

Draft genome of the cotton aphid *Aphis gossypii*

Qingmei Quan^a, Xiao Hu^a, Bohu Pan^{b,c}, Baosheng Zeng^{a,c}, Ningning Wu^{a,c}, Gangqi Fang^{a,c},
Yanghai Cao^a, Xiaoya Chen^{d,e}, Xuan Li^b, Yongping Huang^{a,*}, Shuai Zhan^{a,**}

^a Key Laboratory of Insect Developmental and Evolutionary Biology, CAS Center for Excellence in Molecular Plant Science, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

^b Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Science, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

^c University of Chinese Academy of Sciences, Beijing, 100049, China

^d National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

^e Plant Science Research Center, Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, Shanghai, 201602, China

ARTICLE INFO

Keywords:

Aphis gossypii

Genome assembly

Genome annotation

Comparative genomic analysis

Genome evolution

ABSTRACT

The cotton aphid *Aphis gossypii* Glover is a worldwide agricultural pest that feeds on cotton, melon, and other landscape plants, causing a high level of economic loss. In addition to the common characteristics shared with other aphids, *Ap. gossypii* has evolved multiple biotypes that present substantial differences in host adaption. These intriguing biological features are of interest from both a fundamental and applied perspective. However, the molecular studies of *Ap. gossypii* have been restrained by the lack of a reference genome. Furthermore, in order to establish a platform for the development of novel and sustainable control methods, it is necessary to generate genomic resources for *Ap. gossypii*. Here, we present a 294 Mb draft genome sequence of *Ap. gossypii*, which consists of 4,724 scaffolds with an N50 size of 438 kb. Compared to other aphid species with published genomes, *Ap. gossypii* presents the most compact genome size. A total of 14,694 protein-coding genes were predicted and annotated in the consensus gene set, 98.03% of CEGMA genes and 93.5% of BUSCO genes were captured respectively. Genome-wide selection analyses revealed that significantly evolving pathways in the genus *Aphis* are related to biological processes of detoxification, steroid biosynthesis, and ethylbenzene degradation. The acquisition of the genome of *Ap. gossypii* makes it possible to understand the molecular mechanism of intricate biological traits of this species, and will further facilitate the study of aphid evolution.

1. Introduction

Aphids are a group of tiny sap-sucking insects well known for their flexible reproductive strategies (Moran, 1992). Responding to seasonal changes, aphids exhibit reversible life cycles by shifting between asexual and sexual reproduction. Generally, parthenogenetic generations are successive during spring and summer for rapid colony expansion, while replaced by the sexual generation in late autumn, in which frost-resistant eggs are laid for overwintering (Le Trionnaire et al., 2012; Moran, 1992; Ramos et al., 2003). Unexpectedly, this alternation strategy between asexual and sexual generations has been lost in some aphid species (Vereschagina and Gandrabur, 2014). Furthermore, the evolution of aphids has been shown to be associated with that of their host, providing an ideal model to study the interaction and co-

evolution between herbivores and plants.

Aphis gossypii is an important cosmopolitan pest on Malvaceae and Cucurbitaceae plants (Carletto et al., 2009; Ebert and Cartwright, 1997). Cotton aphids present a similar life-cycle as other aphids species (Fig. 1). During spring and summer, the parthenogenetic population expands rapidly, resulting in a high nutrient consumption of the host plant. The accumulated honeydew produced by large aphid population promotes the growth of molds, which greatly decrease the photosynthetic activity of plants lamina and hence contaminate themselves (Slosser et al., 2002). Moreover, cotton aphids could transmit various types of plant viruses, such as the cucumber mosaic virus and atypical cotton leafroll dwarf virus etc. (Agrofoglio et al., 2017; Schoeny et al., 2017). For decades, the control of *Ap. gossypii* has largely relied on chemical pesticides. However, evolved resistance and uncontrolled

* Corresponding author.

** Corresponding author.

E-mail addresses: yphuang@sibs.ac.cn (Y. Huang), szhan@sibs.ac.cn (S. Zhan).

<https://doi.org/10.1016/j.ibmb.2018.12.007>

Received 11 April 2018; Received in revised form 3 December 2018; Accepted 14 December 2018

Available online 24 December 2018

0965-1748/© 2019 Elsevier Ltd. All rights reserved.

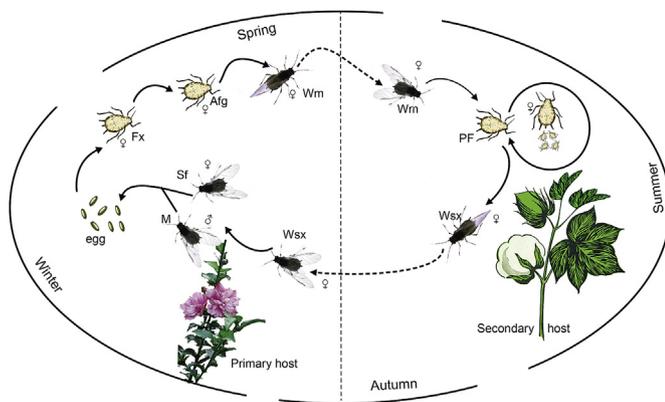


Fig. 1. Life cycle of the cotton aphid, *Ap. gossypii*. Responding to seasonal changes, cotton aphids exhibit complex life cycles shifting between asexual and sexual reproduction. In winter, the sexual phase produces frost-resistant eggs to adapt to cold conditions; this stage mainly occurs on its primary hosts, such as Rose of Sharon, pomegranate, and Chinese prickly ash. In spring, eggs initiate hatching and produce wingless, asexual female offsprings. After several parthenogenetic generations, partial winged migratory aphids are produced to colonize secondary hosts, such as cotton, cucumber, and pumpkin. On these secondary hosts, cotton aphids initiate parthenogenetic generations till the early fall. In the late fall, sexual females and males are produced and oviparous sexual females deposit eggs. Fx, fundatrix; Afg, apterous fundatrigeniae; Wm, winged migrate; PF, parthenogenetic females; Wsx, wing sexuparae; Sf, sexual female, M, male.

pesticide usage had resulted in even more serious population outbreaks (Cao et al., 2008; Herron and Wilson, 2017; Pan et al., 2015).

In addition to the applied research on population control in the field, the cotton aphid has been a popular organism for adaptive evolutionary studies for decades. The recorded plant host spectrum of cotton aphid comprises hundreds of botanical species, making it one of the most polyphagous aphid species. Importantly, this generalist consumer is recognized to be partially subdivided into numerous host-specific races which present a high level of host specialization (Carletto et al., 2009; Wang et al., 2016). The existence of these evident host-associated biotypes made the cotton aphid an ideal model for the study of adaptive evolution.

Recently, most studies of *Ap. gossypii* have focused on reproductive alternation (Kwon and Kim, 2017; Liu et al., 2014), wing polyphenism (Gu et al., 2013; Yang et al., 2014), host specialization (Zhang et al., 2017), and insecticide resistance (Gong et al., 2014, 2017; Pan et al., 2015). However, these studies were somehow limited due to the lack of genomic resources. Although several aphid genomes have been published (International Aphid Genomics, 2010; Mathers et al., 2017; Nicholson et al., 2015; Wenger et al., 2017), the great divergence among aphid species made it impossible to apply these available genomic resources in the study of *Ap. gossypii*. In this study, we provided a high quality draft genome, along with a comprehensive annotation, for the cotton aphid *Ap. gossypii*.

2. Materials and methods

2.1. Sample preparation and genome sequencing

The *Ap. gossypii* samples were collected in 2013, from a cotton field located at the Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, in Shanghai, China. The collected aphids were reared on cotton plants at 25 °C with 70% humidity under a 16:8 h light/dark photoperiod. For genome sequencing, a parthenogenetic cotton aphid line from a single female was used for genomic DNA preparation. Four paired-end libraries with insert sizes of 290 bp, 420 bp, 5 Kb and 10 Kb were constructed (Table 1). All libraries were sequenced on an Illumina HiSeq 2000 platform. We applied a classic high-

Table 1
Statistics of sequencing data for the genome project of *Ap. gossypii*.

Library type	Insert size (bp)	Sequencing data (Gb)	Coverage (X)
Paired-end	290	34.14	116.1
Paired-end	420	34.22	116.4
Mate pair	5,000	5.09	17.3
Mate pair	10,000	3.87	13.2

throughput sequencing technology for the *de novo* assembly of the genome.

2.2. Genome assembly and evaluation

We first trimmed adapter sequences and low-quality bases using *platanus_trim* of the *Platanus* software (Kajitani et al., 2014). Since aphids generally carry symbionts, we further filtered reads that mapped to the sequences of *Buchnera*, the primary endosymbiont of aphids, from the non-redundant nucleotide database of NCBI (NR; release 68). Filtered reads were first assembled into contigs using ALLPATHS-LGv52488 (Gnerre et al., 2011) with default settings. Mate-pair reads of 5-Kb and 10-Kb libraries were used for joining contigs into scaffolds using ScaffMatch v0.9 (Mandric and Zelikovsky, 2015). Gap-Closer v1.12 available in SOAPdenovo (Li et al., 2010) was used to fill the gaps within scaffolds using short reads. We note that most de Bruijn-based assembling approaches cannot clearly differentiate highly similar repeats from heterozygote-caused redundancy. Also, those tandem repeats would be inappropriately merged due to the algorithm of de Bruijn. Thus, a certain fraction of repeats might be inappropriately excluded from the final assembly, causing an underestimated genome size. To assess the integrity and quality of the genome assembly, both core eukaryotic genes mapping (CEGMA v2.4) (Parra et al., 2007) approach and Benchmarking Universal Single-Copy Orthologs software (*insecta_odb9* of BUSCO v3) (Waterhouse et al., 2017) were used to evaluate the completeness of the draft assembly.

2.3. Gene prediction and annotation

We integrated three lines of evidence to predict protein coding genes in the *Ap. gossypii* genome, including *ab initio* features, homology of related species, and RNA-seq data for this species. For *ab initio* predictions, we applied AUGUSTUS v3.1 (Stanke et al., 2006), Genscan v1.0 (Burge and Karlin, 1997), SNAP v20060728 (Korf, 2004), and Maker v2.26 (Cantarel et al., 2008) to generate four independent gene sets. To train AUGUSTUS and SNAP, we performed an iterative strategy. In short, we randomly selected ~20 Mb genomic sequences out of the whole assembly and performed an initial round of prediction using the EST mode of Maker, based on all downloaded expressed sequence tags (ESTs) of *Ap. gossypii* available in the NCBI (downloaded in Feb 2016); resulted gene models were then used to train SNAP. In the subsequent rounds of predictions by Maker, trained SNAP was set as the *ab initio* predictor to yield an improved set of gene models, which were then used to train SNAP and AUGUSTUS further. For homology-based predictions, protein sequences from Uniprot database (the Hemiptera subset; release-2015_12) and five related insect species (the bed bug *Cimex lectularius* (Benoit et al., 2016), the Triatomine bug *Rhodnius prolixus* (Mesquita et al., 2015), the pea aphid *Acyrtosiphon pisum* (International Aphid Genomics, 2010), the Russian wheat aphid *Diuraphis noxia* (Nicholson et al., 2015), and the model insect *Drosophila melanogaster* (FlyBase, 1998)) were aligned to the *Ap. gossypii* genome using TBLASTN with an e-value cutoff of $E < 10^{-5}$; high-scoring pairs (HSPs) were processed by genblast A v1.0.1 (She et al., 2009) and GeneWise v2.2.0 (Birney et al., 2004) to generate gene structures. RNA-seq reads of *Ap. gossypii* were obtained from SRR1257338, SRR1257339, and SRR1257340 (Liu et al., 2014). Two independent

Table 2
Statistics of independent gene sets used in this study.

Gene set	Gene numbers	Average CDS length	Average Gene length	Mean Exon per gene	Average Exon length	
<i>Ab initio</i>	GLEAN	15,467	1,342	7,410	7	192
	AUGUSTUS	14,172	1,050	3,854	5	210
	SNAP	56,425	303	1,954	6	51
RNA-seq	Gen scan	14,048	696	5,593	3	232
	maker	15,321	903	3,129	5	181
	Cufflinks	16,570	906	1,975	3	302
homolog	Trinity	28,833	491	941	2	246
	EST	31,184	567	704	2	284
	<i>R. prolixus</i>	10,810	639	1,164	3	213
	<i>D. melanogaster</i>	8,111	813	1,728	4	203
	<i>Ac. pisum</i>	27,603	429	571	1	429
	<i>C. lectularius</i>	8,891	885	1,911	4	221
	<i>D. noxia</i>	17,025	747	1,452	3	249
	Uniprot	19,683	648	1,194	2	324

approaches were utilized to integrate these transcriptome data into the genome. In the mapping approach, transcript structures were generated by a classic approach of Tophat and Cufflinks (Ver. 2.2.1) (Kim et al., 2013; Trapnell et al., 2010); on the other hand, Trinity v2.3.2 (Grabherr et al., 2011) was used to *de novo* assemble RNAseq reads to transcripts, which were then processed by EXONERATE v2.2.0 (Slater and Birney, 2005) to generate gene models. Finally, we used GLEAN (Elsik et al., 2007) to integrate all above independent gene sets to generate a consensus gene set. Detailed information of the gene sets used in this study is listed in Table 2.

The entire gene set of *Ap. gossypii* was annotated by aligning protein sequences against three public databases, including UniRef (UniRef50; release-2015_12), NCBI RefSeq (the invertebrate subset; release 69), Kyoto Encyclopedia of Genes and Genomes (proteins of KO), using BLASTP with an E-value cutoff of $E < 10^{-5}$. Local run of InterProScan v5.13–52.0 (Jones et al., 2014) with all implemented methods was used for the identification of Interpro and PFAM domains, and the GeneOntology terms (GO terms) in this gene sets. ESTs and assembled transcripts of *Ap. gossypii* were also aligned using BLASTN to support the gene models without homology evidence from other species.

2.4. *Ap. gossypii* orthology and evolution analysis

To investigate genome evolution in hemipteran insects, especially in aphids, we compared the proteome of *Ap. gossypii* with those of eight other hemipterans, including the whitefly *Bemisia tabaci* (Chen et al., 2016), the brown planthopper *Nilaparvata lugens* (Xue et al., 2014), the bed bug *C. lectularius*, the Triatomine bug *R. prolixus*, and four aphid species (pea aphid *Ac. pisum*, green peach aphid *Myzus persicae* (Mathers et al., 2017), Russian wheat aphid *D. noxia* and soybean aphid *Aphis glycines* (Wenger et al., 2017)). Another five representative arthropod species (the termite *Zootermopsis nevadensis* (Terrapon et al., 2014), the body louse *Pediculus humanus humanus* (Kirkness et al., 2010), the western honey bee *Apis mellifera* (Weinstock et al., 2006), the fruit fly *D. melanogaster*, the cotton bollworm *Helicoverpa armigera* (Pearce et al., 2017)) were also used in the orthology analysis. From these 14 collected gene sets (NCBI accession numbers and genome download links were listed in Table S1) we trimmed the very short proteins (< 10aa) and those that contained a large percentage of stop codons (> 20%), resulting in a dataset of 249,148 protein sequences as the input for OrthoMCL-v2.0.9 (Li et al., 2003). We utilized 372 single-copy universal genes to infer the phylogenetic relationships across all examined species. Muscle v3.8.31 (Edgar, 2004) and Gblocks v0.91b (Talavera and Castresana, 2007) were used to perform multiple alignment and identify conserved blocks across species, respectively. The species tree was reconstructed using RAxML v8.2.10 (Stamatakis, 2014)

under the PROTGAMMAJTT model with 100 bootstrap replicates. According to the results of InterProScan, conserved domains were further compared across the five aphid species. InterPro (IPR) domains only present or absent in *Ap. gossypii* were defined as the species-specific or lost IPR domains, respectively.

2.5. d_N/d_S ratios in two *Aphis* species

To explore the genome-wide adaptive evolution in the cotton aphid, we estimated the nonsynonymous-to-synonymous substitution (d_N/d_S) ratios for 8,808 1:1 orthologs between *Ap. gossypii* and *Ap. glycines*, using *M. persicae* as outgroup (Table S2). The ratio of d_N/d_S was calculated using the Codeml program in the PAML package v4.8 (Yang, 2007). To present the evolutionary pattern at the functional level, we mapped these genes to KEGG pathways and calculated the median value of d_N/d_S for each pathway. Rapidly evolving pathways were defined as pathways with significantly higher level of d_N/d_S than other pathways in the same lineage. Significance was determined using the one-tailed Z test. Pathways with less than 5 genes were excluded from the analysis.

2.6. Gene families related to aphid-host interaction

We performed manual annotation and comparisons on genes of potential relationships with environmental adaptation in insects. We applied a classic iterative approach to identify gustatory (GR) and olfactory receptor (OR) genes in all aphid species considered in this analysis (Zhou et al., 2012). Briefly, well characterized protein sequences in *D. melanogaster* (Robertson et al., 2003) and *Ac. pisum* (Smadja et al., 2009) were collected as seeds. TBLASTN ($E < 10^{-5}$) was used to find potential loci across the genome, which were further processed using GeneWise v2.2.0 (Birney et al., 2004) to predict gene structures. This approach was iterated until no novel genes were identified. Ionotropic receptors (IR) were identified based on conserved protein domains of InterProScan. Odorant-binding proteins (OBPs), detoxification genes, and immune related genes in five aphid species were identified using BLASTP with the corresponding gene sets of *D. melanogaster* as seed sequences, the best scored protein sequenced by reciprocal alignments were determined as the potential corresponding genes.

See the Supplementary Methods for detailed information.

3. Results and discussion

3.1. Genome assembly and evaluation

A total of 77.3 Gb of clean reads, representing 263X coverage of the genome, were generated for *de novo* assembling of the *Ap. gossypii* genome. Based on the strategy described above, our *de novo* assembly resulted in a 294 Mb draft genome of *Ap. gossypii*, including 22,569 contigs with an N50 size of 45.6 Kb and 4,724 scaffolds with an N50 length of 438 Kb. The largest contig and scaffold was 471.7 Kb and 5.6 Mb, respectively (Table 3). Compared to the published genomes of other aphid species (International Aphid Genomics, 2010; Mathers et al., 2017; Nicholson et al., 2015; Wenger et al., 2017), the *Ap. gossypii* genome presented the most compact genome size and the lowest GC content (Table 4). We note that the assembled genome size is relatively smaller than the previously estimate based on flow cytometry (Wenger et al., 2017), probably due to the underrepresentation of repeats in de Bruijn-based assembly.

The assembly assessment using CEGMA showed that 94.35% potential conserved genes were completely recovered and 97.18% of them were partially recovered (Table 3). Additionally, the BUSCO approach (Waterhouse et al., 2017) revealed that 94.0% of the universal single-copy orthologs were captured and 91.0% of them were intact. These features suggest that the genome assembly of *Ap. gossypii* presents

Table 3
Features of the genome assembly across five aphid species with published genomes.

	<i>Ap.gossypii</i>	<i>Ap.glycines</i>	<i>D.noxia</i>	<i>M.persicae</i>	<i>Ac.pisum</i>
Assembled size (Mb)	294	303	395	347	542
Number of Scaffolds	4,724	8,397	5,637	4,022	23,924
Scaffold N50 size (bp)	437,960	174,505	397,774	435,781	518,546
Scaffold N90 size (bp)	69,749	23,219	37,041	69,715	43,176
Largest Scaffold length (bp)	5,574,629	1,366,044	2,142,037	2,199,663	3,073,041
Number of Contigs	22,569	35,012	53,013	8,249	60,623
Contig N50 size (bp)	45,572	56,892	12,406	144,275	28,192
Contig size (Mb)	278	303	297	345	500
Largest contig length (bp)	471,691	439,536	147,337	955,056	424,120
CEGMA genome (complete/partial)	94.35%/97.18%	92.74%/93.95%	88.31%/95.16%	95.97%/99.19%	94.35%/99.60%
BUSCO genome (complete/partial)	91.0%/94.0%	91.2%/93.1%	88.1%/93.0%	94.5%/95.9%	94.0%/95.5%

Table 4
Genomic features of aphid species with published genomes.

	<i>Ap. gossypii</i>	<i>Ap. glycines</i>	<i>M. persicae</i>	<i>Ac. pisum</i>	<i>D. noxia</i>
% Repeat	13.55	16.14	18.69	29.44	16.18
% G + C	27.26	27.79	30.03	29.76	29.06
% Coding	7.00	7.85	7.19	13.26	5.55
# Protein-coding genes	14,694	19,182	18,529	36,195	19,097 ^a
Mean CDS length (bp)	1,402	1,240	1,347	1,984	1,804

^a The number of sequences we were able to download from NCBI is 17,476.

relatively complete gene content, which is comparable or superior to the four previously published aphid genome assemblies (Table 3).

3.2. Gene prediction and annotation

Our official gene set of 14,694 protein coding genes (OGS1.0) was based on the non-redundant GLEAN models with at least one type of evidence from either transcriptome or insect homologs. Quality control showed that OGS1.0 is able to recover 98.03% of CEGMA genes and 93.5% of BUSCO genes, respectively. Overall, nearly 99.6% of our predicted genes could be annotated with evidence from selective public databases, and 74.02% genes were supported by transcriptome sequences (Table 5).

3.3. Genome evolution in *Ap. gossypii*

The orthology analyses identified 1,168 (7.9%) single-copy universal genes, as well as 6,747 (45.9%) multi-copy universal genes. We only found 347 *Ap. gossypii* specific genes, which is the smallest set across all examined aphid species and greatly less than that in the pea aphid *Ac. pisum* (4,356) (Fig. 2). We also found 1,110 genes that are universally presented in all examined aphid species and exclusively to other lineages (Fig. 2), which provided an important resource to study the unique biology of aphids.

We next compared the counts of IPR domains across the genomes of five examined aphid species. In *Ap. gossypii*, we identified a number of

Table 5
Feature summaries of the *Ap. gossypii* gene set.

	Annotated genes	% percentage of all genes
UniRef terms	14,134	96.19
RefSeq terms	13,555	92.25
KO terms	8,144	55.42
InterPro domains	10,005	68.09
PFAM domains	9,440	64.24
GO terms	7,231	49.21
EST terms	10,139	69.00
mapped RNAseq reads	10,876	74.02

domains, in which 34 domains were unique, while 123 other domains were missing (Table S3). We also found 157 overrepresented and 272 underrepresented domains, which were specific to *Ap. gossypii* (Table S3). As shown in Fig. S1, we note that the prominent overrepresented domains in *Ap. gossypii* are largely related to processes of oxidation-reduction, methylation, and hydrogenation, while the most prominent contraction of domains are mainly related to carbohydrate metabolic processes and exonuclease activity.

As expected, our phylogenetic analysis clearly clustered the nine species of Hemiptera together. Within the hemipteran clade, five aphid species were further clustered as a single clade and shown to be closely related to the whitefly. In accordance with morphological classification, the two species within the genus *Aphis* (*Ap. gossypii* and *Ap. glycines*) formed a sister clade to other aphid species (Fig. 2).

3.4. Rapidly evolving pathways in the evolution of the genus *Aphis*

We characterized genes and pathways under rapid evolution in *Ap. gossypii*. The pathways with high level of d_N/d_S were significantly occurred in sugar biosynthesis and metabolism pathways, such as the pathways of pentose and glucuronate interconversions, starch and sucrose metabolism, and other types of O-glycan biosynthesis (Fig. 3). Interestingly, we also found rapidly evolving genes in the circadian rhythm pathway (Fig. 3). In addition, we consistently identified pathways under rapid evolution in the genus *Aphis*. Three out of the nine identified pathways seemed associated with detoxification (Fig. 3). Given that both *Ap. gossypii* and *Ap. glycines* are world-wide agricultural pests, the finding of detoxification related pathways are under rapid evolution may provide insights into understanding the resistance to pesticides in *Aphis* aphids (Hanson et al., 2017; Pan et al., 2018; Xi et al., 2015).

3.5. Gene families related to environmental adaptation in *Ap. gossypii*

3.5.1. Chemosensory related genes

It is well known that chemosensory related genes play critical roles in locating host plants and detecting flavors in insect herbivores (Benton et al., 2009; Smadja et al., 2009). We annotated the full sets of chemosensory related genes, including various receptors and binding proteins, and performed comparisons across the other four aphid species. We identified a total of 34 ORs, 50 GRs, 23 IRs, and 12 OBPs in the *Ap. gossypii* genome (Table 6). Compared to other aphid species, the repertoire of chemosensory genes in *Ap. gossypii* presented a reduction in IRs (Table 6), while ORs and GRs were of the second smallest gene number only after *D. noxia*. Previously studies have hypothesized a potential relationship between the host range and the family size of chemosensory genes in aphids (Cao et al., 2014; Nicholson et al., 2015). Among the five aphid species, *Ap. gossypii* is the most highly polyphagous, it has the widest host range and can feed on over 500 different plants (Blackman and Eastop, 2000); *M. persicae* and *D. noxia* are also polyphagous, which can feed on over 400 and 140 plant species,

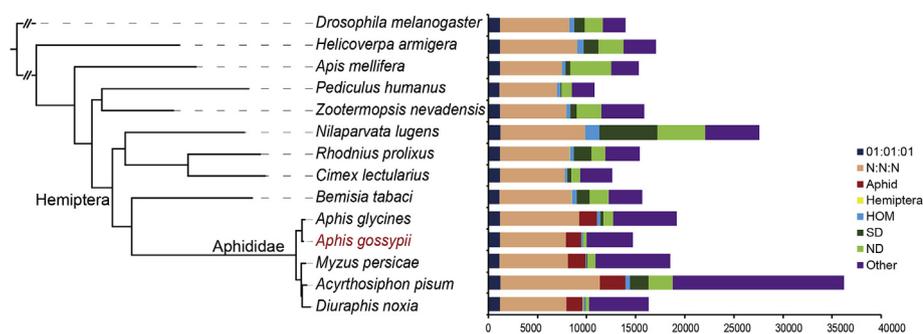


Fig. 2. Phylogeny and orthology comparison between *Ap. gossypii* and other insect species. The maximum-likelihood phylogenetic tree was built using *D. melanogaster* as the outgroup (left). Relationships of ortholog genes were compared among 14 insect species (right). "1:1:1" indicates single-copy genes in all species; "N:N:N" indicates multi-copy genes in all species; "Aphid" indicates aphid-specific genes and presence in all five aphid species; "Hemiptera" indicates hemipteran insects-specific genes; "HOM" indicates partial homology detected with $E < 10^{-5}$ but no orthology grouped; "SD" indicates species-specific duplicated genes; "ND" indicates species-specific genes; "Other" indicates unclassified orthologs.

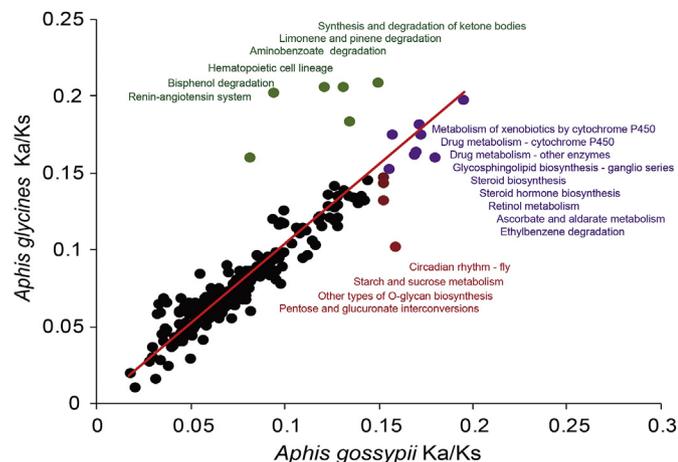


Fig. 3. Pathways under selection in *Ap. gossypii* and *Ap. glycines*. KEGG pathways of significantly higher median values of d_N/d_S are highlighted for *Ap. gossypii* (in red), *Ap. glycines* (in green), and both (in purple). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 6
Statistics of detoxification- and chemosensory-related genes across aphid species.

	Detoxification					Chemosensory			
	P450	ABC	GST	CCE	UGT	OR	GR	IR	OBP
<i>Ap. glycines</i>	64	87	11	23	49	37	64	26	8
<i>Ap. gossypii</i>	62	78	7	20	56	34	50	23	12
<i>M. persicae</i>	67	70	13	22	60	50	56	27	11
<i>D. noxia</i>	55	66	10	20	43	31	30	25	12
<i>Ac. pisum</i>	82	131	20	26	59	79	77	30	14

respectively (Pike et al., 1991; Blackman and Eastop, 2000). By contrast, *Ap. glycines* and *Ac. pisum* are specialists and oligophagous insects which only feed on Glycine/Rhamnus and fabaceae, respectively (Kanvil et al., 2014; Ragsdale et al., 2004). Correspondingly, gene numbers of GRs and IRs were found to be lower in polyphagous insects (*Ap. gossypii*, *M. persicae* and *D. noxia*) than those of specialists and oligophagous insects (*Ap. glycines* and *Ac. pisum*). Thus, this observation indicates a possible relationship between the gene family size of these two chemoreception receptors (GRs and IRs) and the host range of aphids (Cao et al., 2014; Nicholson et al., 2015).

3.5.2. Detoxification gene family

Insect herbivores need to adapt to a variety of plant secondary metabolites and insecticides, which are generally processed by the detoxification system (Ramsey et al., 2010). This system mainly involves several gene families, including cytochrome P450 monooxygenases

(P450s), ATP-binding cassette transporters (ABCs), glutathione-transferases (GSTs), carboxyl/cholinesterases (CCEs), and UDP-glycosyltransferases (UGTs). In total, we identified 62 P450, 78 ABC, 7 GST, 20 CCE, and 56 UGT genes in the genome of *Ap. gossypii* (Table 6). In comparison to other aphids, we found the genome of *Ap. gossypii* encodes the fewest GST genes. More interestingly, the two *Aphis* species (*Ap. gossypii* and *Ap. glycines*) with relatively small genome size occupied more ABC genes than other aphid species, except for *Ap. pisum*. Unexpectedly, the two sister *Aphis* species (*Ap. gossypii* and *Ap. glycines*) showed a similar number of detoxification genes, even though their plant host ranges are highly divergent. This result suggests that there is little concordance between the family size of detoxification related genes and plant host range in aphids. Alternatively, it is possible that the structure and functional divergence of the detoxification proteins or salivary proteins (Li et al., 2004; Nicholson et al., 2015) may influence the process of aphid-plant interactions and hence determine the host range of an aphid species; further research is needed to explore this possible explanation.

3.5.3. Immune system

Symbionts play multiple roles in the biology of aphids (Gauthier et al., 2015; Wilson and Duncan, 2015). Correspondingly, aphids have evolved a specific immune system to maintain the host-bacteria interaction. We characterized the full repertoire of immune related genes in five aphid species using the information of *D. melanogaster* (Early et al., 2017) as references. The insect immune system is generally categorized into three classes, i.e., recognition, signaling, and effector (Early et al., 2017). In the genome of *Ap. gossypii*, we annotated 27,244, and 24 genes in these three classes, respectively (Table 7). Among five analyzed aphid species, *Ap. gossypii* encodes the fewest genes in effecting processes, and the second fewest genes in both signaling and recognition classes. In addition, *Ap. gossypii* encodes an incomplete immune system in comparison with specialists and oligophagous aphids (*Ap. glycines* and *Ac. pisum*). It has been proposed that such gene contraction in the immune system may owe to the acquisition of microbial symbionts in *Ac. pisum* (International Aphid Genomics, 2010). The more substantial reduction in immune-related genes in *Ap. gossypii* may promote the acquisition of symbionts, hence help *Ap. gossypii* adapt to a

Table 7
Statistics of immune-related genes across aphid species.

	<i>D. noxia</i>	<i>M. persicae</i>	<i>Ap. glycines</i>	<i>Ap. gossypii</i>	<i>Ac. pisum</i>
Functional class					
Effector	26	25	32	24	34
Recognition	32	24	35	27	42
Signaling	228	272	295	244	392
Immune process					
IMD	44	68	47	48	99
JAK/STAT	33	39	49	30	60
Toll	42	61	55	52	99

wider range of plant hosts. There is a need for further explorations of the relationship between aphids immune system, their profiles of symbionts and plant host range.

4. Conclusions

In this study, we have presented a high-quality draft genome of *Ap. gossypii*, together with a consensus gene set and comprehensive annotations. Comparative analyses suggest that *Ap. gossypii* encodes the most compact genome in comparison to other published aphid genomes. We have also analyzed and compared gene families with respect to chemoreception, detoxification, and the immune system. These analyses will not only help to understand the complicated biological characteristics in *Ap. gossypii*, but also provide insights into the genome evolution of aphids. Furthermore, we find that rapidly evolving genes in the genus *Aphis* are enriched in pathways related to detoxification, a characteristic likely associated with the widespread development of resistance to chemical control in *Aphis* species.

List of acronyms

CEGMA: core eukaryotic genes mapping; BUSCO: benchmarking universal single-copy ortholog; ESTs: expressed sequence tags; CSP: chemosensory protein; GO terms: gene ontology terms; GR: gustatory receptor; IR: ionotropic receptor; OR: olfactory receptor; OBP: odorant-binding proteins; P450: cytochrome P450 monooxygenases; ABC: ATP-binding cassette transporters; GST: glutathione-S transferase; UGT: UDP-glycosyltransferases, NR database: non-redundant database; KEGG KO terms: Kyoto encyclopedia of genes and genomes KO terms; HSPs: high-scoring pairs.

Funding

This work was supported by the Natural Science Foundation of China (31601901, 31788103, 31571310, 31501877), Key research project of frontier science of Chinese Academy of Sciences (QYZDB-SSW-SMC029, XDB27040205, 173176001000162007), and the Shanghai Yangfan Project (to YHC, 16YF1413700).

Appendix A. Supplementary data

The draft genome of *Ap. gossypii* is available in project accession PRJNA431119 in the NCBI database. Common parameter files of some analyses are available at http://119.78.67.200/Agos_IBMB.html.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ibmb.2018.12.007>.

References

Agrofoglio, Y.C., Delfosse, V.C., Casse, M.F., Hopp, H.E., Kresic, I.B., Distefano, A.J., 2017. Identification of a new cotton disease caused by an atypical cotton leafroll dwarf virus in Argentina. *Phytopathology* 107, 369–376.

Benoit, J.B., Adelman, Z.N., Reinhardt, K., Dolan, A., Poelchau, M., Jennings, E.C., Szuter, E.M., Hagan, R.W., Gujar, H., Shukla, J.N., Zhu, F., Mohan, M., Nelson, D.R., Rosendale, A.J., Derst, C., Resnik, V., Wernig, S., Menegazzi, P., Wegener, C., Peschel, N., Hendershot, J.M., Blenau, W., Predel, R., Johnston, P.R., Ioannidis, P., Waterhouse, R.M., Nauen, R., Schorn, C., Ott, M.C., Maiwald, F., Johnston, J.S., Gondhalekar, A.D., Scharf, M.E., Peterson, B.F., Raje, K.R., Hottel, B.A., Armisen, D., Crumiere, A.J.J., Refki, P.N., Santos, M.E., Sghaier, E., Viala, S., Khila, A., Ahn, S.J., Childers, C., Lee, C.Y., Lin, H., Hughes, D.S.T., Duncan, E.J., Murali, S.C., Qu, J.X., Dugan, S., Lee, S.L., Chao, H., Dinh, H., Han, Y., Doddapaneni, H., Worley, K.C., Muzny, D.M., Wheeler, D., Panfilio, K.A., Jentsch, I.M.V., Vargo, E.L., Booth, W., Friedrich, M., Weirauch, M.T., Anderson, M.A.E., Jones, J.W., Mittapalli, O., Zhao, C.Y., Zhou, J.J., Evans, J.D., Attardo, G.M., Robertson, H.M., Zdobnov, E.M., Ribeiro, J.M.C., Gibbs, R.A., Werren, J.H., Palli, S.R., Schal, C., Richards, S., 2016. Unique features of a global human ectoparasite identified through sequencing of the bed bug genome. *Nat. Commun.* 7, 10165.

Benton, R., Vannice, K.S., Gomez-Diaz, C., Vossell, L.B., 2009. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* 136, 149–162.

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and genomewise. *Genome Res.* 14, 988–995.

Blackman, R.L., Eastop, V.F., 2000. *Aphids on the World's Crops. An Identification and Information Guide*. Wiley, Chichester, UK.

Burge, C., Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94.

Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S., Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196.

Cao, C.W., Zhang, J., Gao, X.W., Liang, P., Guo, H.L., 2008. Overexpression of carboxylesterase gene associated with organophosphorous insecticide resistance in cotton aphids, *Aphis gossypii* (Glover). *Pestic. Biochem. Physiol.* 90, 175–180.

Cao, D.P., Liu, Y., Walker, W.B., Li, J.H., Wang, G.R., 2014. Molecular characterization of the *Aphis gossypii* olfactory receptor gene families. *PLoS One* 9, e101187.

Carletto, J., Lombaert, E., Chavigny, P., Brevault, T., Lapchin, L., Vanlerberghe-Masutti, F., 2009. Ecological specialization of the aphid *Aphis gossypii* Glover on cultivated host plants. *Mol. Ecol.* 18, 2198–2212.

Chen, W., Hasegawa, D.K., Kaur, N., Klot, A., Pinheiro, P.V., Luan, J., Stensmyr, M.C., Zheng, Y., Liu, W., Sun, H., Xu, Y., Luo, Y., Kruse, A., Yang, X., Kongsedolov, S., Lebedev, G., Fisher, T.W., Nelson, D.R., Hunter, W.B., Brown, J.K., Jander, G., Cilia, M., Douglas, A.E., Ghanim, M., Simmons, A.M., Wintermantel, W.M., Ling, K.S., Fei, Z., 2016. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol.* 14, 110.

Early, A.M., Arguello, J.R., Cardoso-Moreira, M., Gottipati, S., Grenier, J.K., Clark, A.G., 2017. Survey of global genetic diversity within the drosophila immune system. *Genetics* 205, 353–366.

Ebert, T.A., Cartwright, B., 1997. Biology and ecology of *Aphis gossypii* Glover (Homoptera: Aphididae). *Southwest. Entomol.* 22, 116–153.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.

Elsik, C.G., Mackey, A.J., Reese, J.T., Milshina, N.V., Roos, D.S., Weinstock, G.M., 2007. Creating a honey bee consensus gene set. *Genome Biol.* 8, R13.

FlyBase, C., 1998. FlyBase: a Drosophila database. *Nucleic Acids Res.* 26, 85–88.

Gauthier, J.P., Outreman, Y., Miezuet, L., Simon, J.C., 2015. Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PLoS One* 10, e0120664.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S., Jaffe, D.B., 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1513–1518.

Gong, Y.H., Yu, X.R., Shang, Q.L., Shi, X.Y., Gao, X.W., 2014. Oral delivery mediated RNA interference of a carboxylesterase gene results in reduced resistance to organophosphorus insecticides in the cotton aphid, *Aphis gossypii* Glover. *PLoS One* 9, e102823.

Gong, Y.H., Ai, G.M., Li, M., Shi, X.Y., Diao, Q.Y., Gao, X.W., 2017. Functional characterization of carboxylesterase gene mutations involved in *Aphis gossypii* resistance to organophosphate insecticides. *Insect Mol. Biol.* 26, 702–714.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., Chen, Z.H., Maucleri, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652.

Gu, S.H., Wu, K.M., Guo, Y.Y., Field, L.M., Pickett, J.A., Zhang, Y.J., Zhou, J.J., 2013. Identification and expression profiling of odorant binding proteins and chemosensory proteins between two wingless morphs and a winged morph of the cotton aphid *Aphis gossypii* Glover. *PLoS One* 8, e73524.

Hanson, A.A., Menger-Anderson, J., Silverstein, C., Potter, B.D., MacRae, I.V., Hodgson, E.W., Koch, R.L., 2017. Evidence for soybean aphid (Hemiptera: Aphididae) resistance to pyrethroid insecticides in the upper midwestern United States. *J. Econ. Entomol.* 110, 2235–2246.

Herron, G.A., Wilson, L.J., 2017. Can resistance management strategies recover insecticide susceptibility in pests?: a case study with cotton aphid *Aphis gossypii* (Aphididae: Hemiptera) in Australian cotton. *Aust. Entomol.* 56, 1–13.

International Aphid Genomics, C., 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8, e1000313.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W.Z., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quin, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., Kohara, Y., Fujiyama, A., Hayashi, T., Itoh, T., 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* 24, 1384–1395.

Kanvil, S., Powell, G., Turnbull, C., 2014. Pea aphid biotype performance on diverse Medicago host genotypes indicates highly specific virulence and resistance functions. *Bull. Entomol. Res.* 104, 689–701.

Kim, D., Pertea, C., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14.

Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., Gerlach, D., Kriventseva, E.V., Elsik, C.G., Graur, D., Hill, C.A., Veenstra, J.A., Walenz, B., Tubio, J.M.C., Ribeiro, J.M.C., Rozas, J., Johnston, J.S., Reese, J.T., Popadic, A., Tojo, M., Raouf, D., Reed, D.L., Tomoyasu, Y., Krause, E., Mittapalli, O., Margam, V.M., Li, H.-M., Meyer, J.M., Johnson, R.M., Romero-Severson, J., VanZee, J.P., Alvarez-Ponce, D., Vieira, F.G., Aguade, M., Romero-Rico, S., Anzola, J.M., Yoon, K.S., Strycharz, J.P., Unger, M.F., Christley, S., Lobo, N.F., Seufferheld, M.J., Wang, N., Dasch, G.A., Struchiner, C.J., Madey, G.,

- Hannick, L.I., Bidwell, S., Joardar, V., Caler, E., Shao, R., Barker, S.C., Cameron, S., Bruggner, R.V., Regier, A., Johnson, J., Viswanathan, L., Utterback, T.R., Sutton, G.G., Lawson, D., Waterhouse, R.M., Venter, J.C., Strausberg, R.L., Berenbaum, M.R., Collins, F.H., Zdobnov, E.M., Pittendrigh, B.R., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. U. S. A* 107, 12168–12173.
- Korf, I., 2004. Gene finding in novel genomes. *BMC Bioinf.* 5, 59.
- Kwon, S.H., Kim, D.S., 2017. Effects of temperature and photoperiod on the production of sexual morphs of *Aphis gossypii* (Hemiptera: Aphididae) in Jeju, Korea. *J. Asia Pac. Entomol.* 20, 53–56.
- Le Trionnaire, G., Jaubert-Possamai, S., Bonhomme, J., Gauthier, J.-P., Guerneq, G., Le Cam, A., Legeai, F., Monfort, J., Tagu, D., 2012. Transcriptomic profiling of the reproductive mode switch in the pea aphid in response to natural autumnal photoperiod. *J. Insect Physiol.* 58, 1517–1524.
- Li, L., Stoekert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189.
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J., 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20, 265–272.
- Li, X.C., Baudry, J., Berenbaum, M.R., Schuler, M.A., 2004. Structural and functional divergence of insect CYP6B proteins: from specialist to generalist cytochrome P450. *Proc. Natl. Acad. Sci. U. S. A* 101, 2939–2944.
- Liu, J.J., Zheng, H.Y., Jiang, F., Guo, W., Zhou, S.T., 2014. Comparative transcriptional analysis of asexual and sexual morphs reveals possible mechanisms in reproductive polyphenism of the cotton aphid. *PLoS One* 9, e99506.
- Mandric, I., Zelikovskiy, A., 2015. ScaffoldMatch: scaffolding algorithm based on maximum weight matching. *Bioinformatics* 31, 2632–2638.
- Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Collella, S., Collin, O., Dalmay, T., Derrien, T., Feng, H., Gabaldon, T., Jordan, A., Julca, I., Kettles, G.J., Kowitzanich, K., Lavenier, D., Lenzi, P., Lopez-Gomollon, S., Loska, D., Mapleson, D., Maumus, F., Moxon, S., Price, D.R.G., Sugio, A., van Munster, M., Uzet, M., Waite, D., Jander, G., Tagu, D., Wilson, A.C.C., van Oosterhout, C., Swarbrick, D., Hogenhout, S.A., 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol.* 18, 27 (vol. 18, 27, 2017).
- Mesquita, R.D., Vionette-Amaral, R.J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F.A., Minx, P., Spieth, J., Bernardo Carvalho, A., Panzera, F., Lawson, D., Torres, A.Q., Ribeiro, J.M.C., Sorgine, M.H.F., Waterhouse, R.M., Montague, M.J., Abad-Franch, F., Alves-Bezerra, M., Amaral, L.R., Araujo, H.M., Araujo, R.N., Aravind, L., Atella, G.C., Azambuja, P., Berni, M., Bittencourt-Cunha, P.R., Braz, G.R.C., Calderon-Fernandez, G., Carareto, C.M.A., Christensen, M.B., Costa, I.R., Costa, S.G., Dansa, M., Daumas-Filho, C.R.O., De-Paula, I.F., Dias, F.A., Dimopoulos, G., Emrich, S.J., Esponda-Behrens, N., Fampa, P., Fernandez-Medina, R.D., da Fonseca, R.N., Fontenele, M., Fronick, C., Fulton, L.A., Gandara, A.C., Garcia, E.S., Genta, F.A., Giraldo-Calderon, G.I., Gomes, B., Gondim, K.C., Granzotto, A., Guarneri, A.A., Guigo, R., Harry, M., Hughes, D.S.T., Jablonka, W., Jacquín-Joly, E., Patricia Juarez, M., Koerich, L.B., Latorre-Estivalis, J.M., Lavore, A., Lawrence, G.G., Lazoski, C., Lazzari, C.R., Lopes, R.R., Lorenzo, M.G., Lugon, M.D., Majerowicz, D., Marcet, P.L., Mariotti, M., Masuda, H., Megy, K., Melo, A.C.A., Missirli, F., Mota, T., Noriega, F.G., Nouzova, M., Nunes, R.D., Oliveira, R.L.L., Oliveira-Silveira, G., Ons, S., Pagola, L., Paiva-Silva, G.O., Pascual, A., Pavan, M.G., Pedrini, N., Peixoto, A.A., Pereira, M.H., Pike, A., Polycarpo, C., Prosdociimi, F., Ribeiro-Rodrigues, R., Robertson, H.M., Salerno, A.P., Salmon, D., Santesmasses, D., Schama, R., Seabra-Junior, E.S., Silva-Cardoso, L., Silva-Neto, M.A.C., Souza-Gomes, M., Sterkel, M., Taracena, M.L., Tojo, M., Tu, Z.J., Tubio, J.M.C., Ursic-Bedoya, R., Venancio, T.M., Walter-Nuno, A.B., Wilson, D., Warren, W.C., Wilson, R.K., Huebner, E., Dotson, E.M., Oliveira, P.L., 2015. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proc. Natl. Acad. Sci. U. S. A* 112, 14936–14941.
- Moran, N.A., 1992. The evolution of aphid life-cycles. *Annu. Rev. Entomol.* 37, 321–348.
- Nicholson, S.J., Nickerson, M.L., Dean, M., Song, Y., Hoyt, P.R., Rhee, H., Kim, C., Puterka, G.J., 2015. The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics* 16.
- Pan, Y., Chai, P., Zheng, C., Xu, H., Wu, Y., Gao, X., Xi, J., Shang, Q., 2018. Contribution of cytochrome P450 monooxygenase CYP380C6 to spirotetramat resistance in *Aphis gossypii* Glover. *Pestic. Biochem. Physiol.* 148, 182–189.
- Pan, Y., Peng, T., Gao, X., Zhang, L., Yang, C., Xi, J., Xin, X., Bi, R., Shang, Q., 2015. Transcriptomic comparison of thiamethoxam-resistance adaptation in resistant and susceptible strains of *Aphis gossypii* Glover. *Comparative Biochemistry and Physiology D-Genomics & Proteomics* 13, 10–15.
- Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067.
- Pearce, S.L., Clarke, D.F., East, P.D., Elfekih, S., Gordon, K.H.J., Jermini, L.S., McGaughran, A., Oakeshott, J.G., Papanikolaou, A., Perera, O.P., Rane, R.V., Richards, S., Tay, W.T., Walsh, T.K., Anderson, A., Anderson, C.J., Asgari, S., Board, P.G., Bretschneider, A., Campbell, P.M., Chertemps, T., Christeller, J.T., Coppin, C.W., Downes, S.J., Duan, G., Farnsworth, C.A., Good, R.T., Han, L.B., Han, Y.C., Hatje, K., Horne, I., Huang, Y.P., Hughes, D.S.T., Jacquín-Joly, E., James, W., Jhangiani, S., Kollmar, M., Kuwar, S.S., Li, S., Liu, N.Y., Maibeche, M.T., Miller, J.R., Montagne, N., Perry, T., Qu, J., Song, S.V., Sutton, G.G., Vogel, H., Walenz, B.P., Xu, W., Zhang, H.J., Zou, Z., Batterham, P., Edwards, O.R., Feyereisen, R., Gibbs, R.A., Heckel, D.G., McGrath, A., Robin, C., Scherer, S.E., Worley, K.C., Wu, Y.D., 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.
- Pike, K.S., Allison, D., Tanigoshi, L.K., Harwood, R.F., Clement, S.L., Halbert, S.E., Smith, C.M., Johnson, J.B., Reed, G.L., Zwer, P.K., 1991. Russian Wheat Aphid-biology, Damage and Management. Pacific Northwest Extension Publication, pp. 1–23 Publ. PNW371.
- Ragsdale, D.W., Voegtlin, D.J., O'Neil, R.J., 2004. Soybean aphid biology in North America. *Ann. Entomol. Soc. Am.* 97, 204–208.
- Ramos, S., Moya, A., Martinez-Torres, D., 2003. Identification of a gene overexpressed in aphids reared under short photoperiod. *Insect Biochem. Mol. Biol.* 33, 289–298.
- Ramsey, J.S., Rider, D.S., Walsh, T., de Vos, M., Gordon, K., Ponnala, L., Roe, B.A., Jander, G., 2010. Comparative analysis of detoxification enzymes in *Acyrthosiphon pisum* and *Myzus persicae*. *Insect Mol. Biol.* 19, 155–164.
- Robertson, H.M., Warr, C.G., Carlson, J.R., 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A* 100, 14537–14542.
- Schoeny, A., Desbiez, C., Millot, P., Wipf-Scheibel, C., Nozeran, K., Gognalons, P., Lecoq, H., Boissot, N., 2017. Impact of Vat resistance in melon on viral epidemics and genetic structure of virus populations. *Virus Res.* 241, 105–115.
- She, R., Chu, J.S.C., Wang, K., Pei, J., Chen, N., 2009. genBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* 19, 143–149.
- Slater, G.S., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* 6, 31.
- Slosser, J.E., Parajulee, M.N., Hendrix, D.L., Henneberry, T.J., Rummel, D.R., 2002. Relationship between *Aphis gossypii* (Homoptera: Aphididae) and sticky lint in cotton. *J. Econ. Entomol.* 95, 299–306.
- Smadja, C., Shi, P., Butlin, R.K., Robertson, H.M., 2009. Large gene family expansions and adaptive evolution for odorant and gustatory receptors in the pea aphid, *Acyrthosiphon pisum*. *Mol. Biol. Evol.* 26, 2073–2086.
- Stamatakis, A., 2014. RaxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34, W435–W439.
- Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577.
- Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P., Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G., Huylmans, A.K., Khalil, S.M.S., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H., Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent, C.S., Elsie, C.G., Goodisman, M.A.D., Liberles, D.A., Roe, R.M., Vargo, E.L., Vilcinskis, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5, 3636.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511 U174.
- Vereschagina, A., Gandrabur, E., 2014. Polymorphism and damage of aphids (Homoptera: Aphidoidea). *Int. J. Biol.* 6, 124–138.
- Wang, L., Zhang, S., Luo, J.Y., Wang, C.Y., Lv, L.M., Zhu, X.Z., Li, C.H., Cui, J.J., 2016. Identification of *Aphis gossypii* Glover (Hemiptera: Aphididae) biotypes from different host plants in north China. *PLoS One* 11, e0146345.
- Waterhouse, R.M., Seppey, M., Simao, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35, 543–548.
- Weinstock, G.M., Robinson, G.E., Gibbs, R.A., Worley, K.C., Evans, J.D., Maleszka, R., Robertson, H.M., Weaver, D.B., Beye, M., Bork, P., Elsie, C.G., Hartfelder, K., Hunt, G.J., Zdobnov, E.M., Amdam, G.V., Bitondi, M.M.G., Collins, A.M., Cristiano, A.S., Lattorff, H.M.G., Lobo, C.H., Moritz, R.F.A., Nunes, F.M.F., Page Jr., R.E., Simoes, Z.L.P., Wheeler, D., Carninci, P., Fukuda, S., Hayashizaki, Y., Kai, C., Kawai, J., Sakazume, N., Sasaki, D., Tagami, M., Albert, S., Baggerman, G., Beggs, K.T., Bloch, G., Cazzamali, G., Cohen, M., Drapeau, M.D., Eisenhardt, D., Emore, C., Ewing, M.A., Fahrback, S.E., Foret, S., Grimmelikhuijzen, C.J.P., Hauser, F., Hummon, A.B., Huybrechts, J., Jones, A.K., Kadowaki, T., Kaplan, N., Kucharski, R., Leboulle, G., Linial, M., Littleton, J.T., Mercer, A.R., Richmond, T.A., Rodriguez-Zas, S.L., Rubin, E.B., Sattelle, D.B., Schlipalius, D., Schoofs, L., Shemesh, Y., Sweedler, J.V., Velarde, R., Verleyen, P., Vierstraete, E., Williamson, M.R., Ament, S.A., Brown, S.J., Corona, M., Dearden, P.K., Dunn, W.A., Elekonich, M.M., Fujiyuki, T., Gattermeier, I., Gempe, T., Hasselmann, M., Kadowaki, T., Kage, E., Kamikouchi, A., Kubo, T., Kucharski, R., Kunieda, T., Lorenzen, M.D., Milshina, N.V., Morioka, M., Ohashi, K., Overbeek, R., Ross, C.A., Schioett, M., Shippy, T., Takeuchi, H., Toth, A.L., Willis, J.H., Wilson, M.J., Gordon, K.H.J., Letunic, I., Hackett, K., Peterson, J., Felsenfeld, A., Guyer, M., Solignac, M., Agarwala, R., Cornuet, J.M., Monnerot, M., Mougel, F., Reese, J.T., Vautrin, D., Gillespie, J.J., Cannone, J.J., Gutell, R.R., Johnston, J.S., Eisen, M.B., Iyer, V.N., Iyer, V., Kosarev, P., Mackey, A.J., Solovyev, V., Souvorov, A., Aronstein, K.A., Bilikova, K., Chen, Y.P., Clark, A.G., Decanini, L.L., Gelbart, W.M., Hetru, C., Hultmark, D., Imler, J.-L., Jiang, H., Kanost, M., Kimura, K., Lazzaro, B.P., Lopez, D.L., Simuth, J., Thompson, G.J., Zou, Z., De Jong, P., Sodergren, E., Csuroes, M., Milosavljevic, A., Osoegawa, K., Richards, S., Shu, C.L., Duret, L., Elhaik, E., Graur, D., Anzola, J.M., Campbell, K.S., Childs, K.L., Collinge, D., Crosby, M.A., Dickens, C.M., Grametes, L.S., Grozinger, C.M., Jones, P.L., Jorda, M., Ling, X., Matthews, B.B., Miller, J., Mizzen, C., Peinado, M.A., Reid, J.G., Russo, S.M., Schroeder, A.J., St Pierre, S.E., Wang, Y., Zhou, P., Jiang, H., Kitts, P., Ruef, B., Venkatraman, A., Zhang, L., Aquino-Perez, G., Whitfield, C.W., Behura, S.K., Berlocher, S.H., Sheppard, W.S., Smith, D.R., Suarez, A.V., Tsutsui, N.D., Wei, X., Wheeler, D., Havlak, P., Li, B., Liu, Y., Sodergren, E., Jolivet, A., Lee, S., Nazareth, L.V., Piu, L.L., Thorn, R., Stolt, V., Newman, T., Samanta, M., Tongprasit, W., Claudiano, C., Berenbaum, M.R., Biswas, S., de Graaf, D.C., Feyereisen, R., Johnson, R.M., Oakeshott, J.G., Ranson, H.,

- Schuler, M.A., Muzny, D., Chacko, J., Davis, C., Dinh, H., Gill, R., Hernandez, J., Hines, S., Hume, J., Jackson, L., Kovar, C., Lewis, L., Miner, G., Morgan, M., Nguyen, N., Okwuonu, G., Paul, H., Santibanez, J., Savery, G., Svatek, A., Villasana, D., Wright, R., Honeybee Genome Sequencing, C., 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443, 931–949.
- Wenger, J.A., Cassone, B.J., Legeai, F., Johnston, J.S., Bansal, R., Yates, A.D., Coates, B.S., Pavinato, V.A.C., Michel, A., 2017. Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochem. Mol. Biol.* <https://doi.org/10.1016/j.ibmb.2017.01.005>. pii: S0965-1748(17)30005-X.
- Wilson, A.C.C., Duncan, R.P., 2015. Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc. Natl. Acad. Sci. U. S. A* 112, 10255–10261.
- Xi, J., Pan, Y.O., Bi, R., Gao, X.W., Chen, X.W., Peng, T.F., Zhang, M., Zhang, H., Hu, X.Y., Shang, Q.L., 2015. Elevated expression of esterase and cytochrome P450 are related with lambda-cyhalothrin resistance and lead to cross resistance in *Aphis glycines* Matsumura. *Pestic. Biochem. Physiol.* 118, 77–81.
- Xue, J., Zhou, X., Zhang, C.X., Yu, L.L., Fan, H.W., Wang, Z., Xu, H.J., Xi, Y., Zhu, Z.R., Zhou, W.W., Pan, P.L., Li, B.L., Colbourne, J.K., Noda, H., Suetsugu, Y., Kobayashi, T., Zheng, Y., Liu, S., Zhang, R., Liu, Y., Luo, Y.D., Fang, D.M., Chen, Y., Zhan, D.L., Lv, X.D., Cai, Y., Wang, Z.B., Huang, H.J., Cheng, R.L., Zhang, X.C., Lou, Y.H., Yu, B., Zhuo, J.C., Ye, Y.X., Zhang, W.Q., Shen, Z.C., Yang, H.M., Wang, J., Wang, J., Bao, Y.Y., Cheng, J.A., 2014. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biol.* 15, 521.
- Yang, X., Liu, X., Xu, X., Li, Z., Li, Y., Song, D., Yu, T., Zhu, F., Zhang, Q., Zhou, X., 2014. Gene expression profiling in winged and wingless cotton aphids, *Aphis gossypii* (Hemiptera: Aphididae). *Int. J. Biol. Sci.* 10, 257–267.
- Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Zhang, Y.C., Lei, H.X., Miao, N.H., Liu, X.D., 2017. Comparative transcriptional analysis of the host-specialized aphids *Aphis gossypii* (Hemiptera: Aphididae). *J. Econ. Entomol.* 110, 702–710.
- Zhou, X., Slone, J.D., Rokas, A., Berger, S.L., Liebig, J., Ray, A., Reinberg, D., Zwiebel, L.J., 2012. Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genet.* 8, e1002930.