Identification and Expression Pattern of Chemosensory Genes in the Transcriptome of *Propsilocerus akamusi*

Chuncai Yan [Equal first author, 1], Xiaoya Sun [Equal first author, 1], Wei Cao [1], Ruoqun Li [1], Cong Zhao [1], Zeyang Sun [1], Wenbin Liu [Corresp., 1], Lina Pan [Corresp. 1]

[1] Tianjin Key Laboratory of Animal and Plant Resistance, Tianjin Normal University, Tianjin, China

Corresponding Authors: Wenbin Liu, Lina Pan
Email address: skylwb@tjnu.edu.cn, skypln@tjnu.edu.cn

Chironomidae is the most ecologically diverse insects in aquatic and semi-aquatic habitats. *Propsilocerus akamusi* (Tokunaga) is a dominant and ubiquitous chironomid species in Eastern Asia and its morphologically unique larvae are also considered as indicator organisms to detect water contamination, potential toxicity and waterborne pathogens. Since few studies to date have focused on the olfactory system of *P. akamusi*, our study aims to elucidate the potential functions of chemosensory genes in *P. akamusi* . In our study, we found that although signals released from male groups might attract female swarmers, it was a completely male-dominated mating process. Sequencing the transcriptome of *P. akamusi* on an Illumina HiSeq platform generated 4.42, 4.46 and 4.53 Gb of clean reads for heads, legs, and antennae, respectively. 27,609 unigenes, 20,379 coding sequences (CDSs), and 8,073 simple sequence repeats were finally obtained. The gene-level differential expression analysis demonstrated variants among three different tissues, including 2,019 genes specifically expressed in heads, 1,540 genes in legs, and 2,071 genes in antennae. Additionally, we identified an assortment of putative olfactory genes consisting of 34 odorant binding proteins, 17 odorant receptors, 32 gustatory receptors, 22 ionotropic receptors, 6 chemosensory proteins as well as 3 sensory neuron membrane proteins; their relative abundances in the above three tissues were also determined by RT-qPCR. Our finding could allow a more plausible understanding of certain olfaction-mediated behaviors in groups of this macroinvertebrate.

1 **Identification and Expression Pattern of**

2 **Chemosensory Genes in the Transcriptome of**

3 ***Propsilocerus akamusi***

5 Chuncai Yan[#], Xiaoya Sun[#], Wei Cao, Ruoqun Li, Cong Zhao, Zeyang Sun, Wenbin Liu* and

6 Lina Pan*

8 Tianjin Key Laboratory of Animal and Plant Resistance, Tianjin Normal University, Tianjin

9 300387, China

11 **Running Title:** Transcriptome sequencing of *P. akamusi*

12 # Note: Chuncai Yan and Xiaoya Sun contributed equally to this work.

14 **Corresponding Author:**

15 Wenbin Liu

16 Tianjin Key Laboratory of Animal and Plant Resistance, Tianjin Normal University, Tianjin

17 300387, China E-mail address: skylwb@ tjnu.edu.cn

19 Lina Pan

20 Tianjin Key Laboratory of Animal and Plant Resistance, Tianjin Normal University, Tianjin

21 300387, China E-mail address: skypln@ tjnu.edu.cn

Abstract

30 Chironomidae is the most ecologically diverse insects in aquatic and semi-aquatic habitats.

31 *Propsilocerus akamusi* (Tokunaga) is a dominant and ubiquitous chironomid species in Eastern

32 Asia and its morphologically unique larvae are also considered as indicator organisms to detect

33 water contamination, potential toxicity and waterborne pathogens. Since few studies to date have

34 focused on the olfactory system of *P. akamusi*, our study aims to elucidate the potential functions

35 of chemosensory genes in *P. akamusi*. In our study, we found that although signals released from

36 male groups might attract female swarmers, it was a completely male-dominated mating process.

37 Sequencing the transcriptome of *P. akamusi* on an Illumina HiSeq platform generated 4.42, 4.46

38 and 4.53 Gb of clean reads for heads, legs, and antennae, respectively. 27,609 unigenes, 20,379

39 coding sequences (CDSs), and 8,073 simple sequence repeats were finally obtained. The gene-

40 level differential expression analysis demonstrated variants among three different tissues,

41 including 2,019 genes specifically expressed in heads, 1,540 genes in legs, and 2,071 genes in

42 antennae. Additionally, we identified an assortment of putative olfactory genes consisting of 34

43 odorant binding proteins, 17 odorant receptors, 32 gustatory receptors, 22 ionotropic receptors, 6

44 chemosensory proteins as well as 3 sensory neuron membrane proteins; their relative abundances

45 in the above three tissues were also determined by RT-qPCR. Our finding could allow a more

46 plausible understanding of certain olfaction-mediated behaviors in groups of this

47 macroinvertebrate.

1. Introduction

49 *Propsilocerus akamusi* (Tokunaga) is one of the most ubiquitous chironomid species

50 emerging from numerous eutrophic lakes in Eastern Asia. Acting as prey species for fish and

51 aquatic birds as well as decomposers of plants, this kind of midge is able to connect aquatic and

52 terrestrial food webs (Zheng et al. 2017). Considering the relatively high density and richness of

53 *P. akamusi* in benthic regions, the presence, absence or quantity of their larvae could be a

54 valuable indicator for water quality issues at the organism, population, community, and

55 ecosystem levels (Hirabayashi et al. 2003). The term, Chironomidae, is derived from a Greek

56 word for "pantomimist" due to a typical posture that adult insects tend to have their first pair of

57 legs held forward and upward during the rest time (Meigen 1803). The up- and outstretched

58 forelegs are thought to assemble antennae and probably act functionally as sensory organs

59 (Armitage 1995). Males will normally aggregate themselves into a great swarm tending to form

60 above tree-tops, objects and even persons for attracting female ones, which is an extremely

61 common swarm-based mating system in nature (Sæther 1996). Plentiful investigations of *P.*

62 *akamusi* have been pronounced but most of them primarily focused on its ecology, behavior,

63 karyotype structure, and toxicological response to cadmium stress (Cao et al. 2014; Hirabayashi

64 et al. 2003; Kiknadze et al. 2009; Zheng et al. 2017). However, limited researches to date have

65 been documented regarding the chemosensory system which actually has a critical role in

66 recognizing chemical cues in the surrounding environment. Since the excessive population of *P.*

67 *akamusi* is problematic to other residents living in the shared community, we assume that a better

68 understanding of the chemosensory system could also enable investigators to seek feasible

69 strategies for pest management. The physiological response to certain chemical stimuli is defined

70 as chemoreception and the roles of which in insects are involved in the process of searching food,

71 locating hospitable ovipostion sites, and socializing with other members. A plethora of

72 chemosensory proteins are categorized into receptors and non-receptor proteins, which are

73 believed to account for chemical communication as well as perception (Leal 2013; Zhao et al.

74 2016).

75   Three major groups of the receptor family have been reported, namely odorant receptors

76 (ORs), gustatory receptors (GRs), and ionotropic receptors (IRs). ORs, classified as 7-pass

77 transmembrane proteins, are capable of discriminating distinct odors through its odorant binding

78 sites primarily in sensory neurons. The perception of these chemical stimuli and the consequent

79 intracellular response are associated with multiple behaviors, such as mating, identifying food

80 resources, or alarming conspecies (Venthur & Zhou 2018). Interface between the surrounding

81 environment and insect taste system could be successfully established by GRs in gustatory

82 neurons since this G protein-coupled receptor is able to detect non-volatile information (Venthur

83 & Zhou 2018). Chemosensory IRs, a novel family evolved from iGluR-related proteins, are

84 thought to accumulate in sensory dendrites and have unique odor-evoked patterns, which means

85 some of their potential ligands are quite different from those for ORs (Rytz et al. 2013).

86   In addition to the above-mentioned receptor families, odorant binding proteins (OBPs),

87 chemosensory proteins (CSPs), and sensory neuron membrane proteins (SNMPs) (Jia et al. 2018;

88 Wang et al. 2017a) are also involved in the peripheral olfactory recognition and referred to as

89 non-receptor proteins. OBP family contains a large group of divergent members which are

90 soluble and highly abundant in the sensillum hemolymph, a hydrophilic environment where

91 these non-receptor proteins function as necessary transporters for volatile chemicals and deliver

92 them to activate ORs (Leal 2013). However, some OBP genes were proved to have significantly

93 higher expression level in the antennae and the rest vary across different tissues (Liu et al. 2020;

94 Zhang et al. 2020). The CSPs own their names due to the fact that these soluble polypeptides are

95 widely expressed in chemosensory tissues of insects, including the antennae, maxillary, labial

96 palps, and proboscis. According the previous reports, some of them have similar functions as

97 OBPs in initiating biochemical recognition while others could perform tasks unrelated to

98 chemosensation, such as development and transition (Liu et al. 2014; Qu et al. 2020; Waris et al.

99 2020; Zheng et al. 2020).

100   Next generation sequencing technology is now routinely applied for identifying candidate

101 genes since it offers a more efficient and cost-effective way than conventional homology-cloning

102 methods. Therefore, this powerful platform has enabled genes involved in chemoreception to be

103 discovered in some insect species without the availability of entire genomes, including

104 *Trichogramma chilonis* (Liu et al. 2018), *Adelphocoris suturalis* (Cui et al. 2017), *Bemisia*

105 *tabaci* (Wang et al. 2017b), *Periplaneta Americana* (Chen et al. 2016), *Aphis gossypii* (Dickens

106 et al. 2014), *Spodoptera littoralis* (Gu et al. 2015), and *Dendroctonus ponderosae* (Andersson et

107 al. 2013).

108   With reference to previous researches, we hypothesize that *P. akamusi* behaviors could be
109   more or less manipulated by chemosensory system. Therefore, the genomic information
110   concerning chemoreception and its proper interpretation could offer us valuable insights on this
111   species, which has still been an uncharted territory. Transcriptomic analysis coupled with
112   quantitative real-time PCR were utilized in our study to screen for chemosensory-related genes
113   and their expression patterns, aiming to elucidate the potential functions of these candidate genes.
114
115

116   2. Materials & Methods

117   *2.1. Insect and tissue collection*

118   Males of *P. akamusi* midges were collected from Tianjin, China, between October 2017 and
119   November 2017 and then transported live to the lab. All samples were transferred within a few
120   hours of eclosion. The antennae, heads (excluding antennae), and legs were excised from 1500
121   male specimens under microscope for transcriptome sequencing. The obtained tissues were then
122   immersed in RNAlater (Ambion, AM7020) and stored at –20°C until processing.
123

124   *2.2. Olfactometer bioassay*

125   The emerging *P. akamusi* males tend to congregate into cloud-like swarms which usually
126   appear in the early evening. They seek to mate when females are around or just pass through,
127   thus starting the life cycle. However, we still face deficient knowledge of the sex differences in
128   mate preferences and wonder whether males will take the first step to attract females or not.
129   A mate selection behavioral test was therefore conducted using a horizontal Y-tube
130   olfactometer consisting of a central glass tube with the length of 115 mm as well as two lateral
131   glass arms with the length of 75 mm. The diameter for both the tube and arms was 22 mm. The
132   test was performed in a dark room with a small incandescent lamp working as the only light
133   source over the olfactometer. One lateral arm of the olfactometer was selected as the

134   experimental arm and the other one is the control. Either the female subjects or the male ones

135   were firstly placed on a piece of filter paper (10 ×10 mm) and this paper was then carefully

136   inserted into the midpoint of the experimental arm. Another piece of paper without any subjects

137   was inserted into the equivalent point in the control arm. After the olfactometer was connected to

138   an assembled shunting device, the air sampler was switched on to pump airflow for 30s and then

139   a test individual with the opposite sex was placed at the entrance of the central glass tube. If this

140   test one might prefer to enter into the experimental arm and stayed there for at least 1min, the

141   chosen arm could be considered as the one with proper pheromones for the insects. However, the

142   midge was allowed to acclimate and make a preference within 5 min. Otherwise, we assumed

143   that no response was given by the insect to the ones with the opposite sex and the trial was ended

144   then. Besides, to prevent light from interfering with the taxis of *P. akamusi,* the experimental and

145   control arms were swapped after completing the behavioral assays of three individuals. Fifty

146   insects were tested for one trial with triplicates.

147

148   ***2.3. RNA preparation and cDNA library construction***

149   The antennae, heads, and legs of 100 male insects were subjected to RNA preparation using

150   TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. After

151   the treatment of DNase I, the mRNA molecules were harvested with Oligo (dT) from the total

152   RNA. The long material of mRNA was then shortened into something compatible for further

153   sequencing with the supplementation of Thermomixer buffer. With TransScript First-Strand

154   cDNA Synthesis SuperMix (TransGen, Beijing, China), the fragmented mRNA next worked as

155   the template for single-strand cDNA which was subsequently synthesized to double strand

156   cDNA. Purification, end reparation, adaptor-ligation and size selection were carried out for final

157   PCR amplification. During the QC (quality control) steps, Agilent 2100 Bioanaylzer and ABI

158   StepOnePlus Real-Time PCR System were used for quantification and qualification of the

159   sample library. The library ended up being sequenced using Illumina HiSeq 4000.

160

161   ***2.4. De novo transcriptome assembly***

162   After sequencing, a preprocessing of the raw data was performed and consisted of a series

163   of steps, including the trimming of adapter sequences, the removal of reads containing more than

164   5% ambiguous bases and the elimination of low quality basis. Here the low quality sequences

165   referred to a situation where a single read comprised over 20% of bases with Q score less than 15.

166   The remaining reads generated by these filtering procedures were defined as high quality clean

167   reads and were consequently used for *de novo* assembly by Trinity v3.0 program (Eisen et al.

168   1998) . This transcript assembly was then clustered into unigenes with the assistance of Tgicl

169   (version: v2.0.6) (de Hoon et al. 2004).

170

171   ***2.5. Functional annotation of unigenes as well as prediction of coding sequences and simple***

172   ***sequence repeats***

173   A homology search of functionally annotated unigenes was performed by BLAST (version:

174   v2.2.23, website: http://blast.ncbi.nlm.nih.gov/Blast.cgi) against NCBI non-redundant protein

175   sequences (NR), non-redundant nucleotide sequences (NT), Clusters of Orthologous Groups of

176   Proteins (COG), Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO) and

177   Swiss-Prot databases (Altschul et al. 1990; Conesa et al. 2005). Besides, InterPro analysis was

178   measured by InterProScan5 (version: v5.11-51.0, website:

179   https://code.google.com/p/interproscan/wiki/Introduction) for providing predictive information

180   of gene families and their critical domains (Quevillon et al. 2005). TransDecoder

181   (https://github.com/TransDecoder/TransDecoder/releases) was utilized to recognize candidate

182   coding regions and peptide sequences within the Trinity-created transcript assembly. Unigenes

183   that could not be aligned to any of the databases mentioned above were later predicted by

184   ESTScan (Saldanha 2004).

185

186   ***2.6. Differentially expressed gene analysis***

187   The clean reads were mapped to unigenes using Bowtie2 (Langmead & Salzberg 2012), and

188 the expression levels of these high-quality unigene sets were computed based on RSEM values

189 (Li & Dewey 2011). Poisson distribution was applied to accurately describe gene expression

190 variation on those with more than 2-fold changes as well as less than 0.001 false discovery rate

191 [FDR] (Audic & Claverie 1997). The performance of hierarchical clustering enabled the

192 differentially expressed genes (DEGs) with similar features to be partitioned into distinct clusters

193 and this analysis was done by having the aid of pheatmap function in R. When two or more

194 relatively homogeneous groups were clustered, the intersection and union DEGs between them

195 were performed. The DEGs were classified according to the GO and KEGG annotation results;

196 FDR was calculated for each p-value. In general, the terms with FDR no more than 0.001 were

197 defined as significantly enriched ones.

198

199 *2.7. Identification of chemosensory genes*

200   To identify candidate chemosensory unigenes for *P. akamusi*, an analysis using the

201 tBLASTn modules was performed with reference to all the publicly available sequences of OBP,

202 OR, CSP, GR, IR, and SNMP from Diptera species. Meanwhile, all the candidates were

203 manually checked by using the BLASTx program.

204

205 *2.8. RT-qPCR analysis*

206   Using previously obtained cDNA from the three body parts of *P. akamusi* (the antennae,

207 heads excluding antennae, and legs) as templates, OBPs, ORs, CSPs, GRs, IRs, and SNMPs

208 were selected for RT-qPCR analysis. Specific primer pairs were designed by Primer5 according

209 to the transcriptome data (Table S1). The RT-qPCR was conducted on the Roche LightCycler

210 480 detector (Stratagene, La Jolla, CA, USA) with the following cycling parameters: 94°C for 30

211 s, 40 cycles of 94 °C for 5 s, 55 °C for 10 s, and 72°C for 10 s. The dissociation curves were

212 carried out as a post-qPCR analysis, during which all the components were firstly denatured for

213   5s at 95°C, followed by cooling to 60°C for 1 min. An increase to 95°C for 30 s then took place,

214   followed by cooling to 50 °C for 30 s. The relative gene expression data were analyzed by

215   normalizing the threshold cycle (Ct) value of each sample to that of endogenous β-tubulin, which

216   was determined with $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001). The gene amplification from

217   each tissue part of *P. akamusi* was actually completed in triplicates and the statistical

218   significance was examined with one-way ANOVA test.

219

220   *2.9 Phylogenetic analysis and classification of Chemosensory Genes*

221   A set of chemosensory gene sequences from different Dipteran species were retrieved from

222   the GenBank database, including *Drosophila melanogaster*, *Aedes aegypti*, *Anopheles gambiae*

223   and *Culex quinquefasciatus*. To visually demonstrate the relationship between candidate

224   chemosensory genes from *P. akamusi* and other dipteran species, they were first joined with

225   alignments by Muscle 3.8.31 (Edgar 2004) with default option and then manually refined by

226   BioEdit v7.2.5 (Hall 1999). The unrooted phylogenetic trees of each family were then generated

227   through running MEGA 7.0 (Kumar et al. 2016) and using neighbor joining (NJ) method with

228   poisson correction, pairwise deletion, and rapid bootstraps (1000 replicates).

229

230   ## 3. Results

231   *3.1. Mate selection behavioral test*

232   To examine whether the male individuals will attempt to mate with female ones, a female

233   insect was placed in the experimental corridor of the Y-tube olfactometer while a male was

234   placed at the open end of the central glass tube. The result showed that the majority of males

235   were relatively active with approximately 58% choosing the female-occupied corridor and 38%

236   choosing the empty side (Fig. 1A). However, this obvious response was not observed when

237   females were given the choice to make a behavioral preference. It turned out that 82% female

238   insects were recorded as non-responders to either corridor and the probability of females

239   selectively orientating to the arm with male subjects was only 8% (Fig. 1B). These results

240   revealed that males tended to seek companions and dominated the mating process probably

241   because of the exposure to airborne pheromones released from the females. We therefore

242   selected male individuals as the main subject for our further investigations and tried to explore

243   how exactly the mutual attraction between the sexes occur during mating.

244

245   *3.2. Transcriptome sequencing and de novo assembly*

246   With the support of high-throughput sequencing platform, a set of raw sequencing reads

247   were constructed from three different tissues of *P. akamusi* males, including 50.20 Mb from

248   heads (excluding antennae), 51.82 Mb from antennae and 55.06 Mb from legs (Table 1). The

249   trimming and elimination of adaptor-polluted and low-quality basis led to the generation of

250   4.42Gb, 4.46Gb, and 4.53 Gb clean reads from heads, legs, and antennae, respectively. The data

251   also showed that the assessed intrinsic quality of the clean reads was sufficient for further

252   analysis (Table 2) and the transcript length distribution was identified as well (Fig. S1, S2). In

253   general, 27,609 unigenes were assembled with a total length, mean length, N50, and GC content

254   of 28,312,956, 1,025 bp, 1,832 bp and 38.23%, respectively (Table 3).

255

256   *3.3. Functional annotation and classification of the unigenes*

257   The unigenes were annotated in accordance to seven functional databases and the result of

258   this comparison illustrated that 17,311 unigenes (62.70%) were markedly matched with

259   published proteins in NR database. Similarly, the total of annotated unigenes against known

260   information from NT, Swiss-Prot, COG, KEGG, GO and InterPro was 7,936 (28.74%), 14,614

261   (52.93%), 7,986 (28.93%), 14,341 (51.94%), 7,501 (27.17%) and 14,752(53.43%), respectively

262   (Table 4).

263   The homology searching against NR database revealed that the annotated sequences of *P.*

264   *akamusi* were partly matched to sequences of *Aedes aegypti* (2,769 matching hits, 16.00%),

265   followed by *Aedes albopictus* (1,870 matching hits, 10.80%), *Culex quinquefasciatus* (1,741

266   matching hits, 10.06%) and *Anopheles gambiae* str. (1,311 matching hits, 7.57%) (Fig. 2).

267   GO analysis, an enrichment tool, rendered 42,572 unigenes to be grouped into three

268   functional categories, i.e. Biological Process (18,955 unigenes), Cellular Component (15,232

269   unigenes), and Molecular Function (8,385 unigenes). There were 25 subcategories in Biological

270   Process with 'cellular process' accounting for 22.2% (4,213 unigenes) followed by 'metabolic

271   process' making up 18% (3,411 unigenes). 19 sub-categories were included in the classification

272   of Cellular Component and among which 'groups of cell' (3,145 unigenes, 20.6%) and 'cell

273   parts' (3,121 unigenes, 20.5%) were the most abundant GO terms. For Molecular Function,

274   sequences were predominately assigned to 'catalytic activity' (3,309 unigenes, 39.5%) and

275   'binding' (3, 267 unigenes, 39%) (Fig. 3A).

276   A more detailed comprehension of gene biochemical function could be further gained via

277   KEGG analysis. The result delineated that a total of 25,377 unigenes were mapped and organized

278   to six functional clusters, including Cellular Process (2,476 unigenes), Environmental

279   Information Processing (2,613 unigenes), Genetic Information Processing (2,531 unigenes),

280   Human Disease (6,293 unigenes), Metabolism (6,405 unigenes), and Organismal System (5,059

281   unigenes) (Fig. 3B).

282   Besides, potential functions of these putative unigenes were predicted with the assistance of

283   COG database which was considered as a useful tool for understanding the orthologous

284   relationships of gene products. The retrieved unigene sets were classified under 25 categories,

285   among which the cluster of 'General function prediction only' (2,338 unigenes, 19%),

286   'Translation, ribosomal structure and biogenesis' (979 unigenes, 8%) and 'Posttranslational

287   modification, protein turnover, chaperones' (941 unigenes, 7.7%) were the most representative

288   three classifications (Fig. 3C). Venn diagram aimed at describing the similarities and differences

289   of unigenes when searched against NR, COG, KEGG, Swissprot and Interpro databases and it

290   turned out that 6,766 unigenes overlapped as the intersection proportion (Fig. 3D).

291

292 *3.4. CDS prediction*

293   17,522 CDSs were screened out from the 27,609 annotated unigenes. Since ESTScan

294 program was capable of constructing CDSs in the remaining unigenes which were not of

295 particular matches with the aforementioned databases, a total of 2,857 CDSs were predicted from

296 these unannotated unigenes. Altogether, 20,379 CDSs were predicted from unigenes with a total

297 length, a mean length, N50 and GC content of 16,667,634 bp, 817 bp, 1,290 bp and 42 %,

298 respectively (Fig. 4, Table 5).

299

300 *3.5. Differentially expressed genes*

301   Differentially expressed genes were examined by comparison among heads, legs, and

302 antennae of male *P. akamusi* midges (Fig. 5A). The gene expression profiling analysis

303 demonstrated that there were 2,281 up-regulated along with 4,865 down-regulated genes in heads

304 against legs. A total of 3,421 were expressed at a markedly higher level while 5,792 genes were

305 at lower levels in antennae vs. heads. Plus, 2,066 genes were expressed with greater numbers

306 while 7,093 genes with relatively smaller numbers in antennae vs. legs. Shared and exclusively

307 expressed genes among three tissues were shown in Fig. 5B. The number of genes differentially

308 expressed among three tissue types was quite similar to each other with 21,103 in heads, 20,980

309 in legs, and 21,750 in antennae. Meanwhile, a subset of 5,620 genes that differentially expressed

310 were identified with a tissue-dependent manner, including 2019 specifically expressed in head,

311 1540 genes expressed in legs and 2071 expressed in antennae.

312

313 *3.6. Candidate odorant binding proteins*

314   Based on functional annotation and tBLASTn results with an E-value of 1E-5 or lower

315 (Table S2), 34 transcripts ranging from 138 to 927 bp were isolated as best candidate OBPs

316 (PakaOBPs) in the *P. akamusi* transcriptome, and 29 of which contained full-length open reading

317   frames (ORFs). The identified PakaOBP transcripts together with corresponding sequence data

318   of OBPs from *A. gambiae*, *C. quinquefasciatus*, *D. melanogaster* and *A. aegypti* were used for

319   obtaining phylogenetic inferences, as depicted in Fig. S3.

320   To determine the transcriptional output of these candidate genes among different tissues, 12

321   PakaOBPs with RPKM>1.2 were selected for RT-qPCR amplification and the expression profile

322   of each gene differed. Five OBPs (PakaOBP1, 5, 8, 9 and 10) had noticeably higher proportion in

323   antennae than either heads or legs. Meanwhile, seven other OBPs (PakaOBP1, 2, 3, 4, 6, 7, and

324   12) were expressed with a markedly greater numbers in legs. Of note, we found a unique

325   antennae-specific expression pattern of PakaOBP9 since this particular gene showed almost no

326   evidence of presence in other tissues (Fig. 6).

327

328   ***3.7. Candidate odorant receptors***

329   Putative OR genes of *P. akamusi* (PakaORs) were represented based on their similarities to

330   known insect ORs and tBLASTn results with an E-value of 1E-5 or lower generated a total of 17

331   PakaORs with length ranging from 312 to 1671 bp (Table S3). Among them, 16 sequences were

332   available as full-length coding ones. Evolutionary distances were evaluated among ORs from the

333   sampled *P. akamusi* and four other Dipteran species, as shown in Fig. S4.

334   RT-qPCR was performed for 6 PakaORs that were relatively abundant in the antennal

335   transcripts with RPKM>1.2. All of these PakaOR genes except for PakaOR4 were highly

336   expressed in antennae compared with heads and legs whereas PakaOR3–6 had a relatively

337   sufficient expression only in heads (Fig. 7). The results of PakaOR expression appeared to be

338   partially consistent with those from the RPKM analysis.

339

340   ***3.8. Candidate gustatory receptors***

341   A total of 32 candidate GR genes (PakaGRs) were identified based on the tBLASTn results

342   with an E-value of 1E-5 or lower, with 20 sequences containing a full-length ORF (Table S4). A

343 phylogenetic tree was constructed by using concatenated orthologous sequences derived from

344 GRs in different Dipteran species (Fig. S5).

345   The relatively abundant transcripts (PakaGR1–10) according to the RPKM analysis were

346 subjected to RT-qPCR detection, aiming to figure out the particular tissues where they preferred

347 to be expressed. It turned out that PakaGR1 and PakaGR2 were mainly expressed in the antennae

348 whereas PakaGR3 had a significant quantity in legs. Meanwhile, a pervasive expression pattern

349 was shown for 10 other PakaGRs, including PakaGR2 and PakaGR3–10 (Fig. 8).

350

351 *3.9. Candidates for other chemosensory genes*

352   Our bioinformatic data facilitated the discovery of 22 transcripts for candidate IR genes

353 (PakaIRs), 6 for candidate CSP genes (PakaCSPs), and 3 for candidate SNMP genes

354 (PakaSNMPs) with the result of tBLASTn being E-value of 1E-5 or lower; of these genes, 19

355 PakaIRs, 2 PakaCSPs, and 3 PakaSNMPs included full length ORFs (Table S5 and S6). The

356 phylogenetic signals of PakaIRs, PakaCSPs, and PakaSNMPs from *P. akamusi* with those from

357 other four species were mapped onto the tree, as described in Fig. S8.

358   RT-qPCR data revealed that 16 PakaIRs were highly expressed in heads and legs (Fig. 9)

359 while CSP1 was distributively expressed at varying levels in all tissues. Additionally, there was

360 an abundance of PakaSNMP1 presented in antennae and legs at transcript level, which was a

361 reliable clue for understanding its role in chemosensory process.

362

4. Discussion

363

364   The Chironomidae chemoreception is barely explored at either genetic or molecular levels,

365 especially in comparison to some other Dipterans, such as Culicidae and Drosophila, which are

366 commonly selected as subjects in researches of chemosensation.

367   Next generation sequencing technology was utilized in our study to yield a diverse array of

368 candidate chemoreception-related genes of *P. akamusi*, a dominant Chironomidae species in

369   freshwater environment. The sequencing data presented six categories of putative chemosensory

370   genes, including 34 PakaOBPs, 17 PakaORs, 32 PakaGRs, 22 PakaIRs, 6 PakaCSPs, and 3

371   PakaSNMPs (Fig. 6-10), which was the first known attempt for the identification of

372   chemosensory genes in *P. akamusi*. However, it turned out that the number of these identified

373   genes was quite tiny when compared to that of genes from other species. For instance, 82 ORs,

374   77OBPs, 30GRs, and 102IRs have previously been collected from *Aedes albopictus*. Strictly

375   speaking, we have to admit that most of our transcripts did not comprise complete ORFs.

376   The genome sequencing in our work facilitated an identification of 34 PakaOBPs, which

377   was fewer than the numbers reported from other species, including *Bombyxmori* (44) (Gong et al.

378   2009), *A. gambiae* (57) (Biessmann et al. 2002; Vogt 2002; Xu et al. 2003), *D. melanogaster* (51)

379   (Hekmat-Scafe 2002), and *Agrotis ipsilon* (33) (Gu et al. 2013). We assume that the less

380   quantitative sufficiency of OBPs in various tissues of *P. akamusi* accompanied with barely

381   detectable levels could be one possible explanation. Besides, the lack of some host-seeking cues

382   in lab could be another factor for this limited number since OBPs will be normally stimulated

383   with the exposure to proper odors in the nature. The tissue-specific patterns of candidate OBPs

384   could provide essential clues for gene function. For example, we observed that PakaOBP9 was

385   specifically expressed in the antennae whereas most PakaOBPs repertoire did not apparently

386   display an antenna-biased expression profile (Fig. 6), suggesting a likely association with certain

387   chemosensory reactions. Besides, PakaOBP1 was abundantly expressed in antennae whereas

388   PakaOBP5 was highly expressed in heads and antennae (Fig. 6). One previous study has reported

389   that OBP21 was highly expressed in a non-olfactory system, like the venom gland of *Apis*

390   *mellifera*, and hypothesized that this molecule might serve as a carrier of potential ligands other

391   than odorants (Li et al. 2013; Pelosi et al. 2018). However, the expression profile of PakaOBPs

392   in a non-olfactory part still remains vague and requires a further exploration to better appreciate

393   the function of this gene.

394   The ORs, belonging to the superfamily of G-protein coupled receptors, are critical

395   recognition elements involved in olfactory sensory system. We have analyzed the whole genome

396   to prospect candidate ORs and finally recognized as many as 11 transcripts in *P. akamusi*, which

397   is actually far fewer than 170 genes found in *Apis mellifera* (Robertson & Wanner 2006) and 60

398   in the parasitoids *Microplitis mediator* (Wang et al. 2015) but greater than 6 discovered in

399   *Cotesia vestalis* (Matsunami et al. 2012b). Most of newly PakaORs transcripts showed a

400   significant enrichment in the antennae whereas PakaOR3–6 were predominantly detected in

401   tissues like heads. The expression pattern of ORs in non-olfactory tissues suggests they might

402   have some undefined physiological functions.

403   GRs, regarded as critical chemoreceptors, usually endow gustatory receptor neurons with

404   the capability to perceive soluble tastes and respond to carbon dioxide. This kind of peripheral

405   receptor comprises more conserved sequences and structures among diverse species when

406   compared to ORs. The *de novo* assembly in our work allowed a total of 32 PakaGRs to be

407   captured, the same number as identified in *A. Albopictus*. Despite the typical antenna-specific

408   expression pattern of GRs in most insect species, our discovery revealed that most PakaGRs

409   were significantly enriched in legs (PakaGR2, 3, 5, 6, 7, 9, and 10) or heads (PakaGR5 and 9).

410   Since *P. akamusi* posturing with forelegs is a common phenomenon in nature, we wonder

411   whether the high abundance of ORs in legs might have potential roles in chemosensory-related

412   perception.

413   IRs are novel families of highly divergent ionotropic glutamate receptors and broader

414   attentions with respect to this group have been constantly received. 66 have been reported in *D.*

415   *melanogaster* (Benton et al. 2009), 21 in *Manduca sexta* (Koenig et al. 2015), 15 in *Cydia*

416   *pomonella* (Matsunami et al. 2012a), and 12 in *Helicoverpa armigera* (Dickens et al. 2015).

417   Although IRs have been reported to exist in various insect genomes or transcriptomes, their

418   function has only been hinted in studies focusing on *Drosophila*. There is a considerable

419   discrepancy between the environmental stimuli recognized by ORs and IRs. In a combinatorial

420   fashion, IRs could provide a strong response to a broad range of odors which, however, only

421   induce weak or even no answer of ORs. We have identified a set of 22 IRs in the genome of *P.*

422   *akamusi* compared to transcriptomes of *D. melanogaster* with 66 IRs (Benton et al. 2009),

423 *Manduca sexta* with 21 IRs (Koenig et al. 2015), *Cydia pomonella* with 15 IRs (Matsunami et al.

424 2012), and *Helicoverpa armigera* with 12 IRs (Dickens et al. 2015). A comprehensive map of

425 their expressions was generated with reference to RPKM values and we found that PakaIR3, 4, 7,

426 8, 9, 10, 12, and 13 showed high transcript levels in heads and legs while none in antennae. Plus,

427 the expression levels of these PakaIRs are higher than those of PakaOBPs and PakaCSPs,

428 indicating possible protein functions in chemosensory processes in these non-olfactory parts.

429

430 **5. Conclusions**

431   In our study, the transcriptome of *P. akamusi* was sequenced by using Illumina Hiseq

432 platform by which 27,609 unigenes, 20,379 CDSs, and 8,073 simple sequence repeats were

433 obtained. The differentially expressed gene analysis showed that there were 2,019 head-specific

434 genes, 1,540 leg-specific genes, and 2,071 antennae-specific genes. Furthermore, candidate

435 olfactory-related genes were identified and their relative abundances in the above tissues were

436 examined by RT-qPCR as well. In general, 34 odorant binding proteins, 17 odorant receptors, 32

437 gustatory receptors, 22 ionotropic receptors, 6 chemosensory proteins as well as 3 sensory

438 neuron membrane proteins were collected.

439

440 **Acknowledgements**

441 This project was supported by the National Natural Science Foundation of China (No. 31702058,

442 31672324, 31801994); Natural Science Foundation of Tianjin (No. 18JCYBJC96300,

443 17JCQNJC14900, 14JCQNJC14600, 18JCQNJC14700) and Tianjin Normal University

444 Foundation (5RL104, 043135202-XB1715, 52XB1003, 52XB1005, 043135202-XK1706,

445 135305JF79).

446

447 **Data availability**

448 All unigenes of *P. akamusi* used in this study were available in the supplementary file (All-

449 Unigene. fa) and GenBank database. The list and accession numbers of candidate olfactory-

450 related genes of *P. akamusi* were listed in the supplementary Table (Table S2-S6).

451

452

453 **References**

454 Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. 1990. Basic local alignment search
455   tool. *Journal of Molecular Biology* 215:403-410. 10.1016/s0022-2836(05)80360-2
456 Andersson MN, Grosse-Wilde E, Keeling CI, Bengtsson JM, Yuen MMS, Li M, Hillbur Y,
457   Bohlmann J, Hansson BS, and Schlyter F. 2013. Antennal transcriptome analysis of the
458   chemosensory gene families in the tree killing bark beetles, Ips typographus and
459   Dendroctonus ponderosae (Coleoptera: Curculionidae: Scolytinae). *BMC Genomics*
460   14:198. 10.1186/1471-2164-14-198
461 Armitage PD. 1995. Behaviour and ecology of adults. In: Armitage PD, Cranston, P. S., Pinder,
462   L. C. V., editor. The Chironomidae Biology and ecology of non-biting midges. London:
463   Chapman and Hall. p 194-224.
464 Audic S, and Claverie J-M. 1997. The Significance of Digital Gene Expression Profiles. *Genome*
465   *Research* 7:986-995. 10.1101/gr.7.10.986
466 Benton R, Vannice KS, Gomez-Diaz C, and Vosshall LB. 2009. Variant Ionotropic Glutamate
467   Receptors   as   Chemosensory   Receptors   in   Drosophila.   *Cell*   136:149-162.
468   10.1016/j.cell.2008.12.001
469 Biessmann H, Walter MF, Dimitratos S, and Woods D. 2002. Isolation of cDNA clones
470   encoding putative odourant binding proteins from the antennae of the malaria-
471   transmitting mosquito, Anopheles gambiae. *Insect Molecular Biology* 11:123-132.
472   10.1046/j.1365-2583.2002.00316.x
473 Cao CW, Niu F, Li XP, Ge SL, and Wang ZY. 2014. Acute and joint toxicity of twelve
474   substituted benzene compounds to Propsilocerus akamusi Tokunaga. *Central European*
475   *Journal of Biology* 9:550-558.
476 Chen Y, He M, Li Z-Q, Zhang Y-N, and He P. 2016. Identification and tissue expression profile
477   of genes from three chemoreceptor families in an urban pest, Periplaneta americana.
478   *Scientific Reports* 6. 10.1038/srep27495
479 Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, and Robles M. 2005. Blast2GO: a
480   universal tool for annotation, visualization and analysis in functional genomics research.
481   *Bioinformatics* 21:3674-3676. 10.1093/bioinformatics/bti610
482 Cui H-H, Gu S-H, Zhu X-Q, Wei Y, Liu H-W, Khalid HD, Guo Y-Y, and Zhang Y-J. 2017.
483   Odorant-binding and chemosensory proteins identified in the antennal transcriptome of
484   Adelphocoris suturalis Jakovlev. *Comparative Biochemistry and Physiology Part D:*
485   *Genomics and Proteomics* 24:139-145. 10.1016/j.cbd.2016.03.001
486 de Hoon MJ, Imoto S, Nolan J, and Miyano S. 2004. Open source clustering software.

487   *Bioinformatics* 20:1453-1454. 10.1093/bioinformatics/bth078 [pii]

488   Dickens JC, Cao D, Liu Y, Walker WB, Li J, and Wang G. 2014. Molecular Characterization of
489   the Aphis gossypii Olfactory Receptor Gene Families. *Plos One* 9:e101187.
490   10.1371/journal.pone.0101187

491   Dickens JC, Gu X-C, Zhang Y-N, Kang K, Dong S-L, and Zhang L-W. 2015. Antennal
492   Transcriptome Analysis of Odorant Reception Genes in the Red Turpentine Beetle (RTB),
493   Dendroctonus valens. *Plos One* 10:e0125159. 10.1371/journal.pone.0125159

494   Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space
495   complexity. *BMC Bioinformatics* 5:113. 10.1186/1471-2105-5-113 [pii]

496   Eisen MB, Spellman PT, Brown PO, and Botstein D. 1998. Cluster analysis and display of
497   genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95:14863-14868.

498   Gong D-P, Zhang H-J, Zhao P, Xia Q-Y, and Xiang Z-H. 2009. The Odorant Binding Protein
499   Gene Family from the Genome of Silkworm, Bombyx mori. *BMC Genomics* 10:332.
500   10.1186/1471-2164-10-332

501   Gu S-H, Zhou J-J, Gao S, Wang D-H, Li X-C, Guo Y-Y, and Zhang Y-J. 2015. Identification
502   and comparative expression analysis of odorant binding protein genes in the tobacco
503   cutworm Spodoptera litura. *Scientific Reports* 5. 10.1038/srep13800

504   Gu S-H, Zhou J-J, Wang G-R, Zhang Y-J, and Guo Y-Y. 2013. Sex pheromone recognition and
505   immunolocalization of three pheromone binding proteins in the black cutworm moth
506   Agrotis ipsilon. *Insect Biochemistry and Molecular Biology* 43:237-251.
507   10.1016/j.ibmb.2012.12.009

508   Hall T. 1999. BioEdit: A user-friendly biological sequence alignment program for Windows
509   95/98/NT. *Nucleic Acids Symposium Series* 41:95-98.

510   Hekmat-Scafe DS. 2002. Genome-Wide Analysis of the Odorant-Binding Protein Gene Family
511   in Drosophila melanogaster. *Genome Research* 12:1357-1369. 10.1101/gr.239402

512   Hirabayashi K, Hanazato T, Ogawara M, Sakuma M, and Nakamoto N. 2003. Long-term
513   investigation of Propsilocerus akamusi (Tokunaga) (Diptera, Chironomidae) from a
514   shallow eutrophic lake, Suwa, in Central Japan : An attempt to forecast the massive
515   emergence of adult midges. *Medical Entomology & Zoology* 54.

516   Jia X, Zhang X, Liu H, Wang R, and Zhang T. 2018. Identification of chemosensory genes from
517   the antennal transcriptome of Indian meal moth Plodia interpunctella. *Plos One*
518   13:e0189889. 10.1371/journal.pone.0189889

519   Kiknadze II, Wang XH, and Istomina AG. 2009. Karyotype of Propsilocerus akamusi (Tokunaga)
520   from China (Diptera : Chironomidae). *Zootaxa* 765:1-8.

521   Koenig C, Hirsh A, Bucks S, Klinner C, Vogel H, Shukla A, Mansfield JH, Morton B, Hansson
522   BS, and Grosse-Wilde E. 2015. A reference gene set for chemosensory receptor genes of
523   Manduca sexta. *Insect Biochemistry and Molecular Biology* 66:51-63.
524   10.1016/j.ibmb.2015.09.007

525   Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis
526   Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33:1870-1874.
527   10.1093/molbev/msw054

528 Langmead B, and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature*
529   *Methods* 9:357-359. 10.1038/nmeth.1923
530 Leal WS. 2013. Odorant reception in insects: roles of receptors, binding proteins, and degrading
531   enzymes. *Annu Rev Entomol* 58:373-391. 10.1146/annurev-ento-120811-153635
532 Li B, and Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with
533   or without a reference genome. *BMC Bioinformatics* 12:323. 10.1186/1471-2105-12-323
534 Li R, Zhang L, Fang Y, Han B, Lu X, Zhou T, Feng M, and Li J. 2013. Proteome and
535   phosphoproteome analysis of honeybee (Apis mellifera) venom collected from electrical
536   stimulation and manual extraction of the venom gland. *BMC Genomics* 14:766.
537   10.1186/1471-2164-14-766
538 Liu J-B, Wu H, Yi J-Q, Song Z-W, Li D-S, and Zhang G-R. 2018. Transcriptome
539   characterization and gene expression analysis related to chemoreception in
540   Trichogramma chilonis , an egg parasitoid. *Gene*. 10.1016/j.gene.2018.07.065
541 Liu YL, Guo H, Huang LQ, Pelosi P, and Wang CZ. 2014. Unique function of a chemosensory
542   protein in the proboscis of two Helicoverpa species. *J Exp Biol* 217:1821-1826.
543   10.1242/jeb.102020 jeb.102020 [pii]
544 Liu Z, Liang X-F, Xu L, Keesey IW, Lei Z-R, Smagghe G, and Wang J-J. 2020. An Antennae-
545   Specific Odorant-Binding Protein Is Involved in Bactrocera dorsalis Olfaction. *Frontiers*
546   *in Ecology and Evolution* 8. 10.3389/fevo.2020.00063
547 Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time
548   quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25:402-408.
549   10.1006/meth.2001.1262 S1046-2023(01)91262-9 [pii]
550 Matsunami H, Bengtsson JM, Trona F, Montagné N, Anfora G, Ignell R, Witzgall P, and
551   Jacquin-Joly E. 2012a. Putative Chemosensory Receptors of the Codling Moth, Cydia
552   pomonella, Identified by Antennal Transcriptome Analysis. *Plos One* 7:e31620.
553   10.1371/journal.pone.0031620
554 Matsunami H, Nishimura O, Brillada C, Yazawa S, Maffei ME, and Arimura G-i. 2012b.
555   Transcriptome Pyrosequencing of the Parasitoid Wasp Cotesia vestalis: Genes Involved
556   in   the   Antennal   Odorant-Sensory   System.   *Plos   One*   7:e50664.
557   10.1371/journal.pone.0050664
558 Meigen JW. 1803. Versuch einer neuen GattungsEintheilung [sic!] der europäischen
559   zweiflügligen Insekten. *Magazin für Insektenkunde Brunswick* 2:259-281.
560 Pelosi P, Iovinella I, Zhu J, Wang G, and Dani FR. 2018. Beyond chemoreception: diverse tasks
561   of   soluble   olfactory   proteins   in   insects.   *Biological   Reviews*   93:184-200.
562   10.1111/brv.12339
563 Qu MQ, Cui Y, Zou Y, Wu ZZ, and Lin JT. 2020. Identification and expression analysis of
564   odorant binding proteins and chemosensory proteins from dissected antennae and
565   mouthparts of the rice bug Leptocorisa acuta. *Comp Biochem Physiol Part D Genomics*
566   *Proteomics* 33:100631. S1744-117X(19)30160-1 [pii] 10.1016/j.cbd.2019.100631
567 Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, and Lopez R. 2005.
568   InterProScan: protein domains identifier. *Nucleic Acids Research* 33:W116-W120.

569   10.1093/nar/gki442

570   Robertson HM, and Wanner KW. 2006. The chemoreceptor superfamily in the honey bee, Apis
571   mellifera: Expansion of the odorant, but not gustatory, receptor family. *Genome Research*
572   16:1395-1403. 10.1101/gr.5057506

573   Rytz R, Croset V, and Benton R. 2013. Ionotropic Receptors (IRs): Chemosensory ionotropic
574   glutamate receptors in Drosophila and beyond. *Insect Biochemistry and Molecular*
575   *Biology* 43:888-897. 10.1016/j.ibmb.2013.02.007

576   Sæther OA, Wang, X. 1996 Revision of the orthoclad genus Propsilocerus Kieffer (=
577   Tokunagayusurika Sasa) (Diptera: Chironomidae) *Ent Scand* 27:441-479.

578   Saldanha AJ. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics*
579   20:3246-3248. 10.1093/bioinformatics/bth349

580   Venthur H, and Zhou J-J. 2018. Odorant Receptors and Odorant-Binding Proteins as Insect Pest
581   Control Targets: A Comparative Analysis. *Frontiers in Physiology* 9.
582   10.3389/fphys.2018.01163

583   Vogt RG. 2002. *Journal of Chemical Ecology* 28:2371-2376. 10.1023/a:1021009311977

584   Wang J, Hu P, Gao P, Tao J, and Luo Y. 2017a. Antennal transcriptome analysis and expression
585   profiles of olfactory genes in Anoplophora chinensis. *Scientific Reports* 7.
586   10.1038/s41598-017-15425-2

587   Wang R, Li F, Zhang W, Zhang X, Qu C, Tetreau G, Sun L, Luo C, and Zhou J. 2017b.
588   Identification and expression profile analysis of odorant binding protein and
589   chemosensory protein genes in Bemisia tabaci MED by head transcriptome. *Plos One*
590   12:e0171739. 10.1371/journal.pone.0171739 PONE-D-16-49859 [pii]

591   Wang S-N, Peng Y, Lu Z-Y, Dhiloo KH, Gu S-H, Li R-J, Zhou J-J, Zhang Y-J, and Guo Y-Y.
592   2015. Identification and Expression Analysis of Putative Chemosensory Receptor Genes
593   in Microplitis mediator by Antennal Transcriptome Screening. *International Journal of*
594   *Biological Sciences* 11:737-751. 10.7150/ijbs.11786

595   Waris MI, Younas A, Ameen A, Rasool F, and Wang MQ. 2020. Expression Profiles and
596   Biochemical Analysis of Chemosensory Protein 3 from Nilaparvata lugens (Hemiptera:
597   Delphacidae). *J Chem Ecol*. 10.1007/s10886-020-01166-6 [pii]

598   Xu PX, Zwiebel LJ, and Smith DP. 2003. Identification of a distinct family of genes encoding
599   atypical odorant-binding proteins in the malaria vector mosquito, Anopheles gambiae.
600   *Insect Molecular Biology* 12:549-560. 10.1046/j.1365-2583.2003.00440.x

601   Zhang F, Merchant A, Zhao Z, Zhang Y, Zhang J, Zhang Q, Wang Q, Zhou X, and Li X. 2020.
602   Characterization of MaltOBP1, a Minus-C Odorant-Binding Protein, From the Japanese
603   Pine Sawyer Beetle, Monochamus alternatus Hope (Coleoptera: Cerambycidae).
604   *Frontiers in Physiology* 11. 10.3389/fphys.2020.00212

605   Zhao Y, Wang F, Zhang X, Zhang S, Guo S, Zhu G, Liu Q, and Li M. 2016. Transcriptome and
606   Expression Patterns of Chemosensory Genes in Antennae of the Parasitoid Wasp
607   Chouioia cunea. *Plos One* 11:e0148159. 10.1371/journal.pone.0148159

608   Zheng R, Xia Y, and Keyhani NO. 2020. Sex-specific variation in the antennal proteome of the
609   migratory locust. *J Proteomics* 216:103681. S1874-3919(20)30049-X

610   [pii]10.1016/j.jprot.2020.103681

611 Zheng X, Xie Z, Wang S, and Lin P. 2017. Determination of the protein expression profiles of
612   Propsilocerus akamusi (Tokunaga) Malpighian tubules response to cadmium stress by
613   iTRAQ coupled LC-MS/MS. *Journal of Proteomics* 164:85-93.

614

615

616

Table 1. Summary of sequencing reads after filtering.

1 Table 1. Summary of sequencing reads after filtering.

Sample	Total Raw Reads(Mb)	Total Clean Reads(Mb)	Total Clean Bases(Gb)	Clean Reads Q20(%)	Clean Reads Q30(%)	Clean Reads Ratio(%)
A	50.2	44.17	4.42	98.09	94.38	87.99
B	55.06	44.55	4.46	97.78	93.6	80.92
C2	51.82	45.28	4.53	98.12	94.35	87.39

Table 2. Quality metrics of transcripts

2 Table 2. Quality metrics of transcripts

Sample	Total Number	Total Length	Mean Length	N50	N70	N90	GC(%)
A	28,143	24,421,248	867	1,576	900	333	36.96
B	27,460	19,032,332	693	1,231	637	263	38.99
C2	25,567	22,046,525	862	1,576	957	326	38.23

3 N50: a weighted median statistic that 50% of the Total Length is contained in transcripts great

4 than or equal to this value. GC (%): the percentage of G and C bases in all transcripts.

Table 3. Quality metrics of Unigenes

1   Table 3. Quality metrics of Unigenes

Sample	Total Number	Total Length	Mean Length	N50	N70	N90	GC(%)
A	21,857	21,364,850	977	1673	1015	389	37.12
B	19,975	16,078,029	804	1375	765	313	38.62
C2	19,355	19,187,584	991	1667	1103	401	38.21
All-Unigene	27,609	28,312,956	1025	1832	1146	396	38.23

2   N50: a weighted median statistic that 50% of the TotalLength is contained in Unigenes great

3   than or equal to this value. GC (%): the percentage of G and C bases in all Unigenes.

Table 4. Summary of functional annotation result.

1   Table 4. Summary of functional annotation result.

Values	Number	Percentage
Total	27609	100.00%
NR	17311	62.70%
NT	7936	28.74%
Swissprot	14614	52.93%
KEGG	14341	51.94%
COG	7986	28.93%
Interpro	14752	53.43%
GO	7501	27.17%
Overall	18544	67.17%

2   Overall: the number of Unigenes which be annotated with at least one functional database.

Table 5. Quality metrics of predicted CDS

1 Table 5. Quality metrics of predicted CDS

Software	Total Number	Total Length	Mean Length	N50	N70	N90	GC(%)
Blast	17,522	15,693,411	895	1,350	912	417	41.83
ESTScan	2,857	974,223	340	339	261	213	44.69
Overall	20,379	16,667,634	817	1,290	834	348	42.00

2 N50: a weighted median statistic that 50% of the TotalLength is contained in CDS great than or

3 equal to this value. GC (%): the percentage of G and C bases in all CDS.

Figure 1

Fig. 1 The movement situation (mean ± SD) of *P. akamusi.*

(A) The movement situation (mean ± SD) of male *P. akamusi* when attracted to female. (B) The movement situation (mean ± SD) of female *P. akamusi* when attracted to male.
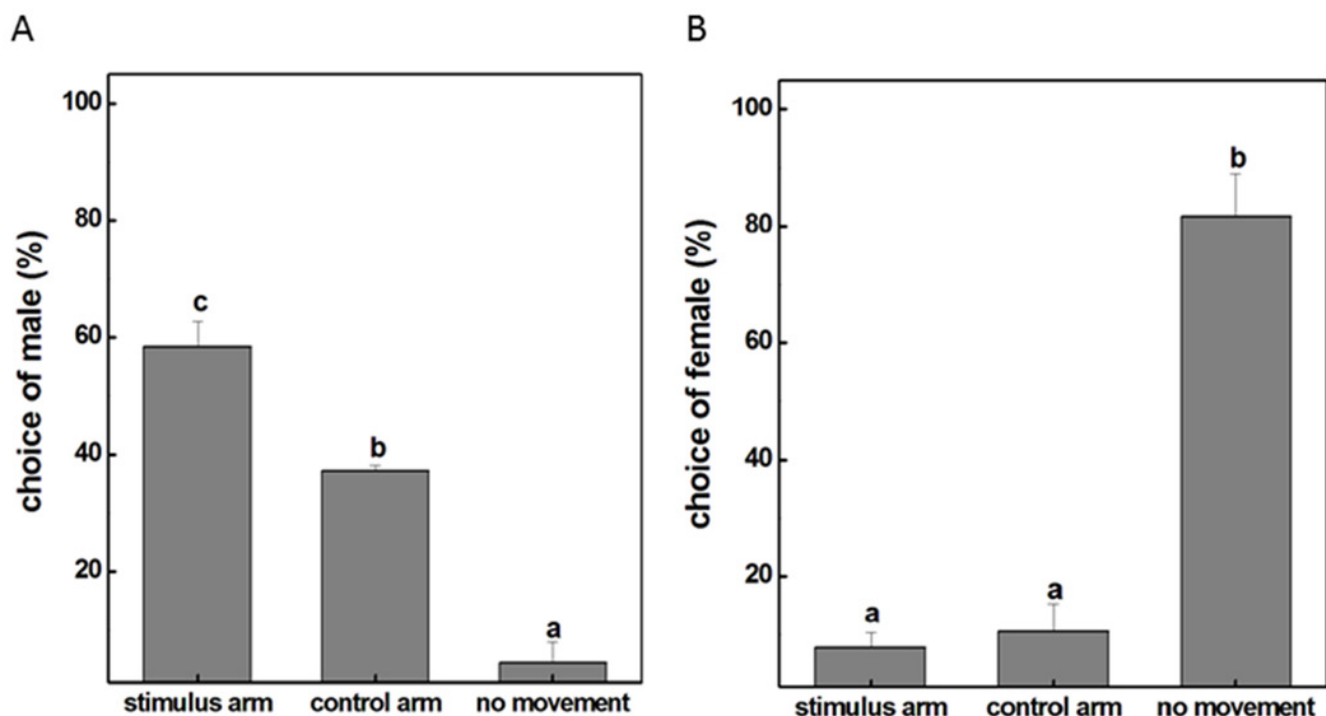
Figure 2

Fig. 2 The homology searching against NR database.
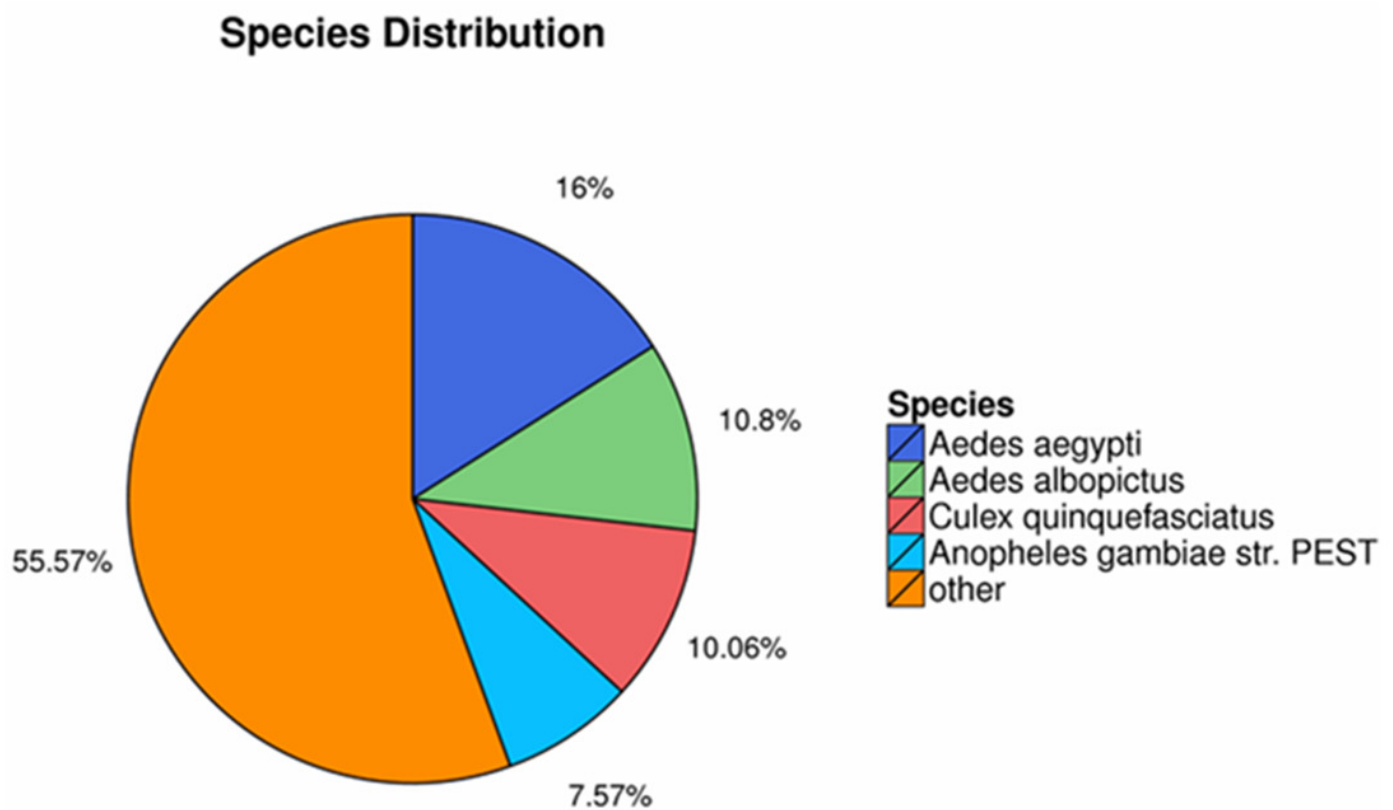
Figure 3

Fig. 3 Functional annotation for Unigenes of *P. akamusi.*

(A) Functional distribution of GO annotation. X axis represents the number of Unigenes. Y axis represents the Gene Ontology functional category. (B) Functional distribution of KEGG annotation. X axis represents the number of Unigenes. Y axis represents the KEGG functional category. (C) Functional distribution of COG annotation. X axis represents the number of Unigenes. Y axis represents the COG functional category. (D)Venn diagram between NR, COG, KEGG, Swissprot and Interpro.
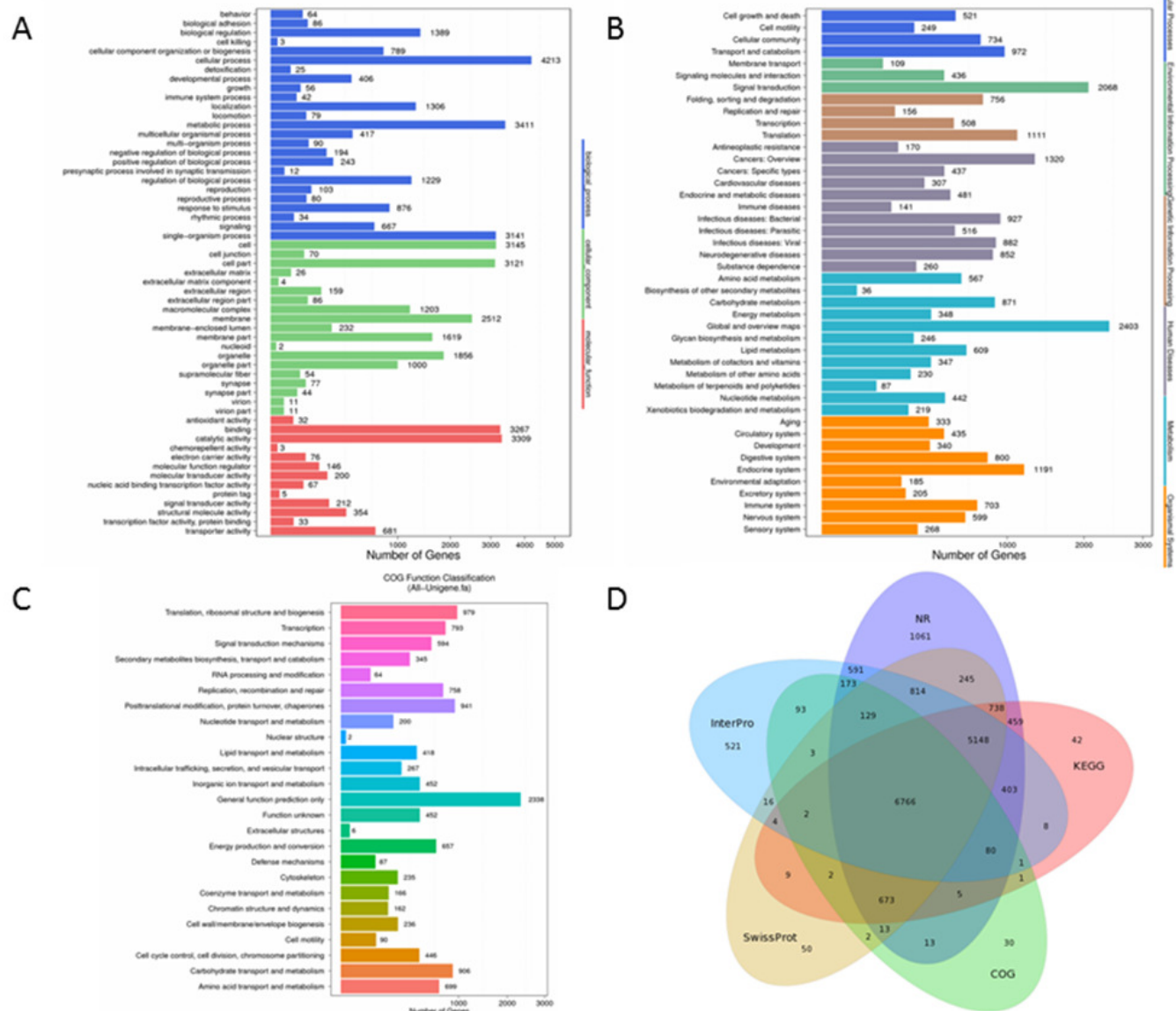
Figure 4

Fig. 4 CDS length distribution of *P. akamusi*.

X axis represents the length of CDS. Y axis represents the number of CDS.
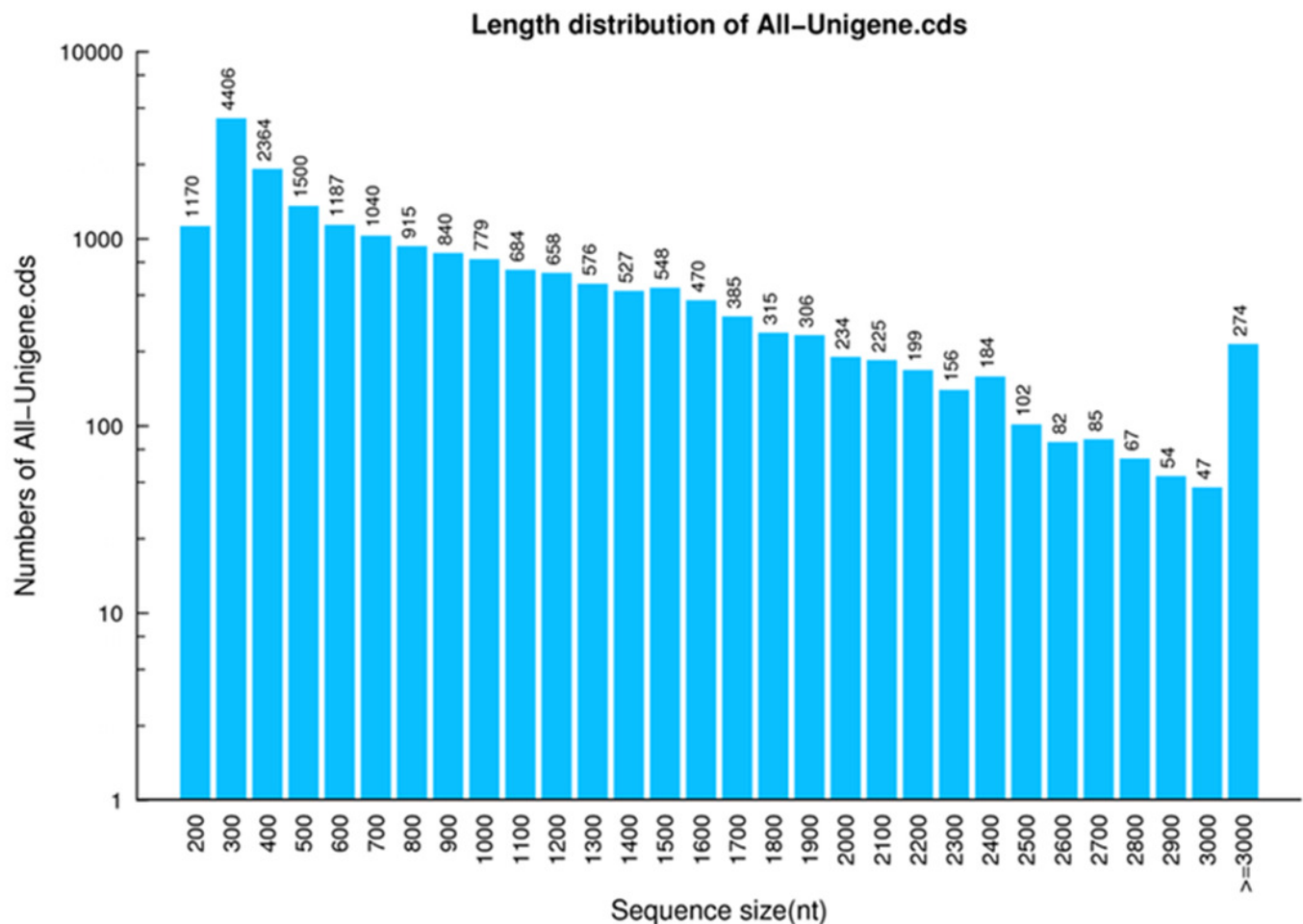
Figure 5

Fig. 5 Differentially expressed genes (DEGs) in *P. akamusi*. (A) Summary of DEGs.

X axis represents comparing samples. Y axis represents the number of DEGs. Red color represents up-regulated DEGs while blue color represents down-regulated ones. Group: A, Heads; B, Leg; C, Antennae. (B) Venn diagram showing intersections and disjunctions of genes expressed in heads, legs and antennae.
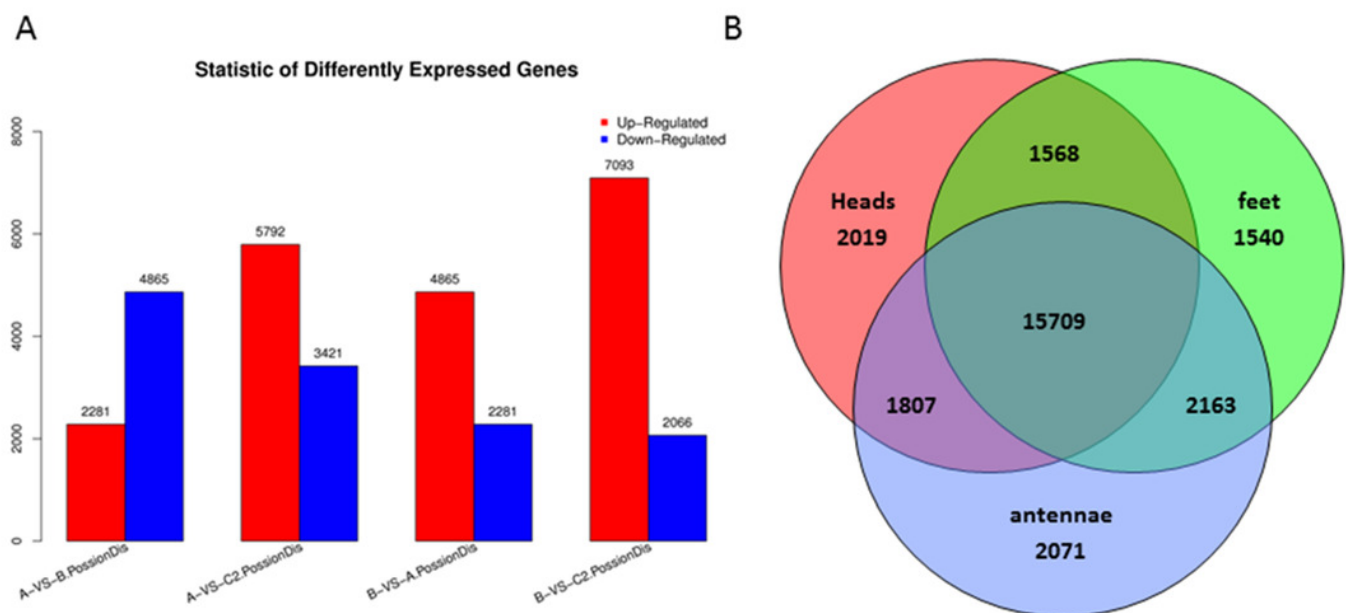
Figure 6

Fig. 6 *P. akamusi* OBPs transcript levels in different tissues measured by RT-qPCR.

" AN " for antennae; " HE " for head; " LE " for leg. The tubulin was used to normalize transcript levels in each sample. The standard error is represented by the error bar, and the different letters (a, b, c, d) above each bar denote significant differences (p<0.05).
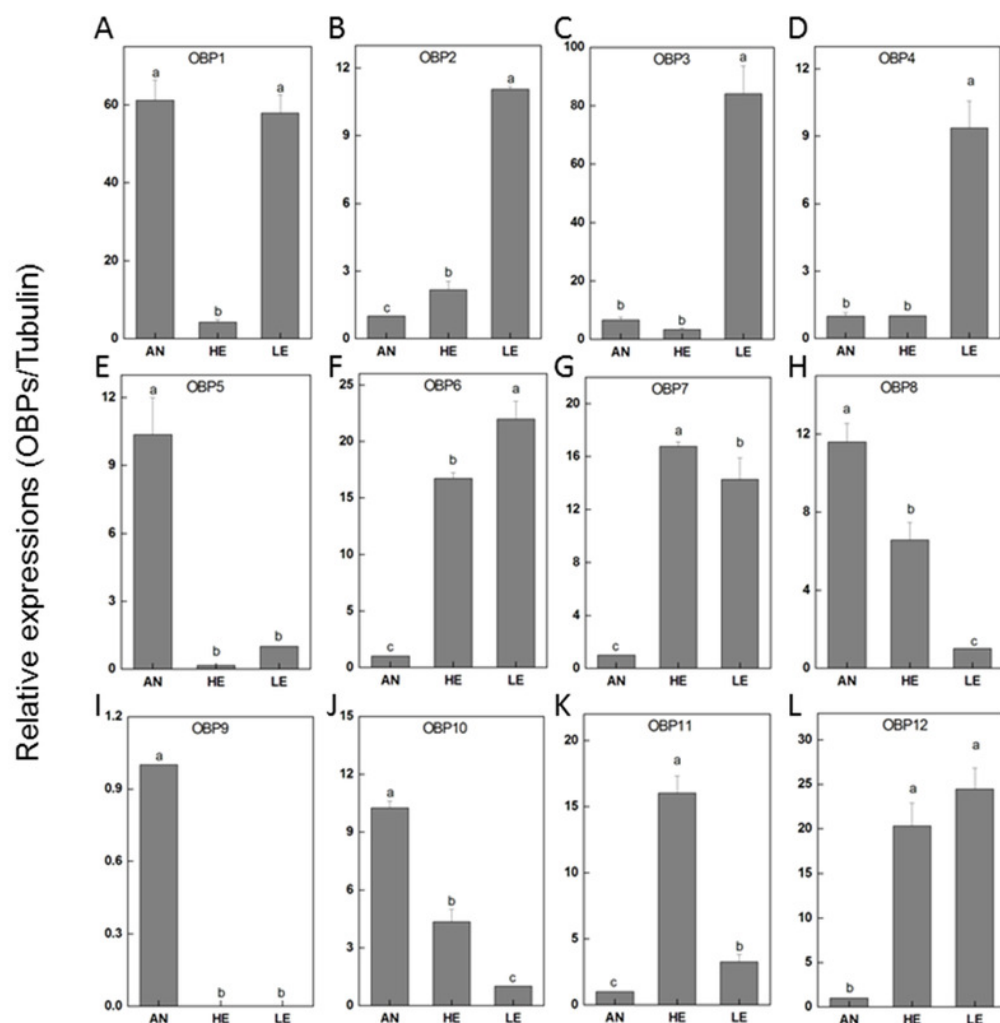
Figure 7

Fig. 7 *P. akamusi* ORs transcript levels in different tissues measured by RT-qPCR.

" AN " for antennae; " HE " for head; " LE " for leg. The tubulin was used to normalize transcript levels in each sample. The standard error is represented by the error bar, and the different letters (a, b, c, d) above each bar denote significant differences (p<0.05).
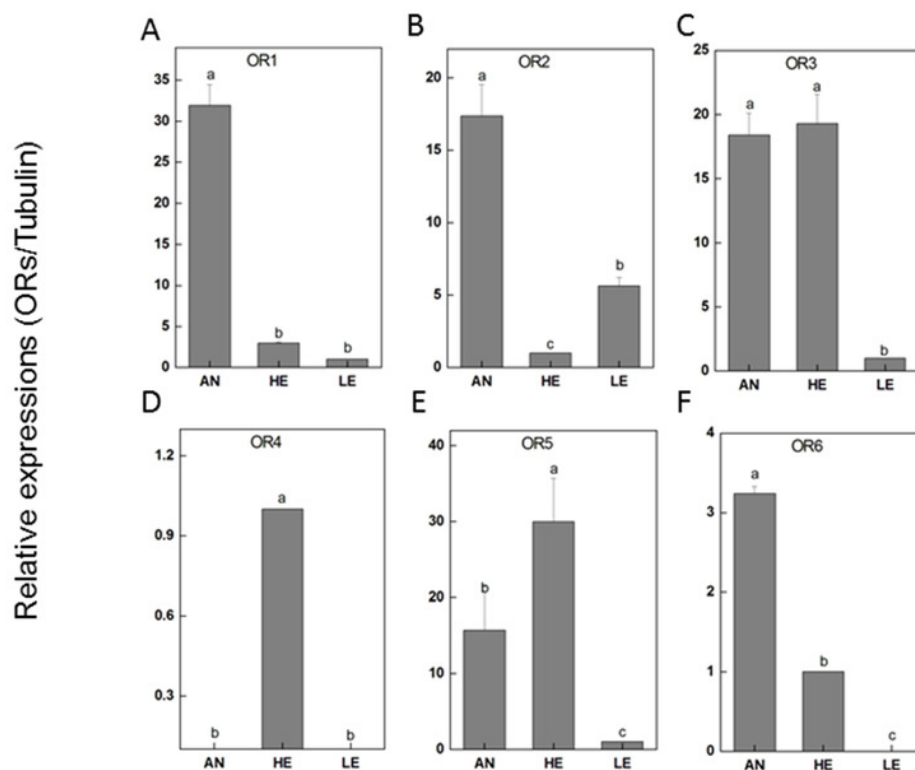
Figure 8

Fig. 8 *P. akamusi* GRs transcript levels in different tissues measured by RT-qPCR.

" AN " for antennae; " HE " for head; " LE " for leg. The tubulin was used to normalize transcript levels in each sample. The standard error is represented by the error bar, and the different letters (a, b, c, d) above each bar denote significant differences (p<0.05).
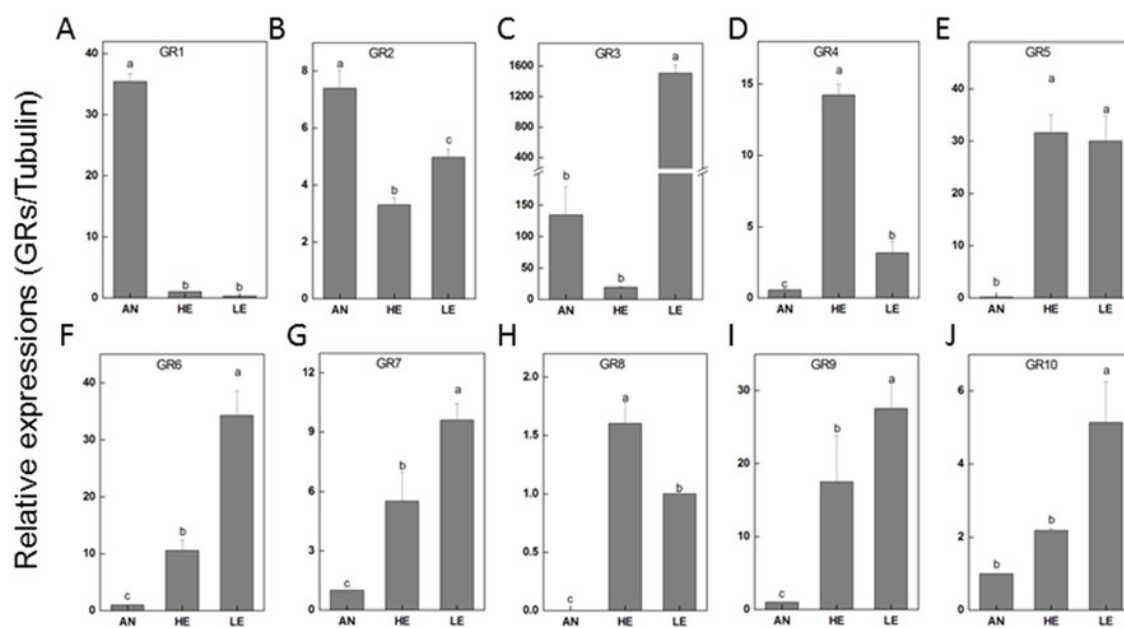
Figure 9

Fig. 9 *P. akamusi* IRs transcript levels in different tissues measured by RT-qPCR.

" AN " for antennae; " HE " for head; " LE " for leg. The tubulin was used to normalize transcript levels in each sample. The standard error is represented by the error bar, and the different letters (a, b, c, d) above each bar denote significant differences (p<0.05).
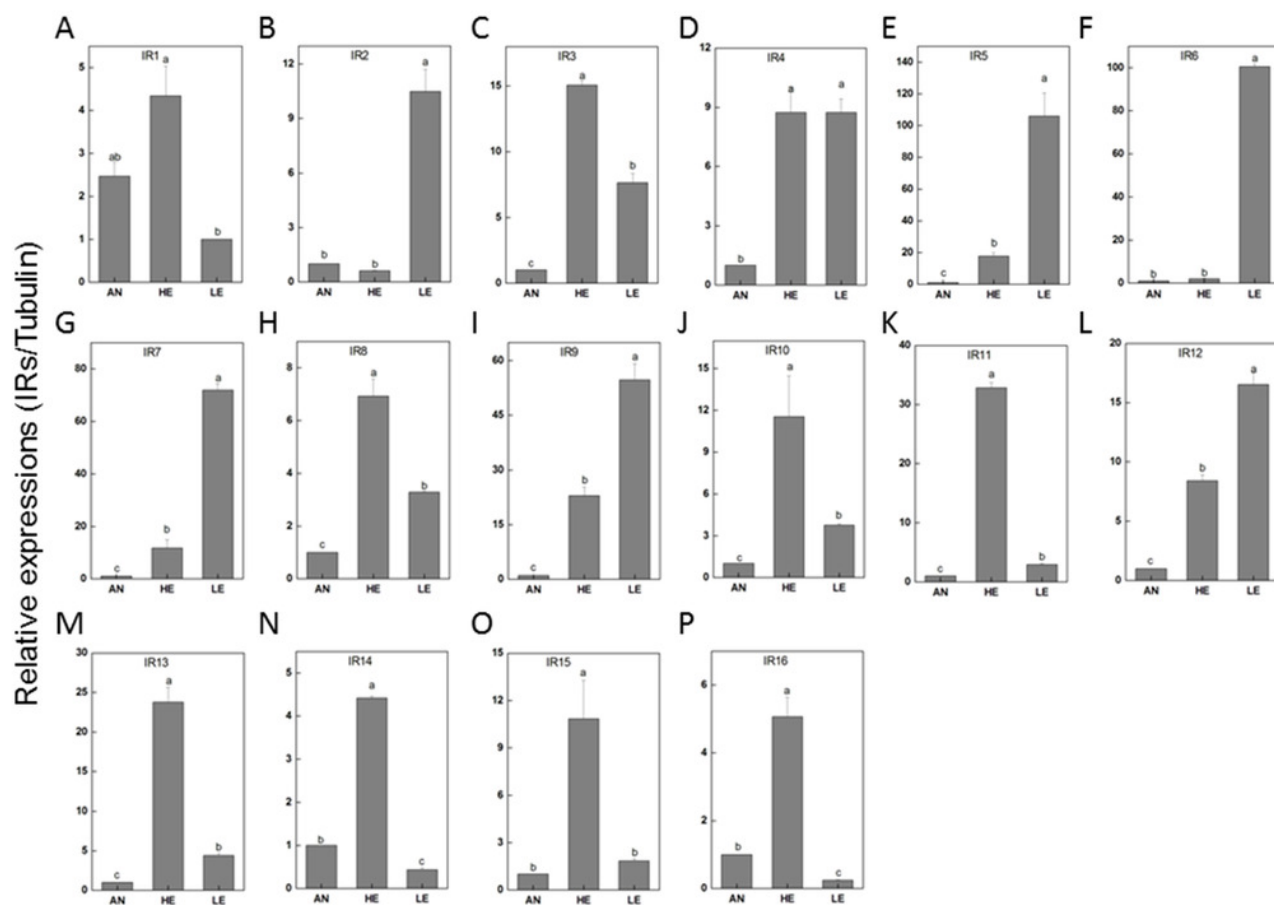
Figure 10

Fig. 10 *P. akamusi* CSP1 and SNMP1 transcript levels in different tissues measured by RT-qPCR.

" AN " for antennae; " HE " for head; " LE " for leg. The tubulin was used to normalize transcript levels in each sample. The standard error is represented by the error bar, and the different letters (a, b, c, d) above each bar denote significant differences($p<0.05$).