Gene Expression Profiling in Winged and Wingless Cotton Aphids, *Aphis gossypii* (Hemiptera: Aphididae)

Xiaowei Yang  
*China Agricultural University, China*

Xiaoxia Liu  
*China Agricultural University, China*

Xiangli Xu  
*China Agricultural University, China*

Zhen Li  
*China Agricultural University, China*

Yisong Li  
*China Agricultural University, China*

See next page for additional authors.

Right click to open a feedback form in a new tab to let us know how this document benefits you. Follow this and additional works at: [https://uknowledge.uky.edu/entomology_facpub](https://uknowledge.uky.edu/entomology_facpub)

Part of the Entomology Commons

Repository Citation
Yang, Xiaowei; Liu, Xiaoxia; Xu, Xiangli; Li, Zhen; Li, Yisong; Song, Dongyan; Yu, Tian; Zhu, Fang; Zhang, Qingwen; and Zhou, Xuguo, "Gene Expression Profiling in Winged and Wingless Cotton Aphids, *Aphis gossypii* (Hemiptera: Aphididae)" (2014). *Entomology Faculty Publications*. 71.  
[https://uknowledge.uky.edu/entomology_facpub/71](https://uknowledge.uky.edu/entomology_facpub/71)

This Article is brought to you for free and open access by the Entomology at UKnowledge. It has been accepted for inclusion in Entomology Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Authors
Xiaowei Yang, Xiaoxia Liu, Xiangli Xu, Zhen Li, Yisong Li, Dongyan Song, Tian Yu, Fang Zhu, Qingwen Zhang, and Xuguo Zhou

Gene Expression Profiling in Winged and Wingless Cotton Aphids, *Aphis gossypii* (Hemiptera: Aphididae)

Notes/Citation Information

© Ivyspring International Publisher. This is an open-access article distributed under the terms of the Creative Commons License (http://creativecommons.org/licenses/by-nc-nd/3.0/). Reproduction is permitted for personal, noncommercial use, provided that the article is in whole, unmodified, and properly cited.

Digital Object Identifier (DOI)
http://dx.doi.org/10.7150/ijbs.7629

This article is available at UKnowledge: https://uknowledge.uky.edu/entomology_facpub/71
Gene Expression Profiling in Winged and Wingless Cotton Aphids, *Aphis gossypii* (Hemiptera: Aphididae)

Xiaowei Yang1*, Xiaoxia Liu1*, Xiangli Xu1, Zhen Li1, Yisong Li1, Dongyan Song2, Tian Yu2, Fang Zhu2, Qingwen Zhang1*, Xuguo Zhou2*

1. Department of Entomology, China Agricultural University, Beijing 100193, China.
2. Department of Entomology, University of Kentucky, Lexington, KY 40546-0091, USA.

* These authors contributed equally to this research.

© Ivyspring International Publisher. This is an open-access article distributed under the terms of the Creative Commons License (http://creativecommons.org/licenses/by-nc-nd/3.0/). Reproduction is permitted for personal, noncommercial use, provided that the article is in whole, unmodified, and properly cited.

Received: 2013.09.10; Accepted: 2014.01.22; Published: 2014.02.19

Abstract

While trade-offs between flight capability and reproduction is a common phenomenon in wing dimorphic insects, the molecular basis is largely unknown. In this study, we examined the transcriptomic differences between winged and wingless morphs of cotton aphids, *Aphis gossypii*, using a tag-based digital gene expression (DGE) approach. Ultra high-throughput Illumina sequencing generated 5.30 and 5.39 million raw tags, respectively, from winged and wingless *A. gossypii* DGE libraries. We identified 1,663 differentially expressed transcripts, among which 58 were highly expressed in the winged *A. gossypii*, whereas 1,605 expressed significantly higher in the wingless morphs. Bioinformatics tools, including Gene Ontology, Cluster of Orthologous Groups, eukaryotic Orthologous Groups and Kyoto Encyclopedia of Genes and Genomes pathways, were used to functionally annotate these transcripts. In addition, 20 differentially expressed transcripts detected by DGE were validated by the quantitative real-time PCR. Comparative transcriptomic analysis of sedentary (wingless) and migratory (winged) *A. gossypii* not only advances our understanding of the trade-offs in wing dimorphic insects, but also provides a candidate molecular target for the genetic control of this agricultural insect pest.

Key words: *Aphis gossypii*, trade-off, migration, digital gene expression, wing polyphenism.

Introduction

Trade-offs between life-history traits have been studied in a variety of organisms [1-3]. Individuals capable of fly benefit from the ability, but it is also costly in energy for flight mechanism and flight behavior. This energy, alternatively, can be used for other aspects, such as body size and reproduction [4]. In the past forty years, trade-offs between flight and reproduction has been widely studied [5-9]. Wing dimorphic insects have been studied extensively and have emerged as a great model system for the study of trade-offs in the past several decades [8-10]. Wing polyphenism has been observed in a wide range of insect species, including Hemiptera, Orthoptera, and Coleoptera, to facilitate their range expansion [8, 10-12]. Wing dimorphic insects display commonly dispersing and non-dispersing morphs [13, 14]. Flight morphs, with long wings, flight muscles and flight capability, need additional energy and time for the migratory flight. Flightless morphs, with reduced or absent wings and underdeveloped flight muscles, show an earlier onset of oviposition and enhanced reproductive output [8, 15-17]. Trade-offs have been
studied at behavioral, morphological, physiological and biochemical levels in wing dimorphic insects, such as crickets, migratory locusts, and aphids [16, 18, 19]. For instance, biochemical aspects and intermediary metabolisms are investigated in a wing dimorphic cricket, *Gryllus firmus* [20-22]. However, the genetic basis of biochemistry and metabolism related to trade-offs remains unclear.

There are trade-offs related to dispersal and reproduction in aphids between two wing morphs. The wing polyphenism in aphids is well-studied [23, 24]. Winged and wingless aphids exist in both sexual and parthenogenetic stages of its life cycle. The migrants in spring, summer and early autumn are parthenogenetic generations, and the migration normally occurs in spring and summer [25, 26]. Wing polyphenism occurs primarily among parthenogenetic females, whereas wing polymorphism, which is genetically determined, has been found only in males [14, 24]. Parthenogenetic females with different wing morphs are descended from the same ancestry. A single genotype can produce various phenotypes as a result of exposure to various environmental conditions [27]. Such phenotypic plasticity is one of the main reasons contributing to the ecological success of aphids and their increasing pest status [26, 28]. Plasticity is not only an adaptive behavioral strategy that has evolved in most insect orders for coping with complex and uncertain environments, but also a key factor that leads to population expansion and infestation. Wing-polyphenic aphids exhibited distinct morphological differences between two wing types of aphids, and showed differences in their life history, particularly in their behavioral and physiological traits [14, 29, 30]. The biochemical basis of wing polyphenism related to energy resources, including glycogen, trehalose, and lipids, was investigated in cotton aphid, *Aphis gossypii* [30] and grain aphid, *Sitobion avenae* [31]. Total lipid and triglyceride were significantly higher in winged morphs, whereas glycogen, soluble protein and water were higher in wingless morphs in both species. Recent genetic studies [32-34], especially the release of a draft genome of pea aphid *Acyrthosiphon pisum* [35], lays the foundation for a better understanding of trade-offs between flight and reproduction at the molecular level. Genome-wide transcriptome analyses provide a new avenue for the future functional characterization study.

Digital gene expression (DGE) is a tag-based RNA sequencing approach that allows millions of RNA reads, including differentially expressed transcripts, to be identified in a sample with or without any annotation information [36]. For aphid wing polyphenism study, DGE has several advantages - a greater sequencing depth, detection of unknown transcripts, and sensitivity [37]. Although external factors such as environmental cues, population density and nutrition status have been studied for years, the genetic basis of trade-offs in the cotton aphids, *A. gossypii*, a major cotton pest worldwide, has yet been examined in details. The objectives of this study are 1) to reveal differential gene expression profiles between winged and wingless *A. gossypii* adults, and 2) to establish a link between transcriptomic profiles and biochemical differences observed in previous studies. Our studies provide molecular basis that underpins trade-offs between flight capability and reproduction in *A. gossypii* and will facilitate the sustainable cotton aphid management through the RNA interference-mediated genetic disruption of its migratory behavior.

**Materials and Methods**

*Aphis gossypii* maintenance and sample preparation

A single *A. gossypii* female adult was collected from a cotton field (40°01′N, 116°17′E) at the China Agricultural University, Beijing, China, in July of 2009. Aphid colonies generated from this individual *A. gossypii* had been established in the laboratory for more than one year before being sacrificed in the subsequent experiments. Specifically, parthenogenetic adults were provisioned with cotton seedlings (three-leaf stage, cultivar: Zhong 35) and maintained in a growth chamber at 60% relative humidity, and 16:8 L: D photoperiod with 24±1˚C in the light and 22±1˚C in the dark [30], [34, 38]. Winged forms were induced by crowding in cotton leaves two weeks after *A. gossypii* was transferred to new host plants. The fourth instars were placed into individual dishes, and observed daily until molting. A total of 50 winged or wingless *A. gossypii* adults were collected separately within two days after molting. Aphids were frozen in liquid nitrogen immediately after the collection. RNA was extracted using the RNeasy Mini kit (Qiagen). Genomic DNA was digested by RNase-Free DNase Set (Qiagen). The quality and quantities of RNAs were determined using a Nanodrop2000 and Agilent bioanalyzer (Agilent).

**DGE library preparation and sequencing**

High-throughput DGE was performed by Solexa/Illumina sequencing technology at the Beijing Genomics Institute, Shenzhen, China. DGE libraries (winged and wingless) were prepared using the Illumina Gene Expression Sample Prep Kit. Briefly, poly(A)+ RNA was enriched from 6 µg of total RNA using oligomagnetic beads. Double-stranded cDNAs were directly synthesized on the poly(A)+
RNA-bound beads and then digested with a 4 base recognition enzyme NlaIII. The fragmentized cDNAs were purified from the magnetic beads, containing 3’ ends and a sticky end. Then the Illumina adaptor 1 (sense: 5’ AACAATCTTTCCCTACACGAccGCCTTCGGATC3’) was ligated to the 5’ ends of these cDNA fragments. Digestion of MmeI and ligation of the Illumina adaptor 2 (sense: 5’ GATCCGGGAAGAGCGGTTTCGCAGGAGTCCCGAG3’) produced short DNA fragments, with two adaptors and 21bp tags started with the CATG sites. Later on, Primer GX1 and Primer GX2 were added for PCR. After 15 cycles of linear PCR amplification, 95bp fragments were purified using 6% TBE PAGE Gel electrophoresis. After denaturation, the single-stranded molecules were attached to the sequencing chip for sequencing via Illumina HiSeq™ 2000 System. During this process, adaptor 1 was used as sequencing primer. Each tunnel of chip (flow cell) generated millions of raw tags with sequencing length of 35 bp [39, 40].

Mapping of DGE tags

Clean tags were generated by removing 3’ adaptor sequences, empty reads (with only 3’ adaptor sequences), low quality tags (tags with unknown nucleotide “N”), tags with erratic lengths, and tags with copy number of 1. A total of 88,851 A. gossypii ESTs were selected at NCBI on Oct, 28, 2010. After eliminating low quality sequences and Phrap assembly (minmatch: 42; minscore: 40, repeat_stringency:0.96), 23,630 unigenes were obtained and used as reference dataset (Supplementary Material: Table S1). Clean tags were aligned to the reference dataset with only one nucleotide mismatch tolerance. The number of unambiguous clean tag corresponding to each transcript was counted and normalized to TPM (number of transcripts per million clean tags) [36, 41].

Identification of differentially expressed transcripts

The differentially expressed transcripts between winged and wingless patterns were identified using an algorithm developed by Audic and Claverie [42]. P value corresponds to differential gene expression test. FDR (False Discovery Rate) is a method of determining the threshold of P Value in multiple tests and analysis through manipulating FDR value [43]. In this study, we considered a transcript as differentially expressed when there was at least a 2-fold difference in expression between the two samples with a FDR value ≤ 0.001.

To functionally annotate these unigenes, distinct sequences were searched by BLASTx against the NCBI non-redundant protein database [44] with a cut-off E-value of 10⁻⁵, and assigned Cluster of Orthologous Groups (COG) and euKaryotic Orthologous Groups (KOG) protein classes with rpstblastn algorithm against COG and KOG databases in NCBI (E-value ≤10⁻⁵) [45, 46]. The BLASTx result was imported into Blast2GO software (version 2.6.4) to map the Gene Ontology (GO) categories [47], and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [48]. The enriched GO terms were developed by Blast2Go using Fisher’s Exact Test with Multiple Testing Correction (FDR<0.05)[47]. The enrichment analysis of pathway was performed to further understand genes biological functions.

qRT-PCR analysis

PCR reactions were performed in a 20 μl final volume containing 2× Power SYBR Green PCR Master Mix buffer (Applied Biosystems), 1 μM of each specific primer and 1 μl of diluted cDNA. PCR amplification was performed on an ABI Prism 7300 Sequence Detection System (Applied Biosystems) under the following conditions: 10 min at 95 °C for polymerase activation, and 40 cycles of 95°C for 15s, 60°C for 30s, and 72°C for 30s for amplification signal collection. A default melting curve was produced to confirm single gene-specific peaks without primer dimerization present. Housekeeping genes, β-actin and 18S rRNA were subjected to the reference gene selection experiment. Eventually, β-actin was used as a constitutive control to normalize the expressions of target genes using a −ΔΔCt method.

To validate DGE results, a total of 20 transcripts were chosen from the top 10 differentially expressed transcripts of winged and wingless libraries, respectively, and three significantly enriched pathways. The gene-specific primers designed for reference and target genes were listed in Supplementary Material: Table S2. Three independent replications were carried out for the qRT-PCR validation study, with approximately 30 winged or wingless females for each replication.

Results

Summary of digital gene expression sequencing

Reference sequences were obtained from 88,851 ESTs, of which 36 low quality (less than 100bp after remove vector) sequences were discarded. 23,630 unigenes, containing 12,507 contigs and 11,123 singletons, with a mean length of 715 bp, were assembled by Phrap. The generated unigenes were subjected to BLAST search and functional annotation using various databases.

In total, 5.30 and 5.39 million raw tags were obtained from winged and wingless adults, respectively.
Distribution details of tags were shown in Supplementary Material: Figure S1. Clean tags were mapped to 9,011 and 10161 unigenes, respectively (Table 1). Sequencing saturation analysis shows the number of detected genes almost ceases to increase when sequencing amount (total tag number) reaches 2 million (Supplementary Material: Figure S2), consequently, 5 million tags are sufficient for the expression profiling of cotton aphid. The total number of tags sequenced in the two samples was 10,479,972. Tags with the same sequence were identified as a distinct tag, and their distributions in different wing morphs were documented. Tags in the two wing forms were compared in total and in distinct (Figure 1). Surprisingly, morph-specific tags (tags detected in one morph only) accounted for a very small proportion (0.85% in winged morph, 2.12% in wingless morph) of the total clean tags, while the percentage of morph-specific tags increased obviously among the distinct tags.

**Differentially expressed transcripts in winged and wingless aphids**

We obtained 1,663 differentially expressed transcripts from a total of 8,308 tag-mapped unigenes. Among them, 58 transcripts were expressed significantly higher in winged aphids, while 1,605 transcripts had a higher expression in wingless morphs (|log2 ratio| ≥ 1; FDR values ≤ 0.001). A total of 334 (20.69%) differentially expressed transcripts had a differential ratio higher than 2 (|log2 ratio| ≥ 2; Figure 2), in which 3 transcripts were detected in winged forms only and 34 transcripts were expressed in wingless forms exclusively (Supplementary Material: Table S3). Top ten differentially expressed transcripts and their ratios are summarized in Table 2.

![Figure 1. Morph-specific sequencing tags in the winged and wingless A. gossypii. Number and percentage of morph-specific tags were documented in total clean tags (A) and distinct clean tags (B), respectively.](http://www.ijbs.com)

![Figure 2. Expression dynamics of differentially expressed transcripts between winged and wingless A. gossypii. Fold Change represents the log2 of the ratio of gene expression level in winged in relation to wingless adult females. Transcript expressed at a constant level (ratio of 1) has a log2 (ratio) of zero, which can be interpreted as “equally expressed among winged and wingless aphids”. Positive number represents highly expressed transcripts in winged adult females. Negative number represents highly expressed transcripts in wingless adult females.](http://www.ijbs.com)
Table 1. Sequencing summary.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Total</th>
<th>%</th>
<th>Distinct</th>
<th>%</th>
<th>Total</th>
<th>%</th>
<th>Distinct</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean Tag</td>
<td>5221975</td>
<td>-</td>
<td>44144</td>
<td>-</td>
<td>5257997</td>
<td>-</td>
<td>61655</td>
<td>-</td>
</tr>
<tr>
<td>Mapped to gene</td>
<td>4781916</td>
<td>91.57</td>
<td>21675</td>
<td>48.80</td>
<td>4345336</td>
<td>82.64</td>
<td>27998</td>
<td>45.41</td>
</tr>
<tr>
<td>Unambiguous tag</td>
<td>1329628</td>
<td>25.4</td>
<td>20748</td>
<td>46.71</td>
<td>2322979</td>
<td>44.18</td>
<td>26749</td>
<td>43.38</td>
</tr>
<tr>
<td>Unknown tag</td>
<td>440059</td>
<td>8.4</td>
<td>22739</td>
<td>51.20</td>
<td>912661</td>
<td>17.36</td>
<td>33657</td>
<td>54.59</td>
</tr>
</tbody>
</table>

Table 2. Top ten differentially expressed transcripts between winged and wingless A. gossypii.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>log2FC*</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG10278</td>
<td>12.5</td>
<td>Flightin-like</td>
</tr>
<tr>
<td>gi</td>
<td>289097293</td>
<td>11</td>
</tr>
<tr>
<td>AG4984</td>
<td>10.1</td>
<td>MPA13 allergen-like</td>
</tr>
<tr>
<td>gi</td>
<td>289151497</td>
<td>8.3</td>
</tr>
<tr>
<td>AG12533</td>
<td>8.1</td>
<td>Phosphate carrier protein, mitochondrial-like</td>
</tr>
<tr>
<td>AG13078</td>
<td>5.7</td>
<td>Adenine nucleotide translocator-like</td>
</tr>
<tr>
<td>AG6703</td>
<td>5.6</td>
<td>PXMP2/4 family protein 4-like</td>
</tr>
<tr>
<td>AG11738</td>
<td>5.3</td>
<td>Tectonin beta-propeller repeat-containing protein-like</td>
</tr>
<tr>
<td>AG12565</td>
<td>3.8</td>
<td>Hypothetical protein LOC100165045 isoform 1</td>
</tr>
<tr>
<td>AG11798</td>
<td>3.5</td>
<td>Very long-chain specific acyl-CoA dehydrogenase</td>
</tr>
<tr>
<td>AG10938</td>
<td>-9.7</td>
<td>Seryl-tRNA synthetase, cytoplasmic-like</td>
</tr>
<tr>
<td>AG2471</td>
<td>-9.3</td>
<td>Hypothetical protein LOC100571294</td>
</tr>
<tr>
<td>gi</td>
<td>289129407</td>
<td>-9.2</td>
</tr>
<tr>
<td>AG9276</td>
<td>-9.1</td>
<td>RNA-binding protein PNO1-like</td>
</tr>
<tr>
<td>AG11819</td>
<td>-9</td>
<td>T-complex protein 1 subunit alpha-like</td>
</tr>
<tr>
<td>gi</td>
<td>289076469</td>
<td>-9</td>
</tr>
<tr>
<td>AG7395</td>
<td>-8.9</td>
<td>Exosome complex exonuclease MTR3-like</td>
</tr>
<tr>
<td>gi</td>
<td>289138983</td>
<td>-8.7</td>
</tr>
<tr>
<td>AG4684</td>
<td>-8.6</td>
<td>N/A</td>
</tr>
<tr>
<td>AG3568</td>
<td>-8.6</td>
<td>N/A</td>
</tr>
</tbody>
</table>

** log2FC* is the fold change of gene expression level between winged and wingless morphs. Positive and negative values, respectively, denote highly and lowly expressed gene expression in winged adult females relative to wingless individuals. Assembly details are provided in Supplementary Material: Table S3.

**Functional analysis of differentially expressed transcripts**

Following assignment of BLASTx top hits, 1,306 differentially expressed transcripts were functionally annotated with four additional databases. Among 8,308 sequenced transcripts, 3,184 transcripts were annotated with GO terms, including 2,318 involved in biological process, 2,547 in molecular function and 1,954 in cellular component. As to 1,663 differentially expressed transcripts, 763 transcripts were assigned with 3,168 GOs and can be classified into 42 secondary level categories (Figure 3). The category of biological process consists of 573 transcripts at 22 secondary levels. Molecular function category has 619 transcripts at 9 secondary levels. Cellular component includes 491 transcripts in 11 secondary categories.

All differentially expressed transcripts were enriched in 16 GO group when compared to all the A. gossypii unigenes (Supplementary Material: Table S4). 7 of them were very deep level (specific) GO terms (Figure 4). Only one term (membrane) was underrepresented. In addition, transcripts were annotated by COG and KOG databases, by which 676 transcripts were categorized into 23 COG functional groups (Figure 5) and 1,144 transcripts into 25 KOG classes (Figure 5). With KOG classification, in addition to the category of general function prediction only, the number of transcripts in category A (RNA processing and modification), O (Posttranslational modification, protein turnover and chaperones), and U (Intracellular trafficking, secretion, and vesicular transport) were 802, 748 and 759, respectively, higher than the number in the other categories.

Pathway analysis of differentially expressed transcripts was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Among 1,663 differentially expressed transcripts, 878 transcripts could be annotated in KO based on sequence homologies and mapped to 211 KEGG pathways (Table 3). Pathway-oriented analysis showed that the top three significantly enriched metabolic pathways in DGEs were ribosome, pyruvate metabolism and proteasome. Differentially expressed transcripts were also enriched in pathways relating to protein synthesis and degradation, lipid metabolism, immunity, RNA transport, and some signaling pathways.
qRT-PCR validation study

qRT-PCR analysis was carried out to validate differentially expressed transcripts identified by DGE. All 20 genes showed the similar expression pattern as indicated by the DGE analysis (Table 4), suggesting that DGE approach can be a powerful tool for the discovery of candidate genes in a high-throughput manner.

Table 3. Top ten enriched metabolic pathways among differentially expressed transcripts.

<table>
<thead>
<tr>
<th>Pathway ID</th>
<th>Description</th>
<th>DET</th>
<th>Reference</th>
<th>P-value</th>
<th>Q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ko03010</td>
<td>Ribosome</td>
<td>49</td>
<td>144</td>
<td>1.35E-14</td>
<td>2.53E-12</td>
</tr>
<tr>
<td>ko00620</td>
<td>Pyruvate metabolism</td>
<td>21</td>
<td>83</td>
<td>8.95E-05</td>
<td>9.47E-03</td>
</tr>
<tr>
<td>ko03050</td>
<td>Proteasome</td>
<td>15</td>
<td>53</td>
<td>0.000239</td>
<td>1.68E-02</td>
</tr>
<tr>
<td>ko03040</td>
<td>Spliceosome</td>
<td>44</td>
<td>268</td>
<td>0.001539</td>
<td>7.15E-02</td>
</tr>
<tr>
<td>ko04141</td>
<td>Protein processing in endoplasmatic reticulum</td>
<td>42</td>
<td>254</td>
<td>0.00695</td>
<td>7.15E-02</td>
</tr>
<tr>
<td>ko00280</td>
<td>Valine, leucine and isoleucine degradation</td>
<td>15</td>
<td>65</td>
<td>0.002402</td>
<td>8.52E-02</td>
</tr>
<tr>
<td>ko00062</td>
<td>Fatty acid elongation in mitochondria</td>
<td>6</td>
<td>16</td>
<td>0.004098</td>
<td>1.24E-01</td>
</tr>
<tr>
<td>ko04612</td>
<td>Antigen processing and presentation</td>
<td>15</td>
<td>75</td>
<td>0.009839</td>
<td>2.60E-01</td>
</tr>
<tr>
<td>ko04146</td>
<td>Peroxisome</td>
<td>23</td>
<td>134</td>
<td>0.011384</td>
<td>2.67E-01</td>
</tr>
<tr>
<td>ko01040</td>
<td>Biosynthesis of unsaturated fatty acids</td>
<td>9</td>
<td>38</td>
<td>0.014408</td>
<td>2.67E-01</td>
</tr>
</tbody>
</table>

"a": Differentially expressed transcripts
"b": Reference sequences were extracted from NCBI (see Materials and Methods for details).
"c": P-value is calculated based on Hypergeometric Test. Pathways with Q value≤0.05 are significantly enriched in differentially expressed transcripts. Pathway ID can be found in KEGG database.

Table 4. Differentially expressed transcripts validated by qRT-PCR analysis.

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
<th>Fold Change</th>
<th>SE DGE</th>
<th>qRT-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG10278</td>
<td>Flightin</td>
<td>5.961.13</td>
<td>1.631.77</td>
<td>159.36</td>
</tr>
<tr>
<td>gi</td>
<td>289097293</td>
<td>Kat</td>
<td>2068.00</td>
<td>11.06</td>
</tr>
<tr>
<td>AG4984</td>
<td>Fabp</td>
<td>1092.00</td>
<td>12.25</td>
<td>2.18</td>
</tr>
<tr>
<td>AG12333</td>
<td>phosphate carrier protein, mitochondrial-like</td>
<td>270.69</td>
<td>129.31</td>
<td>12.41</td>
</tr>
<tr>
<td>AG13078</td>
<td>Ant</td>
<td>52.62</td>
<td>55.20</td>
<td>0.78</td>
</tr>
<tr>
<td>AG11959</td>
<td>very long-chain specific acyl-CoA dehydrogenase, mitochondrial-like</td>
<td>11.63</td>
<td>12.14</td>
<td>2.53</td>
</tr>
<tr>
<td>AG10938</td>
<td>SerRS</td>
<td>-837.00</td>
<td>-3.32</td>
<td>0.21</td>
</tr>
<tr>
<td>AG13124</td>
<td>ribosomal protein S28e-like</td>
<td>11.45</td>
<td>26.07</td>
<td>5.96</td>
</tr>
<tr>
<td>AG13172</td>
<td>ribosomal protein L0</td>
<td>-4.53</td>
<td>-2.03</td>
<td>0.54</td>
</tr>
<tr>
<td>AG12143</td>
<td>ribosomal protein S27A</td>
<td>-4.37</td>
<td>-3.42</td>
<td>1.07</td>
</tr>
<tr>
<td>AG13161</td>
<td>40S ribosomal protein S4-like</td>
<td>-4.15</td>
<td>-1.18</td>
<td>0.17</td>
</tr>
<tr>
<td>AG11336</td>
<td>40S ribosomal protein S21-like</td>
<td>-4.03</td>
<td>-2.46</td>
<td>0.18</td>
</tr>
<tr>
<td>AG13187</td>
<td>ribosomal protein L15</td>
<td>-4.03</td>
<td>-3.76</td>
<td>1.16</td>
</tr>
<tr>
<td>AG8620</td>
<td>acylphosphatase</td>
<td>11.18</td>
<td>26.04</td>
<td>6.96</td>
</tr>
<tr>
<td>gi</td>
<td>289143076</td>
<td>acylphosphatase-2-like</td>
<td>-9.64</td>
<td>-2.06</td>
</tr>
<tr>
<td>AG8620</td>
<td>pyruvate kinase-like</td>
<td>-7.07</td>
<td>-1.87</td>
<td>0.69</td>
</tr>
<tr>
<td>AG10855</td>
<td>malate dehydrogenase, cytoplasmic-like</td>
<td>-6.52</td>
<td>-2.70</td>
<td>0.99</td>
</tr>
<tr>
<td>AG12539</td>
<td>pyruvate dehydrogenase E1 component subunit alpha, mitochondrial-like</td>
<td>-4.16</td>
<td>-2.51</td>
<td>1.21</td>
</tr>
<tr>
<td>AG9726</td>
<td>proteasome activator complex subunit 3-like</td>
<td>-4.88</td>
<td>-1.69</td>
<td>0.18</td>
</tr>
<tr>
<td>AG12585</td>
<td>proteasome inhibitor-like</td>
<td>-4.32</td>
<td>-1.10</td>
<td>0.05</td>
</tr>
</tbody>
</table>

****: Fold change of gene expression level between winged and wingless morphs. SE is the standard error of qRT-PCR. Positive and negative values, respectively, denote highly and lowly expressed gene expression in winged adult females relative to wingless individuals. Morph-specific genes are highlighted in shade. Assembly details are provided in Supplementary Material: Table S3.
Figure 3. Gene ontology classification of transcripts differentially expressed between winged and wingless *A. gossypii*. A total of 763 genes (45.88% of the differentially expressed transcripts) were categorized into three Gene Ontology classes, cellular components, molecular functions, and biological processes, respectively.

Figure 4. Differentially expressed transcripts enriched specific Gene Ontology terms. All 763 differentially expressed transcripts are enriched in 16 GO groups. 7 deep level (specific) GO terms were gained. 6 terms were overrepresented, and one term was underrepresented.

Figure 5. Functional classification of differentially expressed transcripts between winged and wingless *A. gossypii*. Based on Clusters of Orthologous Groups (COG, Figure_5A), 676 genes (40.65% of the Differentially expressed transcripts) were annotated and categorized into 23 functional classes. In comparison, 1144 genes (68.79% of the Differentially expressed transcripts) were annotated and categorized into 25 specific classes within the euKaryotic Orthologous Groups (KOG, Figure_5B). A: RNA processing and modification, B: Chromatin structure and dynamics, C: Energy production and conversion, D: Cell cycle control, cell division, chromosome partitioning, E: Amino acid transport and metabolism, F: Nucleotide transport and metabolism, G: Carbohydrate transport and metabolism, H: Coenzyme transport and metabolism, I: Lipid transport and metabolism, J: Translation, ribosomal structure and biogenesis, K: Transcription, L: Replication, recombination and repair, M: Cell wall/membrane/envelope biogenesis, N: Cell motility, O: Posttranslational modification, protein turnover, chaperones, P: Inorganic ion transport and metabolism, Q: Secondary metabolites biosynthesis, transport and catabolism, R: General function prediction only, S: Function unknown, T: Signal transduction mechanisms, U: Intracellular trafficking, secretion, and vesicular transport, V: Defence mechanisms, W: Extracellular structures, Y: Nuclear structure, Z: Cytoskeleton.
Discussion

DGE analysis

We sequenced wing-polyphenic aphids using the DGE approach and obtained 10.69 million raw base sequences, 10.48 million of which were clean tags. Clean tags were compared in different libraries. Interestingly, compared to the clean tags in non-specific forms, clean tags of morph-specific forms account for a small fraction of the total clean tags (Figure 1A), however, they comprise a higher level of percentage in distinct tags (Figure 1B), indicating that morph-specific tags have a high diversity. This result may suggest that expression level of common genes which maintain the basic growth and development is higher relative to specific genes that play crucial and multiple roles in trait variation. Typically, the reference sequences used by DGE analysis are selected from genome sequence information of evolutionarily related species to the target species, especially when there is no genome information available for the target species. In current DGE analysis, we chose EST sequence of *A. gossypii* as reference sequences, 46,766 tags were mapped to 18,141 transcripts (Table 1).

Differentially expressed transcripts

Through DGE analysis, we obtained a total of 1,663 differentially expressed transcripts, in which 58 transcripts were expressed significantly higher in winged aphids and 1,605 transcripts had a higher expression in wingless *A. gossypii* (Figure 2). This discrepancy is a reflection of the flight-reproduction trade-offs, in which genes associated with reproduction is likely over-numbered the flight capability. Among these differentially expressed transcripts, *Flightin* and *seryl-tRNA synthetase* (*SerRS*) was highly expressed in winged and wingless *A. gossypii*, respectively. Transcripts related to lipid metabolism and energy production were found at higher expression levels in migratory morphs (Table 2, Supplementary Material: Table S2). Differential energy allocation is an important part of trade-offs between flight capability and reproduction. This result is consistent with the previous observation of the significantly higher triglyceride content in winged aphid versus the wingless aphid [30, 31].

Associated with lipid synthesis and degradation, transcripts of 3-ketoacyl-CoA thiolase (*KAT*), very long-chain specific acyl-CoA dehydrogenase, medium-chain specific acyl-CoA dehydrogenase and long-chain-fatty-acid-CoA ligase 1 were shown significantly higher expression levels in winged morphs. Furthermore, 15 highly expressed “mitochondrial like” transcripts in winged morphs may contribute to energy production and thought to be involved in aspects of the citric acid cycle, respiratory chain, ATP synthesis and transport. These results suggest that lipid is the fuel for aphid migration. Similarly, transcripts involved in energy production showed elevated transcriptions in winged morphs of *A. pisum* [34] and the brown planthopper, *Nilaparvata lugens* [40].

Differentially expressed transcripts were significantly enriched in 7 GO terms, and 5 of them, structural molecule activity, ribonucleoprotein complex, translation, translation factor activity/nucleic acid binding, mitotic spindle organization, were directly or indirectly associated with protein translation and assembly. This is consistent with the pathway enrichment analysis. The remaining 2 term, membrane and intracellular non-membrane-bounded organelle, were associated with organelle, in which a series of biological process occurs, including partial protein synthesis processes in endoplasmic reticulum, fatty acid elongate in mitochondria, peroxisome break down very long chain fatty acids.

qRT-PCR validation study confirmed the results from DGE analysis (Table 4). Fully replicated qRT-PCR analysis, to certain extent, complements the lack of DGE replications. All 20 genes subjected to the qRT-PCR analysis showed the similar trends with the DGE results. In comparison to DGE analysis, however, the fold changes in gene expression of *Flightin*, *Fabp*, *Kat* and *SerRS* were less dramatic in qRT-PCR analysis. Such discrepancy is not without precedent [40]. Due to the computation needs (the lowest expression level of any gene in the DGE analysis, by default, is 0.01 even if it is not expressed), the differential gene expression level is typically overestimated in the DGE analysis. In this study, the expression of *Fabp*, *Kat* and *SerRS* were morph-specific, and consequently, the result from the DGE analysis is likely overestimated.

Pathway enrichment analysis

Through the Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation system, differentially expressed transcripts were significantly enriched (Q<0.05) in three pathways, ribosome, pyruvate metabolism and proteasome (Table 3). Of the differentially expressed transcripts 5.58% were in the ribosome pathway. While in the whole reference background, only 1.72% of transcripts were annotated in this pathway. The ribosome is required for protein synthesis, which is critical for all aspects of cell growth and division [49]. They are usually recognized as stable expression genes. However, in the current study, 48 highly expressed transcripts in wingless morph and 1 in winged morph were coded for the ribosomal proteins. Ribosomal genes are typically
highly expressed genes. Previously reported research results showed gene products that are components of ribosomes were over-representation in *A. pism* [34]. One 40S ribosomal protein S20 displayed a higher expression level in brachypterous morph of *N. lugens* than in macropterous adult females [40]. This evidence convinces us of difference in ribosomal proteins between two morphs of *A. gossypii*. Our data indicated that more protein synthesis was detected in the non-dispersing morphs, which results in faster developmental rates and higher fecundity [14, 50].

It is expected that pyruvate metabolism was a differentially expressed transcripts enriched pathway. 21 transcripts in this pathway were screened as differentially expressed transcripts. Pyruvate is a key intersection in several metabolic pathways, and known as the “hub” of carbohydrate, fats and proteins [16, 46]. Pyruvate can be generated from glucose through glycolysis [16]. In recent study, wingless morphs showed significantly higher soluble sugar and glycogen content than winged morphs during adulthood in *A. gossypii* [30] and in *S. avenae* [31]. The result of gene expression was consistent with our previous research in biochemical composition. Most transcripts in this pathway were highly expressed in wingless morphs. Our results suggest that sugar or glycogen metabolism is faster in wingless morphs and those carbohydrates were used for development and production. This is consistent with the contention that sugar and glycogen is not the fuel of aphids for migration [51].

26S proteasome, working with ubiquitin, operates the energy-dependent regulated proteolysis process in eukaryotic cells. The immunotype proteasomes, whose catalytic subunits are replaced by the related counterparts, were discovered recently in vertebrates [52]. The protein degradation process including immunity may be diverse in different morphs. In previous study, soluble protein content were higher in dispersal morphs than in reproduction morphs in *A. gossypii* [30] and *S. avenae* [31]. As such, proteasome for protein degradation and protein turnover increased.

**Genes of interests**

Genes which we are interested in include *Flightin*, fatty acid binding protein gene (*Fabp*), adenine nucleotide translocase gene (*Ant*), abnormal wing discs1 (*Awd1*) and abnormal wing discs2 (*Awd2*) (Supplementary Material: Table S2).

The primary difference between winged and wingless adults is the existence of flight apparatus, including wing and indirect flight muscle. For winged adults, genes expressed in muscle were identified in this study, *Flightin* and *Fabp*. *Flightin* is a multiply phosphorylated, myofibrillar protein in *Drosophila* indirect flight muscles (IFM), and an essential part of the flight muscle contractile mechanism for thick filament assembly and sarcomere stability in *Drosophila* [53-56]. Particularly, *Flightin* exhibits the greatest differential transcript accumulation in wing morphs of *A. pism* [34] and *N. lugens* [40]. Another gene coding FABP in muscle was one of the top ten most differentially expressed genes in winged morphs. FABP will increase the solubility of fatty acids and thus lead to a rapid transport through the cytosol, since the lipid is the preferred fuel for insect migrating [57] In recent research, free fatty acid content was significantly higher in winged than wingless adults before flight behavior in *A. gossypii* [30] and *A. pism* [31].

*ANT* belongs to the mitochondria carrier protein which is embedded in the inner membrane of mitochondria and required for importing ADP into the mitochondrial matrix throughout the ATP synthesis [33]. The gene expression of *Ant* gene is significantly higher in winged morphs than wingless morphs in *A. gossypii* (Table 2). The expression pattern of *Ag Ant* is similar as that of *Mp Ant* [33], suggesting the need of energy for winged morphs during the radiation to diverse food resources and habitats [58].

The *awd1* and *awd2* showed a significantly higher expression level in wingless morphs than winged morphs in *A. gossypii*. The *awd* of *Drosophila* encodes a nucleoside diphosphate kinase (NDPK). Complete loss of function of the *awd* gene causes lethality in *Drosophila* after the larval stage. Mutations in *awd* gene cause abnormal tissue morphology and aberrant differentiation. [59, 60]. The *awd* plays an important role in development and differentiation of insects. They are not involved in the presumptive wing-patterning gene network [61]. Little attention was given to *awd* genes in the previous studies related to wing polyphenism.

**Perspectives and future research**

This study sheds lights on the genetic basis of trade-offs between flight capability and reproduction. Genes involved in lipid metabolism show higher expression levels in winged *A. gossypii* than wingless morphs. In contrast, genes involved in sugar metabolism are elevated in wingless morphs. Our combined results suggest aphids using lipid rather than sugar or glycogen as fuel during migration. Flight ability of alate aphids depends on the energy supply of lipids. KAT, FABP, and very long-chain specific acyl-CoA dehydrogenase seems to be more important for winged morphs. Silencing of these genes using RNA interference (RNAi) may disrupt their energy supply, and compromise their ability to fly. RNAi through injection [62, 63] and feeding [64] methods have been
well established in *A. pisum*. It is germane to investigate the pest control potential of these genes, including *Flightin*, through genetically interrupting the dispersal of *A. gossypii*.

**Supplementary Material**

Table S1-S4, and Fig.S1 – S2. http://www.ijbs.com/v10p0257s1.xlsx

**Abbreviations**

DGE: Digital gene expression; EST: Expressed Sequence Tag; TPM: Number of transcripts per million clean tags; FDR: False Discovery Rate; GO: Gene Ontology; COG: Cluster of Orthologous Groups; KEGG: Kyoto Encyclopedia of Genes and Genomes; KO: KEGG ontology; KOG: euKaryotic Orthologous Groups; qRT-PCR: Quantitative real-time PCR, IFM: indirect flight muscles; SerRS: seryl-tRNA synthetase; KAT: 3-ketoacyl-CoA thiolase; FABP: fatty acid binding protein; NDPK: nucleoside diphosphate kinase; ANT: adenine nucleotide translocase.

**Acknowledgements**

Special thanks go to Timothy Moural (University of Washington State) and Drs. Tracy Lu and John Obrzycki (University of Kentucky) for their comments on an earlier version and Wenbo Chen for his assistance in the data analysis. This research was supported by the National Science & Technology Pillar Program (2012BAD19B05). Xiaowei Yang was supported by a start-up fund to Zhou and College of Agriculture, University of Kentucky. The granting agencies have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Data Deposition**

The accession numbers for DGE libraries derived from sedentary and migratory morphs of cotton aphids, *Aphis gossypii*, are SRX246107 and SRX242773, respectively.

**Competing Interests**

The authors have declared that no competing interest exists.

**References**


