International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 11 Issue 11 November 2022 || PP 08-14

Full-length transcriptome analysis of the tea green leafhoppers *Matsumurasca onukii* based on PacBio Iso-Seq

Wen-Na Chen, Fang Hong, Guo-Fang Jiang *, Zhao-Wei Chen
（College of Oceanology and Food Science, Quanzhou Normal University, Quanzhou 362000, Fujian, China）

Abstract:
This paper uses the next-generation sequencing (Illumina RNA-seq) to correct the third-generation sequencing (PacBio Iso-Seq) method for full-length transcriptome sequencing of the leafhopper *Matsumurasca onukii*, and bioinformatics analysis of the transcriptome data. The results showed that the average length of the full-length transcriptome of the leafhopper was 2 126 bp; after the CD-Hit program to remove redundancy, 26 236 transcripts were obtained with an average length of 2 122 bp. Gene function annotation showed that in the NR, Swiss-Prot, KEGG, KOG, and GO databases, 9 509, 8 602, 6 184, 7 492, and 5 447 transcripts were annotated; among them, 3 761 transcripts were annotated in 5 databases, 8 893 transcripts were annotated in at least one database. In addition, 1 256 microsatellites (SSRs), 455 transcription factors (TFs) and 903 long non-coding RNAs (LncRNAs) were identified or predicted, with the main distribution of CDS lengths ranging from 200-2 000 bp. In this paper, we obtained the full-length transcriptome database of the leafhopper, which provides basic data for molecular biology research of leafhopper insects.
Key words: leafhopper *Matsumurasca onukii*; transcriptome; genome annotation; third-generation sequencing

---

Date of Submission: 10-11-2022 Date of Acceptance: 23-11-2022

I. Introduction

Matsumurasca onukii belongs to the family Cicadellidae of the order Hemiptera (Xu et al., 2021; Liu et al., 2022), formerly known as *Empoasca* (*Matsumurasca*) *onukii*, is the most common and destructive pests on tea plants (Zhang et al., 2017; Yu et al., 2018). Both nymphs and adults, leafhoppers possess stylets that pierce the tea leaves and draw sap from the tea leaves and buds (Hou et al., 2020). This behavior of the leafhopper damages the tea plant, often leading to a syndrome called funnel burning, which significantly reduces the yield and quality of tea leaves (Shao et al., 2021).

At present, high-throughput transcriptome sequencing includes second-generation sequencing and third-generation sequencing according to the base rolling read length. Next-generation sequencing (NGS) has been widely used in the study of gene expression levels (Tang et al., 2019), such as Hemipteran *Adelphocoris lineolatus* (Xiao et al., 2017), *Bemisia tabaci* (Nekkanti et al., 2022)), bed bug *Cimex hemipterus* (Lim & Ab Majid, 2022), leafhopper *Psammotettix striatus* (Yuan & Wei, 2022), Asian citrus psyllid *Diaphorinacitri*(Wu et al., 2021), etc., but NGS reads short in length and cannot span the entire transcript (Koren et al., 2012). In contrast to second-generation sequencing, third-generation transcriptome sequencing captures full-length transcripts without assembly. In recent years, three-generation sequencing has been gradually applied to the study of various Hemiptera, such as the brown planthopper *Nilaparvata lugens* (Zhang et al., 2019), the corn aphid *Rhopalosiphum maidis* (Chen et al., 2019), the white-backed planthopper *Sogatella furcifera* (Chen et al., 2020), etc., increasingly highlighting the advantages of third-generation sequencing technology in obtaining full-length functional genes.

To date, although there are a small number of second-generation transcriptome sequencing studies on leafhoppers in *M.onukii* (Shao et al., 2021), there is no report on third-generation full-length transcriptome sequencing. In this paper, the second-generation sequencing (Illumina RNA-seq) was used to correct the third-generation sequencing (PacBio Iso-Seq) data method to sequence the full-length transcriptome of the leafhopper, and through gene function annotation, CDS prediction, SSR analysis and TFs analysis,the gene data were obtained to provide a theoretical basis for its related gene interaction, transcriptional expression, functional annotation and molecular marker development, and lay a foundation for further control of leafhoppers in *M.onukii* at the molecular level.

II. Materials and Methods

2.1 Insect materials

The leafhoppers *M.onukii* were collected from Qingyuan Mountain, Quanzhou City, Fujian Province, and the adults were healthy and active; after collection, they were quickly frozen with liquid nitrogen and stored in a -80 ℃ low-temperature refrigerator for future use.

2.2 Total RNA extraction and high-throughput sequencing

Total RNA was extracted by RNeasy Plus Mini Kit (Qiagen, Valencia, CA, USA); 1% agarose gel electrophoresis was used to detect RNA degradation and contamination; Nanodrop (NanoDrop products, USA) was used to detect RNA purity (OD260/280); Qubie Accurately quantify RNA concentration; the Agilent 2100 Bioanalyzer accurately detects RNA integrity. PacBio Iso-Seq high-throughput sequencing was performed using a mixed sample of total RNA from the leafhoppers. Illumina RNA-Seq high-throughput sequencing was performed using a single total RNA sample from the leafhopper. The full-length transcriptome sequencing of *M.onukii* was mainly assisted by Wuhan Future Group Technology Co., Ltd.

2.3 PacBio Iso-Seq data processing and correction

SMRTlink 6.0 software processes the high-throughput sequence data obtained above, obtains CCS (cyclic consensus sequence) from subread BAM files and generates CCS.BAM files; CCS.BAM files are divided into full-length reads and non-full-length reads by the pbclassify program, perform isoform clustering on full-length and non-full-length fasta files, and finally use Quiver for fast correction. PacBio Iso-Seq data were corrected by Illumina RNA-seq data and LoRDEC software (Salmela& Rivals, 2014). LoRDEC adopts a mixed error correction mode. First, it reads the reads of RNA-seq sequencing, and uses the reads data to construct a DBG (de Bruijn Graph) graph; then reads the reads of PacBio Iso-Seq in turn, and judges the three generations of data in the constructed DBG graph. Whether there is second-generation data support in the data; correct the data without the second-generation data support in the DBG diagram, output the corrected sequence, and after the CD-hit-est program to remove redundancy, obtain the full-length transcription of the leafhopper.

2.4 Bioinformatics analysis

Functional annotation: Functional annotation was performed on the full-length transcriptome of the leafhopper using the NR, KOG, Swiss-Prot, KEGG and GO databases, respectively ($p< 0.05$).

Transcription factors (TFs) identification: Transcription factor identification was performed through the Animal TFDB 2.0 database (Zhang et al., 2015). Species not in the database were identified by the HMMSEARCH software (Finn et al., 2011) from the protein family database files (PFAM files) of transcription factor families.

Alternative splicing (AS) analysis: *M. onukii* full-length transcriptome AS was analyzed by using the coding genome reconstruction tool (Cogent v3.1, https://github.com/Magdoll/Cogent) and SUPPA (https://github.com/comprna/SUPPA).

Long non-coding RNAs (lncRNAs) prediction: CPC (http://cpc.cbi.pku.edu.cn), CNCI (https://github.com/www-bioinfo-org/CNCI), PLEK (https://bitbucket.org/arrigonialberto/ lncrnas-pipeline) (Li et al., 2014) and 4 methods in the Pfam database to predict lncRNA from the obtained non-redundant transcript sequences, and all four methods are predicted as lncRNA transcript as the final result.

Coding sequence (CDS) prediction and microsatellite sequence SSR analysis: In this paper, ANGEL (https://github.com/PacificBiosciences/ANGEL) software was used to perform CDS and protein prediction for all isoforms (Shimizu et al., 2006). SSR sequences were detected using MISA software (Thiel et al., 2003) to identify SSR markers.

III. Results and Analysis

3.1 Statistics of full-length transcriptome data

The results of PacBio Iso-Seq sequencing showed that the full-length transcriptome database of the leafhopper contains 506 046 CCSs, 439 356 full-length reads, 66 690 non-full-length reads, and 427,399 full-length non-concatemer reads (Table 1). The dataset provided in this study can be found in the Genome Sequence Archive (https://ngdc.cncb.ac.cn/gsa/), the GSA omics raw data archive, with accession number CRA007462.

Table 1 Summary of transcript sequences of *Matsumurasca onukii*

Parameter	Result
Number of reads of insert (CCS)	506046
Number of full-length reads	439356
Number of nonfull-length reads	66690
Number of full-length non-chimeric reads	427399

After clustering and correction with IsoSeq3 polish, we finally got 28 316 consensus sequences with a

total length of 60 221 636 bp and an average length of 2 126 bp, of which the accuracy was greater than 0.99 and the number of full-length sequences supported ≥ 2. There are 27 559 high-quality consensus sequences, with a total length of 58 517 395 bp and an average length of 2 123 bp. The total length of the consensus sequences corrected by the second-generation transcriptome data is 58 470 467 bp, with an average length of 2 122 bp. The corrected sequences of the second generation data were de-redundant and corrected by the CD-hit-est program and then clustered to obtain 26 236 transcripts with an average length of 2 122 bp and 11 006 unigene (single gene) sequences with an average length of 2 296 bp (Table 2).

Table 2 Analysis of clustering results

Parameter	Result
Number of consensus sequences for correction by second generation data	27559
Average consensus sequences read length for correction by second generation data	2122 bp
Total length of consistent transcript sequence for correction by second generation data	58470467 bp
Number of transcripts for redundancy removed by CD-hit	26236
Average length of transcripts for redundancy removed by CD-hit	2112 bp
Number of Unigenes for redundancy removed by CD-hit	11006
Average length of Unigenes for redundancy removed by CD-hit	2296 bp

3.2 Functional Notes

In this paper, 9 666 transcripts were annotated through 5 databases, and the results showed that there were 7 492, 6 184, 9 509, 8 602 and 5 447 transcripts in KOG, KEGG, NR, Swiss-Prot and GO databases, respectively. were annotated (Table 3); of them, 3 761 transcripts were annotated in 5 databases, and 8 893 transcripts were annotated in at least one database (Fig. 1).

Table 3 Statistics of functional annotation unigenes in *Matsumurasca onukii*

Database	Annotated Gene Number	Annotated Gene Ratio (%)
GO	5447	49.49
KEGG	6184	56.19
KOG	7492	68.07
Swiss-Prot	8602	78.16
NR	9509	86.40
Total unigenes	11006	-
Overall annotated	9666	-

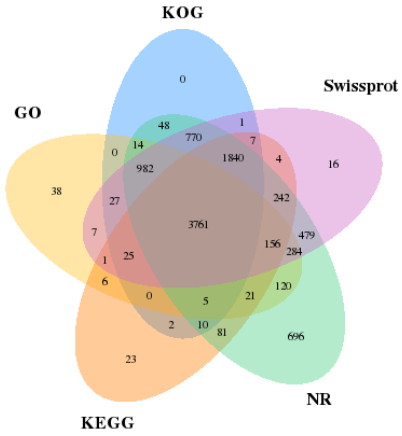

Fig.1 Venn diagram of the results of function annotation of the *Matsumurasca onukii* transcripts

3.3 Transcription factors (TFs) and Alternative splicing (AS)

The results of TFs identification showed that the full-length transcriptome of *M.onukii* contained 455 transcription factors, of which zf-C2H2 (28.15%) and ZBTB (26.13%) were the main transcription factor families.

The results of Cogent and SUPPA analyses showed that the full-length transcriptome of *M. onukii* contains 1 451 ASs, and the number of alternatively spliced genes is 11 006, of which 97.63% of unigenes have no more than 10 isoforms. Interestingly, 50% of unigenes had only one isoform (Fig. 2).

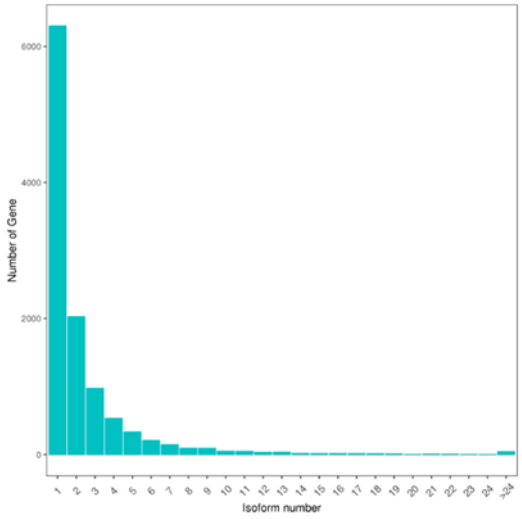

Fig. 2 Alternative splicing (AS) analysis of the *Matsumurasca onukii* unigenes.

3.4 LncRNA prediction

The results of lncRNA analysis of the full-length transcriptome of the leafhopper in *M. onukii* showed that 1 162, 1 104, 1 040 and 1 038 lncRNAs were predicted by CNCI, Pfam, PLEK and CPC, respectively, and 903 lncRNAs were predicted together (Fig. 3).

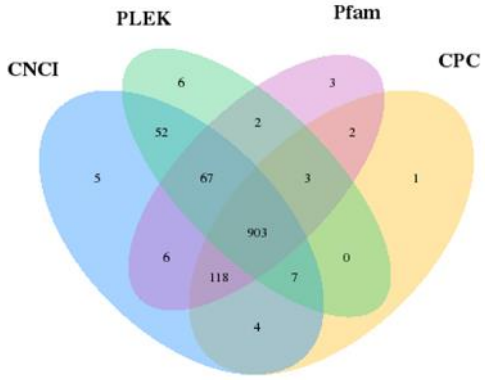

Fig. 3 Venn diagram of lncRNA transcripts identified from CNCI, Pfam, PLEK and CPC

3.5 CDS prediction

The ANGEL pipeline enables ANGEL long reads to determine the protein-coding sequence (CDS) of full-length complementary deoxyribonucleic acid (cDNAs). A CDS is a sequence that encodes a protein product that exactly matches the codons of the protein. In the sequencing results of the full-length transcriptome, the prediction of protein coding regions is helpful for the preliminary analysis of genes, and is also the basis for subsequent protein structure analysis. Using ANGEL software to perform CDS prediction on the obtained full-length transcriptome, the results showed that the main distribution range of CDS length was 200-2 500 bp (Fig. 4).

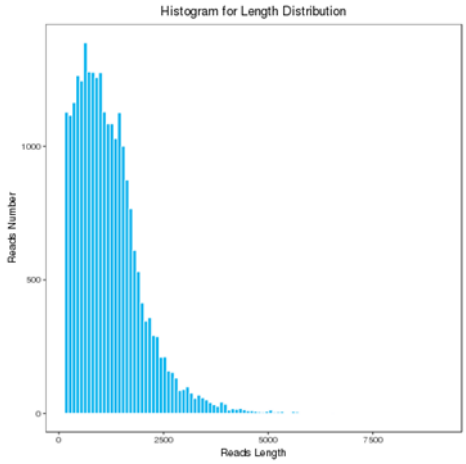

Fig. 4 Number and length distributions of coding sequences of the *Matsumurasca onukii* transcripts

3.6 SSR analysis

Using MISA software to detect SSR, the results showed that a total of 1 255 SSR loci were identified in the full-length transcriptome of the leafhopper *M.onukii*. Among them, Mono-nucleotide motifs were the most (50.6%), followed by Tri-nucleotide repeats (Tri-nucleotides) 25.4%, Di-nucleotide repeats (Di-nucleotide motifs) 14.0%, Tetra-nucleotide repeats (Tetra-nucleotides) 0.5%, Hexa-nucleotide repeats (Hexa-nucleotide motifs) 0.3% and penta-nucleotide repeats (Penta-nucleotides) 0.2% (Fig. 5).

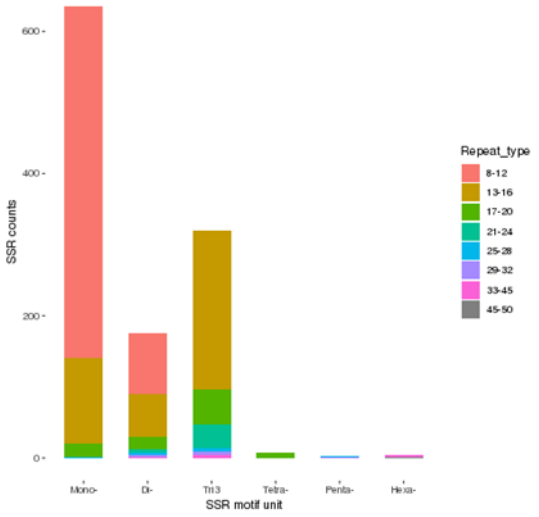

Fig. 5 Frequency of the identified SSR motifs of *Matsumurasca onukii*

IV. Discussion

In this paper, both PacBio Iso-seq and Illumina RNA-seq were used to sequence the full-length transcriptome of the leafhopper *M. onukii*, and perform bioinformatics analysis on the transcriptome data. Dig deep into its genetic data to provide data support for molecular biology research of this species. Compared with the amount of second-generation sequencing data, the total amount of non-redundant full-length transcriptome data obtained by second-generation sequencing combined with third-generation sequencing was 26 236, which was higher than that of other pure second-generation sequencing Hemiptera. Insects, such as soybean aphid *Aphis glycines* 14 861 non-redundant transcripts (Liu et al., 2012), *Lygus alfalfa* 22 143 transcripts (Xiao et al., 2017), bed bugs 24 609 transcripts (Lim & Ab Majid, 2022) et al. In addition, 87.82% of the transcripts of the leafhoppers *M. onukii* were successfully annotated, and the results showed that 32.6% of the genes involved in the biological process played a role in the metabolic process; in terms of the molecular function, 19.2 % were expected to have binding function; in terms of cellular components, 26.6% of genes involved organelles. According to the annotation results of the GO database, by tracing the sub-levels of reproduction and reproductive process, a total of 3 reproductive-related functional genes were screened: PB.3078.1, PB.574.1 and PB.2593.1 (see S1). These genes are all in biological process. A total of 13 genes (see S2) related to Metalloproteinase, 54 genes (see S3) related to heat shock protein

family, and 4 genes related to glycoside hydrolase (Glycosidase) were found in the annotation results. genes (see S4), 21 genes related to acetylcholinesterase (AchE) (see S5) and 17 genes related to sulfotransferase (see S6). At the same time, compared with other insects sequenced by PacBio Iso-seq, such as brown planthopper sequencing obtained 24 891 transcripts (Zhang et al., 2019), corn aphid sequencing obtained 21 114 transcripts (Chen et al. , 2019), 29 700 transcripts were obtained by sequencing the white-backed planthopper *Sogatella furcifera* (Chen et al., 2020), and the number of full-length transcripts obtained in this paper was in the middle, at 26 236.

The main TFs obtained in this paper are mainly zf-C2H2 (28.15%) and ZBTB (26.13%). Previous studies have shown that different TFs may be involved in different metabolic processes and may have multiple different functions (Chen & Rajewsky, 2007). For example, the widespread distribution of the *zf-C2H2* gene suggests a critical regulatory role for the zf-C2H2 domain in activating downstream gene expression in various biological processes (Zhang et al., 2021), and the effects of *zf-C2H2* on early embryonic organs/tissues. Transcriptional regulation and development/differentiation are critical (Mackeh et al., 2018). ZBTB transcription factors are key regulators of T cell development, differentiation, and effector function (Cheng et al., 2021); ZBTB20 is implicated in the generation of subpopulations of corpus callosum projection neurons (CPNs) and the development of astrocyte subtypes guidance (Medeiros de Araújo et al., 2021). The spider molting process is regulated by a series of transcription factors such as ZBTB and zf-C2H2, which prolong its developmental time (Peng et al., 2022).

In addition, this paper obtained 903 lncRNAs of the leafhopper of *M. onukii* through lncRNAs prediction. lncRNAs regulate gene expression at the epigenetic level mainly through transcriptional and post-transcriptional regulation, and exert powerful biological functions by affecting protein localization and telomere replication (Batista & Chang, 2013; Kung et al., 2013; Qureshi &Mehler , 2013). In recent years, a large number of lncRNAs have been discovered from insects. For example, *Rhynchophorusferrugineus* obtained a total of 9 618 lncRNAs (Yang et al., 2020), and 2 914 lncRNAs were identified from the citrus aphid *Aphis citricidus* (Shang et al ., 2021) and *Aedes albopictus* identified a total of 21 414 lncRNAs (Liu et al., 2022). However, the number of lncRNAs found in the leafhopper *M. onukii* is less than 1000, why it is obviously less than the number of lncRNAs in other known insects, which needs to be further explored and confirmed in the future. Growing evidence suggests that lncRNAs play important roles in various biological processes and are critical for regulating gene expression (Chen et al., 2018). LncRNAs play important roles in regulating *Drosophila* development, behavior, stress resistance, sex recognition, and dose compensation (Li et al., 2019a; Choudhary et al., 2021). In addition to regulating growth, lncRNAs also have rapid responses to stress and stimuli (Valluri et al., 2017; Li et al., 2019b). Many studies have confirmed that lncRNAs can regulate a variety of immune responses, including several pathways related to innate immunity, such as bacterial invasion, lncRNAs can promote or weaken the immune response of insects, especially the Toll immune response to bacterial attack (Moure et al., 2022). lncRNA-155 targeting protein tyrosine phosphatase 1B modulates innate immunity against influenza A virus (IAV) infection in mice (Maarouf et al., 2019). The discovery of these lncRNAs will provide certain reference information for further research on the functions of *M. onukii* in terms of growth, development and immunity. This will provide a theoretical basis for further excavating new drug efficacy targets of the leafhopper *M. onukii*, and using molecular design to achieve pest control.

To sum up, based on PacBio Iso-Seq, this paper successfully obtained the full-length transcriptome of the leafhopper *M. onukii*, and successfully annotated its related gene function information. It provides an important basis for the study of behavioral ecology, molecular marker development, and capture of its immune and drug resistance genes.

Acknowledgments

We thank Mr. Ming-Jie Li for his assistance in collecting leafhopper specimens. This study was funded by Quanzhou City Science and Technology Program of China (2018C040R) and the National Natural Science Foundation of China (31970449).

References

[1]. Batista PJ, Chang HY. 2013. Long noncoding RNAs: cellular address codes in development and disease. Cell, 152:1298–1307
[2]. Chen J, Yu YY, Kang K, Zhang DW. 2020. SMRT sequencing of the full-length transcriptome of the white-backed planthopper *Sogatella furcifera*. Peer J, 8: e9320
[3]. Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. Nature Reviews Genetics, 8: 93–103
[4]. Chen W, Zhang X, Li J, Huang S, Xiang S, Hu X, Liu C. 2018. Comprehensive analysis of coding-lncRNA gene co-expression network uncovers conserved functional lncRNAs in zebrafish. BMC Genomics, 19(Suppl 2):112
[5]. Chen W, Shakir S, Bigham M, Richter A, Fei Z, Jander G. 2019. Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). GigaScience, 8(4): giz033
[6]. Cheng ZY, He TT, Gao XM, Zhao Y, Wang J. 2021. ZBTB transcription factors: key regulators of the development, differentiation and effector function of T cells. Frontiers in Immunology, 12: 713294
[7]. Choudhary C, Sharma S, Meghwanshi KK, Patel S, Mehta P, Shukla N, Do DN, Rajpurohit S, Suravajhala P, Shukla JN. 2021. Long non-coding RNAs in insects. Animals (Basel), 11(4):1118
[8]. Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Research,

39:W29–W37
[9]. Hou BH, Tang H, Li JL, Meng X, Ouyang GC. 2020. Susceptibility of selected tea shoots to oviposition by *Empoasca onukii* (Hemiptera: Cicadellidae) and feasibility of egg removal with harvesting. Insects, 11(6): 338
[10]. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, et al. 2012. Hybrid error correction and *de novo* assembly of single- molecule sequencing reads. Nature Biotechnology, 30(7): 693–700
[11]. Kung JTY, Colognori D, Lee JT. 2013. Long noncoding RNAs: past, present, and future. Genetics, 193:651–669
[12]. Li A, Zhang J, Zhou Z. 2014. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved *k*-mer scheme. BMC Bioinformatics, 15(1): 311- 314
[13]. Li MZ, Xiao HM, Kang H, Li F. 2019a. Progress and prospects of noncoding RNAs in insects. Journal of Integrative Agriculture, 18(4):729–747
[14]. Li X, Wang H, Zhang Y, Zhang J, Qi S, Zhang Y, Gao M. 2019b. Overexpression of lncRNA H19 changes basic characteristics and affects immune response of bovine mammary epithelial cells. PeerJ, 7: e6715
[15]. Lim L,Ab Majid AH. 2022. Analysis of transcriptome difference between Blood-Fed and starved tropical bed bug, *Cimex hemipterus* (F.) (Hemiptera: Cimicidae). Insects, 13(4):387
[16]. Liu S, Chougule NP,Vijayendran D, Bonning BC. 2012. Deep sequencing of the transcriptomes of soybean aphid and associated endosymbionts. PLoS ONE,7(9):e45161
[17]. Liu W, Cheng P, Zhang K, Gong M, Zhang Z, Zhang R. 2022. Systematic identification and characterization of long noncoding RNAs (lncRNAs) during *Aedes albopictus* development. PLOS Neglected Tropilal Diseases, 16(4): e0010245
[18]. Liu Y, Liu M, Huang L, Kang Z, Xu Y, Chen Z. 2022. Analysis of diversity and temporal patterns of the insect communities in tea gardens. Journal of Tea Science,42(1):109-119(in Chinese)
[19]. Maarouf M, Chen B, Chen Y, Wang X, Rai KR, Zhao Z, Liu S, Li Y, Xiao M, Chen JL. 2019. Identification of lncRNA-155 encoded by MIR155HG as a novel regulator of innate immunity against influenza avirus infection. Cellular Microbiology, 21(8): e13036
[20]. Mackeh R, Marr AK, Fadda A, Kino T. 2018. C2H2-Type zinc finger proteins: evolutionarily old and new partners of the nuclear hormone receptors. Nuclear Receptor Signaling, 15:1-22
[21]. Medeiros de Araújo JA, Barão S, Mateos-White I, Espinosa A, Costa MR, Gil-Sanz C, Müller U. 2021. ZBTB20 is crucial for the specification of a subset of callosal projection neurons and astrocytes in the mammalian neocortex. Development, 148: dev196642
[22]. Moure UAE, Tan T, Sha L, Lu X, Shao Z, Yang G, Wang Y, Cui H. 2022. Advances in the immune regulatory role of non-coding RNAs (miRNAs and lncRNAs) in Insect-Pathogen interactions. Frontiers in Immunology, 13:856457
[23]. Nekkanti A,Chakraborty P,Ghosh A,Iquebal MA,Jaiswal S,Baranwal VK. 2022. Transcriptomic changes of *Bemisia tabaci* Asia II 1 induced by chillileaf curl virus trigger infection and circulation in its vector. Frontiers in Microbiology, 13:890807
[24]. Peng Y, Lv B, Lei ZY, Peng YD, Chen LJ, Wang Z. 2022. Toxic effects of the combined cadmium and Cry1Ab protein exposure on the protective and transcriptomic responses of *Pirata subpiraticus*. Ecotoxicology and Environmental Safety, 236:113631
[25]. Qureshi IA, Mehler MF. 2013. Long non-coding RNAs: novel targets for nervous system disease diagnosis and therapy. Neurotherapeutics, 10:632–646
[26]. Salmela L,Rivals E. 2014. LoRDEC: Accurate and efficient long read error correction. Bioinformatics, 30(24): 3506–3514
[27]. Shang F, Ding BY, Zhang YT, Wu JJ, Pan ST, Wang JJ. 2021. Genome-wide analysis of long non-coding RNAs and their association with wing development in *Aphis citricidus* (Hemiptera: Aphididae). Insect Biochemistry and Molecular Biology, 139:10366
[28]. Shao E,Song Y,Wang Y,Liao Y,Luo Y,Liu S,Guan X,Huang Z. 2021. Transcriptomic and proteomic analysis of putative digestive proteases in the salivary gland and gut of *Empoasca* (*Matsumurasca*) *onukii* Matsuda. BMC Genomics, 22(1):271
[29]. Shimizu K, Adachi J, Muraoka Y. 2006. ANGLE: a sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. Journal of Bioinformatics and Computational Biology,4(3): 649–664
[30]. Tang D, Zhang GS, Zhao XF. 2019. Bioinformatics analysis of transcriptome sequencing based on next generation sequencing. Journal of Henan University (Medical Science), 38(1):67-76 (in Chinese)
[31]. Thiel T,Michalek W,Varshney RK,Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics, 106(3): 411-422
[32]. Valluri S, Rupam G, Srividya S. 2017. Long non-coding RNAs regulating immunity in insects. Non-coding RNA, 3(1):14
[33]. Wu ZZ, Qu MQ, Chen MS, Lin JT. 2021. Proteomic and transcriptomic analyses of saliva and salivary glands from the Asian citrus psyllid, *Diaphorina citri*. Journal of Proteomics, 238:104136
[34]. Xiao Y,Sun L,Ma XY,Dong K,Liu HW,Wang Q,Guo YY,Liu ZW,Zhang YJ. 2017. Identification and characterization of the distinct expression profiles of candidate chemosensory membrane proteins in the antennal transcriptome of *Adelphocoris lineolatus* (Goeze).Insect Molecular Biology, 26(1):74-91
[35]. Xu Y, Dietrich CH, Zhang YL, Zhang YL, Dmitriev DA, Zhang L, Wang YM, Lu SH, Qin DZ. 2021. Phylogeny of the tribe Empoascini (Hemiptera: Cicadellidae: Typhlocybinae) based on morphological characteristics, with reclassification of the *Empoasca* generic group. Systematic Entomology, 46: 266-286
[36]. Yang H, Xu D, Zhuo Z, Hu J, Lu B. 2020. SMRT sequencing of the full-length transcriptome of *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae). Peer J, 8: e9133
[37]. Yu Y, Zhang J, Huang C, Hou X, Sun X, Xiao B. 2018. Reference genes selection for quantitative gene expression studies in tea green leafhoppers, *Empoasca onukii* Matsuda. PloS ONE, 13(10): e0205182
[38]. Yuan F,Wei C. 2022. Gene expression profiles in malpighian tubules of the vector leafhopper *Psammotettix striatus* (L.) revealed regional functional diversity and heterogeneity. BMC Genomics, 23(1): 67
[39]. Zhang H, Liu T, Liu C, Song S, Zhang X, Liu W, Jia H, Xue Y, Guo A. 2015. Animal TFDB 2. 0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Research, 43(D1): D76–D81
[40]. Zhang ZQ, Zhou C, Xu YY, Huang XQ, Zhang LX. 2017. Effects of intercropping tea with aromatic plants on population dynamics of arthropods in Chinese tea plantations. Journal of Pest Science, 90: 227–237
[41]. Zhang J, Guan W, Huang C, Hu Y, Chen Y, Guo J, Zhou C, Chen R, Du B, Zhu L, et al. 2019. Combining next-generation sequencing and single-molecule sequencing to explore brown plant hopper responses to contrasting genotypes of japonica rice. BMC Genomics, 20(1): 682
[42]. Zhang J, Wei J, Yu H, Dong B. 2021. Genome-Wide Identification, comparison, and expression analysis of transcription factors in Ascidian *Styelaclava*. International Journal of Molecular Sciences, 22(9): 4317

Supplementary Information

S1 Reproductive genes and their coding sequences

Gene name	GeneId	GO annotation	CDS
isolate FS-2020 chromosome 25	PB.574.1	GO:0023052 (signaling)	ATGGTTTCGCCAGAGGCTCTCACCGAGGGGGCACTTAATCACAAACATACAGTTAACGCTTCAATTGACACTGCTGACACCAA AAGAAAAGCGGTGACAACATGTGGAATGGAAACGGCAACAGTGTGTCACCAGCGGCGAGTAAAGTGTCCACATCCACGTTCA ACACTACGGAATCTCCGCTGCCCGTGCTCAAGCTATTGACTCCTGCGTTCCGTGCTCAACTGACCCTGGCCACGAGGAAGTCCA GCTGGAAAAGCCAGGCTCTTAGCTAATCAAAAACTCGCTCAATTCAAAGGACTTTGCCAAAAGAATATTAATCAAGATTTGAC GAGCGGCGCAAAGCAATGCGTCAGAACCAACCCAACTCCGGTGAAGACATCGTCATAGTCTAA
unkown	PB.3078.1	GO:0023052 (signaling)	ATGGAGAGAGCATTCCTCCACAAAAGTAAAACTGCAGGGTTTGGCAATATCGAGGCACGTCGAACAGAGACTTTTGCTGCAG GCTTGGGCAAGAATGAGGTGAAGAAAGAGGAGGTGAGTCGAGTGACAAGGGCCGCATCAGCCAAGGTCGGAATGAAGGCTC AAGACTCAGTTAATCAAGCTGTATCAAATGGTTCTTTAACTGACCAAGCATTGTCATTTAACAACTTTGGAAATGAGGTAGCA CCAGAAAACCCAGAAGCTTTTAAGCAAAGGGAGCGAGTGCGGACTGAGACTGCCAAGTTACTGGAGTTGGCTGATACCTGGC AAAGGAAAAGAGTGAAAAGAATCAAACCTGAAACTATTGCATCCAAAACTGAGGATGTTTACAAGAAAACACCCAGGAGAT TTCTGTGAACAATGAGATTGCATCCGCTACACAATCTAGGGGACGAAAGTCAAGAAGATTACAAATTATTGAACAGAACAAC
unkown	PB.2593.1	GO:0051321 (meiotic cell cycle)	ATGGCGTTGTCGGAGGTTTGCGGAATTGAACATTTACTAGGCAGAAATAGCGCACCTTTCAGACAGGTGGGACATGAACAAG ACTTTAACATGAGCTTACTTAACATTGACACACAATCAACTTTAAACACTAATGGATATCACTTGGGGGAGGAACCTAAAGAC CTATAACAAAATTCCAGAAAACCCTTTCAGAGGGATGGATGCTCTTGACAATAATTATGGGAAAAACATGCTTGATGCCAGAT GCTATGGGTTCCCACCCGTCCTGGACCCGGCTTTGCTCATGCACTTCCGACCTAACAACCCGGAACTGTTCCCCCTGTACGACA AGAACCTGCGCGGTATGCGGCTGCACAACAACATCAGCGTCAACATGCAGAACTGGCTGGACGCTATCAAGCGTGTGCAGGC

S2 Metalloproteinase genes and their coding sequences

gene name	gene ID	annotation	CDS
ADAM10	PB.203.1	K06704 ADAM10, CD156c; disintegrin and metalloproteinase domain-containing protein 10 [EC:3.4.24.81]	ATGAAGCTAGTGTGTCTGTATGTGTGTGGTGTCACTACAATTCGCCCAGATCTCGCAACCGACGCCTGTCCTGCCTAACTACATTCCAGGTGTGTCGGTGAGTGAGT ATCTACAGAGCCTCCGACACTGTCAATCCCACCAGAGACTCATGTGCCAGTGCGGAGTTGGCCAAGAGAGTCAGGGATAGAGAGGTCAACTCGATACAGAGCTCGGAAG CGTGCCAAAGGTAATGACTCGATCATCCCACACTTTCCCAAGTCTAACCACTACAACCCACCTGTGGACAGGAGTCACTACCCCAAGATCATCGTGGGGACGGGCTCCT TATATGTTCACCTACCGAGACTTCGAGATGGGCACTCTGGGCCTCGCCTGGACCGGAGACTTGAAGAATGCCGGGGGAGTCTGCGAGAAGAACGGGCACTACCGAGGCA CAGCGTCGCTACCCTCCCCCGGACCAACCCCCTTGTCACCTCACCCCCAACAGCCTGTGTAGCCCTTCACAGGGTCCGTGCTGTACGACCGAGTGCCAGCTCAAGTTCG TCCGGACCGCTAGCCACGCTGCGCAAGCTGCTCCTGTCCGAAGAGAGCATCGCCACATTTAAGCGCTGGATGTGGGACAGATGGTACTACGCGGCGCTCATCTTCGCAT
ADAM11	PB.1491.1	K06704 ADAM10, CD156c; disintegrin and metalloproteinase domain-containing protein 10 [EC:3.4.24.81]	ATGTTTGGTGCTTTTTGGAATATAGCCTTCCTCATTCTACTCGTACCATACATCGAAAGTGCAAATAGATTAAACGAGTACATCAGACACTATGAGGTGGCAGACTACG CCGTATGAACACAAACGTACAGGCCATGTCGGAGGCTGTGGCATCACAGACTCAGTTAGTCAATGGATGGATCGCATCCAGAATTCAGCAGTGGAAGAAGATACTGATG AAGTTTGAAGTACAGAGAATAAAGATTGACGACGATGCTCCCTGCAAGCCCGGTTACCACGGAGAGTCAAATCCTTTCTGTATGGAGAACATTGATGTGAGCAACTTTC AACAGCAAGTTCAGCAGTTGCAGTGTCGGCAACATCAGTAATGTGCTAGATGCCATTGAAGACAAGAAGAAGCGTAACTGCTTTGAAGTGTCCGAGGGTGCGTTCTGTG ATGCAAGAGTGTTTCCTGACGTCGACGAATGTGCAGAACACAGACAAGCGAAGGCTGTGTGAGCTCGCTTGTCAAAACGGCACAGACACCGAAACCTGCCGCAGCACCT CGTGGGTCAGGAAGCGGCCAGCAGAGACCGGGGCCGAGTCGGGGCCCTGCTCACGGCTACGGCCAAGGCCGGGGTCAGTATTATGCACCTAAAGCGTCAGCGCCACCAC
Adamts4	PB.2197.1	XP_026815892.1 A disintegrin and metalloproteinase with thrombospondin motifs 1 isoform X1 [Rhopalosiphum maidis]	ATGGCCGCTCTGCCTCTTGCCTATCTCTGCCCTTACCTCCTAGTCATCGTCACAGGAATCTCCGCTGCCACAAACACGATACACCATCACATGAACATAGATGATCTCA TCAGCTCCGAGGAGGAGAGACGACGACGATGAAGCGTTCCTGGACGAGGAAGCTGACGAAGACTTGGCAGTGAGGGATGGCTTCCCCCTCCACCTGCCTCCACGACACC GCTCTTACCGACATGGGCAAGTATCTGTTCCGAGAACGCAGGCTCCCAGTCTACGACATCGCTGTCGCCATCACCAAATTAGACATGTGCAGACGGCAATACCCAAATG GGGAAACTGCTCACTCTGGACGCACAGTGTCGCAAGGATAGAGGCACTAGCGCCTGCTTTAAAGACGACAGAGTCTGCGCCCAGCTATTCTGTTTTGATGCTGGCTCGG
MMP1	PB.2352.1	K07763 MMP14; matrix metalloproteinase-14 (membrane-inserted) [EC:3.4.24.80]	ATGCCAGTGTTAAGCAGTTACACTTTCTCACCTTCAAAGATGTTCATCCTCCTGAGGTGTTTGGTTTTGGCTGCTCTGCTAGCGTCTATCAAAGCGAATCCGATACAGA TTCACTCAGAAGAAGAGTGGGTCTGTTCACATCGAGATACGCTTTGAGCGAGGAGAGCACGGCGACGGAGACCCGTTCGACGGACCCGGCGGCACCCTCGCACACGCCT TATGCTTTCAAAGGTTTGCACTACTGGAAACTGACGGAGACAAGTATTGCTGCCGGTTATCCTCGTCTGATCTCCAAGAGCTGGATTGGTCTCCCCGGCAACCTTGATG CCATTCCCGAGGCCAGCAGGCTACTGGTGGTTCGGCTGCAGAGCAGAGACCAAGGGTTACTACACAGGTGACATAACACGTAGTAGTAGTGGCATCCGGGACCAGGGTG
MMP1	PB.2582.1	K07763 MMP14; matrix metalloproteinase-14 (membrane-inserted) [EC:3.4.24.80]	ATGCCAGTGTTAAGCAGTTACACTTTCTCACCTTCAAAGATGTTCATCCTCCTGAGGTGTTTGGTTTTGGCTGCTCTGCTAGCGTCTATCAAAGCGAATCCGATACAGA TTCACTCAGAAGAAGAGTGGGTCTGTTCACATCGAGATACGCTTTGAGCGAGGAGAGCACGGCGACGGAGACCCGTTCGACGGACCCGGCGGCACCCTCGCACACGCCT TATGCTTTCAAAGGTTTGCACTACTGGAAACTGACGGAGACAAGTATTGCTGCCGGTTATCCTCGTCTGATCTCCAAGAGCTGGATTGGTCTCCCCGGCAACCTTGATG CCATTCCCGAGGCCAGCAGGCTACTGGTGGTTCGGCTGCAGAGCAGAGACCAAGGGTTACTACACAGCGGGCCGAAGCTAG
MMP1	PB.2637.1	K07763 MMP14; matrix metalloproteinase-14 (membrane-inserted) [EC:3.4.24.80]	ATGCCAGTGTTAAGCAGTTACACTTTCTCACCTTCAAAGATGTTCATCCTCCTGAGGTGTTTGGTTTTGGCTGCTCTGCTAGCGTCTATCAAAGCGAATCCGATACAGA TTCACTCAGAAGAAGAGTGGTTCTGTTCACATCGAGATACGCTTTGAGCGAGGAGAGCACGGCGACGGAGACCCGTTCGACGGACCCGGCGGCACCCTCGCACACGCCT TATGCTTTCAAAGGTTTGCACTACTGGAAACTGACGGAGACAAGTATTGCTGCCGGTTATCCTCGTCTGATCTCCAAGAGCTGGATTGGTCTCCCCGGCAACCTTGATG CCATTCCCGAGGCCAGCAGGCTACTGGTGGTTCGGCTGCAGAGCAGAGACCAAGGGTTACTACACAGGTGACATAACACGTAGTAGTAGTGGCATCCGGGACCAGGGTG
MMP1	PB.3155.1	K07763 MMP14; matrix metalloproteinase-14 (membrane-inserted) [EC:3.4.24.80]	ATGCCAGTGTTAAGCAGTTACACTTTCTCACCTTCAAAGATGTTCATCCTCCTGAGGTGTTTGGTTTTGGCTGCTCTGCTAGCGTCTATCAAAGCGAATCCGATACAGA TTCACTCAGAAGAAGAGTGGGTCTGTTCACATCGAGATACGCTTTGAGCGAGGAGAGCACGGCGACGGAGACCCGTTCGACGGACCCGGCGGCACCCTCGCACACGCCT TATGCTTTCAAAGGTTTGCACTACTGGAAACTGACGGAGACAAGTATTGCTGCCGGTTATCCTCGTCTGATCTCCAAGAGCTGGATTGGTCTCCCCGGCAACCTTGATG CCATTCCCGAGGCCAGCAGGCTACTGGTGGTTCGGCTGCAGAGCAGAGACTAAGGGTTACTACACAGGTGACATAACACGTAGTAGTAGTGGCATCCGGGACCAGGGTG
LOC124354691	PB.4773.1	KDR17787.1 A disintegrin and metalloproteinase with thrombospondin motifs 18 [Zootermopsis nevadensis]	CATGTTGAAACAGCTGTCTTCGTGGACAGTGACATGTTCCACCTCATGGCTAATAACTTCCCCCAGGACACCGAGAGGGAGGTGGTTCGCTTTGTGCTCGCCATGGTAA AGACACGACGGGCCCCTGGCAGACAACAGCTGTGACCCGACCAGCTACCTCATGTCCCCCACTCTCGGCAGTGGCAAGATCACGTGGTCAGAGTGCAGTAAACACTATC GAAGGAGGCCGGTACACTGCTGGCAGTACAGGGATCATGTTCGCACACCGACAGTGCAACATGCTCCGCCCGGAGAACGGAGGCAACGGCTGTATCGGCACTGATCGCA GGAGTCTGGCAACCGACCAGTGATTGCTACTACAACTGTATCTCACCTGGCTCGGGGGTTCGGCTAGTGGAGAAGCGCAGTTGTCTTCACTGTAACACCTCTGCCAGCA TTTGATCACAATGTTACGATAACAGCAACTTTGGATAACCAAACTTTAGCAGAGATGGTGCAGTACTTGAATAACACAGGGAACCGGAAGGACTCTATAAGAGCAGATG
5_Tge_b3v08	PB.9412.1	EZA54126.1 Zinc metalloproteinase nas-15 [Ooceraea biroi]	ATGTTACTCAAGATGTCGACCCAGCTGGTCTGTCTGCTGCTACTCGCGCTGCTCGCGGGGGTCTCCCCTGTCAGTATCTTCTCCGGGACTACCCGCGACCGCTCTGTCG GGGTGTTGGTCGTATGTCGGGAGGAGGAGTGGAGGCCAAGTTGTGAACCTACAGGGTCGGTGTGTCCAGCACGGTGTCGTTATCCACGAACTGCTCCACGCTCTCGGGT TCCTGGGATAAACCATTTTACTTTTTTTAG

unkown	PB.9628.1	Zinc metalloproteinase nas-4 OS=Caenorhabditis elegans OX=6239 GN=nas-4 PE=2 SV=4	ATCTTTGGTTTCATTACAAGTCGTGGTGTCAAAAAGATGTCATCGCTCCTGGTTTCTCTGCTGCTACTGGGGTTGCTCGCAGGGGTCTCCCCTTATAGCATCTTCCCCC GCAGGTGGCAAACAGACGTTGAACCTTCATAATCCTTCGTGCCTTGACCAAGGAACCATTGTTCATGAACTGCTTCACGCGTTGGGCTTCGAACATCAACAGAGTGCCC
unkown	PB.2810.1	XP_021914861.1 putative zinc metalloproteinase C607.06c [Zootermopsis nevadensis]	ATGGCAACATTTGAGTCTGATATACCATGCCACCATAGAATCACTGTCAAAGATCTTTCCCCGAATGAAGAAGTTGCTTTCTCGTTAGTTCTTATAAAAGGTTCAGTTA ACATTCCAACTGAAAAGAGATATTAGCGATAACTCCGATGAATGTGTGACATTTCACAGCAGTCTTTCAGCGTCAAAAGCTCAAACTATGGGTCAAGAAGAACTGTGGA TTTTTAGCCGACTGGACAAACACTGTAAAGGCAGCGGCACCGCGCAAGACGCGACCCCAGCGCTCCACCGTCACGCTGCTCCACAATCTCAACGTAGAGTGTGTGATAG TGA
unkown	PB.3266.1	XP_008201012.1 PREDICTED: A disintegrin and metalloproteinase with thrombospondin motifs 7 [Tribolium castaneum]	ATGTTACCTGGCCAGATATATTCAGCAGCTCAGCAGTGTAAACTGAATTTCCACATGACAGCTCTGCCTTGTTCGGTCGGACCTTTCTGTGAAAGACTTTATTGTCAAG TTCCTCAATATTGAGTGTGCTTTGGTGTGCCAAAACAGGAAAGGAGAAGTGGTAATGCGCTCTCCCATCGTTGCCGACGGTACTCCTTGCAAAGCAGGGACTCGAAATG TTAATTTTCTATGTGCAATTTCGTGAAGCTAACAAAGGAATACATTACCAGTATGCAGTTCCAGAAATTGACCCCACTTATATTCCACGTTACTTGTGGCAGTACTTGG GATGACAAAGCTCACAGAGATATAAGAAGATCACTCAATGCCAACATGGCTTACTCAGCGTTATCACCAGCAGAAACCAGTGGGATGGGTGCTGGGGGTCAGAACCCAT GACTCCGGCTCACAATGCGGGACAAAGGCACCTTCTTGTCAAACGAGTGAAACTTGCGAGACGGAGCCTAGCTCAGTGAGTACTTGCCGGCAAACATGTCCTGTCCAGC
	PB.6871.1	XP_011256592.1 A disintegrin and metalloproteinase with thrombospondin motifs 1 [Camponotus floridanus]	ATGTCCATTGATAACAAGTGGTTCCACAGGAAAAATTACAAAATGTTGACAACTAGTTTTATCTTGATTGTCATGGCTTTGATTTGCGTGGAATCGTTTCAGATGCACG TCAATACCTGGCCATGTACAGAAAAGTATACTGTATAGGAGAAACCATTCGGACGTGAATGATGTAGGAAACTCGTATCCACATGTTGTTTATAAAGTGTCGGATTCTC TACGACGTAGTTATTACTTTGACCCAGTCCAAATTGTGCACGAGAAAAGTATTTCCAAATTGTTATCCTATTTTAGGCGCGGCATTTACGGGCGGAGCGTGCGGGCGAA TGCGCCCAATTGCGTTGCGTTACACCAGGAAATCCGTCTTTCTGTATAACAAAATCTCCCCCCGCCGAAGGTTCTTCTTGCGGTAACGGTATGATATGTCTCAACGGGA

S3 Heat shock protein genes and their coding sequences

Gene name	GeneId	Annotation	CDS
Hsc70	PB.10077.1	XP_008477019.1 heat shock protein 70 A1-like isoform X1 [Diaphorina citri]	ACCACCTACGCAGACAACCAGCCAGCTGTTACCATCCAGGTGTTCGAGGGAGAGCGAGCCAGGACCAAGGACAACAACCTCCTGGGCACATTTGACCTGACAGGAA
genome assembly	PB.10264.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGACAGTAACGATGTCTCTGGTACCATACTTGATGAACGAGATGATCCGTGATTTCGAGATGCCTTTGTACGACCAGAACTTTGGCTTGGGGTTGTTTCCATCCC
Hsp110	PB.1293.1	XP_021920125.1 97 kDa heat shock protein [Zootermopsis nevadensis]	ATGGCTGCTATGTCAGTTATAGGCATTGATTTCGGTAACGAATCTTGTTATGTTGCTGTTGCAAGAGCGGGTGGTATTGAGACCATCGCGAATGACTACAGTCTTC TCCATTGCTGGCTTAAATGTACTTCGACTGTTCAACGAATCTACTGCAACTGCTCTTGCGTATGGAATATACAAACAAGACTTGCCTGCACCAGAGGAGAAGCCTA TGTTTAGAAGACTCAAAGCTGAGGCAGGATGAAATCCACTCAGTGGAGATCGTAGGAGGGTCCAGTCGTATACCATCAATCAAGTCAATGATAGAGAACATTTTC TTAGCAATCAACAGTGCCTCGCTTGTGGAGAAGCTGCCAGCCTCGGAGAGTGACGTGACAGAAGACCAGTCACAGCCCGACCCCATGGAAGCCTCGCCCAGTCCC TATGCTGAGAGGTTGGCTTCTCTCAAGAGTTCTGGTGAACCAATAAAAACTAGACGATTGGAATTCGAGCTGAGACCAATAGCCTTAGAAGAACTATCTTCATCAC
Hsp110	PB.1732.1	XP_021920125.1 97 kDa heat shock protein [Zootermopsis nevadensis]	ATGGCTGCTATGTCAGTTATAGGCATTGATTTCGGTAACGAATCTTGTTATGTTGCTGTTGCAAGAGCGGGTGGTATTGAGACCATCGCGAATGACTACAGTCTTC TCCATTGCTGGCTTAAATGTACTTCGACTGTTCAACGAATCTACTGCAACTGCTCTTGCGTATGGAATATACAAACAAGACTTGCCTGCACCAGAGGAGAAGCCTA TGTTTAGAAGACTCAAAGCTGAGGCAGGATGAAATCCACTCAGTGGAGATCGTAGGAGGGTCCAGTCGTATACCATCAATCAAGTCAATGATAGAGAACATTTTC TTAGCAATCAACAGTGCCTCGCTTGTGGAGAAGCTGCCAGCCTCGGAGAGTGACGTGACAGAAGACCAGTCACAGCCCGACCCCATGGAAGCCTCGCCCAGTCCC TTTGTGAGTGACAACGAGCGGCAGAGCTTGACCAACATGTTGGAACAGATGGAGAACTGGCTTTACGAGGAGGGAGAGGACTGTACGAGACAGGTGTATGCTGAG CCGACCGACACCAGCGGTGGAGACGCTAAGCAATCAGATCAACCCCCCAATCAGGAAAATATGGATGTAGAATGA
Hsp70	PB.1790.1	AXU24955.1 heat shock protein 70-1 [Cyrtorhinus lividipennis]	ATGCCGTCGAAAATGCCTGCTATTGGAATCGACCTGGGAACGACCTACTCCTGCGTGGGCGTCTGGCAACAGGGCAAAGTGGAGATCATCGCTAACGACCAGGGCA GCCATTGCAGGCCTTAATGTTCTTAGGATCATCAATGAACCAACTGCAGCAGCGTTGGCTTATGGATTGGACAAGAACTTAAAGGGTGAGAGGAATGTGTTGATAT AAAATGGACAAGTCCTCAATCCATGATGTTGTTCTTGTAGGAGGATCCACTAGAATTCCCAAGATCCAAAGTCTTCTGCAGAACTTCTTCTCCGGTAAATCACTGA GTGACCGCCAAGGACAACAGCTCTGGCAAATCAAAGAACATCACCATCCAAAACAACAAAGGAAGACTTTCCAAGGAAGAAATTGACAGGATGGTTGCTGAAGCGC
DNAJA1	PB.2575.1	GO:0031072(heat shock protein binding),GO:0051082(unfolded protein binding)	ATGGTAAAGGAAACTACATTCTATGATATCTTGGGCGTGAAGCCAGGATGTTCTCCAGAAGAATTGAAAAAAGCTTATCGTAAACTGGCTCTCCAATATCACCCTC CAAACCAGGATCAATCAGCTAGGTCCAGGAATGTTGCAACAATTGCAAACACAATGTAAAGATTGTAAAGGACAAGGCGAGCGCATTGATCCACGTGATCGCTGTA ATTCAATTCTTAATCAACTTCCCCAAGACCATTAGTCCTGATGTAATTCCATTGCTCAAGCAATGCTTACCGAGTGTCGAAGAAATTATTGTACCAGATGACGCTC
Hsp90	PB.2671.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGTTGATGTGTAATGGTTACAGTGGAATTGGTATGACAAAGGCTGACCTCGTGAACAACTTGGGAACCATCGCCAAGTCTGGAACTAAGGCTTTCATGGAAGCAC GATAAGAAGCCGAAGATTGAAGATGTGGGTGAGGATGAAGAAGAGGAGGATAAGGAAAAGCCTAAAAAGAAGAAGAAGACGGTAAAGGAGAATTACACAGAGGATC

			CTGGTGAAGAAGTGCTTGGAACTGTTTGACGAGATCGCAGAGGACAAGGAGAGCTACAAGAAGTACTACGAACAGTTCAGCAAGAACCTGAAGCTGGGAATCCACC CTGGACAAGAAGGTGGAGAAGGTGGTCGTGAGCAACAGGCTGGTCGAGTCACCCTGCTGTATCGTCACATCGCAATACGGCTGGACCGCCAACATGGAGAGGATCA
Hsp90	PB.2848.1	K09487 HSP90B, TRA1; heat shock protein 90kDa beta	ATGAATAAATTTTGGTTCTTGGCCTTGGGCCTCTGTCTCTTAGCAGAAATTGTTAGAGCTGAGGATTCGATAGAGGAAAATGTTGGAACTGTTGAAAAAGATCACC GGAACAATTGCCAAATCTGGTACAGCCGAATTTTTGAGTAAAATGCAAAAAGCTGCGGTCGAGGGACAAGATACAAATGATATGATTGGTCAATTTGGTGTTGGAT AAGAAACCTAAAAAAGTTGATAAAACTATATGGGATTGGGAGTTGTTGAATGATGCAAAACCTATTTGGACAATGAAACCTAACGAAATCAGCGACGAACAATATA AAAGAATATAGCACGAATATTAAATTGGGTCTTATTGAGGATTCACAAAATCGTGCGAGGTTGTCAAAACTTTTACGTTTCCGTTCATCGACCAAAGAAATGATGA ATGTTTGGATGGACTGGAAATATGGAAAGATTGGCTTTGTCCAATGCTCATCAAAAGAGCGAGGATCCACAAAAAACTTATTACATGAATCAAAAGAAAACACTCC
HSP90B1	PB.3295.1	K09487 HSP90B, TRA1; heat shock protein 90kDa beta	ATGGGAACGAAAACAACATTTTTATGTTTTATAGCGTTATCATTATGTATAGGTTTATGCAAGGGAGAGGATTTGTTGATGGAAGAATCTACGGTAGAAGGGGATA CTCGGCACCATCGCCAAGTCCGGTACAGCCGAGTTCCTCAGCAAGATGCAGGAGTCCGGCGCCCAGTCCGACATGAACGACATGATCGGTCAATTTGGAGTTGGGT AAGACGAAGAAGGTGGAGAAAACTGTCTGGGACTGGGAGAAAGTCAATGAAAACAAACCTATCTGGACAAGAAAGCCGGCAGAGATTGAAGATAGTGAGTACAACC GAGTACTCTACCAACATCAAGCTGGGACTGATCGAAGACACCACAAACCGAGTGAGGTTGGCCAAGATTCTCCGGTTCCAGTCATCCAAGATGGAGGGCACCACTT TGGACCGGCAACATGGAGAGGCTGGCCGTGAGCAATGCACACCAGAAGGCAGATGACCCGCAGAGGTCGTACTACCTGAGTCAGCGCAAGACCCTGGAGATCAATC
LOC122716684	PB.3605.1	XP_011060405.1 PREDICTED: heat shock protein 83 [Acromyrmex echinatior]	ATGCCGGAAGACGTAACCATGTCGGAAAGCGAATCTCAAACTTTCGCTTTTCAAGCTGAAATAGCTCAACTTATGAGCTTGATTATTAATACATTTTATTCAAATA TCTTTCGTCGTTCGTCCAGATAATGGAGAACCTATTGGGCGTGGTACTAAGATTGTTCTTCATTTTAAAGAAGATCAAACTGAATATTTGGAAGAATCCAAAATCA AAAGCATTGTTGTATGTCCCTCGCAGAGCGCCATTTGACTTATTTGAAAACAAGAAAAAGAGGAACAACATTAAATTATATGTTAGACGAGTTTTTATTATGGACA CAAGTAGCCAACAGTTCGTTCGTCGAAAGAGTTAAAAAACGTGGTTATGAAGTAATTTATATGACTGAACCAATAGATGAATACGTCGTACAACAACTGAAAGAAT ATGCTTTTATTTGAGACTGCGCTTCTTTCATCTGGATTTGGTTTGGAAGATCCTCAAGTTCACGCATCAAGAATTTATAGAATGATCAAACTTGGTCTTGGTATTC
LOC111054336	PB.3616.1	AXU24966.1 heat shock protein 60-2 [Cyrtorhinus lividipennis]	ATGTTCCGCCTACCCTCAGTGTTGAGAACCGCAGCCTCCAGACAACTGTGTCCTTCTATCACAAGGTTCTACGCCAAAGATGTTAGGTTTGGAGCTGAAGTTCGGC GAAGAAATCGCACAAGTAGCTACAATATCTGCCAATGGCGACAAAGAAGTTGGACAGTTGATCTCAGACGCAATGAAAAGGGTCGGGAGAGACGGTGTGATCACAC GCCGCTAATGTGAAGCTGGAAGAGGTTCAGGCCGCCGACTTGGGAGAAGTTGCAGAGGTCCTCATCACCAAAGACGATACACTGCTTCTGAAGGGAAAGGGCAAGA AACAAAGTACTTGCGGCTTCTGGAGACCAAGGATATGATGCGATGAACAATGAATATGTCAATATGATTGAAAAGGGTATCATCGATCCTACAAAGGTTGTAAGGA
Hsp70	PB.4115.1	ACH85201.1 heat shock protein 70 [Trialeurodes vaporariorum]	ATGCCGTCGAAAATGCCTGCAATTGGAATCGACCTGGGTACGACCTACTCCTGCGTCGGCGTCTGGCAACAGGGTAAAGTGGAGATCATCGCCAACGACCAGGGAA GCCATTGCAGGCCTAAATGTTCTTAGGATCATTAATGAGCCTACTGCAGCAGCTTTAGCCTATGGATTAGACAAGAACTTGAAAGGCGAGAGAAATGTACTGATAT AAAATGGACAAGTCTTCTATCCACGATGTTGTTCTTGTTGGAGGTTCTACCAGAATTCCCAAGATTCAGAGTCTTCTGCAGAACTTCTTCTCCGGGAAGTCACTGA GTGACCGCCAAGGACACGAGCTCCGGCAAATCTAAGAACATTATCATCCAAAACAACAAAGGAAGATTTTCCAAAGAAGAAATCGACAGGATGGTTAATGAGGCTC GACTAA
Hsp75	PB.4129.1	XP_011256852.1 heat shock protein 75 kDa, mitochondrial [Camponotus floridanus]	ATGGCCACCACGTTTCGGTTGAGGCACGCTTTTTCATTAGTAAAGTCATTGAAAAGATCGTTTGGCAGCAATACAAGAGTTGTTTTATTGGAAAGATCGCACGGGA GTATCCAATCTTGGTACTATAGCGAGATCTGGATCTAAGGCTTTCATGGAACAAATAAAAGACCAAACAGCAGCAGGCGATGCTTCTAAAATCATAGGACAATTTC CATTATTCCACCGATGTACCATTGAGTATCAAAGCCCTTTTATATTTCCCAGATGGGAAACCTGGATTATTTGACTTAACTCGAGAAATATCCACCGGAGTATCAC GATGAAAAAGATGTATATTATTTGTCCGCTCCAAATCGAGCTTTAGCCGAACAATCACCTTATTATGAATCTTTGAAAAAACGTAATGTCGAAGTATTGTTTTGTT GAACCTAAATTAGCCGATCTTCTCATTCAACAGTTATTCATGAGTGCCATGGTTGGCGCTGGATTGTTTGAAGATCCAAGGGTACTGCTTTCACCGATGAATGAAT
Hsp90	PB.4178.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACCTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTGGAGGAGAGGAAGGTCAAAC GAGTTCAGAGCTCTCCTGTTTATTCCTCGTAGAGCTCCTTTCGACCTGTTTGAGAACAAGAAGCGTAAAAATAACATCAAGTTGTATGTAAGACGAGTATTCATCA ATGACAGAGCCGATCGACGAGTATGTAGTACAGCAGATGAAGGAGTACGACGGCAAGCAGCTGGTGTCCGTCACCAAGGAGGGTCTGGAGCTGCCGGAGGACGATC CACGCAGCCAGGATCCACCGCATGATCAAGCTCGGTCTGGGAATTGACGAGGATGATCCGGTTCCTCAGGAGGATGAGAAGGTCGACGCTGAAATGCCTCCACTGC
Hsp70	PB.4226.1	ACH85201.1 heat shock protein 70 [Trialeurodes vaporariorum]	ATGCCGTCGAAAATGCCTGCAATTGGAATCGACCTGGGTACGACCTACTCCTGCGTCGGCGTCTGGCAACAGGGTAAAGTGGAGATCATCGCCAACGACCAGGGAA GCCATTGCAGGCCTAAATGTTCTTAGGATCATTAATGAGCCTACTGCAGCAGCTTTAGCCTATGGATTAGACAAGAACTTGAAAGGCGAGAGGAATGTACTGATAT GTAAATGTAATGTGA
Hsp90	PB.4250.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACCTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTGGAGGAGAGGAAGGTCAAAC GAGTTCAGAGCTCTCCTGTTTATTCCTCGTAGAGCTCCTTTCGACCTGTTTGAGAACAAGAAGCGTAAAAATAACATCAAGTTGTATGTAAGACGAGTATTCATCA TATGTAGTACAGCAGATGAAGGAGTACGACGGCAAGCAGCTGGTGTCCGTCACAAAGGAGGGTCTGGAGCTGCCGGAGGACGATGCCGAGATGAAGAAGCGAGAGC

			ATGATCAAGCTCGGTCTGGGAATTGACGAGGATGATCCGGTTCCTCAGGAGGATGAGAAGGTCGACGCTGAAATGCCTCCACTGGAAGGAGACGGTGAGGACGCT
Hsp60A	PB.4286.1	ACO57619.1 heat shock protein 60 [Pteromalus puparum]	ATGTACAAGCTACCAACAATTTTGCGTAGCACGGCTCTTCGCCAATTACAGTATCAAACATGCAGTTATGCCAAAGATGTACGCCTTGGAGCTGAAGTTAGAGCTT GAAATTGAACAAGTTGCTAGAATATCTGCGAATGGTGATGAGAAAATTGGAAAGTTAATTTCTCATGCTATCAAAACAGTTGGAAGAGAAGGCGTTATCACTGTAA AACCTTGTTAAATTGGAAGACGTTCAGTTGAGCGATTTGGGGCAAGTCGGTGAAATCATAATCACCAAGGACGACACTCTATTATTAAAAGGTAAAGGTAAAAAA AAAGTGAGCGAAGGTAAACTCGGTTACGACGCGTTAAAGAACGAATATGTCGATATGATTGAAAAAGGAATCATCGATCCCACCAAAGTCGTTCGCACTGCACTTA
Hsp60A	PB.4307.1	AXU24966.1 heat shock protein 60-2 [Cyrtorhinus lividipennis]	ATGTTCCGCCTACCTTCAGTGTTGAGAACCGCAGCCTCCAGACAACTGTGTCCTTCTATCACAAGGTTCTACGCCAAAGATGTTAGGTTTGGAGCTGAAGTTCGGC GAAGAAATCGCACAAGTAGCTACAATATCTGCCAATGGCGACAAAGAAGTTGGACAGTTGATCTCAGACGCAATGAAAAGGGTCGGGAGAGACGGTGTGATCACAC GCCGCTAATGTGAAGCTGGAAGAGGTTCAGGCCGCCGACTTGGGAGAAGTTGCAGAGGCCCTCATCACCAAAGACGATACACTGCTTCTGAAGGGAAAGGGCAAGA
Hsp90	PB.4387.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACCTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTGGAGGAGAGGAAGGTCAAAC GAGTTCAGAGCTCTCCTGTTTATTCCTCGTAGAGCTCCTTTCGACCTGTTTGAGAACAAGAAGCGTAAAAATAACATCAAGTTGTATGTAAGACGAGTATTCATCA AAGGACCAAGTAGCCAACTCTTCCTTCGTAGAGCGAGTGAAGAAGCGCGGATTTGAAGTGGTCTACATGACAGAGCCGATCGACGAGTATGTAGTACAGCAGATGA CCGGTTCCTCAGGAGGATGAGAAGGTCGACGCTGAAATGCCTCCACTGGAAGGAGACGGTGAGGACGCTTCTAGGATGGAGGAGGTTGATTAA
Hsp90	PB.4415.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACTTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTAGAGGAGAGGAAGGTCAAAC CCAGAGTACCTGAACTTCATCAAGGGAGTAGTCGACAGTGAGGATCTTCCTCTGAACATCTCCCGAGAAATGCTTCAACAGAACAAGATCCTGAAGGTTATCCGCA GTGTCCGTCACCAAGGAGGGTCTGGAGCTGCCGGAGGACGATGCCGAGATGAAGAAGCGAGAGGAAGACAAGGCTAAGTACGAGAACCTGTGCAAGGTCATGAAGC CCTCAGGAGGATGAGAAGGTTGACGCTGAAATGCCTCCACTGGAAGGAGACGGTGAGGACGCTTCTAGGATGGAGGAGGTTGATTAA
Hsp90	PB.4482.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACCTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTGGAGGAGAGGAAGGTCAAAC GAGTTCAGAGCACTCCTGTTTATTCCTCGTAGAGCTCCTTTCGACCTGTTTGAGAACAAGAAGCGTAAAAATAACATCAAGTTGTATGTAAGACGAGTATTCATCA GGTCTGGAGCTGCCGGAGGACGATGCCGAGATGAAGAAGCGAGAGGAAGACAAGGCTAAGTACGAGAACCTGTGCAAGGTCATGAAGGACATCCTGGACAAGAAGC GTTGACGCTGAAATGCCTCCACTGGAAGGAGACGGTGAGGACGCTTCTAGGATGGAGGAGGTTGATTAA
Hsp75	PB.4657.1	XP_002425720.1 heat shock protein 75 kDa, putative [Pediculus humanus corporis]	ATGGCGTTCCCCTTAAAATTAGTTTTGAGACTGAAAAACAATAAACTATCATACTTATCTAGACCGATTTATAGTCTATCTTCTCAACAGTACTCTTTACAAGAGA CGCATGTTCATCATTCAGGACACAGGCCTTGGAATGACTAAGGAGGAACTGATTTCAAACTTGGGAACAATCGCTCGCTCCGGATCTAAGGCGTTCCTGCAGCAGC CAGTTTTACCGTTTCGTCGCCAACTCGTACGATCAGCCTCGCTTCACTCTCCACTACAAGGCCGACGTACCTCTGTCTGTTCAGGCCTTGCTGTTCTTCCCTGAGC CCGGCCGGAACGACAATCAGTTTACCGGAATATATCAGCAGACTCAAGACCTCACAGAGAGATATCTACTACCTCTGCGCTCCCAGTCGATCGCTTGCTGAGAGTT CTGGAGATCAACCCTAGACATCCGATAATCAAAAAGTTGAGTTCTCTCTCAGTCACGAATCCTGAACTCGCTGAACTAGTCACAAAACAGTTGTTCTCCAACGCGA
Hsc 70	PB.4719.1	AXU24968.1 heat shock protein 70-2 [Tytthus chinensis]	ATGGCCTTCACCGACACAGAGCGTCTCATCGGAGAAGCTGCCAAGAACCATGTGTCCATGAACCCCAACTACACTATCTTCGACGCTAAGTGGTGGATTGGTCGCA GGTACCTTCGATGTGTCCATCCTGACGATCGAGGATGGTATCTTTGAGGTGAAGTCGACAGCAGGAGACACTCACTTGGGAGGAGAGGACTTTGACAACAGGATGC
Hsp90	PB.4752.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACCTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTGGAGGAGAGGAAGGTCAAAC GAGTTCAGAGCTCTCCTGTTTATTCCTCGTAGAGCTCCTTTCGACCTGTTTGAGAACAAGAAGCGTAAAAATAACATCAAGTTGTATGTAAGACGAGTATTCATCA ACCTGTGCAAGGTCATGA
Hsc 70	PB.4798.1	AXU24968.1 heat shock protein 70-2 [Tytthus chinensis]	ATGGCCTTCACCGACACAGAGCGTCTCATCGGAGAAGCTGCCAAGAACCATGTGTCCATGAACCCCAACTACACTATCTTCGACGCTAAGTGGTGGATTGGTCGCA GGTACCTTCGATGTGTCCATCCTGACGATCGAGGATGGTATCTTTGAGGTGAAGTCGACAGCAGGAGACACTCACTTGGGAGGAGAGGACTTTGACAACAGGATGC
Hsp70	PB.4847.1	AXU24955.1 heat shock protein 70-1 [Cyrtorhinus lividipennis]	ATGAGGGTTCTCTCTTCGAAGTCAAGAGCTCTCCGAAGACTCCGCACAGCTGCTGAACGGGCCAAGAGGACTCTTTCATCCAGCACCGAGGCCAACATTGAAATTC ATCATCGAGCGCAACGCCAGAATTCCTTGCAAGCAGAGTCAAACCTTCACCACCTACGCAGACAATCAGCCAGCCGTTACCATCCAGGTGTTCGAGGGAGAGCGAC TTGAGATGGCTGGAGAACAACTCACTAGCCGATAAGGAGGAGTTGGAACACCGACTGCAAGAGTTGCAGAAGAAATGCTCTCCCTTCATGTCAAAGCTGCATGGAC
Hsp70	PB.4985.1	ACH85201.1 heat shock protein 70 [Trialeurodes vaporariorum]	ATGCCGTCGAAAATGCCTGCTATTGGAATCGACCTGGGGACGACCTACTCCTGCGTGGGCGTCTGGCAACAGGGCAAAGTGGAGATCATCGCCAACGACCAGGGCA GCCATTGCTGGCCTAAATGTTCTTAGGATCATTAATGAGCCTACAGCGGCAGCTTTAGCCTATGGATTAGACAAGAATTTGAAGGGCGAGAGGAATGTATTGATAT TTCTCTGGGAAGTCACTGAACCTCTCCATCAACCCAGATGAAGCAGTTGCCTATGGAGCCGCTGTGCAAGCTGCCATCCTGAACGGAGACCAGAGCTCAGCCATCC AGGATGGTCAATGAAGCTGAGAAGTACAAGGAAGAGATGACAAGCAGAGGGAGAAGATTGCTGCAAGAAATCAATTGGAGAGTTACATCTTCAGCGTGA

LOC114325611	PB.5207.1	Heat shock protein 70 A1 OS=Anopheles albimanus OX=7167 GN=HSP70A1 PE=3 SV=1	ATGCCAGGGAAGATCCCAGCGATCGGTATCGACTTAGGAACGACCTACTCCTGCGTGGGCGTCTGGCAACAGGGAAAAGTGGAAATTATCGCCAACGACCAGGGA GCTATTTGTGGCTTGAATGTTCTAAGAATCATAAACGAGCCCACTGCTGCTGCCTTAGCGTACGGGCTGGACAAAAACCTTAAAGGAGAGAGAAATGTTTTGATTT AAACTAGATAAAAATGCAATACATGATATTGTTATGGTCGGTGGATCTACAAGGATTCCAAAAATTCAGAGTCTTCTTCAAAACTTCTTTTCTGGGAAGCAGCTAA GTATCTGCAAAAGATAACAGCTCTGGAAAATCCAAAAACATCACCATAAAAAACGACAAAGGGCGACTATCTAAGGAAGAGATCGACAGGATGGTCAATGAGGCAC
Hsp70	PB.5213.1	ACF74975.2 70 kDa heat shock protein [Trialeurodes vaporariorum]	ATGCCGTCGAAAATGCCTGCTATTGGAATCGACCTGGGGACGACCTACTCCTGCGTCGGCGTCTGGCTACAGGGTAAAGTGGAGATCATCGCCAACGATCAGGGCA GCCATTGCAGGCCTAAATGTTCTTAGGATCATTAATGAGCCTACTGCGGCAGCTTTAGCCTATGGATTAGACAAGAACTTGAAGGGCGAGAGGAATGTATTGATAT
Hsp90	PB.5637.1	AIY24626.1 heat shock protein 90 [Graminella nigrifrons]	ATGCCTGAAGATATCGATATGAAGGAAGAAGTCGAGACCTTCGCCTTCCAGGCAGAGATCGCTCAGTTGATGAGCTTGATCATCAACACCTTCTACTCCAACAAGC TTCACCGTTCGCACTGACCCCGGCGAGCCCCTGGGTCGCGGAACCAAGATTGTCCTTTTCATCAAGGAGGATCAAGCGGAGTTCTTGGAGGAGAGGAAGGTCAAAC GAGTTCAGAGCTCTCCTGTTTATTCCTCGTAGAGCTCCTTTCGACCTGTTTGAGAACAAGAAGCGTAAAAATAACATCAAGTTGTATGTAAGACGAGTATTTATCA CCTGACCACTCCATCATGGACACCCTGAGACAGAAGGCCGACGTGGACAAGAACGACAAGGCCGTCAAGGACTTGGTCATGTTGCTATTCGAGACTGCTCTCCTG
HSP70-3	PB.5860.1	ADG03465.1 heat shock protein 70 [Bemisia tabaci]	ATGCCTGGCAAAGCACCTGCTATCGGGATCGACCTGGGCACCACCTACTCGTGCGTGGGCTGCTGGCAACAGGGCAAGGTGGAGATCATCGCCAACGACCAAGGAA ACGATCGCTGGACTGAATGTTCTCAGAATCATCAACGAGCCGACAGCTGCAGCGCTGGCTTACGGGCTTGACAAAAATCTGCAAGGTGAAAAGAATGTTTTGATTT AAAATGGACAAGGCGTCAATACACGACGTAGTTCTGGTGGGAGGATCTACAAGAATCCCGAAGATTCAAACTCTACTTCAAAGCTTCTTTGCTGGAAAGGCACTCA GTAACTGCCAAGGATAATAGCTCGGGTAGATCTAAGAATATCACTATTCAGAACAACAAGGGACGCCTGAGTAAGGAGGAGATCGACCGTATGGTGAACGAAGCAC
Hsp70	PB.5948.1	Heat shock protein 68 OS=Drosophila melanogaster OX=7227 GN=Hsp68 PE=1 SV=1	ATGGACAAGTCTTCAATCCACGATGTTGTTCTGGTCGGAGGTTCGACCAGAATCCCCAAGATCCAAAGTCTTCTACAGAATTTCTTCTCCGGGAAGTCATTGAACC
DNAJA3	PB.6282.1	GO:0031072(heat shock protein binding),GO:0051082(unfolded protein binding)	ATGGCCTCTAGGAGTTTAGTTAAATATTTCGGCTTAAAAAATTGGAACAATGTGAATTTTTCCTCCACAGGGAAGTTTTCTGTATGCTTTCGGAACAAATGCAGTC GGTGGCAGAAAGCAGGGCTTCTCCTCACAAAACTGGCAGTTCCATTCGACCATTAACCCTGAGGACCTCTTCCGCAAGATCTTTGGTGATGAAGGATTCAACAGCA ATGCCTGTGGGGACCAAGGAGCTGTTCATCACGTTCCGAGTTGACAGGTCAGATTACTTCAAGAGAGATGGTCCCGATGTTCACACGGATGCAGAGATCTCACTGT
AHSA1	PB.6448.1	Activator of 90 kDa heat shock protein ATPase homolog 1 OS=Homo sapiens OX=9606 GN=AHSA1 PE=1 SV=1	ATGGCCAAGTGGGGAGAAGGAGATCCACGTTGGATTGTTGAGGAACGCCCTGATGCAACCAATGTAAACAACTGGCATTGGACGGAGAAGAATGCTTGCGCTTGGT AAAGGGATGATTCTACCAAAAAAGGAAGATGAACAGAGTTCTCAAGACAAACAAATCACTAACTTATCTTCTGGTTTTAACAAAAAGATGCAAGTGAACCAGCAG TACTACTGGGACCCCATCAAGCGTACATTTGGCTTTGGCTCATTCTTGTGA
Hsp70	PB.704.1	XP_008477022.1 heat shock protein 70 A1-like [Diaphorina citri]	ATGGTTCTGACTAAGATGAAGGAGACAGCTGAGGCTTACCTTGGAGGAACAGTTAGAGATGCAGTCATTACAGTTCCAGCTTATTTCAATGACTCTCAGAGACAGC ATTGATGCTTTATACGAAGGTATTGATTTCTACACGAAAGTCTCACGAGCCAGGTTTGAAGAACTCTGCTCAGATTTGTTCCGTTCTACACTCCATCCAGTAGAA CGAGCCATGTCCAAGGACAACAACCTCCTGGGCACATTTGACCTGACAGGAATACCACCAGCTCCTCGCGGCGTACCCAAGATTGACGTCACGTTTGACCTTGACC
Cryaa	PB.7205.1	Heat shock protein beta-1 OS=Gallus gallus OX=9031 GN=HSPB1 PE=2 SV=1	ATGGCTGACAGTGGTATCAAACGCAATATTCCCATCAAGTTGGGTGATTTCAGCGTCATCGACACTGAATTCTCCAACATTCGTGAACGTTTTGACGCCGAAATGA AGCAAAGACGGAGTCCTGACCGTGGAAGCTCCATTACCAGCCATTGGAACAGGAGAGAAACTCATCCCAATCGCGCACCAATAA
genome assembly	PB.7397.1	GO:0031072(heat shock protein binding),GO:0051082(unfolded protein binding)	CTTCAATCTCAACATCTGAAAATGGCTGATACCAAGTTGTACGAAGTTCTTGGCGTTAGTAGAAGTGCTTCCGATTCAGAAATTAAGAAGGCATATCACAAACTAC TTAATGTGCCAGTCGTGTCACGGAAACGGCATAAAGGTGACGGTTAGGCGGTTAGGAGCGGGCATGATGCAACACGTTCAGAGTCGTTGCAACGATTGCCAGGGAC CATTACAAAAATCCTTTTGAAAAAGGCAATTTGTATATAAAATTCAAAGTAGATTTCCCAGAAAATCACTTTGTTGATGAAGCCAAATTAAAAGAACTGGAAGCAC
Hsp70	PB.7649.1	Heat shock protein 70 B2 OS=Anopheles albimanus OX=7167 GN=HSP70B2 PE=3 SV=1	ATGAACGGAGACCAGAGCTCAGCCATCCAAGATGTCCTGCTGGTGGATGTCACTCCGCTGTCCCTTGGCATCGAGACTGCGGGTGGAGTGATGGCCAAGATCATCC GAGAGTTACATCTTCAGCGTTAAGCAGGCGTTGGAAGAGTGTGATGACAAAATGAACGACAGCGACAAATCTACTGTGAAGAGTGAATGTGAGGAGACATTGAGAT
Cryaa	PB.8266.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCTCTGGTACCATACTTGATGAACGAGATGATCCGTGATTTCGAGATGCCTTTGTACGACCAGAACTTTGGCTTGGGGTTGTTCCCATCCGACATGATCGTAC ACCAACAAGCCTGCCATCAAGCCGAAGAAATCTGATAACGTAGAGAAGATGCAGTCCTAA
Hspa41	PB.839.1	XP_006561225.1 97 kDa heat shock protein isoform X2 [Apis mellifera]	ATGGCTGCAATGTCCGTAATTGGAATCGACTTTGGAAATCAAAGTTGTTATGTGGCTGTAGCCAGATCTGGTGGCATTGAAACCATAGCAAATGATTATAGTCTTC CGTATAGCGGGTTTGAACGTTCTTCGATTATTCAATGAAACTACTGCGATTGCCCTATGTTACGGTATTTATAAACAAGATTTGCCAGCTCCAGAAGGAGCTCCTA TGTCTTGAAGATTCGAAATTGAAACCAGAAGATATCCACTCGGTAGAAATAGTTGGTGGATCGAGTCGAGTTCCCGCTATCAAGAAATTAATTGAAGAAGTATTTC ATTTTTACTGTATCGTCGGCAGTGCTTGTAGAAAAACGCGAGTCCACGCAACAAGAGAAAGAAGGTGAACAACAGAACATGGAAGTCGAAAATAATCAGGACAAAA AAACAAGTGTTCACGGAACATCTATCCCATCTCAAATCCATAGGCGAACCGATCAAAGAGAGAAAACAAGAGTACGTTGGAAGTGAACAAGCACTTGACGAATTAT GAGTAA
HSPB1	PB.8464.1	BAN20225.1 small heat shock protein hsp20 family [Riptortus pedestris]	ATGTCGGTGGAAGAAACTAGAAATGTTTCAACGTTCGAGCAGCGAACATCTAGGCGAGTAGATTCTCAATCTCGATGGAAGGAACTTCGTGATAAGTTCGACAAAC CTGACCGTGGAGGCGCCCCTGCCTGCCCTGGGTGCTCCGGAGAAGATGATCCCCATCGCCCAGAACTAA
CRYAB	PB.8644.1	AIU47037.1 heat shock protein	ATGGCGGACAGTGAGAAGCGGAACATCCCCATCAAGCTGGGAGACTTCAGCGTGATCGACACAGAGTTCAGCAACATTCGCGAGAGGTTCGACGCCGAGATGAGGA

		[Phenacoccus solenopsis]	GGAGTGCTGACCGTGGAGGCGCCCCTGCCTGCCCTGGGTGCTCCGGAGAAGATGATCCCCATCGCCCAGAACTAA
Cryaa	PB.9903.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCTCTGGTACCATACTTGATGAACGAGATGATCCGTGATTTCGAGATGCCTTTGTACGACCAGAACTTTGGCTTGGGGTTGTTCCCATCCGACATGATCGTAC ACCAACAAGCCTGCCATCAAGCCGAAGAAATCTGATAACGTAGAGAAAATGCAGTCCTAA
unkown	PB.9962.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCACTGGTTCCGTACATCGTGAATGAGCTCCTGGGAGAGCTGGACCGGCCGGCGTACGACCACCACTTCCTCCACCCCTACGAGCGACTGAGGTCCCCGGCTC ACCACCGACAAGAAAAAGGAGAACAAGGAGAACGGAGTCAATAAGATGGAGAGTTAA
unkown	PB.10139.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCTCTTGTTCCCTTGATTGTCAACGAGTTGCTGAGTGAGCTGGACAGGCCCACTTACGACAGCCACTTCTTCCACCCGTACGAGAGAATGATGTCTCCACCCT GAGAAGAAGCAGAAGCTGGAAAACGGCCACTCCAACCAGATGGAGAGTTAG
unkown	PB.10345.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCTCTTGTTCCCTTGATTGTCAACGAGTTGCTGAGTGAGCTGGACAGGCCCACTTACGACAGCCACTTCTTCCACCCGTACGAGAGAATGATGTCTCCACCCT
unkown	PB.10365.1	AYP00110.1 small heat shock protein 20.1 [Laodelphax striatella]	ATGTCTCTGGTACCATACTTGATGAACGAGATGATCCGTGATTTCGAGATGCCTTTGTACGACCAGAACTTTGGCTTGGGGTTGTTCCCATCCGACATGATCGTAC
unkown	PB.10551.1	AIU47044.1 heat shock protein [Phenacoccus solenopsis]	ATGGCTCTTGCTGGTGCTGCCAAGAAGTTGATCCCTATGTTTGACAGAGTTCTCATTCAAAGAGCAGAAGCAATCACAAAAACCAAAGGAGGAATAGTTATTCCCC
unkown	PB.3053.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTTTAAATACGTAAACCAGTTTGAAACATCATTTGTAGTTTTACTCCATGTTACAAGCGGTGCCAGAGCGGTACCCGCAGGCGCTCTTTCTGGCCTCACTTTAA TATCAGCTGCCTGAGGATGTAGACGTGGAGTCACTGCAGTCCAGCCTCACTGCGGACGGGGTACTGCAGCTGATTGCTGTCAAGAGGACTGAGACCGAGAAGGAGC
unkown	PB.3369.1	XP_021923360.1 activator of 90 kDa heat shock protein ATPase homolog 1 [Zootermopsis nevadensis]	ATGATTCTACCAAAAAAGGAAGATGAACAGAGTTCTCAAGACAAACAAATCACTAACTTATCTTCTGGTTTTAACAAAAAGATGCAAGTGAACCAGCAGAACAAA TGGGACCCCATCAAGCGTACATTTGGCTTTGGCTCATTCTTGTGA
unkown	PB.4105.1	GO:0031072(heat shock protein binding),GO:0051082(unfolded protein binding)	TTCAAGACGGAGGGACTACCGCAACGCGTACGACGAGGACGAGGGACACGGGCCCTCACGCGTGCAGTGCGCAACCAACTAGACCCCACGCAATGGGCTGGGCCCA
unkown	PB.444.1	Heat shock protein stil homolog OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812 GN=stil PE=3 SV=1	ATGGATGCAGAACAAAGAGTACGCTCTGATGCTGAAATTAAACACGATTTAGCATTGGAGTGCAAAACTTCAGGAAATGAAAAATTTAACAAGAATGATTATAAAC ATTGCTGAGTGGTCTGATATTGATGACGGTGATGGTGATGATGATTTGCCTCCTCCATCACAGTCCCTCCGTGTTGGATCAGGTTTTCGAACAGAAAGATTACCAA TTCCGCAAGGCGGCTGAATGCTACAAAGCCGCCAGAGAGCGGGTAGGAACTAAGCTTACTGTAAGTGCCCTGCATGAAGCCCTCTCACTTAATTATCGTAACGTTA
unkown	PB.7772.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCTCTGCGCATTGTTCTTCGCCCCAAAGTCATCGACCAAATCAGACCTCTGGTTAATAAGGGACAATATGTTCAACGAAGCTCAACGAGATCGTACCATAGAC TCACTGCAGTCCAGCCTCACTGCGGACGGGGTACTGCAGCTGATTGCTGTCAAGAGGACTGAGACCGAGAAGGAGGTGAGGAGCATCCCCATCACACGGAGCGACC
unkown	PB.9475.1	ABD98776.1 putative small heat shock protein [Graphocephala atropunctata]	ATGTCTCTGCGCATTGTTCTTCGCCCCAAAGTCATCGACCAAATCAGACCTCTGGTTAATAAGGGACAATATGTTCAACGAAGCTCAACGAGATCGTACCATAGAC

S4 Glycoside hydrolase genes and their coding sequences

Gene name	GeneId	Annotation	CDS
Nfu_g_1_008143	PB.5135.1	Myogenesis-regulating glycosidase OS=Homo sapiens OX=9606 GN=MYORG PE=1 SV=2	ATGGCTTCCCTGACGTGGATTCTGCTGCTTTTCATAGGTGGCAGTGATATTGTTTTTGCAAATGCAGAAAAGATATTTACCGTGGGAAAAGACTTCCAAATAAAAGT GGTTTTATGTCTATGTCCACAGTACTACTCCTCTGTTCATAGACGCCAACCACCACCTGCCTAAATCTCTTTGCTTGGAAGCCAAACAGGCTCCACCGTACTTAGCT GTTTGATACCTTGCAGAAGAAGGGATTCTTGGTGAAGGACCAGAACGGGAATGTAGTGAGTGAATGGTGGGATGGAACTGCTGGAGTGGTGGACTTCACCAACCCAC AAGGAACTCTTTATAAGATGGCTTCAAGCCAATACCTTCATGCCGTCCATCCAGTTCTCTGCTTCTCCGTGGGATTTTGATGATGAGACTGTAGAAATCTCTAAGAA
Nfu_g_1_008143	PB.6342.1	Myogenesis-regulating glycosidase OS=Homo sapiens OX=9606 GN=MYORG PE=1 SV=2	ATGATTCAACACCCAATCTGGTCAACATGGGTCAGAGACAAGTACGATGTCAATGACACCGATGTCCGAGCTTTTGCTAAAGCTATTGTCGACCACGGTTTCAACAA CAGCGAAAATTGACGAGCAACCTAACGCGTTCTCCAAGGCCTACGTCAGCACCGTGGCCAAATTCAACAACATAGTCGAAGTAAGATCTGCTTCCAATAGTCAAGAC CACTGCACATTCTATTTGGTCAGAATACATGCTGGGAGAGGATATTCTGGTGGCTCCAGTGGTGGAGAAGGGGGCGGTGACCAGGGACATCTACCTGCCTCAGGGCA
unkown	PB.2023.1	Myogenesis-regulating glycosidase OS=Homo sapiens OX=9606 GN=MYORG PE=1 SV=2	ATGGATGTAGAGCAGGACAAGGAGAAGGAGGAGGGGGTGGAGGAGACGAGCGCACTGTTGGGTGACGTGGACACGATGCAGTCCACCGTGAGACTCAACCGGAGGAA ACAAAAAGCCGAAAGAATACAAGCTGCAGGGTCTTGTGGGATTGCTGTTCCTCACAGTCGTGTTTATGGTTGGATTCGCTCATGTCTTTTACCATCAACAAGTTTTA TGTCACTGGTCAGGTCGGCAGACACAGGTGGGGGGGAGTGTTGCGCCGGTACTTTATCAATTCCCGGGGTGTTGCCATTTCCGTGGACCATGAGACTCCTTTGTACC

			ACGCTGGACAAGACCGTGGAGACTTTGCACAGACGAGGGTTCCGGATAGCGCTTACAGTGCAGCCGTTCATCGGTACCGAGTCCAAGAACTTCGCCCCGGCAGTGCC TCATCATCCCTACAGTGCTCACCTTCGGCATCATCGGCTACCCCTTCCTGATGCCGGGCCCGGTCGGCGGCGACTACGTCATCAAGGACCCCCAGCGCTTCTCCACA CGAGGTGTACCTGCCCGCGGGCGTGTGGAAGGACGGCATCGACGGCAGCCATCGCAAGGGCTCCCGCTGGATTCACCACTATAAAGTTCCTCTCGAGAGAATCGCAC
unkown	PB.3560.1	Pseudouridine-5'-phosphate glycosidase OS=Clostridium botulinum (strain Langeland / NCTC 10281 / Type F) OX=441772 GN=psuG PE=3 SV=1	ATGATTGTAAGAAAACTTCACAGGCTGTCAAAGTTCATTGAAATCAGCTCTGAGGTAAAACATGCCCTGGAGTCAAACAAAGCAGTTGTTGCACTGGAAAGTACAAT TCTGCAGTGGTGTCAAATCCATCCTGGACATTGCACGTACTCTGGAGTATCTGGAGACACAAGGTGTGTGTGTGGCGACGTACGGTGCGACTAAGCAGTTCCCTGCC ATTTTCAAAATCTAAAGAACAAGAGCGTGTATTTGCCATACCAAGGGCCTCAAAACCACAACCAGTGGTAGTTGGCGCCTCAATCGTAGATTCAGTAACTATTCCTC ATGGAACACTGTGCAAAATATGGAATACCTGTGATATTTGAGCCGACGGATGTAGCCAAGGCTGCCAAGCCATTCCACAACAGTCTTTGGAAGTCTCTTCACCTCAT CTCAGCACTTCTCCTCGTCTCCGCAGTCAGCTGACTACACCCTTGTCAACTCTACCGTCAGGAGATCCTTTCTGTGA

S5 AchE genes and their coding sequences

Gene name	GeneId	Anotation	CDS
unkown	PB.2154.1	Acetylcholinesterase OS=Homo sapiens OX=9606 GN=ACHE PE=1 SV=1	GTCGTCAACATGTTCTCTGTACGTCTAGTCTTCAGCTCTCTGGTTCTGGCAGTCTTCTCCAGTCTAGTACTTACCATAGGTGAGGGAACAGTGCAGAATGTGACAGTGGTGAC CAGCAGGTTTCCTGACACTGGGCTCTGCTGTGGCTCCGGGTAACCTGGGACTGCACGACACACTGCTTGCTCTGCAGTGGGTTCAGGACGAGATCTCCAGCTTCGGGGGAGAC GCTGCAGCAAGCCCTCGCCCGCGGCAAACCAATCCTCACCGGCTTCACTCGAGACGAGGGCATCATCAAGTTCATTGACGATGCGTGGACGTTAGCCGAGAGTAACCTGGCCG CGGTGGACTAACTTCGCTAAAACTGGAGTTCCCGCCGAGGACTGGAGCACTGTCAGCCAAGGCGGATATCTACTCATAGACTCCTCACCGAGCATCATGAATGTAACAGAGTTC
unkown	PB.3228.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	ATGTGCTCCAGGAATTCCCAATTAGCTTTCGTAAGTTTGTTGTGGTTACTATTGACATCATTTAATTACTCCTCAGCAGAAACAATCACAGTGCAAATCTCTGATGGAGAAAT TTACACTGGGGACCAGTGAGGTTCCAGGCAATGCTGGAATGTTTGATCAAGTACTCGCCATGAAATGGGTTCAGAGTGAGATCTCCAAGTTTGGAGGAGACCCAAACAGAGTG CTCCGTGCCTTTCCTCATGGGGGCCAACAGAGACGAAGGCACAGTGTCATTTCTTCTAGGAGCTGCGAGTCAGACGAAGGTCAGTGAGTTCGCCACTAACATGCAATATCTTC ACGAACTTTATTAAATACCAAACTCCAACTCCTTCATCAGAATCTCTTCTCCAGAACTTGAGTTGGACTCCAGTGCAGAACAACTCCCTCCACTACTTGGAGATCACTGACAA
Ache	PB.3546.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	ATGCCGATGAGGGGAAACCTATGCTCTTCGACGAGGTACCGGACGCTTAGGGTGCTCTGCTTCGCGGTGCTCCTCTACCTGGACGCGATCGAGTGTTCACCCTCCCATAGGAG ACACCAAATTGAGCGAAGACTGCCTTTACGTGAACGTGGTCACCCCACGTCCAAGGCCTACTAACGCCGCCGTCATGGTATGGATATTCGGTGGTGGATTTTATTCGGGTTCG TTGTCCGCACGATCGTAACAATTTGCGCGAAGTGATCGATTGTCTGATCAAGAAAGACCCGGGGGTTTTAGTGAACAACGAATGGGGGACGATGGGCATTTGCGAATTCCCAT TACATGTACTACTATAAGCACAGATCCGTCAATAATCCTTGGCCGAGGTGGACAGGGGTAATGCACGCCGACGAAATAAATTACATTTTCGGAGAGCCTTTGAACCAATCGAA TAGGGCTACATCATTGCTCGTACGACACGCGCAGACTTTCCTACGGGGTCTAG
ESTF	PB.4514.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	ATGTTGTGTCTAAACCACTCTGTTAGGTTCTTAGTATTCACTACAAGTGTCTTTTGGTTACTAGTGACCTCAAACAGTTCAGTCACTGCAGACGAACTTACTATACAAATATC TAGGTTTCCTGACTCTGGGGACCTCTGATATGCCCCGGCAATGCCGCGGTGTTTGATCAAATCCTCGCCATGAAGTGGGTCCAGAAGGAGATCTCCAAGTTTGGAGGAGATCCG TATCCTCATGGGCAACAACAGAGACGAGGGCAGTGTCGGGTATCGAGTTACACAGTTTGATCAGGTAACAAAGAATATTGACTATTACATGCGTGATTTTATCCTTCCTGAAG AACTTCATGAAATATCGAGATCCGACTCCTTCTTCTGACTCTCTTCTCCAGAACGTGAGTTGGACTCCAGTTCAGGACAATGTTCTCTCATACCTAGAGATCTCTGACAATCT
unkown	PB.5988.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	ATGACATGCCTAAATATATTTACTGTGTTGTTGGCTTCGCTATCAATAACAACATGCAGTTTGGCTAGTACAGATACAGTCACAGTGCAAATCTCTGATGGAGAAATCCTTGG TGGGAACCACAGGAGTTCCAGGAAATGCTGCTCTATTTGATCAAATACTCGCCTTGAAGTGGGTTCAGAAAGAGATCTCCAAGTTTGGAGGAGACCCGAACAGAGTGACGCTT GCCTTTCCTCATGGGCACCAACAGAGACGAGGGCTCAGTCATGTTAGCATTGGGAGTGACCGAGACGAACGTGACTGAGTTTGCTGCCAGTATGGACCTCTTGATCCCACCTG ACCAACTTTATCAAATACGGAACTCCAACTCCTTCATCAGAATATCTTCTCCAGAGCGTGAGTTGGACTCCAGTGCAGAACAACTCCCTCCACTACTTGGAGATCTCTGACAA
LOC124221656	PB.6017.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	GTGGTTTGTAAAATGTCATTTCTAACACTAGTCCTGTTAATATGTAGTTTCAACACCGTGTTATGCCAAAACCTCGATGTCCCCTTGGCTACAAAGAAGGGGCCTATTGTCGGC GGTCTAAGGTGGCCCCTGGGAACCTAGGCCTACACGACACTCTGCTGGCCCTGCAGTGGGTACAGGGCGAGATCTCCGTCTTCGGGGGAGACCCGACACAAGTCACACTCTTC GGCAAGAGGCAGAGCAATGATCGCTGGGCACAACTTGGACGAAGGAATTCTCAAATTTGAATACGACTATTGGAAAGACGCAGAGAACAATCTGGCTGCTTTCATCCCTTACAC GCCAAAACTGGCTCTCCAGGCAGTGACTGGTCCACGATAGATCAAGGAGAATACCTTGTCATAGACTCCACAATCTCTTCCATGAACGTGTCTGCGTTCGAAGATCGCTACTAC
unkown	PB.6683.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	AAACTGACTCTACTATCGAGGAGTTGTGTACATGTTAGCTCAAACATGTGTACAAGTAGCCCCGTTACATTGTTGGCAATACCTACGAGTTTCTGTATTTGGATATTAGTGGC TTGCTGAAGGAGTGGTCGTGGTCACTATCAACTACAGGCTTGGAGCACTCGGATTCCTAACCCTAGGGACCAGTGAGGTGCCAGGTAATGCTGGAATGTTTGATCAAATACTC GGACAACCCCTTCCTGCCAGCGACTCCTGCAGAGCTCATCGCCCAAGGGCAAGGATCTGCCGTGCCTTTCCTCATGGGTACCAACAAAGACGAAGGAACTGTAGCTTTTGGTA AACCAAACTGGGTTTGCTCCAAACAGCAGAGAAATGATAACTTGGAACAAGATGGTGAAGCTTTTCACGAACTTCATTAAATACCAGAATCCAACACCTTCTTCAGACTCTCT
LOC100494782	PB.6599.1	Acetylcholinesterase-1 OS=Trittame loki OX=1295018 PE=1	ATGAGAGGACCTGCGATGGTTTGGGTGGCAGCCGTGGCTGGAACACTATTGAGTATTGTGTCTGCACAAGTATGGCCACATGTGACGATATCCAGTGGAAACATCATTGGACGC CTTTAGGATTTTTGACTCTTGGGACATCCGAGGCTCCTGGAAACGCAGCACTTTACGATGCAATCCTGGCTCTAAAGTGGGTCCAAACAAACATCGCACAGTTCGGCGGAGAC

		SV=1	GAACCAGGGCTCCAACGTGCCCATCCTCAGATGA
unkown	PB.6798.1	Acetylcholinesterase-1 OS=Trittame loki OX=1295018 PE=1 SV=1	ATGGCTCTGGATGATCTCGTCAACGCGCAGAACAACGTCAGTGTTGGCTTTCAGGACTCGGTGACCGGCAACTGGATAGCGTGGAGAGCCAATGTGGATCAAGGGTTGGTTTCA CAGGCGGTTTCACAGGAACCGACCCTCTGGCCCAGAAGTACAACACTACATTTATCGGCCACGGGTCGGATCTTCAGTACCTGTTCGTCCCTGACGGTCTGACAGCGCTGCCTC
unkown	PB.3677.1	KO1049 ACHE; acetylcholinesterase [EC:3.1.1.7]	ATGGCATTTCAAACAGTAGTAGTGTTAATATTTGTAATCTTTAACAACGTGATCTGCCAAACCACCGATGTTCCCTTGGCTACAAAGAAGGGGCCTATTGTCGGGGTACGTTAC TGGCTCCTGGGAACCTAGGCCTACACGACAATCTGCTGGCCCTGCAGTGGGTACAGGACGAGATCTCCGTCTTCGGGGGAGACCCGACACAAGTCACACTCTTCGGCGAGAGC CAAACCAATGATCGCAGGCCACAACTTGGACGAGGGAATTCTCAAATTCGAAGAAGACTACTGGAAAGACGCCGAGAACAACTTGGCTGCGCTCATCCCTTACAATGCTCAAC GGTTCTCCAGGCAGTGACTGGTCCACGATAGATCAAGGAGAATACCTTGTCATAGACTCCACAATCTCTTCCATGAATGTGTCTGCATTCGAAGATCGCTACTACAGCTTCTG
unkown	PB.6010.1	Acetylcholinesterase OS=Homo sapiens OX=9606 GN=ACHE PE=1 SV=1	ATGATGGGAATAAGCTCGATTTACCTGATACATCTTTTAATTCTTGGATTTTGCGTGATCAATAATTTTGGCCAGGATGCTCCTCCGATAGTGGAGATCGACAAAGGTAAAAT GTCTAGGAGCTTTTGGGTTCTTATCATTGAATCATCCTAACGCAACCGGAAACACCGCCTTGAAAGATATAACCTTGGCGTTGAAATGGGTGAGGAATAATATTAATAAATTTC AAGGAATTTTAGTCAAGTCCCATTGATGATGGGCTATATGAAAGATGAAGGGATATTCTTTTCCGGAGCCGAACTTATGAAGATTCGCTTTGGAGACATCCCCGAGAATGAAA GGTAGACAATTGGACATCAATAGAGATCTACGAATGCATTCGCGCTTGAGCAATTCGAGGGCTAGATTTTTTGGAATTCTCATGCGCAATAATGCCGAAGTATCAACGGCTTG
unkown	PB.6721.1	Acetylcholinesterase OS=Culex pipiens OX=7175 GN=ACHE1 PE=2 SV=2	ATGGTCAATGCGTTTGTATTTCCGTTGCTGTTAGTGTTAGTGTTGTGCAGCTGTTGGCAACAGTTTAACAGATGACTCTCCCGAGGTGACAATCACAAATGGCGTCATCAG CTTTGGGCACGTCGGAAGCGCCTGGAAACGCTGGCTTGAAGGACCAGCGACTGGCGTTGATTTGGGTTCAAAACGAAATATCCAAGTTTGGAGGAGACCCGGCCAAAGTAACC CTCCCTTATACCTTTTCTCATGGGTTTCAATAGGGATGAAGGACTCATTGGACTGCCAGAGCTACCTGCAGCAGAGGCAAATGTTACAAGCCTTATGACTTATTTCATCCCAC AACTTTGTAATTAATGGGAGTCCGACGGTAGGGGCTGACGATCTACTGCAGAATGTAGACTGGCTCCCTGTGAATCAGACCTCTACCCGTTACATGTGGATTGACAAGGATCT

S6 Sulfotransferase genes and their coding sequences

Gene name	GeneId	Anotation	CDS
PIP	PB.3627.1	GO:0008146 (sulfotransferase activity)	ATGGATTTTCGTCAAAGAAGGTTGCTGATGCCTAAAAGGACGAGTGAGTTGGTGGCACTGATGGCGATCTCGAGCACGCTCTTCCTGTTCCTACACACGAGGGAGTTGCACAC TCATCAATTTCACTCACTTCGGACTGCCGGAGCCCATCTACGTGAACTTGGTGCGGGACCCGGTGGAGAGGGTGATCTCTTGGTACTATTATGTCAGAGCTCCCTGGTACTAC CCCTGTGAGCGAGGAAGTCAAGGACTTGGTCAGACGTAACTTCACCAGAGAGATCGACTTCTACCTGTTCTGTCGGCAGAGACTGCACAGACAGCTGTTAGCTCTGCGCCTGC
LOC122371912	PB.8342.1	GO:0008146(sulfotransferase activity)	GGCAGTGTGTGCCCGCCCTCTCTACCTACCACACGTACGATGAAGGAGATAACGTTCACGCCGCTAACTGACGACCTGTCAGAGGAGCTGCGAGTGTCCTGCAAGCCGGGACT AGCAGATCTGGACCAAGAAACCAAAAATAATCTATGTGACACGAAATCCAATGGATGCAGCGATCTCTTACTACCACCATCACAGACTTTGGAATGGTTATGTTGGATCTTAT CAAGCTGTGGACTAAAGAGAACCTCAGAGGCACCACCTATACTGAAGACGATTTTTACTTTTAA
LOC124368504	PB.7537.1	GO:0008146 (sulfotransferase activity)	ATGCCGGCCGCGGTGGACTACACCCCCGTGGGTGGTGACGCGGAGAAGCTTCTAAATGAACACTTCATCACAGTTTTCCGCCACGGCTATATCCGGTGCCAGGGGACGGTACT GCAATGTGAAGGACACTTGCTTGTCCTACTTCCACCACTGTCAGCTGCTCGAAGGATACAAGGGCAGTTTCCACGACTTTGCTACACTTTTCATCAACGACGCATTGTGTTTT
LOC124355184	PB.697.1	GO:0008146 (sulfotransferase activity)	ATGGCTGCTGTCACAGTGCGCTGGCTACGACGTTGTGTCACCTCTCTACTGCTCGCACTCTTCTGTCTACCCCTGGCTGTGCTCATATTCGACACGCTGCCACCGCGCGCGCA TAGTGAGACATCCTCTGGAGCGTTTGCTGTCTGCATACCGCGACAAGATACAGTACGCTTTGCCCAACACGCACCATGCAAAGCTGGGGCAGAAGATCATCCAGATGTATCGC TTACACGCTGGACCGATACTTGGAGCTGGCAGCAAACGAAAACTCTGCGAAAGCTGTGCTGGTCTGA
isolate FS-2020 chromosome 28	PB.9108.1	GO:0008146 (sulfotransferase activity)	ATGATGGACGAGGAGAAGCAACCCTCATCACTGGAGATAACCAACATTGACGAAGAGTTGAACAGGCAACTGCTGGAGGACTTTACGGGTGAGAGGACCGGGTTCTTGCGGGT AGAACCTGCTGCATGTTGGATGCAAGGTGGTGTACGTTGCAAGGAACCCGAAGGACGTGGCCGTGTCCTTCTACCATCTCAACCGACTGATTCGCACTCAAGGTTACCGAGGC GATCAATGAGAACCTGAGGAAGACAGACTTAAGGTTCCCCGGCATGTTCGAAGCATAA
genome assembly	PB.6232.1	GO:0008146 (sulfotransferase activity)	ATGTTCGCAATGGATTCTAGTCTCTGGCGCCAAAGGCGTCTTGGTATACCGAAAAGAACTATCGAGTTGATCGCCCTGATGGCGATATCCAGTTCTTTGTTTCTTTTTCTACA CCATGGTGAACGACTTCGTCGAACCTTCGGTCTACATAAAACACGTGTGTTTTACTAATTTCACCAAATTCAAACTACCAACACCAATCTACATAAATGTAGTGCGGGACCCG TTGGGGTCAACTCAACTCGTTCACAAAAATCAACCGAAACTTCTTCAAACCGCCCGTCAGCGAGGAGGTGAAAGACATGGTCCGGGCCAATTTCACTCGTGAAATGGAGTTTT
TPST	PB.3299.1	K01021 TPST; protein-tyrosine sulfotransferase [EC:2.8.2.20]	ATGGGTCGAGGGGGGCGCAAGTTGGTGCTGCTGGTGGCTGTGGTGGTGGGCTTCCTGGTGCTCACATACAAGCTGCAGCTCTGCGGTCCCCCCGACAGACAAGCCATCATGGT CCAAGTTCATCTTCATGGTGCGGGACGGCCGGGCGACAGTGCATTCCATCATCTCTAGACAGGTGACCATCACAGGTTTCGACCTGACGAACCATCGCCAGTGCATGCAGAAG CGCAGACAACACGCGGCGCATACGTGACCAGAACTCACTCTGGGAGCAGCGTGCCAGCAACCTCAACCTGCCTGCGGAGCAAGTTGACGACATGGCCACCACCACACACGACT
unkown	PB.8193.1	K01021 TPST; protein-tyrosine sulfotransferase [EC:2.8.2.20]	ATGAATGCCTCAACCCACGTAGTGCCTTGCCAGCGATGTGGCGGCTATAAGGAGGTGGTGTGGTCGCGGAGGGGACGCAGCAAGCTGACGTTGTTGATGGCTGTGCTGATATT TCATGTTCCCCAACGCCAAGTTCCTCTGGGTCGTGCGAGACCCCCGCGCCACCGTCTACTCCATCATATCCAGGCAGATACACGTGAGCAAGTTCAACCTGGACAGCTACTCA

			CCTGCAGTCGGAGTTGGGTTTTTAG
unkown	PB.5763.1	GO:0001733 (galactosylceramide sulfotransferase activity)	ATGATAACAGTCATGCGGCAGGTTATGTGCAGGTATCCTCTCGCCAGCGGCATTTTCGTGTTGGGGTTCACATTTCTCACGTTAAATCTTATTTCAGACAGGAATGTTTACCA TGTACGAGTCCATCTATGATTACTACAAGTTTCAGAGGGTTTTTAATTTGTCGTTTGAAAATCTGTTGAGTTCGCCAGAGATTCCGGAGAGGTTTAACAACAGATATTTTGAC CAGGTTACATTTCCGGTGTGTTCAAAGTATGACAATGAAGGGGTTTGGCCACACGCAGGCGTATAAACTACGCGACGAGAGCGACAGACAATGTTTCTACGCAGCGCGAGGAG
unkown	PB.1788.1	GO:0008146 (sulfotransferase activity)	ATGAGGTCGAGAAGATGTCTTCAGTTGGGTTCATTTTTCAAATGTATGTTATTCGTCAGCTTTACGGTAGTCTATTGCTACTTTGCCTACATCACGCTGCCCTCCACCAAACC TCGCCCGAGACCCGCTTCACCGTCTTCTCTCCGCGTATCGCAACAAAATCGAACGAGTCCACTCTCAGTACTACCGCAGACTCGCTAGGGCTATCGTGATTCGGTACCGAGGC CGACTCGTGA
unkown	PB.7033.1	GO:0008146 (sulfotransferase activity)	ATGGATCTGATATTCGAGGAGCTGCCGGAAAATGACATTGCGTCACGTGTACACAAGTTGATCAAGCCACATCCTCACCTTAATGATCTCATCATAGAGAAGAACACTGGGAT ACATGTACTCGTACTCGGGTCCTCTAGATCTGCTGCTAGAGTCTTTCTACCAGGAGAGAAACATCTTTGGTTCCTACTGGAAACACGTTCTTGACTTCTGGGACCTGAGGCAC
unkown	PB.9094.1	GO:0008146 (sulfotransferase activity)	ACAGTGAAACAGTCGCAACACAACATGGAGGATGTAGAGAAGCAAGAAGTGCAGCAGGGACAGCGATCTGAGGAAACGACAGCTGAGGAAGTGGCTCCTGAGGAAGTGGCTGC AGGGGGCGGAGCAGGAGGACACGGTAGGCCAGGTGGAAGTTGCACCCTCTCCTCGCTACATCAAGTCTCATCTACCCCTGGAGCTGCTGCCTGCCCAACTGTGGACTCAAAAG AACTGGGAAGGCTGGAGAATGGCGGACTTTGATGACGCCGGCTCTTGCAGAGAAGTTTGATAAATGGACCAAAGAGAAACTTGCAGGCACGGACTTCAGCAATGTCTTCATGA
unkown	PB.3924.1	GO:0008146 (sulfotransferase activity)	ATGTTGATGGCTTCGGCTATATCCACAAGGAGGTTGGATGCTCCTTTCACGCCAAGAGGAAGAGTTTGTAATCCTCCGAAGAACAAAGAGCCTGAGGTGACCCCCCTGGAGCC CGGAGAACATGAACTCCCCTCGTCACTTCAAAACACACCTTCCCTTCAGTCTGCTGCCTCCCTCACTGCTGGACCAATGCAAGGTGATATATGTCGCGAGGAATCCGCTTGAC AGGAAATAGTGAGTTTTCCGAAGAGTTGGAAGCAGAAGCAAACAAGTGGATCTCTGAAAATCTCTTGTCTCTAGATCTTATTGATTTCAAATTTCCCAGCAAAGAGTAA
unkown	PB.5910.1	GO:0005525(GTP binding),GO:0008146 (sulfotransferase activity)	ATGTTAGCAACAGTTTTCCAGAATGGAACTAAGCCTACCCTAGGCCTAGGAAATGACATTTATTCATTTTTAAAATTATATTCTTCAAGAAGCAAACCTTGGCCCCACTGCTA GAGGCAGGCTCGATCCCAAGGTCCTACGCCTGCTGTATCTCTATCCGTACAAGCACACGATACTCATCTTGAACAAGGTGGACACTGTGAAGAAGAAACGTTACTTGTTAGAG GGTGACGTGTCCGAACCTGCGCTGTGAGAAGATGGTGTTGAAGCGGATCAGACTGATCGCCCAGGAGACAGAGGCGGAACTACGCAACACATTCCTCACACAAGTGCTGGTCA
unkown	PB.7074.1	GO:0008146 (sulfotransferase activity)	ATGGCGGTGCAGTTTGAGCGACTGACCGACCCGGTGAGCATAGAGTTGCAGAGCTTGTGCAAGGGAGCGATGCCGGCGGGTGAGGTGAGGGTACTCCCTGGAGGCACTGTCAT ACCACTTCAGAACATGGAACAACTACTCCGGTCCTCAAGACCTCTTCCTTAAAGCATTCGTTCAAGACAGAGTGGTTTTCAGTCCGTTTTGGAGACATGTATTATCATTTTGG
unkown	PB.7406.1	GO:0008146 (sulfotransferase activity)	GTGGCACAGCAGTTCACACTGTTGATGCTTAGATCGTGTGTCGTCTGCTGTGTTATTGTTTTTACAGTGGCAACTTACGCCTTGCAGGACAGGATGGCGGTGCAGTTTGAGCG CAGTCGAACTGCTGCCGAAGCAGATCTGGACCAAAAAACCAAAGATCATATACGTATTCAGGAACGCCAAAGACGCAGCGATTTCGTATTTTCACCACTTCAGAACATGGAAC GATGTACGACGAATGGACAAGCAGAAATCTCGAGGGCACGGACTTCCCGAAACCTCACTGA
unkown	PB.2829.1	GO:0008146 (sulfotransferase activity)	ATGATGTCTGTACGATGCAGCTACGCTTGCAAGATTGCTTGTTTAGTAGCTTTTATCATCGGTCTCAAACTTGCGTCTTATCTCTCGAGGCAAGACATGATGTCTCTCAGTTC GAGGAGAGCCTCTGGGCACACCGGCAGTACGAGCTCTCAGGAGTCTACTCACCTGCAGATACAGTGTGCTAGAGAAGTATCTAGAACGCAGCAAAGAGTACACTTGGCTGTTC GCCAGACATGTTCAAGTTCTTCGGACTGTCGTATCACGAGTCTGTGCGCAACTTTGTGGACACACACACCACCAAGACGAGCAAGAACAGATACAGCACGTACCGCAACTCCA