



Database tool

WaspBase: a genomic resource for the interactions among parasitic wasps, insect hosts and plants

Longfei Chen^{1,#}, Kun Lang^{1,2,#}, Shoudong Bi^{2,#}, Jiapeng Luo¹, Feiling Liu¹, Xinhai Ye¹, Jiadan Xu¹, Kang He¹, Fei Li^{1,*}, Gongyin Ye^{1,*} and Xuexin Chen¹

¹Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, Zhejiang University, 866 Yuhangtang Road, Hangzhou, 310058, China and ²Anhui Agricultural University, 130 Changjiangxilu, Hefei, Anhui, 230036, China

*Corresponding author: Tel/Fax: +86-571-88982679; Email: lifei18@zju.edu.cn

Correspondence may also be addressed to Dr Gongyin Ye. Tel/Fax: +86-571-88982679; Email: chu@zju.edu.cn

#These authors contributed equally.

Citation details: Chen, L., Lang, K., Bi, S. *et al.* WaspBase: a genomic resource for the interactions among parasitic wasps, insect hosts and plants. *Database* (2018) Vol. 2018: article ID bay081; doi:10.1093/database/bay081

Received 24 March 2018; Revised 27 June 2018; Accepted 10 July 2018

Abstract

Insect pests reduce yield and cause economic losses, which are major problems in agriculture. Parasitic wasps are the natural enemies of many agricultural pests and thus have been widely used as biological control agents. Plants, phytophagous insects and parasitic wasps form a tritrophic food chain. Understanding the interactions in this tritrophic system should be helpful for developing parasitic wasps for pest control and deciphering the mechanisms of parasitism. However, the genomic resources for this tritrophic system are not well organized. Here, we describe the WaspBase, a new database that contains 573 transcriptomes of 35 parasitic wasps and the genomes of 12 parasitic wasps, 5 insect hosts and 8 plants. In addition, we identified long non-coding RNA, untranslated regions and 25 widely studied gene families from the genome and transcriptome data of these species. WaspBase provides conventional web services such as Basic Local Alignment Search Tool, search and download, together with several widely used tools such as profile hidden Markov model, Multiple Alignment using Fast Fourier Transform, automated alignment trimming and JBrowse. We also present a collection of active researchers in the field of parasitic wasps, which should be useful for constructing scientific networks in this field.

Database URL: <http://www.insect-genome.com/waspbase/>

Introduction

Insects are the most widely distributed animal species on earth. Most insects are herbivores that cause huge yield losses when feeding on crops. Insects such as houseflies and mosquitos are vectors of pathogens that cause disease in humans and domesticated animals (1). To combat these insect pests, many methods have been developed, and some of which are used in agriculture. Insecticides are one of the main methods of pest control in agriculture. Unfortunately, overuse of insecticides causes serious environment pollution and food safety problems (2). Therefore, alternative, environment-friendly pest control methods should be developed.

Biological control is an environment-friendly pest control method. Parasitic wasps are well-known biological control agents (3, 4) as they are effective natural enemies of many economically important insect pests. Parasitic wasps are a group of hymenopteran insects that lay eggs in or on the bodies of hosts (5). The wasp larvae feed on the host until pupation and eventually kill the host (6). However, pest control using parasitic wasps has some apparent disadvantages such as wasp development lagging behind pest outbreaks and low-control efficiencies. Understanding the antagonistic interactions between parasitic wasps and their hosts is an important task to improve control efficiencies (7). At present, the genomes of 34 parasitic wasps have been deposited in public databases such as National Center for Biotechnology Information (NCBI). In addition, the genomes of six hosts of these wasps and eight plants that are damaged by these insect hosts are available. Among these species, five parasitic wasps (4,8–11), six insect hosts (12–18) and six plants (19–24) were publicly reported.

Though these data can be retrieved from NCBI, they are not well organized and thus have not been fully explored. Here, we collected the genome and transcriptome data of 34 parasitic wasps, 9 insect hosts and 8 plants from NCBI, i5k workspace@NAL (25) and InsectBase (7). Then, we constructed a database, which we named WaspBase, to serve as an integrated genomic resource for a tritrophic system of wasps, hosts and plants.

Data resources

Genomes

We collected the genome data of 12 parasitic wasps from the NCBI including *Ceratosolen solmsi*, *Copidosoma floridanum*, *Cotesia vestalis*, *Diachasma alloeum*, *Fopius arisanus*, *Microplitis demolitor*, *Macrocentrus cingulum*, *Nasonia giraulti*, *Niphoparmena longicornis*, *Nasonia vitripennis*, *Orussus abietinus* and *Trichogramma pretiosum* (Figure 1) (8, 9, 11). The gene annotation files were

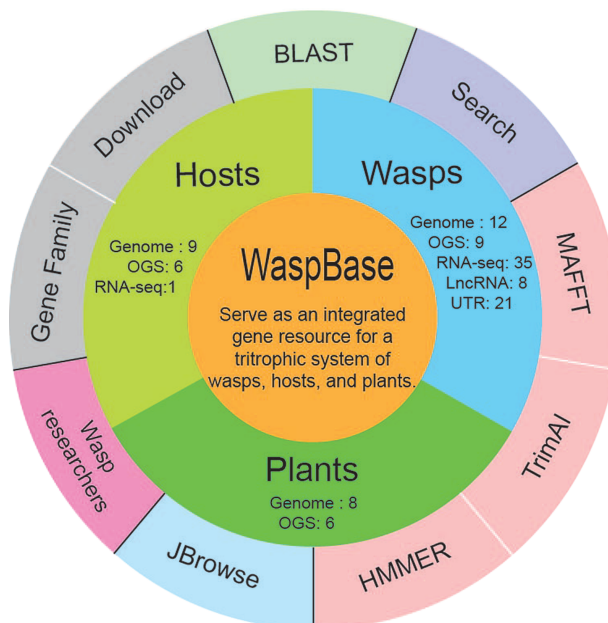


Figure 1. The design of WaspBase. The diagram shows the data and software used in WaspBase.

obtained for nine parasitic wasps including *C. solmsi* (8), *C. floridanum* (10), *D. alloeum*, *F. arisanus* (4), *M. demolitor* (11), *M. cingulum*, *N. vitripennis* (9), *O. abietinus* and *T. pretiosum*. We then focused on these nine parasitic wasps with gene annotation information. There are nine insect hosts for these nine parasitic wasps, of which five have genome data and five have annotated genomes (12, 13). These five insect pests damage eight crops all of which have genome data, but six have annotation information (Figure 2). So, we collected a final genome data of nine parasitic wasps, five insect hosts and six plants (Table 1). The references reporting the interactions between parasitic wasps, insect hosts and plants were given in supplementary table S1.

OGS

The General Feature Format version 3 (Gff3) files containing annotation information were downloaded with the genome data, and the official gene sets (OGSs) were extracted from the genome based on the annotation in the Gff3 file. Then, the nucleotide sequences and protein sequences of annotated genes were produced (Table 2).

Transcriptomes

The raw data of 34 samples of parasitic wasps were downloaded from the NCBI SRA (Sequence Read Archive) database (<https://www.ncbi.nlm.nih.gov/sra>). We assembled 22 transcriptomes using Trinity and TopHat-Cufflinks with default parameters (26, 27). Together with 21 other



Figure 2. The parasitic wasps, hosts and plants included in the WaspBase. Dashed line: parasitic wasps parasitize plants but not insect hosts. Not dashed line: parasitic wasps parasitize the insect hosts or the insects damage plants.

available transcriptomes, we collected a final transcriptome dataset of 573 RNA-Seq samples from 35 parasitic wasps (Table 3).

lncRNA

Long non-coding RNAs (lncRNAs) are transcribed RNA molecules >200 nucleotides in length that are not protein

coding (28, 29). We predicted lncRNAs of eight parasitic wasps using a previously reported pipeline (30). In total, we predicted 49 607 lncRNAs from eight parasitic wasps.

UTR

We developed a pipeline to predict untranslated regions (UTR) from the transcriptomes and genomes using

Table 1. The genome data in the WaspBase

	Species name	Accession ID	Source	
Wasps	<i>Ceratosolen solmsi</i>	GCF_000503995.1	NCBI	
	<i>Copidosoma floridanum</i>	GCF_000648655.1	NCBI	
	<i>Cotesia vestalis</i>	GCA_001675545.1	NCBI	
	<i>Diachasma alloenum</i>	GCF_001412515.1	NCBI	
	<i>Fopius arisanus</i>	GCF_000806365.1	NCBI	
	<i>Macrocentrus cingulum</i>	-	InsectBase	
	<i>Microplitis demolitor</i>	GCF_000572035.2	NCBI	
	<i>Nasonia giraulti</i>	GCA_000004775.1	NCBI	
	<i>Nasonia longicornis</i>	GCA_000004795.1	NCBI	
	<i>Nasonia vitripennis</i>	GCF_000002325.3	NCBI	
	<i>Orussus abietinus</i>	GCF_000612105.1	NCBI	
	<i>Trichogramma pretiosum</i>	GCF_000599845.2	NCBI	
	Hosts	<i>Bactrocera dorsalis</i>	GCF_000789215.1	NCBI
		<i>Calliphora vicina</i>	GCA_001017275.1	NCBI
<i>Ceratitis capitata</i>		GCF_000347755.2	NCBI	
<i>Helicoverpa armigera</i>		GCF_002156985.1	NCBI	
<i>Helicoverpa zea</i>		GCA_002150865.1	NCBI	
<i>Heliobis virescens</i>		GCA_002382865.1	NCBI	
<i>Manduca sexta</i>		-	InsectBase	
<i>Musca domestica</i>		GCF_000371365.1	NCBI	
Plants		<i>Brassica oleracea</i>	GCF_000695525.1	NCBI
		<i>Citrus maxima</i>	GCA_002006925.1	NCBI
	<i>Gossypium hirsutum</i>	GCF_000987745.1	NCBI	
	<i>Malus domestica</i>	GCF_000148765.1	NCBI	
	<i>Nicotiana tabacum L</i>	GCF_000715135.1	NCBI	
	<i>Pyrus x bretschneideri</i>	GCF_000315295.1	NCBI	
	<i>Vitis aestivalis</i>	GCA_001562795.1	NCBI	
	<i>Zea mays</i>	GCF_000005005.2	NCBI	

TransDecoder-V5.3.0 (<https://github.com/TransDecoder/TransDecoder>), identifying the UTR sequences of 21 parasitic wasps.

Gene families

We used manual annotation by Blastp against known genes (e-value = 10^{-5}), GO annotation and phylogenetic analysis to identify the members of a gene family. We obtained the information of 25 gene families that have been widely studied, including those related to chemoreception, the immune system and detoxification (Figure 3). We also provided a web server for phylogenetic analysis of selected gene members, and we use ClustalW2 (31) to construct a phylogenetic tree by the neighbor-joining clustering method. The bootstrap value was set as 500. The Newick Utilities V1.6 (32) was used to display the phylogenetic tree.

Database construction

Database system implementation

WaspBase was developed on an Apache HTTP (Apache 2.4.25) server in a Linux (RedHat 4.8.2) operating system.

The web pages were written using PHP (PHP 5.6.30), html language, Cascading Style Sheets and JavaScript. All data are stored in the MySQL (MySQL 5.7.17) environment. The Apache server handles queries from web clients through PHP scripts to perform searches.

Search function

WaspBase provides search function using keywords, gene ID, gene names, annotation keywords, KEGG ID, KEGG annotation (33), Pfam ID or Pfam annotation (34). Once a gene is searched for, all related gene information was presented in the result webpages. The genes from parasitic wasps, insect hosts and plants were given in the searched results.

Tools module

The tools module contains Basic Local Alignment Search Tool (BLAST) (35), profile hidden Markov model (HMMER), Multiple Alignment using Fast Fourier Transform (MAFFT), automated alignment trimming (TrimAl) and JBrowse (36).

Table 2. The protein and nucleotide dataset in the WaspBase

	Species name	Accession ID	Source
Wasps	<i>Ceratosolen solmsi</i>	GCF_000503995.1	NCBI
	<i>Copidosoma floridanum</i>	GCF_000648655.1	NCBI
	<i>Diachasma alloeum</i>	GCF_001412515.1	NCBI
	<i>Fopius arisanus</i>	GCF_000806365.1	NCBI
	<i>Macrocentrus cingulum</i>	-	InsectBase
	<i>Microplitis demolitor</i>	GCF_000572035.2	NCBI
	<i>Nasonia vitripennis</i>	GCF_000002325.3	NCBI
	<i>Orussus abietinus</i>	GCF_000612105.1	NCBI
	<i>Trichogramma pretiosum</i>	GCF_000599845.2	NCBI
Hosts	<i>Bactrocera dorsalis</i>	GCF_000789215.1	NCBI
	<i>Ceratitis capitata</i>	GCF_000347755.2	NCBI
	<i>Helicoverpa armigera</i>	GCF_002156985.1	NCBI
	<i>Manduca sexta</i>	-	InsectBase
	<i>Musca domestica</i>	GCF_000371365.1	NCBI
Plants	<i>Brassica oleracea</i>	GCF_000695525.1	NCBI
	<i>Gossypium hirsutum</i>	GCF_000987745.1	NCBI
	<i>Malus domestica</i>	GCF_000148765.1	NCBI
	<i>Nicotiana tabacum L</i>	GCF_000715135.1	NCBI
	<i>Pyrus x bretschneideri</i>	GCF_000315295.1	NCBI
	<i>Zea mays</i>	GCF_000005005.2	NCBI

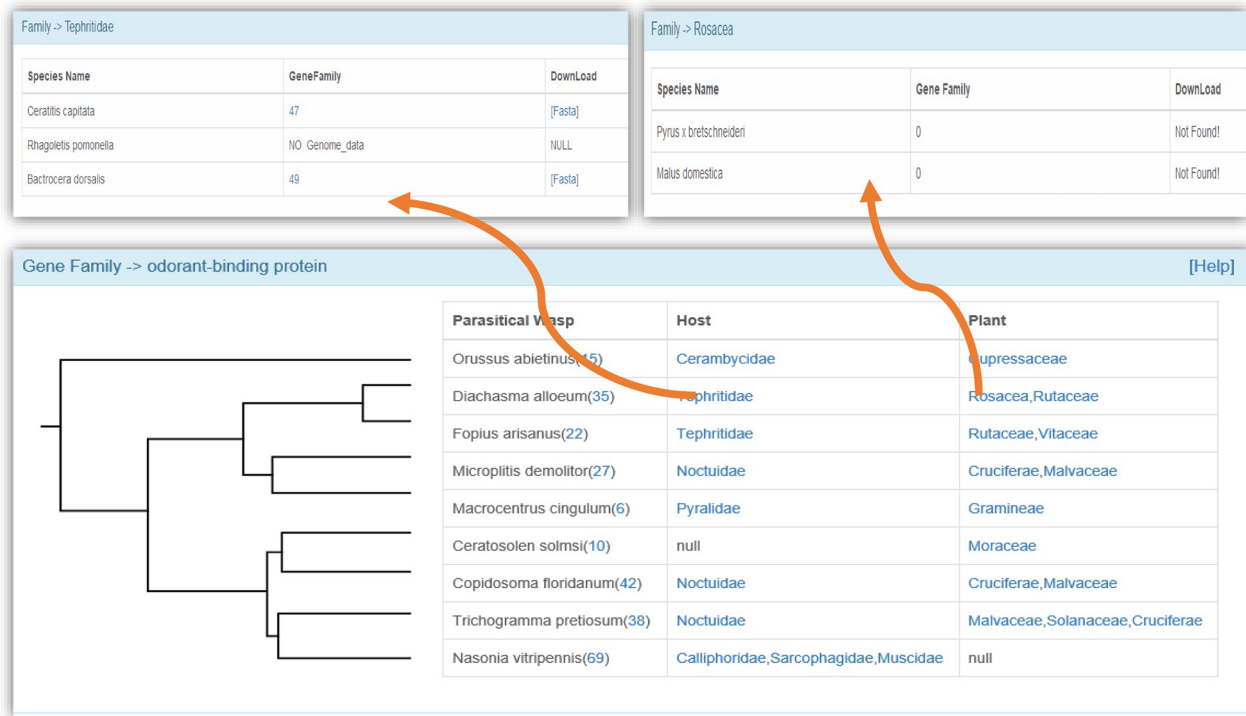


Figure 3. The identified gene families in the WaspBase.

BLAST (35) is provided using the Web-based BLAST server 2.6.0+. The data used for nucleotide BLAST (BLASTN, TBLASTN) searches include 12 insect genomes and 9 insect OGSs. The protein data used for amino

acid BLAST (BLASTP, TBLASTX, BLASTX) searches contain nine insect protein sequences. In the BLAST results webpage, users can choose to display top 5 hits, top 10 hits or all hits. The top five BLAST hits are used as default.

Table 3. The transcriptome data in the WaspBase

Species name	Assembly	SRA
<i>Aenasius bambawalei</i>	Trinity	SRR2966926
<i>Anastatus japonicus</i>	Trinity	SRR4034898
<i>Anisopteromalus calandrae</i>	Trinity	SRR2910690,SRR2910691
<i>Asobara tabida</i>	Not assembled	-
<i>Biorhiza pallida</i>	Trinity	ERR1353142,ERR1354102,ERR1354103,ERR1354104,ERR1354105,ERR1354106,ERR1354107,ERR1354108,ERR1354109,ERR1354110,ERR1354111,ERR1354112,ERR1354113,ERR1354114,ERR1354115,ERR1354116,ERR1354117,ERR1354118,ERR1354119,ERR1354354
<i>Ceratosolen solmsi</i>	TopHat-Cufflinks	SRR974922,SRR974923,SRR974924,SRR974925,SRR974926,SRR974927,SRR974928,SRR974929
<i>Copidosoma floridanum</i>	TopHat-Cufflinks	SRR1864696,SRR1864697
<i>Cotesia glomerata</i>	Not assembled	-
<i>Cotesia rubecula</i>	Not assembled	-
<i>Cotesia vestalis</i>	Not assembled	-
<i>Diachasma alloeum</i>	TopHat-Cufflinks	SRR2040481,SRR2041626
<i>Diachasmimorpha longicaudata</i>	Not assembled	SRR3336273,SRR3336336,SRR3336337
<i>Diadromus collaris</i>	Not assembled	SRR4294717,SRR1022346
<i>Fopius arisanus</i>	TopHat-Cufflinks	SRR1560649,SRR1560650,SRR1560651,SRR1560653
<i>Leptopilina bouleardi</i>	Not assembled	ERR1109367,ERR1109368,ERR1109369,ERR1109370,ERR1109371,ERR1109372,ERR1109373,ERR1109374,ERR1109375,SRR559221,SRR559222
<i>Leptopilina clavipes</i>	Trinity	SRR921610
<i>Leptopilina heterotoma</i>	Trinity	SRR559223,SRR559224
<i>Lysiphlebus fabarum</i>	Not assembled	-
<i>Macrocentrus cingulum</i>	TopHat-Cufflinks	SRR2968845,SRR2968846
<i>Megastigmus spermotrophus</i>	Trinity	SRR1805073,SRR1805097,SRR1805105,SRR1805115
<i>Microctonus aethiopoidea</i>	Not assembled	-
<i>Microplitis bicoloratus</i>	Not assembled	-
<i>Microplitis demolitor</i>	TopHat-Cufflinks	SRR955015,SRR955076,SRR955374,SRR955397
<i>Nasonia giraulti</i>	TopHat-Cufflinks	SRR3457435,SRR3457436,SRR3457437,SRR3457438,SRR3457439,SRR3457457,SRR1566028,SRR1566029,SRR1566030,SRR1566031,SRR1566032,SRR1566033,SRR1264518,SRR1264519,SRR1264521,SRR1264522,SRR1264523,SRR1264524,SRR1264525,SRR1264526,SRR1264527,SRR1264529,SRR1264530,SRR1264531
<i>Nasonia longicornis</i>	Not assembled	-
<i>Nasonia vitripennis</i>	Not assembled	-
<i>Orussus abietinus</i>	TopHat-Cufflinks	ERR1333211,SRR1850925,SRR1850924,SRR921626
<i>Ostrinia furnacalis</i>	Trinity	DRR018822,DRR018823,DRR018824,DRR018825,DRR018826,DRR018827,DRR030133,DRR030134,DRR030135,DRR030136,DRR030137,DRR030138,DRR030139,DRR030140,DRR030141,DRR030142,SRR1032037,SRR1032038,SRR1226611,SRR1265986,SRR1560699,SRR1560709,SRR1560711,SRR1565323,SRR1640337,SRR1640339,SRR1640341,SRR3189772,SRR3204354,SRR3204356,SRR3204357,SRR3374123,SRR3374124,SRR3374125
<i>Psytalia concolor</i>	Trinity	SRR1593901,SRR1593902
<i>Psytalia lounsburyi</i>	Trinity	SRR1593906,SRR1593907,SRR1593908
<i>Pteromalus puparum</i>	Not assembled	-
<i>Spalangia endius</i>	Trinity	SRR2954670,SRR2954673,SRR2954678,SRR2954681,SRR2954683,SRR2954686,SRR2954688,SRR2954692,SRR2954704,SRR2954706,SRR2954708,SRR2954710,SRR1038395
<i>Telenomus podisi</i>	Trinity	SRR1274857,SRR1274858
<i>Trichogramma chilonis</i>	Trinity	SRR3756972,SRR3756974,SRR3756975,SRR3756979
<i>Trichogramma pretiosum</i>	TopHat-Cufflinks	SRR1826957,SRR1826958
<i>Venturia canescens</i>	Trinity	ERR791800

Figure 4. The Download page of WaspBase. The genomes, transcriptomes and OGSs of parasitic wasps and insect hosts are provided together for the convenience of download.

User can also adjust other parameters such as similarity percentage and BLAST score. Links of the BLAST hits were given to directly connect to NCBI for full annotation information. All sequence can be downloaded.

Multiple sequence alignment (MSA) is important for evolutionary analyses. MAFFT (37) is a widely used program for MSA analysis because of its high performance. WaspBase provides a web server of MAFFT and uses TrimAl to trim the aligned sequences (38). To use MAFFT web server, users need to input the sequences in FASTA format with either the default parameters or the customized parameters. To use TrimAl, users need to input the aligned sequences at the TrimAl webpage. The trimmed sequences are showed at TrimAl result webpage. If the number of sequence is more than four, a phylogenetic tree can be constructed using the abovementioned method.

A web server of HMMER is provided to search sequence homologs and to make sequence alignments. It uses probabilistic models called profile hidden Markov models (profile HMMs) (39). To use HMMER, users input the protein sequences at the HMMER webpage. After running the HMMER, the protein sequences are used to search against the Pfam database and the results of protein domain information will be showed at the HMMER result webpage.

Genome visualization

JBrowse is a well-known browser that displays genome annotations by integrating the databases and interactive web pages (36). We used JBrowse in WaspBase to provide

interactive views of annotations along with the genome scaffolds. The genome data and the Gff3 files required for JBrowse are stored in a MySQL database using prepare-refseqs.pl, flatfile-to-json.pl, add-bam-track.pl and add-track-json.pl provided by BioPerl. In WaspBase, JBrowse visualizes the annotations and transcriptomes as tracks on the browser for Coding Sequence and coverage of the transcriptome reads. Pop-up balloons in the gene model track display links to gene sequences of interest.

Wasp researchers

To construct a scientific network in the field of parasitic wasp research, we performed reference mining of parasitic wasp studies, which yielded 189 references. Based on publications in the last 5 years, we collected a list of active researchers studying parasitic wasps.

Download

All data can be downloaded, including genomes, transcriptomes, UTR, Gene families and lncRNA. For the convenience of downloading, the gene data of parasitic wasps, insect pests and plants are provided for download at the same webpage (Figure 4).

Conclusions

We constructed WaspBase for parasitic wasps and their corresponding insect hosts and plants. WaspBase provides conventional functions of search, download, domain

analysis and phylogenetic analysis, JBrowse display of annotations and other functions described herein. In addition to genomes and transcriptomes, WaspBase also provides lncRNA, UTR and gene family information. A typical feature of WaspBase is that we integrated the gene information of parasitic wasps, their insect hosts and plants targeted by insect pests. Thus, gene data of the tritrophic system in food chains (parasitic wasp–insect pest–plant) were analyzed together, which should be useful for studying cross-species regulation in parasitism and convergent evolution analysis among wasps, hosts and plants.

Future plan

1. As the cost of sequencing has been significantly reduced in recent years, the genomes of an increasing number of parasitic wasps will be sequenced. We plan to update WaspBase periodically to keep the database up-to-date.
2. Genome annotation is still a time-consuming task and significantly lags behind genome sequencing. We noticed that a number of parasitic wasp genomes are not annotated at present though their genome sequences have been uploaded in the NCBI genome database. We will annotate these genomes using OMIGA (Optimized Maker-Based Insect Genome Annotation) (40), a genome annotation pipeline that we developed.
3. It is important to understand cross-species regulation mechanisms and convergent evolution in parasitism. To this end, we will carry out a systematic analysis of more gene families from the OGSs of ‘wasps–insects–plants’, which should be useful to improve control efficiencies in biological control.

Supplementary data

Supplementary data are available at *Database* Online.

Funding

National Key Research and Development Program (2017YFD0200900, 2016YFC1200600 to F.L.); National Science Foundation of China (NSFC) (31772238, 31701785 to K.H.).

Conflict of interest. None declared.

References

1. Miles, A., Harding, N.J., Botta, G. *et al.* (2017) Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature*, **552**, 96–.
2. Kim, K.H., Kabir, E. and Jahan, S.A. (2017) Exposure to pesticides and the associated human health effects. *Sci. Total Environ.*, **575**, 525–535.
3. Santos, M.A.B., de Macedo, L.O., de Souza, J.B. *et al.* (2017) Larvae of *Ixodiphagus* wasps (Hymenoptera: Encyrtidae) in *Rhipicephalus sanguineus sensu lato* ticks (Acari: Ixodidae) from Brazil. *Ticks Tick Borne Dis.*, **8**, 564–566.
4. Geib, S.M., Liang, G.H. and Murphy, T.D. (2017) Whole genome sequencing of the braconid parasitoid wasp *Fopius arisanus*, an important biocontrol agent of pest tephritid fruit flies. *G3 (Bethesda, Md.)*, **7**, 2407–2411.
5. Vinson, S.B. (1976) Host selection by insect parasitoids. *Annu. Rev. Entomol.*, **21**, 109–133.
6. Whitfield, J.B. (1998) Phylogeny and evolution of host-parasitoid interactions in Hymenoptera. *Annu. Rev. Entomol.*, **43**, 129–151.
7. Yin, C., Shen, G., Guo, D. *et al.* (2016) InsectBase: a resource for insect genomes and transcriptomes. *Nucleic Acids Res.*, **44**, D801–D807.
8. Xiao, J.H., Yue, Z., Jia, L.Y. *et al.* (2013) Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol.*, **14**, R141.
9. Werren, J.H., Richards, S., Desjardins, C.A. *et al.* (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*, **327**, 343–348.
10. Bonasio, R., Zhang, G., Ye, C. *et al.* (2010) Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Science*, **329**, 1068–1071.
11. Burke, G.R., Walden, K.K., Whitfield, J.B. *et al.* (2014) Widespread genome reorganization of an obligate virus mutualist. *PLoS Genetics*, **10**, e1004660.
12. Papanicolaou, A., Schetelig, M.F., Arensburger, P. *et al.* (2016) The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol.*, **17**, 192.
13. Pearce, S.L., Clarke, D.F., East, P.D. *et al.* (2017) Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. *BMC Biol.*, **15**, 63.
14. Scott, J.G., Warren, W.C., Beukeboom, L.W. *et al.* (2014) Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol.*, **15**, 466.
15. Andrade, G.S., Pratisoli, D., Dalvi, L.P. *et al.* (2011) Performance of four *Trichogramma* species (Hymenoptera: Trichogrammatidae) as biocontrol agents of *Heliothis virescens* (Lepidoptera: Noctuidae) under various temperature regimes. *J. Pest Sci.*, **84**, 313–320.
16. Oatman, E.R. and Platner, G.R. (1978) Effect of mass releases of *Trichogramma pretiosum* (Hymenoptera: Trichogrammatidae) against lepidopterous pests on processing tomatoes in Southern California, with notes on host egg population trends. *J. Econ. Entomol.*, **71**, 896–900.
17. Cao, X.L. and Jiang, H.B. (2015) Integrated modeling of protein-coding genes in the *Manduca sexta* genome using RNA-Seq data from the biochemical model insect. *Insect Biochem. Mol. Biol.*, **62**, 2–10.
18. Frederickx, C., Dekeirsschieter, J., Verheggen, F.J. *et al.* (2014) Depth and type of substrate influence the ability of *Nasonia vitripennis* to locate a host. *J. Insect Sci.*, **14**, 58.
19. Parkin, I.A., Koh, C., Tang, H. *et al.* (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biol.*, **15**, R77.

20. Li,F., Fan,G., Lu,C. *et al.* (2015) Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.*, **33**, 524–530.
21. Velasco,R., Zharkikh,A., Affourtit,J. *et al.* (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat. Genet.*, **42**, 833–839.
22. Sierro,N., Battey,J.N., Ouadi,S. *et al.* (2014) The tobacco genome sequence and its comparison with those of tomato and potato. *Nat. Commun.*, **5**, 3833.
23. Wu,J., Wang,Z., Shi,Z. *et al.* (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.*, **23**, 396–408.
24. Jiao,Y., Peluso,P., Shi,J. *et al.* (2017) Improved maize reference genome with single-molecule technologies. *Nature*, **546**, 524–527.
25. Poelchau,M., Childers,C., Moore,G. *et al.* (2015) The i5k Workspace@NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Res.*, **43**, D714–D719.
26. Haas,B.J., Papanicolaou,A., Yassour,M. *et al.* (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, **8**, 1494–1512.
27. Trapnell,C., Roberts,A., Goff,L. *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, **7**, 562–578.
28. Lakhota,S.C. (2017) From heterochromatin to long noncoding RNAs in *Drosophila*: expanding the arena of gene function and regulation. In: Rao M. (eds). *Long Non Coding Rna Biology*. Advances in Experimental Medicine and Biology, Springer, Singapore, **1008**, 75–118.
29. Cho,J. (2018) Transposon-derived non-coding RNAs and their function in plants. *Front. Plant Sci.*, **9**, 600.
30. Xiao,H., Yuan,Z., Guo,D. *et al.* (2015) Genome-wide identification of long noncoding RNA genes and their potential association with fecundity and virulence in rice brown planthopper, *Nilaparvata lugens*. *BMC Genomics*, **16**, 749.
31. Larkin,M.A., Blackshields,G., Brown,N.P. *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.
32. Junier,T. and Zdobnov,E.M. (2010) The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics*, **26**, 1669–1670.
33. Du,J., Yuan,Z., Ma,Z. *et al.* (2014) KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol. Biosyst.*, **10**, 2441–2447.
34. Finn,R.D., Coghill,P., Eberhardt,R.Y. *et al.* (2015) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
35. Johnson,M., Zaretskaya,L., Raytselis,Y. *et al.* (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.*, **36**, W5–W9.
36. Buels,R., Yao,E., Diesh,C.M. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.
37. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
38. Capella-Gutierrez,S., Silla-Martinez,J.M. and Gabaldon,T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
39. Mistry,J., Finn,R.D., Eddy,S.R. *et al.* (2013) Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.*, **41**, e121.
40. Liu,J., Xiao,H., Huang,S. *et al.* (2014) OMIGA: Optimized Maker-based Insect Genome Annotation. *Mol. Genet. Genomics*, **289**, 567–573.