GMR

RNA-Seq-based transcriptome and the reproduction-related genes for the aphid *Schlechtendalia chinensis* (Hemiptera, Aphididae)

P. Liu[1,2,3], Z.X. Yang[1,2], X.M. Chen[1,2] and P. Yang[1,2]

[1]Research Institute of Resource Insects, Chinese Academy of Forestry, Kunming, Yunnan, China
[2]Key Laboratory of Breeding and Utilization of Resource Insects of State Forestry Administration, Kunming, Yunnan, China
[3]Yunnan Forestry Technological College, Kunming, Yunnan, China

Corresponding author: Z.X. Yang
E-mail: yzx1019@163.com

Genet. Mol. Res. 16 (1): gmr16019448
Received October 18, 2016
Accepted December 5, 2016
Published March 22, 2017
DOI http://dx.doi.org/10.4238/gmr16019448

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License

ABSTRACT. Chinese galls form on sumac plants (*Rhus chinensis*) and are used for medicinal and chemical purposes, due to their richness in tannins. The galls are formed by aphids, the most prominent of which is *Schlechtendalia chinensis*, which forms horn-shaped galls on the winged rachis of *R. chinensis*. *S. chinensis* has a complex life cycle, including both *R. chinensis* and certain mosses as hosts, as well as the existence of both sexual and asexual reproduction (cyclical parthenogenesis). Previous studies have shown that the alternate occurrence of sexual and asexual reproduction relies on many environmental factors, such as temperature, photoperiod, and host-plant. However, the sexual and asexual modes of reproduction are poorly understood on the molecular level. We aimed to

identify genes that respond to changes in temperature that may be related to the reproduction process. We compared the transcriptome of two samples of *S. chinensis*, which had been reared at different temperatures. Using gene ontology analysis, a total of 51 evolutionary conserved genes related to reproduction determination in insects were identified. Of these genes, *S. chinensis* harbors 42 genes. When we analyzed aphids that had asexually or sexually produced offspring, eight out of these 42 genes were identified and expressed differently in two temperature conditions. This is the first report on genes associated with reproduction determination in *S. chinensis*, which has a complex life cycle. Genes, expressed differently in response to different temperature conditions will be helpful to understand the mechanism of aphid reproductive determination.

Key words: RNA-seq; Transcriptome; Reproduction-related gene; Horned gall aphid; *Schlechtendalia chinensis*

INTRODUCTION

Chinese galls are abnormal growths of plant tissue induced by some aphids belonging to the tribe Fordini (Eriosomatinae, Aphididae). Among these aphids, eleven species form galls on *Rhus* trees. These species are mainly distributed in East Asia, especially in southwest China. The galls have been used for medicinal and chemical purposes, as they are rich in tannins that account for about 60-70% of total dry weight. The galls are most often used as a source of tannic, gallic, and pyrogallic acids (Zhang, 1987). *Schlechtendalia chinensis*, the best-studied gall-producing aphid, induces sealed horn-shaped galls on the winged rachis of *R. chinensis*. The horned galls account for about 75% of Chinese gall production (Tang and Cai, 1957).

S. chinensis has a complex life cycle that includes two different hosts. In the early spring, alate sexuparae fly to the trunk of their primary host-plant *R. chinensis*, to asexually produce sexual females and males that mate to produce fundatrices ovoviviparously. The fundatrices crawl to and feed on the tender leaves and initiate gall formation (Zhang and Zhong, 1983). The fundatrices reproduce parthenogenetically for three generations within a gall. Usually each gall is induced by a single fundatrix. After the gall matures and dehisces in autumn, the third alate fundatrigeniae migrate from the gall to their secondary host; mosses. There, they asexually produce larvae that live in the tender stem of mosses and excrete wax to envelop themselves for overwintering. The following spring, the larvae molt and form alate sexuparae, which then fly back to *R. chinensis* and start the next life cycle (Zhang et al., 1999; Shao et al., 2013; Liu et al., 2014).

The alternation of sexual and asexual reproduction (cyclical parthenogenesis) is mainly influenced by temperature, which plays an important role for aphid polymorphism (Dixon, 1973; Lamb, 1992; Artacho et al., 2011). Zhang et al. (1993) found that all overwintering larvae of *S. chinensis* developed into alate sexuparae followed by production of males and females when they were reared at 4.5°C. However, all larva developed to apterous larvae followed by parthenogenetic reproduction on mosses when they were reared at 22°C (Zhang et al., 1993). By contrast, the larva developed into alate sexuparae and apterous larvae at different ratios, when reared at 4.5°-18°C (Zhang et al., 1993).

RNA sequencing (RNA-Seq) is important for microarray and quantitative real-time polymerase chain reaction (qRT-PCR) techniques (Malone and Oliver, 2011; Ozsolak and

Milos, 2011). It covers a broad range of transcript abundances and has been applied to a wide range of species including human, mouse, yeast, and some plant species (Nagalakshmi et al., 2008; Sultan et al., 2008; Koenig et al., 2013). Herein, we applied RNA-Seq to construct cDNA libraries to analyze the global transcriptome of *S. chinensis*. This was done to identify differentially expressed transcripts between two samples that differed only with respect to the temperature at which they had been reared. In addition, we selected some genes related to sex determination from other insects, to compare and analyze with our unigenes. A preliminary analysis of sex determination of *S. chinensis* was performed.

MATERIAL AND METHODS

Host-plant and aphid production

In August 2013, the winter host moss, *Plagiomnium maximoviczii* (Lindb.) T. Kop., was cultivated in plastic trays (30 x 20 x 3 cm, with 2 cm loose loam) in a greenhouse at Kunming, Yunnan Province, Southwest China. Conventional management was carried out, in order to avoid disease and pest infestation.

In the same autumn, mature horned galls were collected from the field and moved to the moss nursery. Alate fundatrigeniae were allowed to migrate from the gall dehiscence to the mosses and to asexually produce larvae on the tender stems. After a few days, when the larvae were observed to excrete white wax to envelop themselves, all trays were transferred to an incubator kept at 75% relative humidity and 13L:11D photoperiod. Three constant temperatures, 7.5, 18 and 22°C, respectively, were used (Zhang et al., 1993). About 1000 aphid individuals were included in each treatment.

After 40 days, 150 aphid larvae were collected from each treatment and stored at -70°C in Trizol (Invitrogen, Carlsbad, California, USA). The rest of the aphids were left to overwinter until next spring. At this point, their eclosion rates at the three different temperatures were measured. The aphid samples reared at 7.5° and 18°C were named LA (low temperature) and HA (high temperature), respectively. And the aphids stored in Trizol from the LA and HA treatments were used for further RNA analysis.

RNA-Seq, clustering, and transcriptome assembly

Total aphid RNA was isolated using the Trizol kit (Invitrogen) and submitted to Novogene (Beijing, China) for RNA-Seq analysis. RNA purity was examined using a NanoPhotometer® spectro photometer (IMPLEN, Westlake Village, CA, USA) and degradation and contamination was investigated on a 1% agarose gel. RNA concentration was measured using Qubit® RNA Assay Kit in a Qubit® 2.0 Flurometer (Life Technologies, Carlsbad, CA, USA). Sequencing libraries were generated using NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, Beverly, MA, USA) following the manufacturer recommendations. Index codes were used to attribute sequences to each sample. Finally, products were purified using the AMPure XP system (Beckman Coulter, Beverly, MA, USA) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina, USA) according to the manufacturer instructions. After cluster generation, the library preparations were sequenced on an Illumina

HiSeq 2000 platform and paired-end reads were generated. Finally, the transcriptome assembly was accomplished using Trinity (Grabherr et al., 2011) with min_kmer_cov set to 2 and all other parameters set to default.

Gene annotation and functional classification

Gene function was annotated using the following databases: Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family), KOG (euKaryotic Orthologous Groups), Swiss-Prot (a manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database), and GO (Gene Ontology). The GO enrichment analysis of differentially expressed genes (DEGs) was implemented using GOseq R package based on Wallenius non-central hyper-geometric distribution (Young et al., 2010), which can adjust for length biases in the DEGs. KEGG (Kanehisa et al., 2008) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, organism, and ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). KOBAS (Mao et al., 2005) software was used to test the statistical enrichment of DEGs in the KEGG pathways.

Gene expression analysis

Quantification of transcript levels was estimated by RSEM (Li and Dewey, 2011). For each sample: 1) clean data were mapped back onto the assembled transcriptome; 2) a read count for each gene was obtained from the mapping results. A differential expression analysis of two samples was performed using the DESeq R package (1.10.1). DESeq provides statistical routines for determining differential expression in digital gene expression data, using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini-Hochberg approach for controlling false discovery rates (Benjamini and Hochberg, 1995). Genes with an adjusted $P \leq 0.05$ found by DESeq were assigned as differentially expressed.

qRT-PCR analysis

Total RNA of each sample was reverse-transcribed in a 20-mL reaction system according to the protocol provided with the M-MLV first strand kit (Invitrogen, Carlsbad, California, USA). A total of 12 genes, identified as being differently expressed in this study, were selected for qRT-PCR analysis. EvaGreen (Bio-RAD, California, USA) was used as DNA-binding fluorescent dye, and tubulin was used as internal standard. The absolute concentrations of target genes in two samples were determined using the standard curve quantitative method and the statistical analyses were performed using the least significant difference test at $P = 0.01$ using SPSS statistical software.

RESULTS

Offspring differentiation and eclosion rates of overwintering larvae

The results of offspring differentiation and eclosion rates of overwintering larvae

reared at different temperatures indicated that 1) at 7.5°C, 241 alate sexuparae migrated, 28 remained in the mosses, and the eclosion rate was 89.6%. All emigrants produced sexual male and female. 2) At 18°C, only 6 alate sexuparae migrated and 107 remained in the mosses. The eclosion rate was 5.3%. 3) At 22°C, no alate sexuparae migrated, instead all continued reproducing parthenogenetically in the mosses.

Transcriptome sequencing and assembly

The Illumina sequencing data of *S. chinensis* generated 126,726,344 raw reads. After removing adapter sequences, ambiguous nucleotides, and low-quality sequences, 122,039,510 clean reads remained. The mean read length was 666 bp. All clean reads were assembled into 84,280 unigenes that ranged from 201 to 23,871 bp with an N50 length of 1073 bp.

Functional annotation and pathway assignment

All assembled unigenes were annotated against the sequences in Nr, Nt, KO, SwissProt, Pfam, GO, and KOG databases. In the Nr protein database, 24,948 genes were annotated, which accounted for 29.6% of all assembled unigenes. In all databases, 4403 (5.22%) unigenes were annotated and 34,753 (41.23%) unigenes were annotated in at least one database (Table 1). According to GO, 28,484 non-redundant unigenes were classified into three major functional categories (biological process, cellular component, and molecular function) and 54 subcategories (Figure 1).

Table 1. Annotation statistics of all assembled unigenes from Nr, Nt, KO, SwissProt, Pfam, GO, and KOG.

Annotated	Number of Unigenes	NPercentage (%)
Nr	24,948	29.6
Nt	10,702	12.69
KO	11,747	13.93
SwissProt	19,630	23.29
PFAM	26,170	31.05
GO	28,484	33.79
KOG	16,346	19.39
all Databases	4,403	5.22
at least one Database	34,753	41.23
Total Unigenes	84,280	100

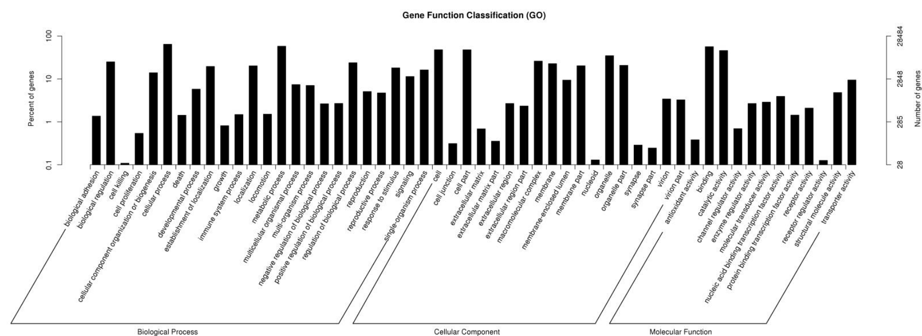

Figure 1. GO categories of 28,484 non-redundant unigenes. The right y-axis indicates the number of genes in each category, whereas the left y-axis indicates the percentage of a specific gene subcategory within each main category.

To classify orthologous gene products, 16,346 (19.39%) non-redundant unigenes were subdivided into 26 KOG classifications. Among these, the cluster of ‘general function prediction’ (2700, 16.52%) represented the largest group, followed by ‘signal transduction’ (1859, 11.37%) and ‘post-translational modification, protein turnover, chaperone’ (1853, 11.34%), whereas ‘unnamed protein’ (5, 0.03%) was the smallest group (Figure 2).

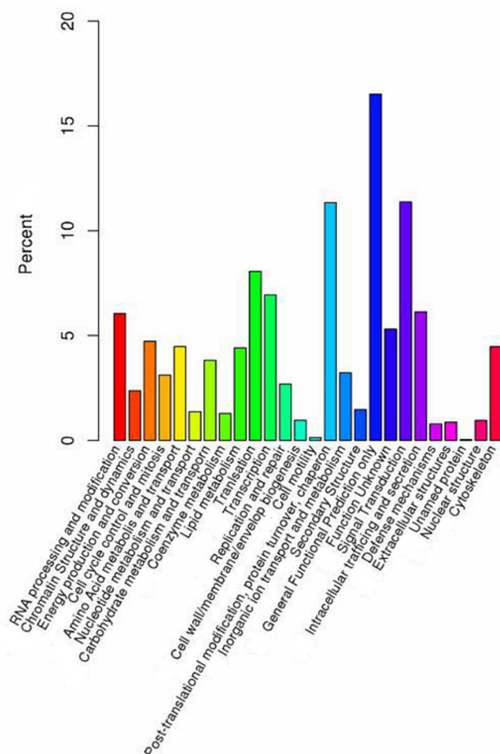

Figure 2. KOG classification of putative proteins.

According to the KEGG pathways of gene participation, 12,722 unigenes were assigned to five specific categories, including cellular processes, environmental information processing, genetic information processing, metabolism, and organism systems (Table 2). Metabolism pathways contained most unigenes (4831, 37.68%), with the majority being involved in carbohydrate metabolism, amino acid metabolism, and energy metabolism.

Different expression analysis of genes

Among all the annotated unigenes, using the filter criteria q-value <0.005 combined with an absolute fold change >2, 623 unigenes were found to be differently expressed between LA and HA. These included 459 unigenes that were down-regulated and 164 unigenes that were up-regulated (Figure 3). The top three enriched GO terms were structural constituent of cuticle, oxidation-reduction process, and oxidoreductase activity in the degree of enrichment (Figure 4).

Table 2. KEGG classification of unigenes.

KEGG category	KEGG subcategory	No. of unigenes
Cellular processes	Cell communication	300
	Cell growth and death	367
	Cell motility	171
	Transport and catabolism	730
Environmental information processing	Membrane transport	162
	Signal transduction	935
	Signaling molecules and interaction	96
Genetic information processing	Folding, sorting and degradation	920
	Replication and repair	252
	Transcription	553
	Translation	1216
Metabolism	Amino acid metabolism	776
	Biosynthesis of other secondary metabolites	78
	Carbohydrate metabolism	965
	Energy metabolism	768
	Glycan biosynthesis and metabolism	243
	Lipid metabolism	559
	Metabolism of cofactors and vitamins	418
	Metabolism of other amino acids	283
	Metabolism of terpenoids and polyketides	134
	Nucleotide metabolism	417
	Xenobiotics biodegradation and metabolism	190
Organismal systems	Circulatory system	165
	Development	142
	Digestive system	291
	Endocrine system	473
	Environmental adaptation	150
	Excretory system	210
	Immune system	398
	Nervous system	285
	Sensory system	75

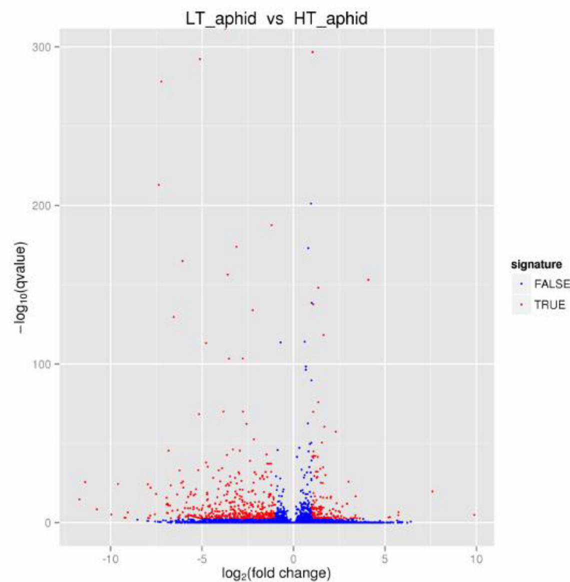

Figure 3. Analysis of differential gene expression between the LA and HA treatment groups. Blue dots represent individual genes that did not differ in expression, whereas red dots represent individual genes with differential expression among the treatment groups.

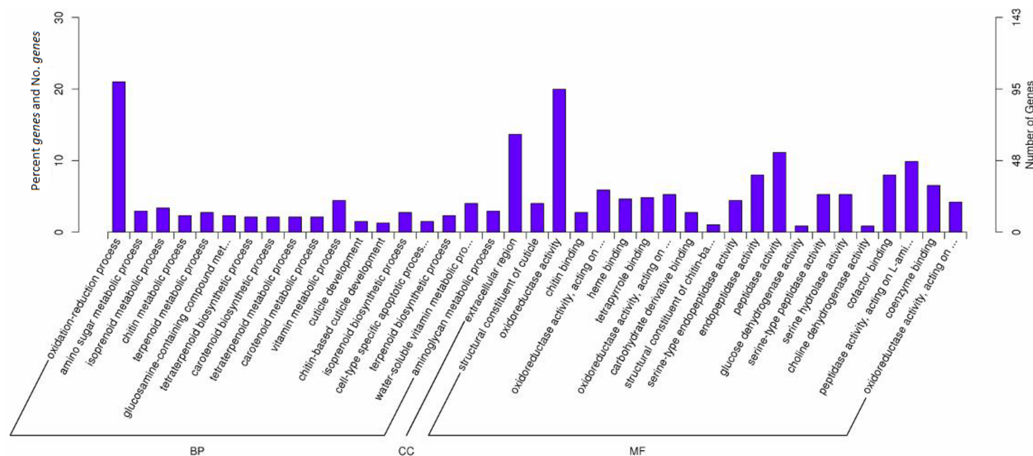

Figure 4. Enriched GO terms of the genes that were differentially expressed between the LA and HA treatment groups. BP: biological process; CC: cellular component; MF: Molecular Function.

In our analysis, we identified 51 genes with orthologs in *Acyrthosiphon pisum* (*Nudel*, *orb*, *bicC*, *kelch*, *nanos*, *putative kelch-like*, *lodestar*, *gld2*, *gle1*, *pop2*, *cbp20*, *histone H2B.3*, *histone H1*, *suv4-20*, *uhrf1*, *cyclin J*, *clasp1*, *drp 1*, *arl6ip1*, *lsd1*, *six 4*, and *cbp 1*) and *Drosophila melanogaster* (*Sxl*, *tra*, *tra-2*, *ix*, *dsx*, *sis-a*, *gro*, *dpn*, *fl*, *snf*, *da*, *DK*, *vir*, *doa*, *Rbp1*, *emc*, *os*, *run*, *her*, *STAT*, *SR*, *vitelline*, *yp1*, *yp2*, *yp3*, *ovo*, *fru*, *sex peptide*, and *esg*) that are likely to have functions in *S. chinensis* reproduction determination. We found that 42 genes were expressed in wintering larvae, which have asexual or sexual individuals as offspring. Eight genes (*nudel*, *gld2*, *histone H2B.3*, *SR*, *yp1*, *yp3*, *ovo*, and *esg*) were differently expressed between LA and HA (Table 3).

qRT-PCR validation

The gene expression profiles were validated by qRT-PCR and the results were closely consistent with the RNA-Seq results. The discrepancies in expression observed in a few genes between the qRT-PCR and RNA-Seq results may have been caused by a sensitivity bias between the two methods or by the use of different statistical methods and threshold values in the qRT-PCR and RNA-Seq.

DISCUSSION

Most aphids constantly rearrange their reproductive modes between sexual and asexual reproduction as an adaptation to the changing environment (Dixon, 1973; Lamb, 1992; Vorburger et al., 2003; Tagu et al., 2005; Gilabert et al., 2009; Sandrock et al., 2011). In our study, most *S. chinensis* larvae reared at low temperature (7.5°C) developed into alate sexuparae (~89.6%) that migrate to the host-plant for sexual reproduction. By contrast, when larvae were reared at a high temperature (18°C), only 5.3% developed into alate sexuparae, whereas the rest continued to reproduce parthenogenetically in the mosses. Moreover, no larvae developed into alate sexuparae when they were reared at 22°C (Table 1). This indicates

Table 3. Description of the 42 transcripts/unigenes with predicted roles in reproduction determination.

Similarity	Annotations	q value	|Fold-change|
*nudel**	Proteolysis//serine-type endopeptidase activity	4.11E-06	11.31
orb	Regulation of translation//synaptonemal complex assembly//regulation of neuronal synaptic plasticity	0.83174	1.55
bicC	RNA binding	0.14841	1.42
Kelch	Developmental process involved in reproduction//anatomical structure formation involved in morphogenesis//organelle organization//ovarian nurse cell to oocyte transport//cellular process involved in reproduction in multicellular organism acyl-carrier-protein biosynthetic process	0.99863	1.26
Nanos	Protein phosphorylation//DNA binding//ATP binding//protein binding//protein kinase activity//helicase	0.17256	2.03
Lodestar	Activity//binding//nucleic acid binding Transferase activity	0.99863	1.17
*Gld2**	Taurine metabolic process//poly(A)+ mrna export from nucleus//leukotriene metabolic process//glutathione metabolic	4.33E-05	24.18
Gle1	Process//prostaglandin metabolic process//viral transcription Nucleic acid binding//nucleus	0.99863	1.41
Pop2	Nucleotide binding//zinc ion binding//nucleic acid binding	0.99863	1.78
cbp20	Nucleosome assembly//DNA-dependent transcription, initiation//DNA binding//protein heterodimerization	0.99863	1.01
*histone H2B.3**	Activity//sequence-specific DNA binding Nucleosome assembly//DNA binding//nucleus//chromosome//nucleosome	8.63E-11	2.76
histone H1	Cellular macromolecule metabolic process//lysine catabolic process//protein methylation//primary metabolic process	0.010951	1.75
Suv4-20	Organ regeneration//maintenance of DNA methylation//regulation of transcription, DNA-dependent//protein ubiquitination//lens development in camera-type eye//chromatin modification	0.99863	1.17
uhrf1	Protein folding//DNA replication initiation//response to stress Carbohydrate biosynthetic process	0.99863	1.22
cyclin J	Small gtpase mediated signal transduction//GTP catabolic process	4.31E-99	1.57
clasp1	DNA repair//DNA replication//regulation of transcription, DNA-dependent//DNA recombination	0.99863	1.24
Drp-1	Peptidoglycan biosynthetic process//nitrogen compound metabolic process//glutamine	0.06549	1.56
Arl6ip1	Biosynthetic process	0.99863	1.17
Lsd-1	RNA-dependent DNA replication, regulation of transcription, DNA-dependent//amino acid transport	0.99863	1.27
pol protein	Phosphorylation//regulation of transcription, DNA-dependent//serine family amino acid metabolic process//protein	0.99863	1.00
Six 4	Phosphorylation//cytokinesis Positive regulation of mrna splicing, via spliceosome//regulation of alternative mrna splicing, via spliceosome	0.99863	1.89
Sxl	Regulation of transcription from RNA polymerase II promoter	0.99863	3.91
tra-2	DNA repair//regulation of transcription, DNA-dependent//translational elongation//ribosome biogenesis//sex differentiation	0.33145	1.87
Ix	Phosphorylation//angiogenesis//regulation of cell division//negative regulation of Wnt receptor signaling pathway//regulation	0.99863	1.08
Dsx	Of establishment of cell polarity//regulation of protein localization//positive regulation of gastrulation Pantothenate biosynthetic process//beta-alanine metabolic process//regulation of transcription, dna-dependent	0.99863	1.15
Gro	Mrna splicing, via spliceosome//regulation of transcription from RNA polymerase II promoter//proteolysis Lipid transport//viral reproduction Transport//pathogenesis//ion channel inhibitor activity	0.0046532	1.33
dpn	Protein phosphorylation	0.045305	8.44
snf	Mrna export from nucleus//mrna splicing, via spliceosome//insulin receptor signaling pathway//termination of RNA	0.99863	1.09
da	Polymerase II transcription//mrna 3'-end processing//pathogenesis	0.0047589	1.81
vir	Negative regulation of transcription, DNA-dependent	0.99863	1.21
doa	Imaginal disc-derived female genitalia development//spermathecum morphogenesis//regulation of transcription,	0.99863	1.70
Rbp 1	DNA-dependent//oogenesis//scab formation//sensory organ precursor cell fate determination//negative regulation of transcription from RNA polymerase II promoter//regulation of transcription from RNA polymerase II promoter//positive	0.99863	1.52
Emc	Regulation of compound eye retinal cell programmed cell death//lymph gland crystal cell differentiation//R7 cell	0.18251	1.66
Run	Development//embryonic crystal cell differentiation//signal transduction//antennal development//positive regulation of transcription from RNA polymerase II promoter//defense response//compound eye cone cell differentiation//regulation of	0.99863	2.5
	Apoptotic process		
	Signal transduction//regulation of transcription, DNA-dependent Transport//DNA repair//intracellular signal transduction//pathogenesis//DNA binding//ion channel inhibitor activity//nucleic acid binding//zinc ion binding Catalytic activity		
STAT	Serine-type endopeptidase inhibitor activity	0.39153	2.42
*SR**	Glycerolipid metabolic process//glycosphingolipid metabolic process// photosynthesis, light reaction//histidine biosynthetic process//glycosaminoglycan catabolic process//DNA repair//galactose metabolic process//photosynthesis// DNA	0.0013114	3.58
*yp1**	Binding//ATP binding//imidazoleglycerol-phosphate dehydratase activity//zinc ion binding//nucleic acid binding//nuclease	6.56E-07	3.52
*yp3**	Activity//metal ion binding//beta-galactosidase activity	5.92E-06	2.79
*ovo**	Metabolic process//DNA binding//protein binding//zinc ion binding//binding //oxidoreductase activity	0.00049941	2.92
	G-protein coupled receptor signaling pathway//folic acid biosynthetic process//tetrahydrobiopterin biosynthetic process//oxidation-reduction process		
Fru	Glycerolipid metabolic process//glycosphingolipid metabolic process// photosynthesis, light reaction//histidine biosynthetic process//glycosaminoglycan catabolic process//DNA repair//galactose metabolic process//photosynthesis	0.99863	1.04
sex peptide receptor		0.99863	1.03
*esg**		0.00049941	3.89

*****Genes expressed differently in two samples based on the filter criteria qvalue < 0.005 & |Fold-change| > 2.

that the rearing temperature not only affected the eclosion rate of sexuparae, but also the reproduction mode of the offspring.

Based on RNA-Seq technology, 84,280 unigenes with a mean length of 666 bp, ranging from 201 to 23,871 bp were assembled. A comparison of the assembled unigenes

to other insect species using BLAST analysis and functional annotation indicated that we had obtained an expansive and diverse gene expression set. Of the DEGs, 476 unigenes were annotated, while and 147 genes were not. The larvae that were reared under different temperature conditions, but same photoperiod and humidity, had diverged by adapting to the specific ambient conditions with respect to the different reproduction profiles observed in the two treatments. This raises the possibility that at least some of the 147 non-annotated unigenes were new genes that may be related to the ability of *S. chinensis* to respond to changes in temperature. In this study, almost all aphids reared at 7.5°C gave birth to male and female aphids, whereas aphids reared at 18°C reproduced parthenogenetically. We do not have any data on the unclassified 147 genes, and future studies should therefore address whether these genes are linked to aphid reproduction.

We found that there was no significant enrichment GO term related to reproduction. However, of 27 enriched unigenes involved in reproduction function, 21 were down-regulated and six were up-regulated unigenes. This means that 21 unigenes had a relatively higher expression in LA and only six unigenes had a relatively higher expression in HA.

The growth, development, reproduction, and other physiological and behavioral responses of insects cannot be completed without the participation of hormones. Estrogens and androgens are sex hormones that promote the maturity of reproductive organs and the development of secondary sexual characteristics of males and females. The estrogen and androgen metabolic process terms, which were absent in HA, were enriched in six unigenes found in LA. Similarly, there were 18 highly expressed unigenes in LA that were enriched for the terms insemination, sperm competition, sperm storage, spermatogenesis, oogenesis, sexual reproduction, sex differentiation, and ovarian follicle cell development in LA.

Aphids have the ability to produce three generations within one viviparous female, where the mature embryos developing inside the maternal abdomen carry the first developmental stage of the third generation (telescoping of generations) (Stevens, 1905). Therefore, two reproductive modes are present in these viviparous females: The second generation (all females) is produced parthenogenically inside the first generation of overwintering females, and the third generation (males and females) is produced sexually inside the second generation of females (**Figure S1**). This could explain why some enrichment terms related to sex differentiation and spermatogenesis were seen in LA but not in HA, which continue to reproduce parthenogenically.

In insects, some genes have specific functions in males and females whose functions are not controlled via transcriptional regulation but instead by sex-specific alternative splicing (Bell, 1988; Burtis and Baker, 1989; Hoshijima et al., 1991). Alternative splicing allows for protein diversity by producing more than one mature mRNA from a single pre-mRNA. In *D. melanogaster*, the mRNA precursors of *Sxl*, *tra*, and *DSX* are the same in males and females. However, after sex-specific splicing, they are processed into female and male-specific mRNAs. Since *Sxl* has several exons and introns, it can be spliced in different ways to result in different mRNAs. In honeybees (Lattorff et al., 2005; Jarosch et al., 2011), worker reproduction is regulated by the queen, brood pheromones, and worker policing. However, workers can evade these control mechanisms and activate their ovaries to produce diploid female offspring parthenogenetically. And the behaviours were influenced by alternative splicing of a gene homologue (*thelytoky*) to the transcription factor that controls worker sterility, which is located on chromosome 13. In our study, the expressions of *Sxl*, *tra*, and *DSX* at different temperatures in *S. chinensis* showed no differentiation. However, the downstream genes of *Sxl*, *tra*, and *DSX*

as well as the sex differentiation genes *ovo*, *yp1*, *yp3*, and *esg* were differentially expressed. Using RNA-Seq technology we are able to observe differential expression of some genes, but we can gain no further information. To further our understanding, our next step will be to identify the sex-specific splicing, especially related to the three genes (*Sxl*, *tra*, and *DSX*) that were differentially expressed in response to the two temperature conditions. In conclusion, this study will not only be helpful to identify the genes associated with reproduction determination in *S. chinensis*, but also to understand the mechanism of aphid reproductive determination.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge Prof. Kirst King-Jones for his helpful comments. Research supported by the Natural Sciences Foundation of China (#U1402263; #31370651; #31372266), the National High-tech R&D Program of China ("863" Program, #2014AA021802), the China Scholarship Council, and the Grant for Essential Scientific Research of Chinese National Non-profit Institute (#CAFYBB2014ZD005).

REFERENCES

Artacho P, Figueroa CC, Cortes PA, Simon JC, et al. (2011). Short-term consequences of reproductive mode variation on the genetic architecture of energy metabolism and life-history traits in the pea aphid. *J. Insect Physiol.* 57: 986-994. http://dx.doi.org/10.1016/j.jinsphys.2011.04.013

Bell G (1988). Uniformity and diversity in the evolution of sex. In: The evolution of sex: An examination of current ideas (Michod RE and Levin BR, eds.). Sinauer Associates, Sunderland, 126-138.

Burtis KC and Baker BS (1989). *Drosophila* doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* 56: 997-1010. http://dx.doi.org/10.1016/0092-8674(89)90633-8

Benjamini Y and Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Ser. B*, (57): 289-300.

Dixon AFG (1973). Biology of aphids. Edward Arnold, London.

Gilabert A, Simon JC, Mieuzet L, Halkett F, et al. (2009). Climate and agricultural context shape reproductive mode variation in an aphid crop pest. *Mol. Ecol.* 18: 3050-3061. http://dx.doi.org/10.1111/j.1365-294X.2009.04250.x

Grabherr MG, Haas BJ, Yassour M, Levin JZ, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644-652. http://dx.doi.org/10.1038/nbt.1883

Hoshijima K, Inoue K, Higuchi I, Sakamoto H, et al. (1991). Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila. Science* 252: 833-836. http://dx.doi.org/10.1126/science.1902987

Jarosch A, Stolle E, Crewe RM and Moritz RFA (2011). Alternative splicing of a single transcription factor drives selfish reproductive behavior in honeybee workers (Apis mellifera). *Proc. Natl. Acad. Sci. USA* 108: 15282-15287. http://dx.doi.org/10.1073/pnas.1109343108

Kanehisa M, Araki M, Goto S, Hattori M, et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36: D480-D484. http://dx.doi.org/10.1093/nar/gkm882

Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* 110: E2655-E2662. http://dx.doi.org/10.1073/pnas.1309606110

Lamb RJ (1992). Developmental rate of *Acyrthosiphon pisum* (Homoptera, Aphididae) at low temperatures: implications for estimating rate parameters for insects. *Environ. Entomol.* 21: 10-19. http://dx.doi.org/10.1093/ee/21.1.10

Lattorff HMG, Moritz RF and Fuchs S (2005). A single locus determines thelytokous parthenogenesis of laying honeybee workers (*Apis mellifera capensis*). *Heredity (Edinb)* 94: 533-537. http://dx.doi.org/10.1038/sj.hdy.6800654

Li B and Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323. http://dx.doi.org/10.1186/1471-2105-12-323

Liu P, Yang ZX, Chen XM and Foottit RG (2014). The effect of the gall-forming aphid *Schlechtendalia chinensis* (Hemiptera: Aphididae) on leaf wing ontogenesis in *Rhus chinensis* (Sapindales: Anacardiaceae). *Ann. Entomol. Soc. Am.* 107: 245-250. http://dx.doi.org/10.1603/AN13118

Malone JH and Oliver B (2011). Microarrays, deep sequencing and the true measure of the transcriptome. *BMC Biol.* 9: 34. http://dx.doi.org/10.1186/1741-7007-9-34

Mao X, Cai T, Olyarchuk JG and Wei L (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21: 3787-3793. http://dx.doi.org/10.1093/bioinformatics/bti430

Nagalakshmi U, Wang Z, Waern K, Shou C, et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344-1349. http://dx.doi.org/10.1126/science.1158441

Ozsolak F and Milos PM (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* 12: 87-98. http://dx.doi.org/10.1038/nrg2934

Sandrock C, Razmjou J and Vorburger C (2011). Climate effects on life cycle variation and population genetic architecture of the black bean aphid, *Aphis fabae*. *Mol. Ecol.* 20: 4165-4181. http://dx.doi.org/10.1111/j.1365-294X.2011.05242.x

Shao SX, Yang ZX and Chen XM (2013). Gall development and clone dynamics of the galling aphid *Schlechtendalia chinensis* (Hemiptera: Pemphigidae). *J. Econ. Entomol.* 106: 1628-1637. http://dx.doi.org/10.1603/EC13114

Stevens NM (1905). A study of the germ cells of *Aphis rosae* and *Aphis oenotherae*. *J. Exp. Zool.* 2: 313-333. http://dx.doi.org/10.1002/jez.1400020302

Sultan M, Schulz MH, Richard H, Magen A, et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321: 956-960. http://dx.doi.org/10.1126/science.1160342

Tagu D, Sabater-Muñoz B and Simon JC (2005). Deciphering reproductive polyphenism in aphids. *Invertebr. Reprod. Dev.* 48: 71-80. http://dx.doi.org/10.1080/07924259.2005.9652172

Tang J and Cai BH (1957). Studies on the Chinese gallnuts of Meitan, Kweichow. *Acta Entomol. Sin* 7: 131-140.

Vorburger C, Lancaster M and Sunnucks P (2003). Environmentally related patterns of reproductive modes in the aphid *Myzus persicae* and the predominance of two ‘superclones’ in Victoria, Australia. *Mol. Ecol.* 12: 3493-3504. http://dx.doi.org/10.1046/j.1365-294X.2003.01998.x

Young MD, Wakefield MJ, Smyth GK and Oshlack A (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11: R14. http://dx.doi.org/10.1186/gb-2010-11-2-r14

Zhang CX, Xu H and Tang J (1993). Influence of temperature on polymorphism of the overwintering generation of Chinese horned gall aphid, *Schlechtendalia chinensis* (Bell). *Entomol. Sin.* 36: 497-499.

Zhang GX and Zhong TS (1983). The economic insect fauna of China. Science Press, Beijing.

Zhang GX, Qiao GX, Zhong TS and Zhang WY (1999). Fauna Sinica, Insecta Vol. 14 Homoptera, Mindaridae and Pemphigidae. Science Press, Beijing.

Zhang ZH (1987). The processing and utilization technology in Chinese gallnuts. China Forestry Press, Beijing.

Supplementary material

Figure S1. The ‘telescoping of generations’ phenomenon in *S. chinensis*.