

108 residual genomic and rRNA contaminants during mRNA purification. The quantity and quality

109 (purity and integrity) of mRNA were assessed using a NanoDrop ND-2000 UV spectrophotometer

110 (Thermo Fisher Scientific).

111 For mRNA sequencing library preparation, mRNA was first enriched and purified with oligo (dT)-

112 rich magnetic beads and then broken into short fragments, followed by paired-end sequencing on

113 an Illumina Hiseq 4000 platform.

114

115 **Data processing and *de novo* assembly**

116   Sequencing adaptors and low-quality sequences were trimmed using Trimmomatic v 0.36

117 (Bolger, Lohse & Usadel, 2014) with default parameters. The clean data were assembled with

118 Trinity (release 2.11.0) with 100 GB memory and a path reinforcement distance of 50 (Grabherr

119 et al., 2011). The redundancy of all assembled transcripts was removed using CD-HIT v. 4.8.1

120 (under the cd-hit-est mode, with default parameters(Li & Godzik, 2006)). TransDecoder v5.5.0

121 (https://github.com/TransDecoder/TransDecoder) was then utilised for nucleotide sequence

122 translation and longest ORF selection.

123

124 **Orthology assignment and phylogenetic matrix construction**

125   Both classical pipeline (OrthoFinder) and single-copy gene selection method (Benchmarking

126 Universal Single-Copy Orthologs; BUSCO) were used in orthology assignment among the 60

127 selected taxa. Orthogroups were first identified using OrthoFinder v 2.2.7(Emms & Kelly, 2015)

128 with default settings (BLASTP E value $\leq$ 1e−5 and MCL inflation parameter of 1.5). Single-copy

129 genes in arthropods were identified in our datasets with BUSCO v4.1.4 with default settings (E

130 value $\leq$ 1e−6) based on hidden Markov model profiles, where BUSCO dataset arthropoda_odb10

131   (https://busco-data.ezlab.org/v4/data/lineages) was used as reference(Seppey, Manni & Zdobnov,

132   2019). For each BUSCO and each taxon, the longest hit of duplicated BUSCO homologous genes

133   was retained for further analysis.

134   Putative orthogroup filtering was based on the gene occupancy threshold, which means that

135   an orthogroup (or BUSCO) was selected if it could be found in more than or equal to the threshold

136   number of taxa. For example, a 50% gene occupancy threshold would select orthogroups that were

137   present in $\geq$ 50% of the included taxa. We selected two thresholds of 100% and 90% gene

138   occupancy to obtain information on most species and to minimise the computational burden.

139   Protein sequences of each orthogroup were aligned with MAFFT v 7.305b (maxiterate was set to

140   1000 in globalpair mode) prior to concatenation(Katoh & Standley, 2013). Then, Aliscore v2.0

141   was used to perform each orthogroup's multiple sequence alignment (MSA) for ambiguous or

142   randomly aligned sections' identification, followed by Alicut v2.2 for error section's

143   trimming(Kück, 2009). After filtering with 100% and 90% gene occupancy thresholds, two raw

144   matrices were constructed using custom Python scripts. Finally, the matrices were trimmed using

145   MARE (v 0.1.2) to select optimised data subsets from the supermatrices for phylogenetic

146   inference(Meyer, Meusemann & Misof, 2011).

147

148   **Best partition schemes finding**

149   The best-fit partitioning schemes and models of evolution for phylogenetic analyses were

150   searched and estimated with PartitionFinderV2.0.0(Lanfear et al., 2017), where amino acid

151   substitution models were restricted to LG, WAG, JTT, and BLOSUM62. The corrected Akaike

152   Information Criterion (AICc) model was selected and was set under greedy search.

153

154   **Phylogenetic tree inference**

155   **Maximum-Likelihood tree**

156   Tree searches were performed for the two supermatrices with a ML approach, using IQ-TREE

157   (v1.6.12) with the above best partitioning schemes. Statistical support was derived from 100 non-

158 parametric slow bootstrap replicates. Replicates to perform SH-like approximate likelihood ratio

159 test were set to 1000 and unsuccessful iterations to stop were set to 300. The initial tree searches

160 were set from a completely random tree(Nguyen et al., 2015). The full command was: 'iqtree -s

161 MSAmatrix.phy -alrt 1000 -b 100 -t RANDOM -spp partition_ file. nex-nstop 300.'

162

163 **Bootstrap support inference**

164 Bootstrapping analyses were applied with RAxML-NG v1.0.0(Kozlov et al., 2019), and the

165 autoMRE bootstrap convergence test was set for a sufficient number of replicates. The bootstrap

166 support was then mapped onto the phylogenetic trees using RAxML v8.2.11(Stamatakis, 2014).

167

168 **Topology testing**

169   To evaluate support for the different hypotheses concerning the relationship proposed by

170 previous studies among the four major classes within Myriapoda, topology tests were run for each

171 dataset using the corresponding best partition scheme with IQ-TREE (v1.6.12), where the AU-test,

172 weighted KH-test, and weighted SH-test were included, and all tests performed 100,000

173 resamplings using the resampling of estimated log-likelihoods (RELL) method(Nguyen et al.,

174 2015). Six proposed hypotheses representing the two most controversial phylogenetic

175 relationships (Edafopoda and Dignatha) of four major classes in Myripoda were compared in our

176 topology tests; detailed information is shown in Figure 1.

177 Hypothesis Eda.1(topology: ((CHI,DIP), (SYM, PAU));).

178 Hypothesis Eda.2(topology: (CHI, (DIP, (SYM, PAU)));).

179 Hypothesis Eda.3(topology: (DIP, (CHI, (SYM, PAU)));).

180 Hypothesis Dig.1(topology: ((CHI, SYM), (DIP, PAU));).

181 Hypothesis Dig.2(topology: (CHI, (SYM, (DIP, PAU)));).

182 Hypothesis Dig.3(topology: (SYM, (CHI, (DIP, PAU)));).

183

184 **Identification of LIT genes**

185 A modified version of the phylogenetically informed annotation tool pipeline named PIA2

186 (https://github.com/xibalbanus/PIA2) was applied for the identification of visual opsins in this

187 study, and the parameters were set default(Pérez-Moreno et al., 2018). In this pipeline, 111 genes

188 from the LIT, a collection of genes that underlie the function or development of light-interacting

189 structures in metazoans, representing 13 different parts in visual pathways (photoreceptor

190 specification, retinal determination network, phototransduction, rhabdomeric, phototransduction,

191 ciliary, retinoid pathway, vertebrate, retinoid pathway, invertebrate, melanin synthesis, pterin

192 synthesis, ommochrome synthesis, heme synthesis, crystallins, diurnal clock, and opsin), was

193 taken as reference (Speiser et al., 2014). We applied the pipeline to protein sequences of each

194 species.

195

196 **Identification of positively selected genes**

197 Evidence of positive selection was indicated by estimating the ratios of nonsynonymous

198 substitutions (Ka or dN) and synonymous substitutions (Ks or dS), also called substitution rates

199 (Ka/Ks or dN/dS value). The coding sequence of each identified LIT genes was aligned between

200 a pair of taxa separately with MAFFT v 7.305b with default settings. And then the substitution rate

201 was calculated using the KaKs_calculator with the following settings, method of calculation:

202 GMYN, genetic code table: The Echinoderm and Flatworm Mitochondrial Code(Wang et al.,

203 2010).

204

205 **Results**

206 **Transcriptome assembly and phylogenomic dataset construction**

207 NGS technologies have empowered phylogenomic analyses in the last few decades. It has

208 dramatically increased the size of datasets applied to phylogenetic questions. Within the

209 framework of combining NGS technologies and phylogenomic techniques, we decided to re-

210 investigate Myriapoda phylogeny with three newly sequenced species. Combined with published

211 data from two outstanding studies on Myriapoda phylogeny, the data from a total of 60 species (39

212 from Myriapoda, 8 Chelicerata, 11 Crustacea, and 2 Onychophora) were used in this study (Table

213 1)(Fernández, Edgecombe & Giribet, 2018; Szucsich et al., 2020). Except for the four species with

214 published genome data, raw reads of the remaining 56 species were trimmed and assembled *de*

215 *novo*. Orthology assignments of the 60 species were mainly based on BUSCO results, and more

216 than 70% of the ortholog gene set (BUSCO dataset: arthropoda_odb10, comprising 1013 single-

217 copy protein-coding genes or ortholog group, OG) were identified in 56 species (details in

218 Supplementary Table 1). Additionally, 786 of the BUSCO orthology assignments were confirmed

219 using OrthoFinder (details in Supplementary Table 1).

220 Concatenated supermatrices were compiled using a threshold of percentage gene occupancy of

221 100% and 90% (Figure 2)(González et al., 2015). We found that 32 OGs were represented in all

222 60 species (100% gene occupancy), of which 20 were confirmed by OrthoFinder. There were 505

223 OGs represented in more than 54 species (90% gene occupancy), and 369 OG assignments were

224 confirmed by OrthoFinder. Thus, two datasets comprising 20 and 369 OGs were obtained (

225 Supplementary Table 1). After MSA, identification, and removal of ambiguously aligned sections

226 in each dataset, two phylogenomic supermatrices on the amino acid levels were constructed, which

227 are hereafter referred to as OCC100 and OCC90. Matrix OCC100 included 20 gene partitions and

228 spanned 8,401 aligned sites with 0.395 overall information content. Matrix OCC90 included 369

229 gene partitions and spanned 129,085 aligned sites with 0.318 overall information content.

230

231 **Phylogenetic tree inference and topology analysis**

232   We constructed Maximum-Likelihood (ML) trees based on the best partition schemes and

233 best-fitting substitution models schemes with matrices OCC100 and OCC90. Three main results

234 were found from the inferred trees. All the analyses recovered Myriapoda as the monophyletic

235 sister group of Pancrustacea with high support (Figure 3). As for the relationships among the four

236 myriapod classes: Symphyla (SYM), Chilopoda (CHI), Diplopoda (DIP), and Pauropoda (PAU),

237 we found a sister group relationship of CHI+DIP, and another sister group relationship of

238 PAU+SYM (Figure 3). Both were highly supported by the bootstrap result (PAU+SYM: 100%,

239 DIP+CHI: 100%). The three newly sequenced species (*Scolopendra* sp., *Epanerchodus* sp., and

240 *Skleroprotopus* sp.) were positioned in the expected clades.

241   A variety of groupings of the Myriapoda classes have been proposed, where two hypotheses,

242 Edafopoda and Dignatha, received the most attention (Figure 1). Edafopoda is a grouping of

243 PAU+SYM which has been supported by shared genetic sequences (Figure 1). However, in

244 Dignatha, the PAUs were positioned with the DIPs. In this study, all trees inferred were congruent

245 with the unrooted quartet topology with CHI+DIP and PAU+SYM (Hypothesis Eda.1, Figure 1).

246 We conducted three types of topology tests—AU test, KH test, and weighted SH test—on the

247 quartet topology of Edafopoda and Dignatha, where four different phylogenomic datasets were

248 applied. The results consistently supported the topology Hypothesis Eda.1, which is the only

249 topology to not be rejected in any test (Table 2). Almost all hypotheses derived from Dignatha

250 were rejected with high significance, especially in the phylogenomic matrix OCC90 (Table 2).

251 When comparing the results from the two phylogenomic matrices, we found that all testing results

252 of matrix OCC100 were consistent with and covered by that of matrix OCC90. Under matrix

253 OCC90, the datasets that were different in outgroup selection (PAN or CHE) exclusively showed

254 divergence when rejecting Hypothesis Eda.2, where all three topology tests on the datasets with

255 CHE were not rejected, which was completely opposite to the results of datasets with PAN (Table

256 2). In other words, we found that the sister group of Edafopoda (PAU+SYM) received less support

257 from CHI than the clade of DIP+CHI, but it cannot be completely denied.

258

259 **Outgroup dependence of myriapod phylogeny inference**

260   Despite the quartet topology of CHI+DIP and PAU+SYM being recovered in our analyses,

261 the relationships among the four major classes in Myriapoda varied across phylogenomic datasets,

262 with dependence on outgroup selection proposed in previous studies(Fernández, Edgecombe &

263 Giribet, 2018; Szucsich et al., 2020). Given that the phylogeny inference was sensitive to outgroup

264 choice, we conducted topology tests on datasets with different clades of outgroups: one with only

265 CHE as outgroup, and the other with only PAN (Supplementary Figure 2-5). ML tree inference of

266   the former resulted in a sister group relationship of CHI+DIP and another sister group relationship

267   of PAU+SYM, which was congruent with the results inferred from the datasets with the full taxon

268   sampling (Figure 3). However, we found that the ML tree inferred from the latter datasets resulted

269   in a sister group relationship of CHI and DIP, with SYM as a sister to this clade, followed by PAU.

270   Although these quartet topology results were also obtained in previous studies, negligible support

271   could be obtained from bootstrapping analyses(Szucsich et al., 2020).

272

273   **LIT genes' identification in Myriapoda**

274   Using the PIA2 pipeline, we identified 2,001 transcripts in 39 Myriapoda species as putative

275   components involved in the development of light-interacting structures, including 96 LIT genes,

276   which are important components of 11 visual pathways (Figure 4). A total of 13 visual pathways

277   were compiled in the pipeline, and two were absent in this study (Table 3, retinal determination

278   network and opsin synthesis) which involved 10 LIT genes. In addition, the other five absent LIT

279   genes were *Gq_gamma*, *RBP3*, *Dat, TYR*, and *reflectin_1a*. We investigated the completeness of

280   the amino acid level of each ortholog by calculating the ratio of the length of the identified peptide

281   to the target reference peptide, as depicted in Figure 4; the lighter the cell-filling colour, the more

282   incomplete the transcript. As shown in Figure 4, the distribution patterns of the identified LIT

283   genes among the four classes are very similar; LIT genes from prc (photoreceptor specification),

284   reti (retinoid pathway, invertebrate), heme (heme synthesis), and crys (crystallins) were rarely

285   identified in Myriapoda (shown as a large blank area in Figure 4). We also found that LIT genes

286   from reti (retinoid pathway, invertebrate, 0), ommo (ommochrome synthesis, 2), and clock (diurnal

287   clock, 2) were rarely identified in PAU (Table 3). As for the common LIT genes among these four

288   classes, we found that 62 LIT genes could be identified in at least one of the four classes (Figure

289   5c), and three (*GC*, *TH*, *KF*) could be identified in all 39 myriapods investigated. The following

290   three genes were separately involved in three different pathways, *ctrans*: phototransduction in

291   ciliary, *mel*: melanin synthesis, and *ommo*: ommochrome synthesis. We compared the LIT genes

292   co-identified between DIP and CHI, and PAU and SYM according to the visual pathways in which

293   the genes participated (Figure 5b and Figure 5a). We found that clock (diurnal clock), crys

294   (crystallins), ommo (ommochrome synthesis), and reti (retinoid pathway, invertebrate) were more

295   abundant in the sister group of DIP and CHI.

296

297   **Selection tests on LIT genes**

298   To test whether genes associated with the evolution of light interactions in Myriapoda have

299   undergone potentially adaptive changes, Ka/Ks calculations were conducted. A total of 23,832

300   aligned LIT gene pairs were calculated, including 8,717 pairs from CHI&DIP, 2,048 from

301   CHI&SYM, 1,220 from CHI&PAU, 1,158 from DIP&PAU, 1,981 from DIP&SYM, 278 from

302   PAU&SYM, 4,241 from CHI&CHI, 3,981 from DIP&DIP, 47 from PAU&PAU, and 161 from

303   SYM&SYM. Positive selection was detected in 27 pairs, indicated by Ka/Ks$\geq$1, five pairs from

304   CHI&DIP, one from CHI&PAU, three from CHI&SYM, two from DIP&PAU, two from

305   DIP&SYM, seven from CHI&CHI, and seven from DIP&DIP. As depicted in Figure 5, no

306   evidence of positive selection for LIT genes was found in PAU&PAU, SYM&SYM, PAU&SYM.

307   Values of Ka/Ks in the range of 0.5 to 1.0, which indicates relaxed selection, were observed in 395

308   pairs, covering all classes combinations (details in Supplementary Table 2). The remaining 23,435

309   pairs had Ka/Ks values ranging from 0.0002 to 0.5, representing 98% of the pairs we calculated,

310   which means that most of the genes in the four major classes were under purifying selection.

311   Positive selection was detected in the following 14 LIT genes: *clot*, *Cnga1*, *CSAD*, *DAGK*, *DDC*,

312   *Galpha_it*, *GC*, *Gprk1*, *Gq_alpha*, *Pde6abc*, *PKC*, *PLC*, *RBP1*, *RDH8*, *timeless*, and *trp*, which

313   cover the clock, ctrans, mel, pter, retv, and rtrans visual pathways (Figure 6). In addition, the

314   transient receptor potential protein trp, which encodes a component of the rhabdomeric

315   phototransduction pathway, was identified and positively selected in PAU&CHI, SYM&CHI, and

316   SYM&DIP.

317

318

319   **Tables and Figures**

320 **Table 1. Taxon sampling**

321 Species included in this study SRA accession numbers(Brewer & Bond, 2013; Fernández et al., 2014; Sharma et al., 2014;

322 Fernández, Edgecombe & Giribet, 2016, 2018; Szucsich et al., 2020), information collection, and data sources are indicated.

323

324 **Table 2. Results of topology tests**

325 Results of approximately unbiased (AU), weighted Kishino-Hasegawa (KH), and weighted Shimodaira-Hasegawa (SH) tests

326 comparing historically proposed hypotheses of the inner relationships of Myriapoda. A total of 100,000 RELL replicates were

327 performed for each test, plus signs (+) denote the 95% confidence sets (not rejected), minus signs (-) denote significant exclusion

328 (rejected).

329

330 **Table 3. Distribution of LIT genes' identification**

331 Statistical results of LIT gene identification. Sum: the sum of transcripts from a specific class that was identified as LIT genes

332 involved in a specific visual pathway. Max: the maximum quantity of transcripts from a species among a specific class that was

333 identified as the LIT genes involved in a specific visual pathway. Mean: ratio of the sum and species quantity of a specific class.

334

335 **Figure 1. Hypotheses on relationships of the major myriapod lineages Chilopoda, Diplopoda, Symphyla**

336 **and Pauropoda**

337 Hypothesis Eda.1, Hypothesis Eda.2 and Hypothesis Eda.3 are three quartet topologies derived from Edafopoda, which grouping

338 the PAU and SYM as a sister clade; Hypothesis Dig.1, Hypothesis Dig.2 and Hypothesis Dig.3 are three quartet topologies derived

339 from Dignatha, which grouping the PAU and DIP as a sister clade.

340

341 **Figure 2. Schematic of the two supermatrices used in this study.**

342 Matrix OCC100 was based on the blue section (100% gene occupancy) where 32 BUSCOs were included, and matrix OCC90 was

343 based on the purple and blue section (>90% gene occupancy) where 505 BUSCO were included.

344

345 **Figure 3. Best ML tree on matrix OCC90**

346 Best Maximum-Likelihood tree inferred with IQ-TREE derived from matrix OCC90 (60 taxa, alignment length: 129,085 amino

347 acid positions, 369 gene partitions), and rooted with Onychophora, where all the topology were consistent with the best ML tree

348 inferred from matrix OCC100. Statistical support was derived from 1,000 non-parametric bootstrap replicates, trees were converged

349 after 700 replicates.

350

351 **Figure 4. LIT genes identified in four major subgroups of Myriapoda**

352 Tree structure on the left of the figure was the best ML tree in this study. The colorful cells represent the completeness on amino-

353 acid level of each ortholog by calculating the ratio of the length of the identified peptide and the target reference one provided in

354 the PIA2. The lighter of the cell filling color, the more incomplete the transcript. Abbreviations of the visual pathways are following,

355 PRC: Photoreceptor Specification, RTRANS: Phototransduction, Rhabdomeric, CTRANS: Phototransduction, Ciliary, RETV:

356 Retinoid Pathway, Vertebrate, RETI: Retinoid Pathway, Invertebrate, MEL: Melanin Synthesis, PTER: Pterin Synthesis, OMMO:

357 Ommochrome Synthesis, HEME: Heme Synthesis, CRYS: Crystallins, CLOCK: Diurnal Clock.

358

359 **Figure 5. Comparison of the LIT genes in four main subgroups of Myriapoda**

360 Co-identified LIT genes grouping by visual pathways between PAU and SYM (a), DIP and CHI (b). Number of distinct LIT genes

361 identified among each monophyletic subgroup of Myriapoda.

362

363 **Figure 6. Positive selection support for LIT genes by Ka/Ks calculation**

364 Subgroup distribution of the positively selected genes, the bigger the bubbles, the more pairs found under positive selection, and

365 the smallest bubbles mean the support was one.

366

367 Supplementary Table 1. Ortholog group assignment

368 Supplementary Table 2. Ka/Ks results of LIT genes

369 Supplementary Figure 1. ML tree on matrix OCC100

370 Supplementary Figure 2. ML tree on matrix OCC90 without outgroups from Pancrustacea

371 Supplementary Figure 3. ML tree on matrix OCC100 without outgroups from Pancrustacea

372 Supplementary Figure 4. ML tree on matrix OCC90 without outgroups from Chelicerata

373 Supplementary Figure 5. ML tree on matrix OCC100 without outgroups from Chelicerata

374

375 **Discussion**

376   In this study, we first performed phylogenomic analyses on Myriapoda with three newly

377 sequenced members by integrating phylogenetic tree inference and topology testing. Our results

378 showed that CHI+DIP and PAU+SYM were the best quartet topologies for interrelationships

379 among the four major classes. This is consistent with the recent study by Szucsich et al., but in

380 conflict with the earlier results published by Fernández et al. and with morphological

381 evidence(Fernández, Edgecombe & Giribet, 2018; Szucsich et al., 2020). We provided an

382 innovative point in taxon sampling, which was the addition of three newly sequenced Myriapoda

383 species with high-quality sequencing, two of which were members of DIP and one of CHI. It is

384   particularly worth mentioning that the taxon sampling of the PAU class in this study was

385   represented by two species (*Pauropus huxleyi* and *Acopauropus ornatus*), instead of just one

386   specie in previous studies which may increase the risk of mispositioning PAU in quartet topology.

387   Previous studies have shown that, in molecular studies, the more extensive a taxon sample

388   collection, the more convincing the phylogeny results(Fernández, Edgecombe & Giribet, 2018;

389   Szucsich et al., 2020).

390   There seems to be no end to the debate about the interrelationships among the four main

391   Myriapoda classes, though phylogenomic analyses do provide robust evidence for phylogeny,

392   where novel research points were being summoned. We then turned our study's focus on the

393   common environmental and ecological habits and conditions of most species of PAU and SYM.

394   The pauropods (PAU) inhabit a variety of soil types, but sometimes found in plant litter and

395   decaying logs, and they cannot burrow themselves, which makes them be confined to crevices and

396   tunnels already present(Adis et al., 2002). Besides, they are small-sized with little mobility, which

397   can be the reason for that they rarely appear outside the Amazonian sampling sites(Battirola et al.,

398   2018; Hilgert et al., 2019; Santos-Silva et al., 2019). Though symphylans (SYM) most often are

399   true soil-dwellers, they can live in many different habitats: in leaf litter, in the upper humus layer,

400   and in pure soil, both in upper layers and in the mineral subsoil(Adis et al., 2002). In summary,

401   most species of PAU and SYM are small-sized, soil-dwelling and blind, while most of CHI and

402   DIP are eyed. It cannot be denied that these characters are shared with some orders of DIP and

403   CHI. The order Geophilomorpha (comprise over 1,254 species) and the family Cryptopidae (order

404   Scolopendromorpha, comprise over 184 species) of the class CHI (comprise over 4,142 species)

405   are both live underground and blind(Adis et al., 2002; Borja & Rincón, 2017), which accounts for

406   a third of the CHI. In class DIP, order Polydesmida, order Platydesmida, order Glomeridesmida

407   separately comprise over 3,500, 50 and 30 species, which accounts for almost a half of the DIP

408   (comprise over 7,753 species) (Adis et al., 2002; Borja & Rincón, 2017). However, almost all

409   species of PAU (comprise over 835 species) and SYM (comprise over 197 species) were blind and

410   soil-dwelling(Adis et al., 2002; Borja & Rincón, 2017). It could be due to biases in species

411   richness, but we insisted that visual capability was a good innovation point, and made a preliminary

412   exploration. With the help of analytical pipeline developed for whole-genome wide identification

413   of visual genes(Speiser et al., 2014; Pérez-Moreno et al., 2018), we identified visual genes for each

414   of the 39 Myriapoda species separately, and compared the distribution of positively selected genes

415   among the four major classes. As our results showed, positive selection was detected between

416   species from CHI and DIP, CHI and PAU, CHI and SYM, DIP and PAU, DIP and SYM, but none

417   was found between PAU and SYM. Both, the LIT gene identification and the positive selection,

418   indicated that the components of the rhabdomeric phototransduction pathway, which are employed

419   by the photoreceptors found in the eyes of many invertebrates, received the most attention (Figure

420   5 and Figure 6). The majority of the components of the rhabdomeric phototransduction pathway

421   are responsible for conferring light sensitivity to photoreceptors from the retinas of

422   cephalopods(Mitchell & Mayeenuddin, 1998; Kishigami et al., 2001; Murakami & Kouyama,

423   2008), including Gq protein, r-opsins, and transient receptor potential protein. The positive

424   selection pressure on *trp* in PAU&CHI, SYM&CHI, and SYM&DIP could reflect adaptive

425   changes in the machinery during the rhabdomeric phototransduction pathway. In this respect, the

426   positive selection signatures on the components of the rhabdomeric phototransduction pathway in

427   Myriapoda could be regarded and further analysed from a broader perspective. Besides, blind

428   species account for more than a third in both CHI and DIP, thus LIT genes' patterns between eyed

429   CHI and blind CHI, eyed DIP and blind DIP, warrants further study.

430   Although phylogenomic data covering 39 myriapods have been published, the acquisition of

431   more sufficient data is still expected, especially transcriptomic data from species in PAU and

432   SYM. To address the interrelationships of the four major classes of Myriapoda, we consolidated

433   most of the available data for Myriapoda phylogeny analyses, and conducted a series of

434   phylogenomic analyses, which provided strong evidence for the PAU+SYM topology of

435   Edafopoda. To find other evidence for Edafopoda, we identified visual genes that detected positive

436 selection pressure among the four major classes of Myriapoda. In short, our analyses offered more

437 information to further promote the study of interspecific evolutionary relationships among

438 Myriapoda.

439

Conclusion

441   For the highly disputed interrelationships of Myriapoda, our best phylogenetic tree involved

442 39 species favored the hypothesis Edafopoda, which was supported by a series of topology tests

443 we conducted and consistent with plenty of previous studies. The commonness of living habits

444 was investigated among the four major classes, and we made a preliminary exploration by LIT

445 analyses. Though weak evidence was found to support the monophyly of PAU and SYM, we think

446 it is a good research point, which needs further study.

Acknowledgments

448 This work was supported by the Natural Science Foundation of the Higher Education Institutions

449 of Anhui Province(KJ2018ZD041)and the Key Program in the Youth Elite Support Plan in

450 Universities of Anhui Province(gxyqZD2020045).

Authors' contributions

452 DY conceived and designed the experiments. DY and WJJ contributed to paper writing. ZHF and

453 MUN conducted the experiment. WJJ and BY contributed to the data analysis.

Disclosure statement

455 No potential conflict of interest was reported by the author(s).

Availability of data and materials

457   The data are presented in the manuscript and the supporting materials. The raw reads data are

458   submitted to the Short Read Archive (SRA) and BioProject accession number

459   PRJNA758760(https://www.ncbi.nlm.nih.gov/bioproject/PRJNA758760).

460

461   **References**

462   Adis J, Foddai D, Golovatch SI, Hoffman RL, Minelli A, de Morais JW, Pereira LA, Scheller U,

463   Schileyko AA, Würmli M. 2002. Myriapoda at "Reserva Ducke", central Amazonia/Brazil.

464   *Amazoniana* 17:15–25.

465   Bäcker H, Fanenbruck M, Wägele JW. 2008. A forgotten homology supporting the monophyly

466   of Tracheata: The subcoxa of insects and myriapods re-visited. *Zoologischer Anzeiger*

467   247:185–207. DOI: 10.1016/j.jcz.2007.11.002.

468   Battirola LD, Golovatch SI, Pinheiro TG, Batistella DA, Rosado-Neto GH, Chagas A, Brescovit

469   AD, Marques MI. 2018. Myriapod (Arthropoda, Myriapoda) diversity and distribution in a

470   floodplain forest of the Brazilian Pantanal. *Studies on Neotropical Fauna and Environment*

471   53:62–74. DOI: 10.1080/01650521.2017.1397978.

472   Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence

473   data. *Bioinformatics* 30:2114–2120. DOI: 10.1093/bioinformatics/btu170.

474   Borja R, Rincón B. 2017. *Reference Module in Life Sciences*.

475   Brewer MS, Bond JE. 2013. Ordinal-level phylogenomics of the arthropod class Diplopoda

476   (millipedes) based on an analysis of 221 nuclear protein-coding loci generated using next-

477   generation sequence analyses. *PLoS ONE* 8. DOI: 10.1371/journal.pone.0079935.

478   Dong Y, Sun H, Guo H, Pan D, Qian C, Hao S, Zhou K. 2012. The complete mitochondrial

479   genome of Pauropus longiramus (Myriapoda: Pauropoda): Implications on early

480   diversification of the myriapods revealed from comparative analysis. *Gene* 505:57–65. DOI:

481   10.1016/j.gene.2012.05.049.

482   Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome

483   comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*

484   16:157. DOI: 10.1186/s13059-015-0721-2.

485   Fernández R, Edgecombe GD, Giribet G. 2016. Exploring Phylogenetic Relationships within

486   Myriapoda and the Effects of Matrix Composition and Occupancy on Phylogenomic

487   Reconstruction. *Systematic Biology* 65:871–889. DOI: 10.1093/sysbio/syw041.

488   Fernández R, Edgecombe GD, Giribet G. 2018. Phylogenomics illuminates the backbone of the

489   Myriapoda Tree of Life and reconciles morphological and molecular phylogenies. *Scientific*

490   *Reports* 8:83. DOI: 10.1038/s41598-017-18562-w.

491   Fernández R, Laumer CE, Vahtera V, Libro S, Kaluziak S, Sharma PP, Pérez-Porro AR,

492   Edgecombe GD, Giribet G. 2014. Evaluating topological conflict in centipede phylogeny

493   using transcriptomic data sets. *Molecular Biology and Evolution* 31:1500–1513. DOI:

494   10.1093/molbev/msu108.

495   Giribet G, Edgecombe GD. 2019. The Phylogeny and Evolutionary History of Arthropods.

496   *Current Biology* 29:R592–R602. DOI: 10.1016/j.cub.2019.04.057.

497   González VL, Andrade SCS, Bieler R, Collins TM, Dunn CW, Mikkelsen PM, Taylor JD,

498   Giribet G. 2015. A phylogenetic backbone for Bivalvia: An RNA-seq approach.

499   *Proceedings of the Royal Society B: Biological Sciences* 282:20142332. DOI:

500   10.1098/rspb.2014.2332.

501   Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,

502   Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma

503   F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length

504   transcriptome assembly from RNA-Seq data without a reference genome. *Nature*

505   *Biotechnology* 29:644–652. DOI: 10.1038/nbt.1883.

506   Hilgert M, Akkari N, Rahmadi C, Wesener T. 2019. The Myriapoda of Halimun-Salak National

507   Park (Java, Indonesia): Overview and faunal composition. *Biodiversity Data Journal* 7.

508   DOI: 10.3897/BDJ.7.e32218.

509   Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:

510   Improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.

511   DOI: 10.1093/molbev/mst010.

512   Kishigami A, Ogasawara T, Watanabe Y, Hirata M, Maeda T, Hayashi F, Tsukahara Y. 2001.

513   Inositol-1,4,5-trisphosphate-binding proteins controlling the phototransduction cascade of

514   invertebrate visual cells. *Journal of Experimental Biology* 204:487–493. DOI:

515   10.1242/jeb.204.3.487.

516   Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: A fast, scalable

517   and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*

518   35:4453–4455. DOI: 10.1093/bioinformatics/btz305.

519   Kück P. 2009. ALICUT: a Perlscript which cuts ALISCORE identified RSS. *Department of*

520   *Bioinformatics, Zoologisches Forschungsmuseum A. Koenig (ZFMK), Bonn, Germany,*

521   *version 2*.

522   Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. Partitionfinder 2: New

523   methods for selecting partitioned models of evolution for molecular and morphological

524   phylogenetic analyses. *Molecular Biology and Evolution* 34:772–773. DOI:

525   10.1093/molbev/msw260.

526   Li W, Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein

527   or nucleotide sequences. *Bioinformatics* 22:1658–1659. DOI:

528   10.1093/bioinformatics/btl158.

529   Marek PE, Bond JE. 2006. Rediscovery of the world's leggiest animal. *Nature* 441:707. DOI:

530   10.1038/441707a.

531   Meyer B, Meusemann K, Misof B. 2011. MARE: MAtrix REduction—a tool to select optimized

532   data subsets from supermatrices for phylogenetic inference. *Bonn (Germany): Zentrum fuur*

533   *molekulare Biodiversit{ä}tsforschung (zmb) am ZFMK*.

534   Mitchell J, Mayeenuddin LH. 1998. Purification, G protein activation, and partial amino acid

535   sequence of a novel phospholipase C from squid photoreceptors. *Biochemistry* 37:9064–

536   9072. DOI: 10.1021/bi972768a.

537   Miyazawa H, Ueda C, Yahata K, Su ZH. 2014. Molecular phylogeny of Myriapoda provides

538   insights into evolutionary patterns of the mode in post-embryonic development. *Scientific*

539   *Reports* 4. DOI: 10.1038/srep04127.

540   Moore J. 2006. An Introduction to the Invertebrates: Chelicerata and Myriapoda.

541   10.1017/CB:181–191.

542   Murakami M, Kouyama T. 2008. Crystal structure of squid rhodopsin. *Nature* 453:363–367.

543   DOI: 10.1038/nature06925.

544   Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective

545   stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology*

546   *and Evolution* 32:268–274. DOI: 10.1093/molbev/msu300.

547   Pérez-Moreno JL, DeLeo DM, Palero F, Bracken-Grissom HD. 2018. Phylogenetic annotation

548   and genomic architecture of opsin genes in Crustacea. *Hydrobiologia* 825:159–175. DOI:

549   10.1007/s10750-018-3678-9.

550   Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW.

551   2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding

552   sequences. *Nature* 463:1079–1083. DOI: 10.1038/nature08742.

553   Regier JC, Zwick A. 2011. Sources of signal in 62 protein-coding nuclear genes for higher-level

554   phylogenetics of arthropods. *PLoS ONE* 6. DOI: 10.1371/journal.pone.0023408.

555   Santos-Silva L, Golovatch SI, Pinheiro TG, Chagas-Jr A, Marques MI, Battirola LD. 2019.

556   Myriapods (Arthropoda, myriapoda) in the Pantanal of Poconé, Mato Grosso, Brazil. *Biota*

557   *Neotropica* 19. DOI: 10.1590/1676-0611-bn-2018-0631.

558   Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing genome assembly and annotation

559   completeness. *Methods in Molecular Biology* 1962:227–245. DOI: 10.1007/978-1-4939-

560   9173-0_14.

561 Sharma PP, Kaluziak ST, Pérez-Porro AR, González VL, Hormiga G, Wheeler WC, Giribet G.

562   2014. Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic

563   signal. *Molecular Biology and Evolution* 31:2963–2984. DOI: 10.1093/molbev/msu235.

564 Speiser DI, Pankey MS, Zaharoff AK, Battelle BA, Bracken-Grissom HD, Breinholt JW, Bybee

565   SM, Cronin TW, Garm A, Lindgren AR, Patel NH, Porter ML, Protas ME, Rivera AS, Serb

566   JM, Zigler KS, Crandall KA, Oakley TH. 2014. Using phylogenetically-informed

567   annotation (PIA) to search for light-interacting genes in transcriptomes from non-model

568   organisms. *BMC Bioinformatics* 15:350. DOI: 10.1186/s12859-014-0350-x.

569 Stamatakis A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of

570   large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.

571 Szucsich NU, Bartel D, Blanke A, Böhm A, Donath A, Fukui M, Grove S, Liu S, Macek O,

572   Machida R, Misof B, Nakagaki Y, Podsiadlowski L, Sekiya K, Tomizuka S, Von Reumont

573   BM, Waterhouse RM, Walzl M, Meng G, Zhou X, Pass G, Meusemann K. 2020. Four

574   myriapod relatives – but who are sisters? No end to debates on relationships among the four

575   major myriapod subgroups. *BMC Evolutionary Biology* 20. DOI: 10.1186/s12862-020-

576   01699-0.

577 Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: A Toolkit Incorporating

578   Gamma-Series Methods and Sliding Window Strategies. *Genomics, Proteomics and*

579   *Bioinformatics* 8:77–80. DOI: 10.1016/S1672-0229(10)60008-3.

580 Zwick A, Regier JC, Zwickl DJ. 2012. Resolving Discrepancy between Nucleotides and Amino

581   Acids in Deep-Level Arthropod Phylogenomics: Differentiating Serine Codons in 21-

582   Amino-Acid Models. *PLoS ONE* 7:e47450. DOI: 10.1371/journal.pone.0047450.

583

Figure 1

Hypotheses_on_relationships_of_the_major_myriapod_lineages_Chilopoda,_Diplopoda,_Symphyla_and_Pauropoda

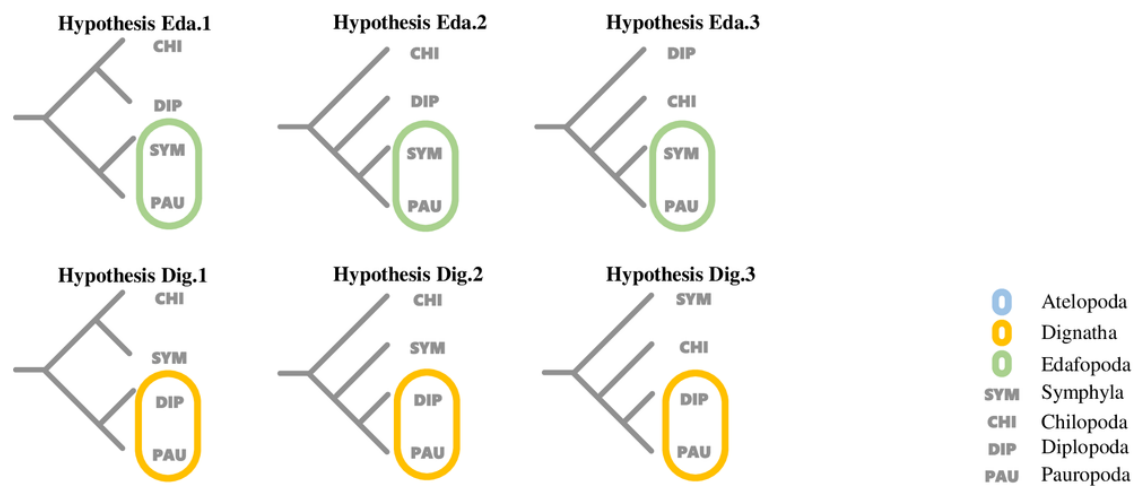
Hypothesis Eda.1, Hypothesis Eda.2 and Hypothesis Eda.3 are three quartet topologies derived from Edafopoda, which grouping the PAU and SYM as a sister clade; Hypothesis Dig.1, Hypothesis Dig.2 and Hypothesis Dig.3 are three quartet topologies derived from Dignatha, which grouping the PAU and DIP as a sister clade.

	Atelopoda
	Dignatha
	Edafopoda
SYM	Symphyla
CHI	Chilopoda
DIP	Diplopoda
PAU	Pauropoda

Figure 2

Schematic_of_the_two_supermatrices_used_in_this_study

Matrix OCC100 was based on the blue section (100% gene occupancy) where 32 BUSCOs were included, and matrix OCC90 was based on the purple and blue section (>90% gene occupancy) where 505 BUSCO were included.

Figure 3

Best_ML_tree_on_matrix_OCC90

Best Maximum-Likelihood tree inferred with IQ-TREE derived from matrix OCC90 (60 taxa, alignment length: 129,085 amino acid positions, 369 gene partitions), and rooted with Onychophora, where all the topology were consistent with the best ML tree inferred from matrix OCC100. Statistical support was derived from 1,000 non-parametric bootstrap replicates, trees were converged after 700 replicates.

Tree scale: 0.1 ⊢————⊣

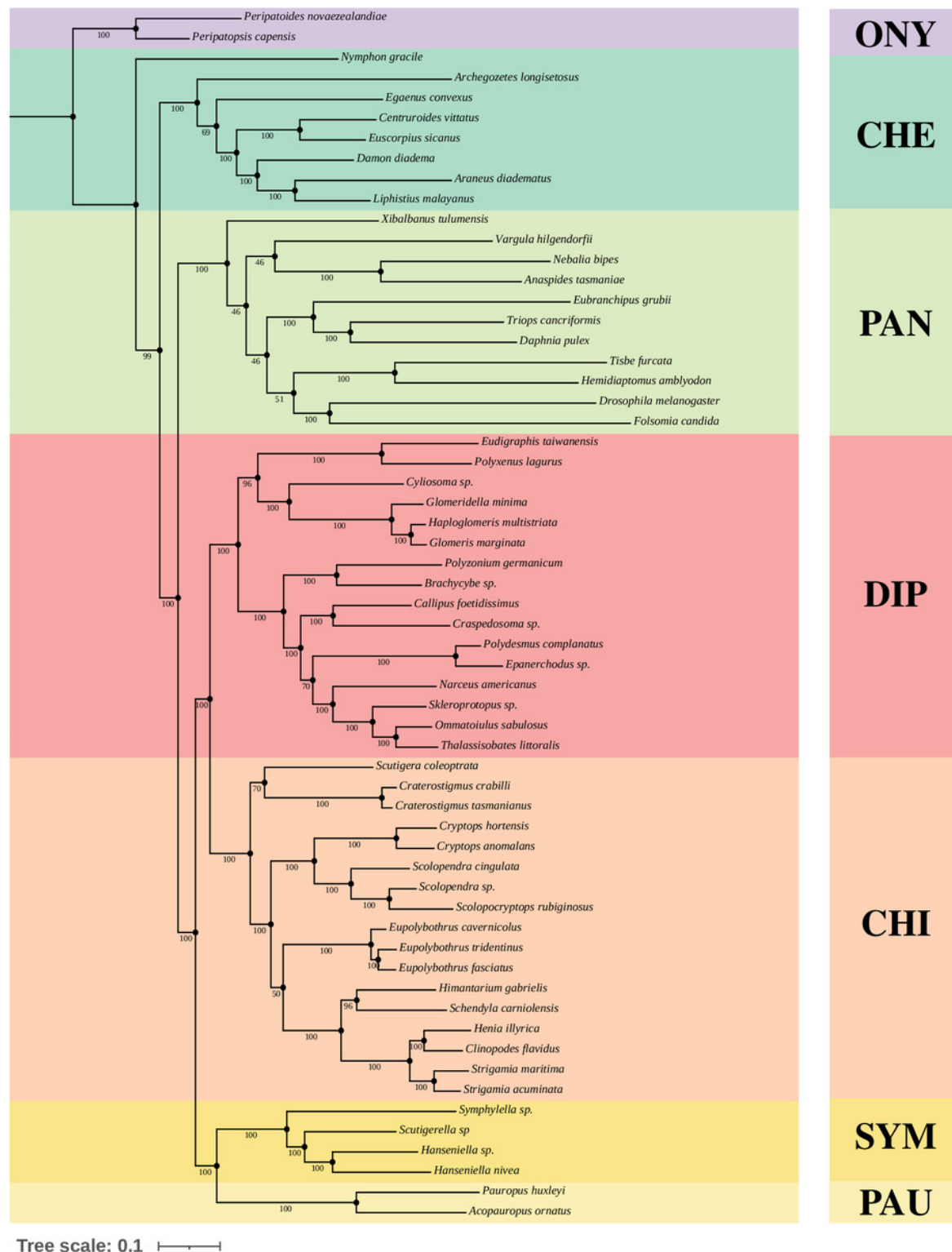



Figure 4

LIT_genes_identified_in_four_major_subgroups_of_Myriapoda

Tree structure on the left of the figure was the best ML tree in this study. The colorful cells represent the completeness on amino-acid level of each ortholog by calculating the ratio of the length of the identified peptide and the target reference one provided in the PIA2. The lighter of the cell filling color, the more incomplete the transcript. Abbreviations of the visual pathways are following, PRC: Photoreceptor Specification, RTRANS: Phototransduction, Rhabdomeric, CTRANS: Phototransduction, Ciliary, RETV: Retinoid Pathway, Vertebrate, RETI: Retinoid Pathway, Invertebrate, MEL: Melanin Synthesis, PTER: Pterin Synthesis, OMMO: Ommochrome Synthesis, HEME: Heme Synthesis, CRYS: Crystallins, CLOCK: Diurnal Clock.

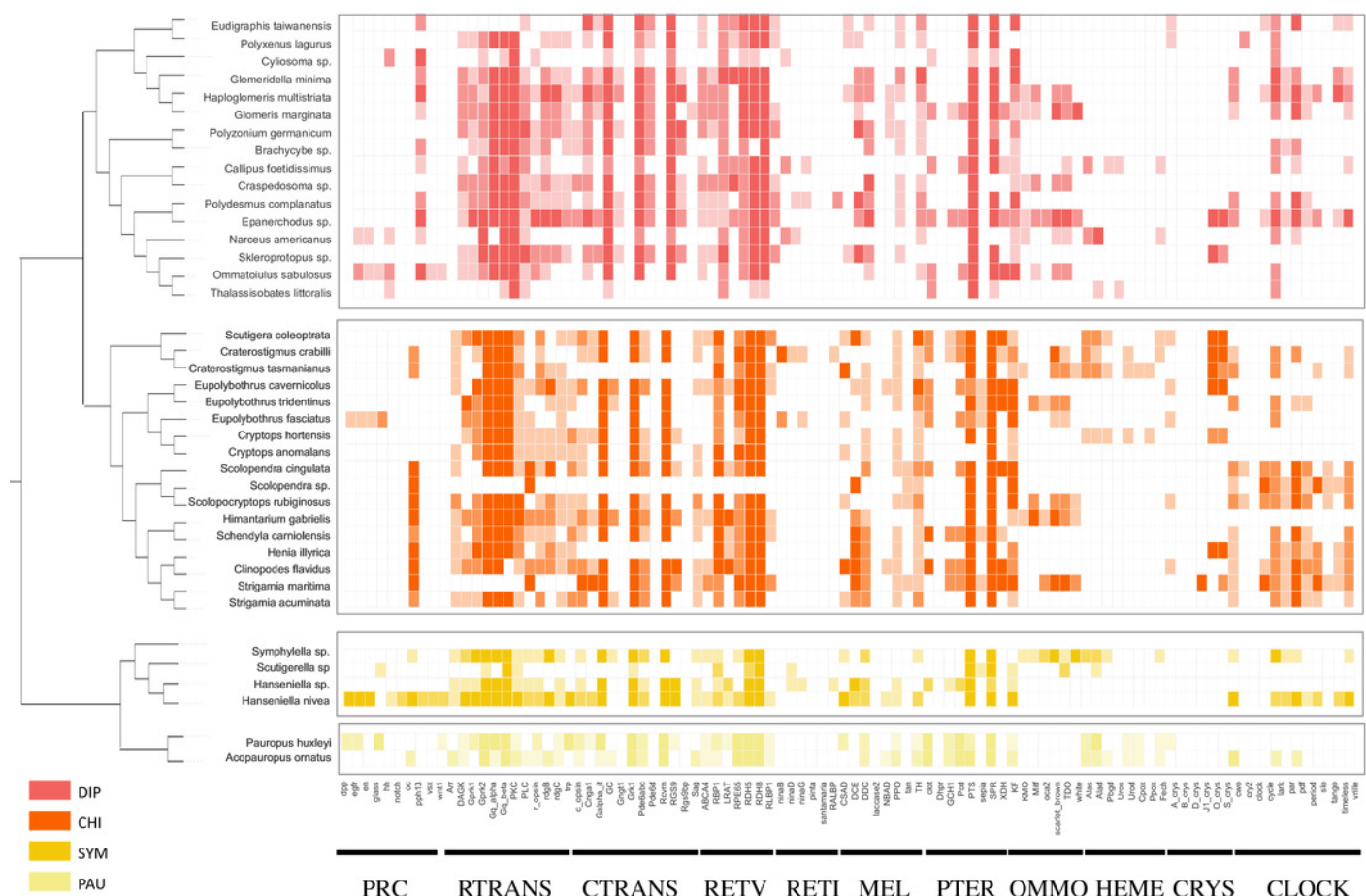



Figure 5

Comparison_of_the_LIT_genes_in_four_main_subgroups_of_Myriapoda

Co-identified LIT genes grouping by visual pathways between PAU and SYM (a), DIP and CHI (b). Number of distinct LIT genes identified among each monophyletic subgroup of Myriapoda.

a

b

c

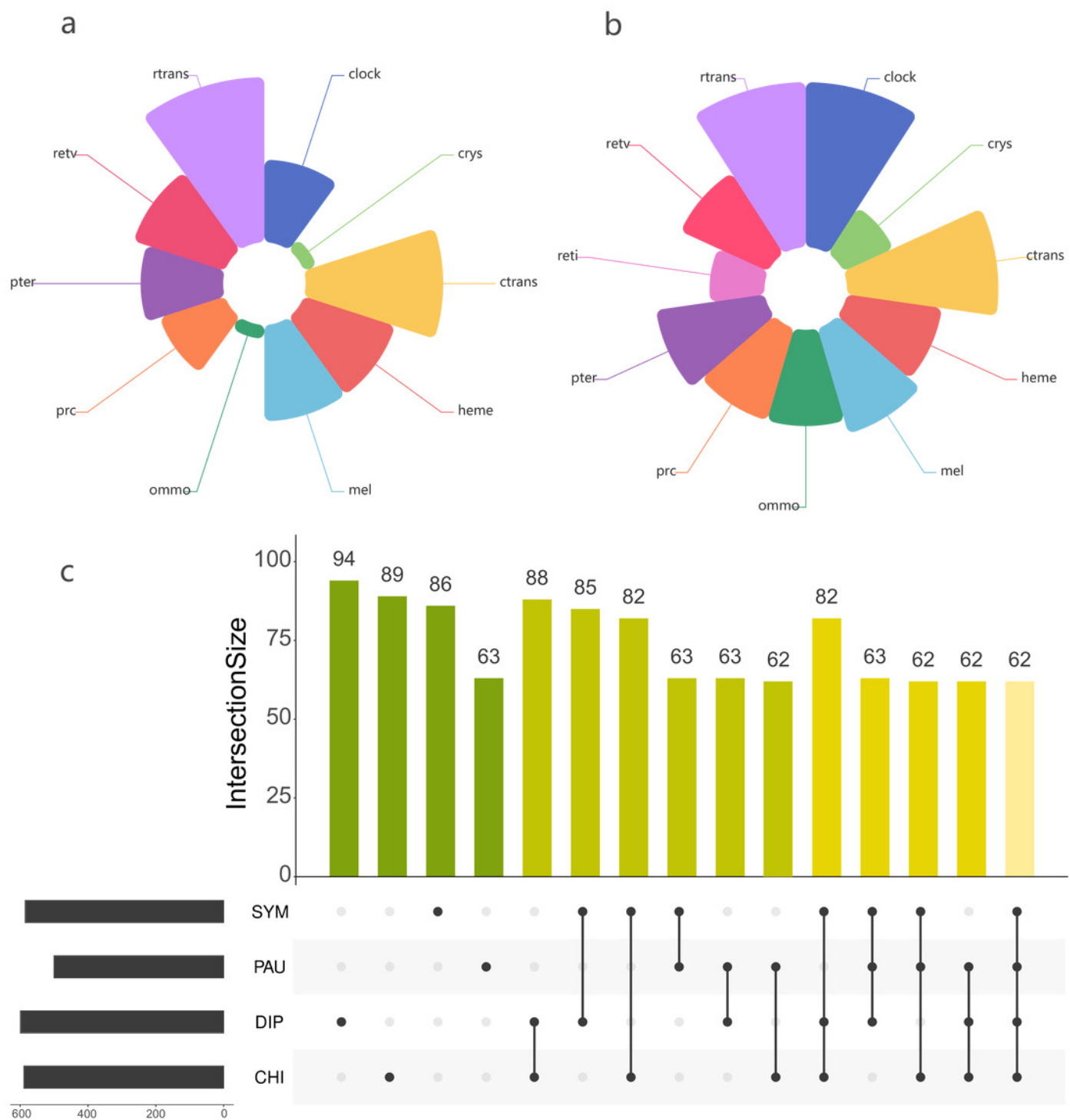
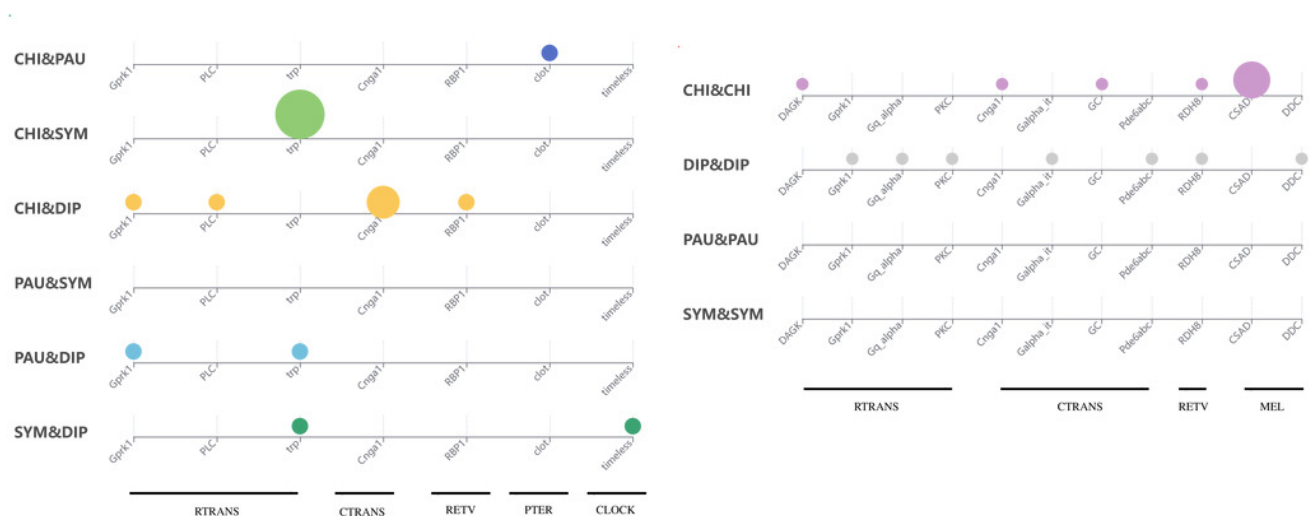



Figure 6

Positive_selection_support_for_LIT_genes_by_KaKs_calculation

Subgroup distribution of the positively selected genes, the bigger the bubbles, the more pairs found under positive selection, and the smallest bubbles mean the support was one.

Taxon_sampling

Species included in this study SRA accession numbers, information collection, and data sources are indicated.

Taxonomy	Clade Alias	Species	DataType	Source	SRA #	Species Alias	
Myriapoda, Chilopoda	CHI	*Eupolybothrus cavernicolus*	Transcriptome	Fernández et al. 2014	ERR338470	Spe01	*Eupolybothrus cavernicolus*
Myriapoda, Chilopoda	CHI	*Cryptops hortensis*	Transcriptome	Fernández et al. 2016	SRR1153457	Spe02	*Cryptops hortensis*
Myriapoda, Chilopoda	CHI	*Scutigera coleoptrata*	Transcriptome	Fernández et al. 2014	SRR1158078	Spe03	*Scutigera coleoptrata*
Myriapoda, Chilopoda	CHI	*Craterostigmus crabilli*	Transcriptome	Fernández et al. 2014	SRR3232915	Spe04	*Craterostigmus crabilli*
Myriapoda, Chilopoda	CHI	*Strigamia maritima*	Genome	Chipman et al. 2014	--	Spe05	*Strigamia maritima*
Myriapoda, Chilopoda	CHI	*Scolopendra sp.*	Transcriptome	This study		Spe06	*Scoropendra sp.*
Myriapoda, Chilopoda	CHI	*Craterostigmus tasmanianus*	Transcriptome	Szucsich et al., 2020	SRR2774008	Spe31	*Craterostigmus tasmanianus*
Myriapoda, Chilopoda	CHI	*Henia illyrica*	Transcriptome	Szucsich et al., 2020	SRR3485986	Spe32	*Henia illyrica*
Myriapoda, Chilopoda	CHI	*Clinopodes flavidus*	Transcriptome	Szucsich et al., 2020	SRR1653181	Spe33	*Clinopodes flavidus*
Myriapoda, Chilopoda	CHI	*Himantarium gabrielis*	Transcriptome	Szucsich et al., 2020	SRR1653198	Spe34	*Himantarium gabrielis*
Myriapoda, Chilopoda	CHI	*Strigamia acuminata*	Transcriptome	Szucsich et al., 2020	SRR3485997	Spe35	*Strigamia acuminata*
Myriapoda, Chilopoda	CHI	*Schendyla carniolensis*	Transcriptome	Szucsich et al., 2020	SRR3485996	Spe36	*Schendyla carniolensis*
Myriapoda, Chilopoda	CHI	*Eupolybothrus fasciatus*	Transcriptome	Szucsich et al., 2020	SRR3485981	Spe37	*Eupolybothrus fasciatus*
Myriapoda,	CHI	*Eupolybothrus*	Transcriptome	Szucsich et al., 2020	SRR3485982	Spe38	*Eupolybothrus*

Chilopoda		*tridentinus*					*tridentinus*
Myriapoda, Chilopoda	CHI	*Cryptops anomalans*	Transcriptome	Szucsich et al., 2020	SRR3485978	Spe39	*Cryptops anomalans*
Myriapoda, Chilopoda	CHI	*Scolopendra cingulata*	Transcriptome	Szucsich et al., 2020	SRR1653235	Spe40	*Scolopendra cingulata*
Myriapoda, Chilopoda	CHI	*Scolopocryptops rubiginosus*	Transcriptome	Szucsich et al., 2020	SRR1653236	Spe41	*Scolopocryptops rubiginosus*
Myriapoda, Diplopoda	DIP	*Glomeris marginata*	Transcriptome	Fernández et al. 2016	SRR3233211	Spe11	*Glomeris marginata*
Myriapoda, Diplopoda	DIP	*Narceus americanus*	Transcriptome	Fernández et al. 2016	SRR3233222	Spe12	*Narceus americanus*
Myriapoda, Diplopoda	DIP	*Eudigraphis taiwanensis*	Transcriptome	Fernández et al. 2016	SRR3458640	Spe13	*Eudigraphis taiwanensis*
Myriapoda, Diplopoda	DIP	*Cyliosoma sp.*	Transcriptome	Fernández et al. 2016	SRR3458641	Spe14	*Cyliosoma sp.*
Myriapoda, Diplopoda	DIP	*Brachycybe sp.*	Transcriptome	Brewer & Bond 2013	SRR945430	Spe15	*Brachycybe sp.*
Myriapoda, Diplopoda	DIP	*Epanerchodus sp.*	Transcriptome	This study		Spe16	*Epanerchodus sp.*
Myriapoda, Diplopoda	DIP	*Skleroprotopus sp.*	Transcriptome	This study	SRR1145732	Spe17	*Skleroprotopus sp.*
Myriapoda, Diplopoda	DIP	*Callipus foetidissimus*	Transcriptome	Szucsich et al., 2020	SRR3485975	Spe42	*Callipus foetidissimus*
Myriapoda, Diplopoda	DIP	*Craspedosoma sp. [AD-2016]*	Transcriptome	Szucsich et al., 2020	SRR3485977	Spe43	*Craspedosoma sp. [AD-2016]*
Myriapoda, Diplopoda	DIP	*Haploglomeris multistriata*	Transcriptome	Szucsich et al., 2020	SRR3485985	Spe44	*Haploglomeris multistriata*
Myriapoda,	DIP	*Glomeridella minima*	Transcriptome	Szucsich et al., 2020	SRR3485983	Spe45	*Glomeridella minima*

Diplopoda							
Myriapoda, Diplopoda	DIP	*Ommatoiulus sabulosus*	Transcriptome	Szucsich et al., 2020	SRR1653222	Spe46	*Ommatoiulus sabulosus*
Myriapoda, Diplopoda	DIP	*Thalassisobates littoralis*	Transcriptome	Szucsich et al., 2020	SRR1653242	Spe47	*Thalassisobates littoralis*
Myriapoda, Diplopoda	DIP	*Polydesmus complanatus*	Transcriptome	Szucsich et al., 2020	SRR3485993	Spe48	*Polydesmus complanatus*
Myriapoda, Diplopoda	DIP	*Polyxenus lagurus*	Transcriptome	Szucsich et al., 2020	SRR3485994	Spe49	*Polyxenus lagurus*
Myriapoda, Diplopoda	DIP	*Polyzonium germanicum*	Transcriptome	Szucsich et al., 2020	SRR3485995	Spe50	*Polyzonium germanicum*
Myriapoda, Pauropoda	PAU	*Pauropus huxleyi*	Transcriptome	Fernández et al. 2017	SRR6145369	Spe10	*Pauropus huxleyi*
Myriapoda, Pauropoda	PAU	*Acopauropus ornatus*	Transcriptome	Szucsich et al., 2020	SRR3485973	Spe51	*Acopauropus ornatus*
Myriapoda, Symphyla	SYM	*Scutigerella sp*	Transcriptome	Fernández et al. 2014	SRR3458649	Spe07	*Scutigerella sp*
Myriapoda, Symphyla	SYM	*Hanseniella sp.*	Transcriptome	Fernández et al. 2016	SRR6217953	Spe08	*Hanseniella sp.*
Myriapoda, Symphyla	SYM	*Symphylella sp.*	Transcriptome	Fernández et al. 2017	SRR6144316	Spe09	*Symphylella sp.*
Myriapoda, Symphyla	SYM	*Hanseniella nivea*	Transcriptome	Szucsich et al., 2020	SRR3485984	Spe52	*Hanseniella nivea*
Chelicerata	CHE	*Liphistius malayanus*	Transcriptome	Sharma et al. 2014	SRR1145736	Spe18	*Liphistius malayanus*
Chelicerata	CHE	*Centruroides vittatus*	Transcriptome	Sharma et al. 2014	SRR1146578	Spe19	*Centruroides vittatus*
Chelicerata	CHE	*Damon diadema*	Transcriptome	Szucsich et al., 2020	SRR3485979	Spe25	*Damon diadema*
Chelicerata	CHE	*Archegozetes longisetosus*	Transcriptome	Szucsich et al., 2020	SRR1653174	Spe26	*Archegozetes longisetosus*

Chelicerata	CHE	*Araneus diadematus*	Transcriptome	Szucsich et al., 2020	SRR3485974	Spe27	*Araneus diadematus*
Chelicerata	CHE	*Egaenus convexus*	Transcriptome	Szucsich et al., 2020	SRR3485980	Spe28	*Egaenus convexus*
Chelicerata	CHE	*Euscorpius sicanus*	Transcriptome	Szucsich et al., 2020	SRR1653192	Spe29	*Euscorpius sicanus*
Chelicerata	CHE	*Nymphon gracile*	Transcriptome	Szucsich et al., 2020	SRR1653221	Spe30	*Nymphon gracile*
Onychophora	ONY	*Peripatopsis capensis*	Transcriptome	Szucsich et al., 2020	SRR1145776	Spe23	*Peripatopsis capensis*
Onychophora	ONY	*Peripatoides novaezealandiae*	Transcriptome	Szucsich et al., 2020	SRR3485992	Spe24	*Peripatoides novaezealandiae*
Crustacea	PAN	*Daphnia pulex*	Genome		--	Spe20	*Daphnia pulex*
Crustacea	PAN	*Folsomia candida*	Genome		--	Spe21	*Folsomia candida*
Crustacea	PAN	*Drosophila melanogaster*	Genome		--	Spe22	*Drosophila melanogaster*
Crustacea	PAN	*Eubranchipus grubii*	Transcriptome	Szucsich et al., 2020	SRR1653190	Spe53	*Eubranchipus grubii*
Crustacea	PAN	*Triops cancriformis*	Transcriptome	Szucsich et al., 2020	SRR1653248	Spe54	*Triops cancriformis*
Crustacea	PAN	*Nebalia bipes*	Transcriptome	Szucsich et al., 2020	SRR1653215	Spe55	*Nebalia bipes*
Crustacea	PAN	*Anaspides tasmaniae*	Transcriptome	Szucsich et al., 2020	SRR1653173	Spe56	*Anaspides tasmaniae*
Crustacea	PAN	*Hemidiaptomus amblyodon*	Transcriptome	Szucsich et al., 2020	SRR1653196	Spe57	*Hemidiaptomus amblyodon*
Crustacea	PAN	*Tisbe furcata*	Transcriptome	Szucsich et al., 2020	SRR1653244	Spe58	*Tisbe furcata*
Crustacea	PAN	*Vargula hilgendorfii*	Transcriptome	Szucsich et al., 2020	SRR1811940	Spe59	*Vargula hilgendorfii*
Crustacea	PAN	*Xibalbanus tulumensis*	Transcriptome	Szucsich et al., 2020	SRR1653240	Spe60	*Xibalbanus tulumensis*

Results_of_topology_tests

Results of approximately unbiased (AU), weighted Kishino-Hasegawa (KH), and weighted Shimodaira-Hasegawa (SH) tests comparing historically proposed hypotheses of the inner relationships of Myriapoda. A total of 100,000 RELL replicates were performed for each test, plus signs (+) denote the 95% confidence sets (not rejected), minus signs (-) denote significant exclusion (rejected).

Occ100	Pau-test		Pkh-test		Psh-test	
	ExcludeCHE	ExcludePAN	ExcludeCHE	ExcludePAN	ExcludeCHE	ExcludePAN
Hypothesis Eda.1	0.6200 +	0.2110 +	0.5300 +	0.1320 +	1.0000 +	0.5310 +
Hypothesis Eda.2	0.2950 +	0.9320 +	0.2330 +	0.8680 +	0.6130 +	1.0000 +
Hypothesis Eda.3	0.0101 -	0.1630 +	0.0341 -	0.1010 +	0.1490 +	0.4840 +
Hypothesis Dig.1	0.0088 -	0.1210 +	0.0925 +	0.0832 +	0.1630 +	0.1440 +
Hypothesis Dig.2	0.0805 +	0.0372 -	0.1380 +	0.0194 -	0.3380 +	0.1360 +
Hypothesis Dig.3	0.5410 +	0.0302 -	0.4700 +	0.0177 -	0.8100 +	0.0398 -

Occ90	Pau-test		Pkh-test		Psh-test	
	ExcludeCHE	ExcludePAN	ExcludeCHE	ExcludePAN	ExcludeCHE	ExcludePAN
Hypothesis Eda.1	1.0000 +	0.7770 +	1.0000 +	0.7770 +	1.0000 +	1.0000 +
Hypothesis Eda.2	3.62E-38 -	0.2230 +	0.0000 -	0.2230 +	1.00E-05 -	0.5570 +
Hypothesis Eda.3	2.85E-50 -	4.15E-87 -	0.0000 -	0.0000 -	0.0000 -	0.0039 -
Hypothesis Dig.1	9.72E-121 -	4.44E-110 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -
Hypothesis Dig.2	2.21E-91 -	5.15E-55 -	0.0000 -	0.0000 -	0.0000 -	0.0000 -
Hypothesis Dig.3	1.09E-37 -	3.04E-11 -	0.0000 -	0.0000 -	1.00E-05 -	0.0000 -

Distribution_of_LIT_genes_identification

Statistical results of LIT gene identification. Sum: the sum of transcripts from a specific class that was identified as LIT genes involved in a specific visual pathway. Max: the maximum quantity of transcripts from a species among a specific class that was identified as the LIT genes involved in a specific visual pathway. Mean: ratio of the sum and species quantity of a specific class.

	CHI			DIP			PAU			SYM		
	unique	sum	max	unique	sum	max	unique	sum	max	unique	sum	max
rdn	0	0	0	0	0	0	0	0	0	0	0	0
prc	8	20	8	10	35	8	6	6	5	10	12	9
rtrans	12	179	12	12	170	12	12	24	12	12	44	12
ctrans	11	139	11	12	153	11	10	19	10	10	31	10
retv	7	102	7	7	99	7	7	13	7	7	24	7
reti	4	8	4	5	18	4	0	0	0	5	10	5
mel	8	116	8	8	100	8	7	13	7	8	23	8
pter	8	81	8	8	56	8	6	12	6	8	18	8
ommo	7	48	7	7	53	7	1	2	1	7	13	7
heme	7	27	7	8	22	7	7	14	7	7	8	7
crys	5	34	4	5	22	4	1	2	1	1	4	1
clock	12	114	12	12	87	10	6	6	6	11	20	10
opsin	0	0	0	0	0	0	0	0	0	0	0	0

	CHI				DIP				PAU				SYM			
	unique	sum	max	mean	unique	sum	max	mean	unique	sum	max	mean	unique	sum	max	mean
rdn	0	0	0	0.00	0	0	0	0.00	0	0	0	0.00	0	0	0	0.00
prc	8	20	8	1.18	10	35	8	2.19	6	6	5	3.00	10	12	9	3.00
rtrans	12	179	12	10.53	12	170	12	10.63	12	24	12	12.00	12	44	12	11.00
ctrans	11	139	11	8.18	12	153	11	9.56	10	19	10	9.50	10	31	10	7.75
retv	7	102	7	6.00	7	99	7	6.19	7	13	7	6.50	7	24	7	6.00
reti	4	8	4	0.47	5	18	4	1.13	0	0	0	0.00	5	10	5	2.50
mel	8	116	8	6.82	8	100	8	6.25	7	13	7	6.50	8	23	8	5.75
pter	8	81	8	4.76	8	56	8	3.50	6	12	6	6.00	8	18	8	4.50
ommo	7	48	7	2.82	7	53	7	3.31	1	2	1	1.00	7	13	7	3.25
heme	7	27	7	1.59	8	22	7	1.38	7	14	7	7.00	7	8	7	2.00

crys	5	34	4	2.00	5	22	4	1.38	1	2	1	1.00	1	4	1	1.00
clock	12	114	12	6.71	12	87	10	5.44	6	6	6	3.00	11	20	10	5.00
opsin	0	0	0	0.00	0	0	0	0.00	0	0	0	0.00	0	0	0	0.00