Differential transcriptome profiles of heads from foragers: comparison between Apis mellifera ligustica and Apis cerana cerana

Zhiguo Li, Fang Liu, Wenfeng Li, Shaowu Zhang, Dong Niu, Haisheng Xu, Qihua Hong, Shenglu Chen, Songkun Su

Apidologie (2012) 43:487–500
© INRA, DIB and Springer-Verlag, France, 2012
DOI: 10.1007/s13592-012-0119-z

Original article

Differential transcriptome profiles of heads from foragers: comparison between *Apis mellifera ligustica* and *Apis cerana cerana*

Zhiguo Li[1], Fang Liu[1], Wenfeng Li[1], Shaowu Zhang[1,2], Dong Niu[1], Haisheng Xu[1], Qihua Hong[1], Shenglu Chen[1], Songkun Su[1]

[1]College of Animal Sciences, Zhejiang University, Hangzhou 310058, China
[2]ARC Centre of Excellence in Vision Science, Research School of Biology, College of Medicine, Biology and Environment, The Australian National University, Canberra, Australia

Received 27 March 2011 – Revised 9 November 2011 – Accepted 6 January 2012

Abstract – Given the differences in ecology and biology between *Apis cerana cerana* and *Apis mellifera ligustica*, we first used the Illumina–Solexa deep sequencing technology to describe the differences in the heads of *A. cerana cerana* and *A. mellifera ligustica* foragers at the gene expression level. We obtained over 3.6 million clean tags per sample and found about 2,370 differentially expressed genes related to metabolism, development, and signal transduction between *A. cerana cerana* and *A. mellifera ligustica*. Also, the many antisense transcripts found in our study indicated that they may represent novel paths involving gene expression regulation in honeybees. Our results indicated that differences in head expression profiles relate to sets of genes, and there existed significant enrichment of 22 pathways in differentially expressed genes. We conclude that the deep sequencing method provides us a better insight into differences at the molecular level between species within the genus *Apis*.

Apis cerana cerana / ***Apis mellifera ligustica*** / **foraging behavior** / **Illumina–Solexa sequencing** / **differential gene expression**

1. INTRODUCTION

The honeybee is a highly eusocial insect and is known for several noticeable social behaviors (Winston 1987; Seeley 1995). For example, adult worker honeybees conduct tasks in the hive when young and forage when older, thus demonstrating an age-based division of labor (Robinson 1992). The division of labor is also flexible, and colonies respond to changing internal and external conditions by adjusting the ratios of workers engaged in the various tasks (Robinson 1992). Overall, therefore, the process of behavioral development in honeybee is affected by environmental, social, and genetic factors (Robinson et al. 1997). Honeybees are also known for their typical dance language by which foragers can communicate with each other and find food sources successfully (von Frisch 1967). Behavioral changes related to age in adult honeybees are associated with corresponding changes in the expression of different genes in the brain (Ben-Shahar et al. 2002, 2003; Nelson et al. 2007). Honeybees have, therefore, long been recognized as a model organism for the study of social interactions through investigations of their genome (Weinstock et al. 2006).

Our knowledge of honeybee behaviors stems mainly from investigations on the western honeybee, *Apis mellifera ligustica*, which has been studied widely up to now. *Apis cerana*

Corresponding author: S. Su,
susongkun@zju.edu.cn
Manuscript editor: Marina Meixner

cerana is a species that is bred locally in China. There are about two million *A. cerana cerana* colonies in China (Chen 2001), and *A. cerana cerana* is widely used for commercial beekeeping in the mountain areas of South China, mainly because of its resistance to diseases, wasps, and bee mites (Su et al. 2005). By changing its flight behavior, *A. cerana cerana* can escape from wasp predation, while *A. mellifera ligustica* cannot (Tan et al. 2007). In terms of mite resistance, *A. cerana cerana* is more effective and successful in both removing mites and causing damage to mites than *A. mellifera ligustica* (Fries et al. 1996). However, studies on differences between *A. mellifera ligustica* and *A. cerana cerana* at the molecular level are limited. Here, differential gene expression in the heads of *A. cerana cerana* and *A. mellifera ligustica* foragers was analyzed using Illumina–Solexa's digital gene expression (DGE) tag profile system.

Advances in our understanding of transcriptome profiles have recently become possible with the development of novel technologies in the field of deep sequencing (Cloonan and Grimmond 2008; Morozova and Marra 2008; Hegedùs et al. 2009; Wang et al. 2009; Wilhelm and Landry 2009; Wu et al. 2010; Liu et al. 2011), which have overcome the limitations of tag-based methods in the detection of transcripts with alternative splicing in regions remote from the 3′-end and novel transcripts ('t Hoen et al. 2008). This is the first report of a deep sequencing study of differential expression in the heads of *A. cerana cerana* and *A. mellifera ligustica* foragers, and our results give further insight into the differences in behavior and physiology between *A. cerana cerana* and *A. mellifera ligustica* at the gene expression level.

2. MATERIAL AND METHODS

2.1. Sample collection of heads from foragers

Ninety individual bees were taken from 3 *A. cerana cerana* colonies and 3 *A. mellifera ligustica* colonies, respectively, in Huajiachi campus of Zhejiang University, Hangzhou, China. The *A. mellifera ligustica* and *A. cerana cerana* samples were collected on the same day in the early afternoon, when the honeybees forage frequently. Thirty foragers returning with pollen loads from each colony were collected and directly put into liquid nitrogen (Toma et al. 2000) to minimize the possible degradation of RNA. The samples of two types of honeybees used for creating tag libraries consisted of the pooled heads of 3 colonies (90 individual bees) from each respective species. The pooled heads used for RNA extraction were isolated in liquid nitrogen using sterile scalpels to avoid the contamination of samples and stored at −80°C until further investigation.

2.2. RNA extraction and cDNA synthesis

Total RNA was isolated using the TRIZOL reagent (Invitrogen, Carlsbad, CA, USA) according to the protocol of the manufacturer. The integrity of RNA was demonstrated by 1.0% denaturing gel electrophoresis. Six micrograms total RNA, the mRNA of which was purified by magnetic oligo(dT) beads, was extracted from the heads of *A. cerana cerana* and *A. mellifera ligustica* foragers. Oligo(dT) was also used for guiding the synthesis of first-strand cDNA through reverse transcription, which was followed by the synthesis of double-strand cDNA using the M-MLV RTase cDNA Synthesis Kit (TaKaRa, Dalian, China) following the manufacturer's instructions.

2.3. Preparation of tags and the Illumina–Solexa sequencing

Preparation of the tag library: Endonuclease *Nla*III was used to recognize and cut the cDNA on the CATG sites. Fragments other than the 3′ cDNA fragments attached to magnetic oligo(dT) beads were washed away, and an Illumina adapter 1 was ligated to the free 5′-end of the digested bead-bound cDNA fragments (Hegedùs et al. 2009). The Illumina adapter 1 contains a restriction site for *Mme*I, which cuts 17 bp downstream from the *Nla*III site. Thus, tags with adapter 1 were created. The 3′-ends of tags were ligated to the second adapter (Illumina adapter 2), thereby producing a library of 21-bp-long tags starting from the CATG that *Nla*III recognizes. The adapter-ligated cDNA tags were amplified using PCR linear amplification. After 12 cycles of PCR linear

amplification, 85-bp fragments were purified by 6% TBE PAGE gel electrophoresis. Samples from *A. cerana cerana* and *A. mellifera ligustica* were run on the same lane of a flow cell simultaneously. Post melting, the single-chain molecule was added and fixed to the flow cell according to the protocol of manufacturer, thus converting each molecule into a single-molecule cluster of the sequencing templates through amplification in situ. Then nucleotides labeled with four-color fluorescence were added, and sequencing by synthesis was carried out (Hall 2007). The original image data generated via sequencing were indexed with standard Illumina index in order to be recognized. The image data were converted into sequence data through base calling. Therefore, each channel would generate millions of raw reads, and the length of each read was 35 bp (Li et al. 2008; Mardis 2008).

2.4. Tag mapping and statistical analysis

The original sequence data were called raw reads. Clean tags are tags used to analysis after removing the 3′-adaptor sequence, empty tags (tags containing adaptors without tag sequence), low-quality tags (tags with unknown base N), and tags with a single copy number resulting from sequencing errors from raw reads. Hence, a mass of clean tags containing CATG were obtained, and we mapped the clean tags onto the database of reference tags and the honeybee genome. Since the honeybee genome was constructed from the genome sequence of just *A. mellifera*, the database of reference tags was assembled from all reference sequences, mRNA, and expressed sequence tags (ESTs) found in GenBank. We took all tags including the CATG site in a gene, and not only the 3′-end, as the reference tags. Therefore, a database of all possible CATG+17-nt reference tags was created. Perfect match (100% identity and coverage) and only 1-bp mismatch were both considered in the present study.

Clean tags mapped to reference sequences from multiple genes called ambiguous clean tags that were discarded from the tag mapping. The remaining clean tags that can only be mapped to reference sequences from single genes were defined as unambiguous clean tags. The number of unambiguous clean tags was determined and then normalized to transcript per million clean tags (TPM) in order to compare the expression level of each gene between *A. cerana cerana* and *A. mellifera ligustica* directly and more accurately ('t Hoen et al. 2008; Morrissy et al. 2009).

Differentially expressed genes between *A. mellifera ligustica* and *A. cerana cerana* were screened according to an algorithm (Audic and Claverie 1997). Therefore, the *P* value was used for the identification of differentially expressed genes, and the *P* value threshold was set by controlling the false discovery rate (FDR) in multiple testing according to Benjamini and Yekutieli (2001). A FDR <0.001 and $|\log_2 \text{ratio}| \geq 1$ (ratio: *A. mellifera ligustica*/*A. cerana cerana*) were used to judge the significant difference of expressed genes and assess genes upregulated or downregulated between *A. cerana cerana* and *A. mellifera ligustica* in the study.

2.5. Gene Ontology and KEGG Orthology analysis

Functional annotation of differentially expressed genes was carried out using the Gene Ontology terms (GO) database (http://www.geneontology.org) to determine their categories (biological process, cellular component, and molecular function). Given each pathway in the Kyoto Encyclopedia of Genes and Genomes (KEGG), reference pathway maps are assigned a KEGG Orthology (KO) identifier (Kanehisa et al. 2006). Significant enrichment of pathways in differentially expressed genes between *A. cerana cerana* and *A. mellifera ligustica* was also analyzed. The enriched *P* values that were used to discover statistically significantly enriched GO terms and KO terms were computed based on the hypergeometric test:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where *N* stands for the total number of all genes with GO or KO annotation, *n* stands for the total number of differentially expressed genes within *N*, *M* stands for the total number of annotated genes in each GO or KO term, and *m* stands for the total number of differentially expressed annotated genes in *M*. The resulting *P* values were corrected via Bonferroni, and the threshold for GO enrichment analysis of differentially expressed

genes was determined as corrected P value <0.05. The FDR <0.05 was used for KO enrichment analysis of differentially expressed genes.

3. RESULTS

3.1. Quality evaluation of DGE tags

Perfect match tags accounted for 75% of all tags mapping to genes in *A. cerana cerana*, and the rate was 95% in *A. mellifera ligustica*. Low-quality tags in *A. cerana cerana* only accounted for 1.24% (48,980) of total tags (3,937,152). For *A. mellifera ligustica*, low-quality tags only accounted for 1.21% (56,233) of total tags (4,633,191). Besides, single-copy tags and various impurities only occupied 6.61% and 6.27% of the total tags in *A. cerana cerana* and *A. mellifera ligustica*, respectively. In general, the proportion of low-quality tags to the raw reads keeps within 2% and the total number of single-copy tags and various impurities account for <15% of raw reads, which reveal that the sample preparation and the test data were in perfect condition. The data derived from the only sequencing run of the two types of bees in the study meet the aforementioned criteria entirely.

To test whether there was a positive correlation between the number of identified genes (distinct tag number) and capacity of sequencing (total tag number), an analysis of sequencing saturation was conducted for *A. cerana cerana* and *A. mellifera ligustica*. When the number of total tags reached about 3.5 million in *A. cerana cerana*, the percentage of genes identified no longer increased. The same pattern was shown in *A. mellifera ligustica*, no further genes were found when the total tag number reached about three million. The results further revealed that the sample preparation and the test data were reasonable and comprehensive.

3.2. Analysis of the expression level and distribution of clean tags

With Illumina's DGE assay, we sequenced DGE libraries produced from the heads of *A. cerana cerana* and *A. mellifera ligustica* foragers and obtained ~3.6 million clean tags of *A. cerana cerana* and ~4.2 million clean tags of *A. mellifera ligustica*. After making the difference comparative analysis of clean tags between *A. mellifera ligustica* and *A. cerana cerana*, we found differential expression of 40,886 clean tags with an estimated FDR of <0.001. Through analysis of public, specific tags of *A. cerana cerana* and *A. mellifera ligustica*, we found that tags, which expressed at a high level, were public tags existing in both *A. cerana cerana* and *A. mellifera ligustica*, whereas some scarce tags existed only in either *A. cerana cerana* or *A. mellifera ligustica* (Table I).

For *A. cerana cerana*, the total set of all clean tags could be separated into five categories (Figure 1a). There were 33,323 kinds of unambiguous clean tags, which accounted for 29.15% of total distinct clean tags (114,330); the counts of unambiguous clean tags totaled 1,411,154, which accounted for 38.90% of total clean tags (3,628,091). Similarly, for *A.*

Table I. Statistical analysis of public, specific tags between *A. cerana cerana* and *A. mellifera ligustica.*

	Total tag number				Distinct tag number			
	A. cerana cerana		*A. mellifera ligustica*		*A. cerana cerana*		*A. mellifera ligustica*	
Total clean tag	3,628,091	100%	4,286,250	100%	114,330	100%	115,789	100%
A. cerana cerana and *A. mellifera ligustica*	2,325,018	64.08%	3,014,064	70.32%	95,972	83.94%	100,124	86.47%
A. cerana cerana only	1,303,073	35.92%	–	–	18,358	16.06%	–	–
A. mellifera ligustica only	–	–	1,272,186	29.68%	–	–	15,665	13.53%

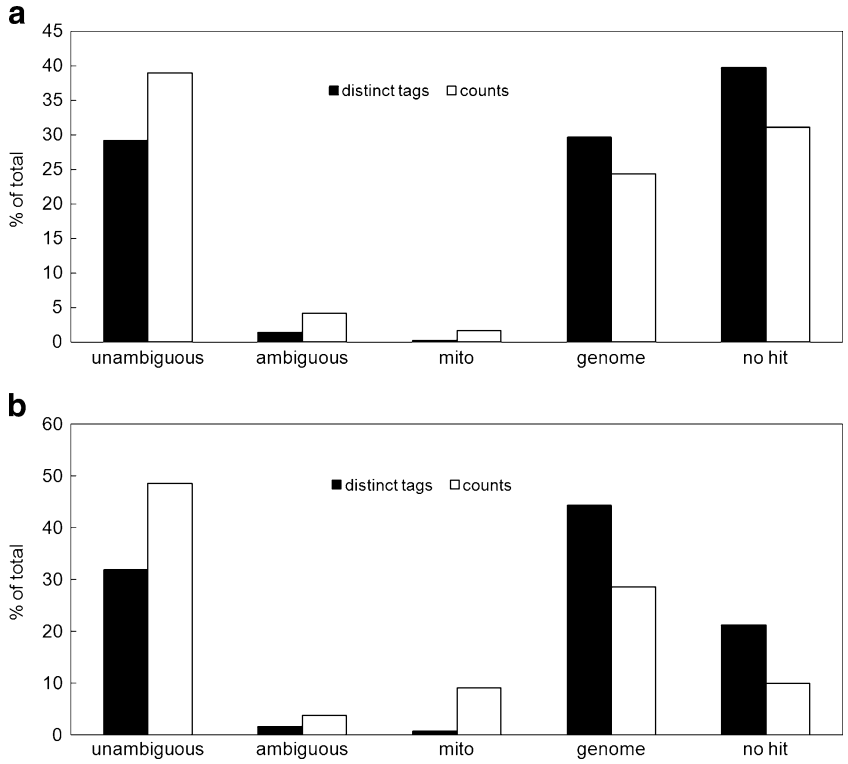

Figure 1. Categorization and abundance of clean tags in *A. cerana cerana* (**a**) and *A. mellifera ligustica* (**b**), respectively. The *black bar* in each category represents the number of distinct clean tags that accounts for the percentage of the total distinct clean tags, and the *open bar* in each category represents the counts of clean tags that account for the percentage of the total clean tags: *unambiguous* unambiguous clean tags, *ambiguous* ambiguous clean tags, *mito* clean tags mapping to the mitochondrion genome sequence, *genome* just the genome of *A. mellifera ligustica*, *no hit* unknown tags with no hits in the mapping to genome and the database of reference tags.

mellifera ligustica, the total set of all clean tags could also be divided into five categories (Figure 1b). There were 36,978 kinds of unambiguous clean tags, which accounted for 31.94% of total distinct clean tags (115,789); the counts of unambiguous clean tags totaled 2,080,793, which accounted for 48.55% of total clean tags (4,286,250). The distinct clean tags represent types of transcripts, and the total clean tags correspond to transcript abundance. Therefore, the differences in kinds and counts of unambiguous clean tags reflect differences in types of transcripts and transcript abundance between *A. mellifera ligustica* and *A. cerana cerana*, respectively.

3.3. Analysis of difference in gene abundance between *A. cerana cerana* and *A. mellifera ligustica*

After tag mapping, a great many annotated genes were created, and both the sense and the antisense sequences were included in the data collection, which was used for further analysis of differentially expressed genes and differentially expressed pathways. About 2,370 genes relating to metabolism, development, and signal transduction that were expressed differentially at significant levels between *A. mellifera ligustica* and *A. cerana cerana* (FDR <0.001, 2× difference) and the top 25 most differentially

expressed annotated genes with more than eightfold difference between *A. cerana cerana* and *A. mellifera ligustica* were listed in Table II.

There were 244 genes expressed significantly in *A. mellifera ligustica*, which were not expressed in *A. cerana cerana* (Supplementary Table I). Similarly, there were 125 genes expressed significantly in *A. cerana cerana*, which were not found in *A. mellifera ligustica* (Supplementary Table II). In addition, we found 1,287 genes which were upregulated in the heads of *A. cerana cerana*, and the number of upregulated genes in the heads of *A. mellifera ligustica* was 1,083. About 4,946 nondifferentially expressed genes existed in both *A. mellifera ligustica* and *A. cerana cerana* (Figure 2). The Pearson correlation coefficient between *A. cerana cerana* and *A. mellifera ligustica* was 0.904, suggesting that the data used in the gene expression profiles of *A. cerana cerana* and *A. mellifera ligustica* were reliable and stable in the study. We assumed that the smaller the FDR and the greater the fold difference in expression, the higher was the confidence in the difference of the gene expression profile. Using the more stringent criteria of both FDR <0.0001 and existence minimum of more than eightfold expression difference ($P< 10^{-4}$), our results showed that 125 genes were expressed significantly in *A. cerana cerana* and 250 genes were expressed significantly in *A. mellifera ligustica*, 6 of which were also expressed slightly in *A. cerana cerana*.

3.4. Functional annotation and classification of differentially expressed genes

Significant enrichment of six GO terms within differentially expressed genes between *A. cerana cerana* and *A. mellifera ligustica* was calculated (Figure 3). The results showed two functional categories in regard to biological process and cellular component. However, there were no GO nodes with a P value <0.05 for the differentially expressed genes related to molecular function. Five GO terms were associated with mitochondrial function, indicating that these differentially expressed genes may play an important role in influencing the energy metabolism in honeybee. Furthermore, the differentially expressed genes correlated with energy metabolism were upregulated in *A. mellifera ligustica* compared with *A. cerana cerana*. Four genes encoding troponin C and troponin T, one of which was not expressed in *A. cerana cerana*, were expressed at significantly higher levels in *A. mellifera ligustica* than in *A. cerana cerana*. Schippers et al. (2006) found that increases in troponin T may contribute to enhanced flight performance (rate of foraging delivery and average weight of food collected per trip). The differences in troponin expression, coupled with the differences in mitochondrial genes mentioned above, suggest that there may be significant differences in energy metabolism during flight in the two species. This hypothesis is also supported by the fact that the honey stomach of *A. mellifera ligustica* is larger than that of *A. cerana cerana* and the amount of nectar and pollen collected by *A. mellifera ligustica* is always more than that of *A. cerana cerana*. For each foraging trip, the average size of the pollen loads carried by *A. mellifera ligustica* and *A. cerana cerana* are 19.3 and 16.17 mg, respectively (Chen 2001). The only GO term, response to stress, to which differentially expressed genes annotated were upregulated in *A. cerana cerana* compared with *A. mellifera ligustica*, revealed difference in stress resistance between the two species.

3.5. Analysis of differentially expressed pathways

A total of 1,670 differentially expressed genes with 213 annotated pathways were discovered in the study, and 22 pathways exhibited significant differences between *A. cerana cerana* and *A. mellifera ligustica*. Genes involved in 16 pathways were upregulated in *A. mellifera ligustica*, compared with *A. cerana cerana*. On the contrary, only six pathways were upregulated in *A. cerana cerana* (Figure 4). A pathway with respect to caprolactam degradation involved seven differentially expressed genes, and only one gene was

Table II. A list of the 25 most differentially expressed annotated genes with more than eightfold difference between *A. cerana cerana* and *A. mellifera ligustica.*

Gene ID	TPM—*A. cerana cerana*	TPM—*A. mellifera ligustica*	Gene description
gi|94158717|ref|NM_001040206.1|	0.01	730.01	*Apis mellifera* odorant binding protein 21 (Obp21), mRNA
gi|58585187|ref|NM_001011622.1|	0.01	291.63	*Apis mellifera* major royal jelly protein 6 (Mrjp6), mRNA
gi|58585117|ref|NM_001011589.1|	0.01	145.58	*Apis mellifera* odorant binding protein 4 (Obp4), mRNA
gi|94158708|ref|NM_001040205.1|	0.01	53.43	*Apis mellifera* odorant binding protein 16 (Obp16), mRNA
gi|94158728|ref|NM_001040221.1|	0.01	45.49	*Apis mellifera* odorant binding protein 3 (Obp3), mRNA
gi|94158812|ref|NM_001040222.1|	0.01	22.16	*Apis mellifera* odorant binding protein 20 (Obp20), mRNA
gi|94158710|ref|NM_001040207.1|	0.01	21.46	*Apis mellifera* odorant binding protein 17 (Obp17), mRNA
gi|58585125|ref|NM_001011593.1|	0.01	14.46	*Apis mellifera* odorant binding protein 6 (Obp6), mRNA
gi|94158667|ref|NM_001040208.1|	0.01	11.9	*Apis mellifera* odorant binding protein 15 (Obp15), mRNA
gi|94158841|ref|NM_001040226.1|	0.01	5.83	*Apis mellifera* odorant binding protein 11 (Obp11), mRNA
gi|58585107|ref|NM_001011580.1|	12.68	5,580.4	*Apis mellifera* major royal jelly protein 2 (Mrjp2), mRNA
gi|58585251|ref|NM_001011653.1|	0.01	4.2	*Apis mellifera* troponin C type IIa (TpnCIIa), mRNA
gi|58585141|ref|NM_001011601.1|	0.55	215.11	*Apis mellifera* major royal jelly protein 3 (Mrjp3), mRNA
gi|58585169|ref|NM_001011610.1|	1.1	328.96	*Apis mellifera* major royal jelly protein 4 (Mrjp4), mRNA
gi|58585137|ref|NM_001011599.1|	112.18	13,312.34	*Apis mellifera* major royal jelly protein 5 (Mrjp5), mRNA
gi|58585121|ref|NM_001011591.1|	0.55	46.19	*Apis mellifera* odorant binding protein 2 (Obp2), mRNA
gi|58585089|ref|NM_001011574.1|	815.3	20,413.65	*Apis mellifera* glucose oxidase (LOC406081), mRNA
gi|58585181|ref|NM_001011619.1|	3.86	92.86	*Apis mellifera* hyaluronoglucosaminidase (LOC406146), mRNA
gi|58585147|ref|NM_001011600.1|	0.55	10.97	*Apis mellifera* hexamerin 70b (HEX70b), mRNA
gi|58585123|ref|NM_001011588.1|	0.55	8.4	*Apis mellifera* odorant binding protein 5 (Obp5), mRNA
gi|94400892|ref|NM_001040258.1|	6.06	77.22	*Apis mellifera* troponin T, skeletal muscle (TpnT), mRNA
gi|94400898|ref|NM_001040256.1|	24.26	293.5	*Apis mellifera* troponin I (TpnI), mRNA
gi|94158853|ref|NM_001040236.1|	6.06	71.16	*Apis mellifera* alpha glucosidase 2 (AGLU2), mRNA
gi|58585153|ref|NM_001011607.1|	14.06	0.01	*Apis mellifera* melittin (Melt), mRNA
gi|58585217|ref|NM_001011638.1|	27.29	1.63	*Apis mellifera* defensin 2 (Def2), mRNA

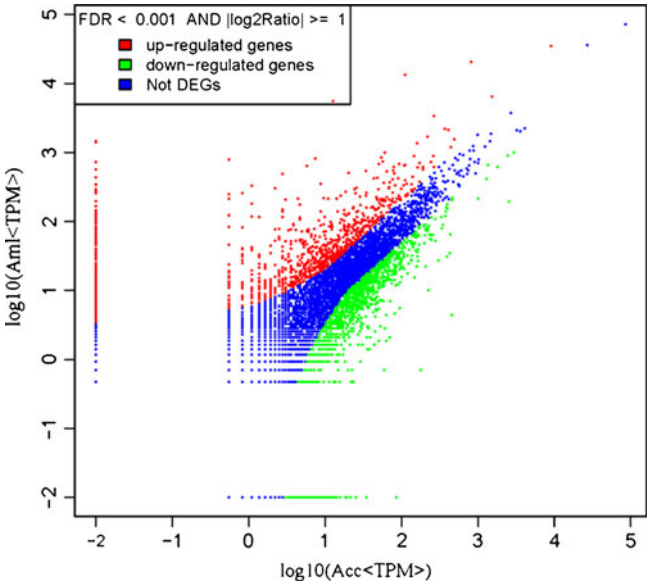

Figure 2. Gene expression level of *A. cerana cerana* (Acc) vs *A. mellifera ligustica* (Aml). *TPM* a standardized value indicating the total amount of transcript copies per one million clean tags. The ratio was calculated as the number of normalized clean tags in *A. mellifera ligustica* divided by the number of normalized clean tags in *A. cerana cerana*: the value of $\log_2$ ratio is more than or equal to 1, which means that genes are upregulated in *A. mellifera ligustica*; the value of $\log_2$ ratio is less than or equal to −1, which means that genes are upregulated in *A. cerana cerana*. Nondifferentially expressed genes are shown in *blue*, upregulated genes and downregulated genes are shown in *red* and *green*, respectively.

upregulated in *A. cerana cerana*. However, the remaining 21 differentially expressed pathways involved more than 1 gene upregulated in *A. cerana cerana* and *A. mellifera ligustica*, re-

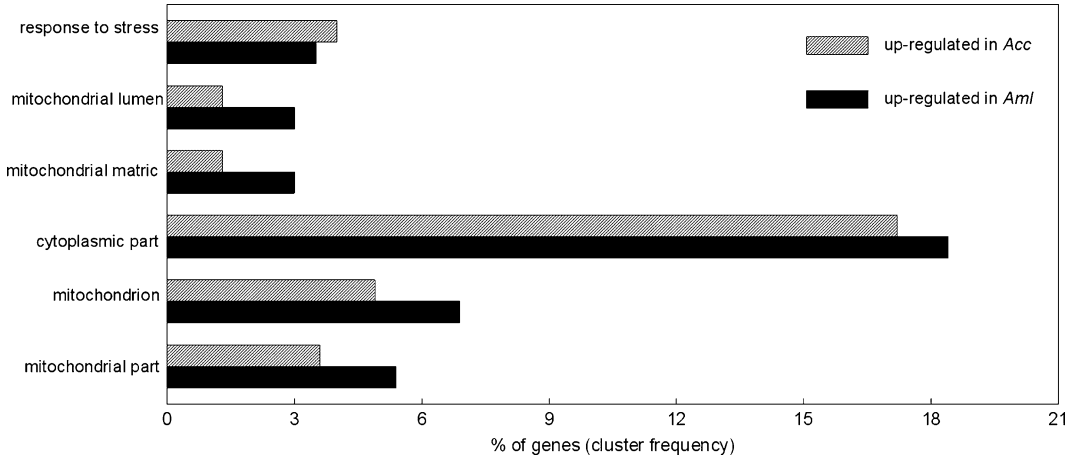

Figure 3. Significant enrichment of GO terms within differentially expressed genes between *A. cerana cerana* (Acc) and *A. mellifera ligustica* (Aml). The *abscissa* represents the percentage of genes annotated to each specific GO term and the *ordinate* represents the GO terms. All the processes are shown with *P* value <0.05.

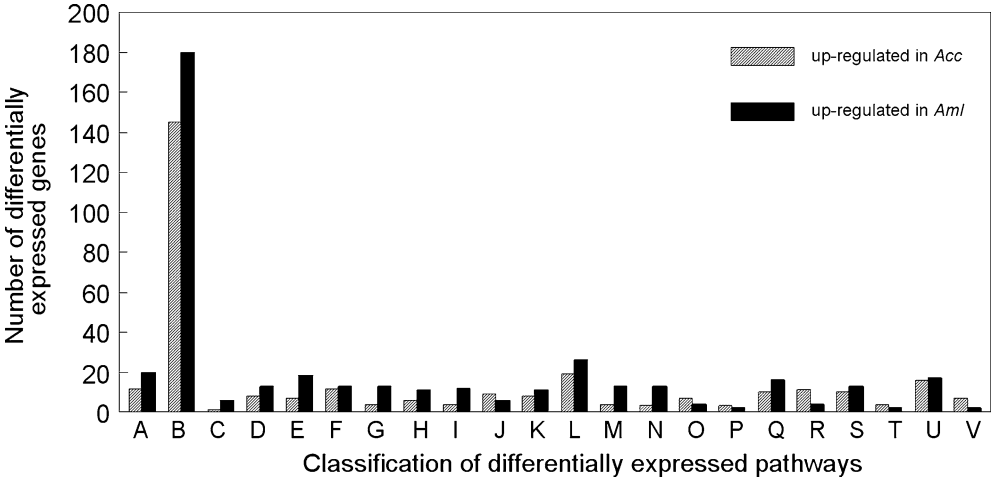

Figure 4. Classification of differentially expressed genes with pathway annotation. Out of 213 pathways, 22 pathways (listed using the *capital letters*) have *P* values <0.05. The number of differentially expressed genes within each special pathway in *A. cerana cerana* (Acc) and *A. mellifera ligustica* (Aml) is represented by the *ordinate*. *A* gamma-Hexachlorocyclohexane degradation; *B* metabolic pathways; *C* caprolactam degradation; *D* starch and sucrose metabolism; *E* drug metabolism—other enzymes; *F* PPAR signaling pathway; *G* metabolism of xenobiotics by cytochrome P450; *H* fatty acid metabolism; *I* valine, leucine, and isoleucine degradation; *J* 1- and 2-methylnaphthalene degradation; *K* benzoate degradation via CoA ligation; *L* Alzheimer's disease; *M* drug metabolism 698–cytochrome P450; *N* retinol metabolism; *O* glycosaminoglycan degradation; *P* vitamin B_6 metabolism; *Q* limonene and pinene degradation; *R* mTOR signaling pathway; *S* aminoacyl-tRNA biosynthesis; *T* glycosphingolipid biosynthesis—globo series; *U* tyrosine metabolism; *V* type II diabetes mellitus.

spectively. Metabolic pathways involved 325 differentially expressed genes, 180 of which were upregulated in *A. mellifera ligustica*; however, no corresponding pathway map was found in the KEGG database in our study. Twenty-one differentially expressed genes were involved in the pathways of starch and sucrose metabolism, of which 13 genes were highly upregulated in *A. mellifera ligustica*. The results showed that the foragers from *A. mellifera ligustica* may be better able to get energy from starch and sucrose for flight because the flight distance of *A. mellifera ligustica* is farther than that of *A. cerana cerana* (Chen 2001).

The target of rapamycin (TOR), which is a kinase involved in nutrition and energy-sensing, is responsible for organismal growth (Patel et al. 2007). Foragers of both *A. cerana cerana* and *A. mellifera ligustica* expressed TOR; however, genes related to the TOR signaling pathway were upregulated in *A. cerana cerana*. High levels of TOR activity are associated with higher metabolic activity and shorter lifespan (Stanfel et al. 2009); the significant difference of TOR expression found between *A. cerana cerana* and *A. mellifera ligustica* is possibly due to their different genetic background. The findings revealed that some sophisticated biological processes underlying the molecular mechanism of metabolism and energy expenditure in the honeybee may exist, which need further research.

3.6. Annotation of antisense genes

Previous studies demonstrated that sense–antisense transcription is an important means of gene expression regulation (Katayama et al. 2005; Yelin et al. 2003). More than 60% of sense–antisense transcriptional units are transcribed from both DNA strands in the human genome (Yelin et al. 2003). Our results showed that the antisense transcripts were also expressed at considerable levels both in *A. cerana cerana*

and *A. mellifera ligustica*, which were not reported in previous studies. For *A. cerana cerana*, the number of genes with only antisense transcripts was 142 (Supplementary Table III), which accounted for 2.07% of all expressed genes. The corresponding values in *A. mellifera ligustica* were 156 and 2.21% (Supplementary Table IV).

Additionally, the sense–antisense transcriptional units were expressed both in *A. cerana cerana* and *A. mellifera ligustica*, though the abundance of antisense transcripts was much lower than that of sense transcripts in terms of the same genes in most cases as a previous study described (’t Hoen et al. 2008). A total of 3,890 sense–antisense transcriptional units that were transcribed from both DNA strands accounted for about 56.8% of all expressed genes in *A. cerana cerana*. For *A. mellifera ligustica*, there were 4,208 sense–antisense transcriptional units that were transcribed from both DNA strands accounted for about 59.5% of all expressed genes.

Our results further elucidated that the genes encoding major royal jelly proteins (MRJPs), odorant binding proteins (OBPs), and troponin-like proteins also exhibited antisense transcriptional expression in honeybees. As mentioned above, *mrjp7*, *obp7*, and *obp12* belonged to the sense–antisense units, which indicate that they were transcribed from both DNA strands. However, the abundance of antisense transcripts with respect to *mrjp7*, *obp7*, and *obp12* was much lower than that of sense transcripts. Therefore, the three proteins were mainly produced from sense transcripts, and the role of low abundant antisense transcripts will need further research. Besides, antisense transcripts of *mrjp7* and *obp12* were expressed both in *A. mellifera ligustica* and *A. cerana cerana*, while the antisense transcript of *obp7* was expressed only in *A. mellifera ligustica*.

4. DISCUSSION

Among the differentially expressed genes, we observed that gene family encoding odorant binding-like proteins were expressed at a significantly higher level in *A. mellifera ligustica* than in *A. cerana cerana*. They consisted of 12 genes, 9 of which were not expressed in *A. cerana cerana*, and encoded 12 different kinds of odorant binding-like proteins in *A. mellifera ligustica*, such as OBP3, OBP8, OBP11, and so on. OBPs are an important component of the insect chemosensory system and have a potential influence on chemosensation (Xu et al. 2005). Previous studies have reported that *A. mellifera* had few genes for gustatory receptors, but many genes for odorant proteins (Weinstock et al. 2006). Both *A. mellifera ligustica* and *A. cerana cerana* have few genes for gustatory receptors, which corresponded with the findings from previous studies. In contrast to previous studies, however, genes for odorant proteins were expressed at significantly higher levels in *A. mellifera ligustica* than in *A. cerana cerana* in our study. Two possible explanations can be forwarded to explain this difference. Firstly, our experiments were executed in July when there were few nectar resources but abundant resin and gum resources from different kinds of trees in the environment. Foraging for gum and resin, which are the raw materials of propolis, is a well-known behavior of *A. mellifera ligustica*, but not of *A. cerana cerana*. It is, therefore, possible that the collection of gum and resin from trees in the summer contributed to the increased expression of genes encoding OBPs in *A. mellifera ligustica*. Secondly, the food consumption and colony population of *A. mellifera ligustica* are both larger than that of *A. cerana cerana*, and therefore, the foraging rate of *A. mellifera ligustica* is likely to be stronger.

In addition, among the six genes encoding MRJPs that are registered in the honeybee genome database, *mrjp1*, *mrjp2*, *mrjp3*, *mrjp4*, and *mrjp5* were expressed at significantly higher levels in *A. mellifera ligustica* than in *A. cerana cerana*. *mrjp6*, not found in *A. cerana cerana*, was only expressed in *A. mellifera ligustica* at a higher level and is mostly closely related to *mrjp5* (Albert and Klaudiny 2004). *mrjp1* and *mrjp2*, expressed in hypopharyngeal glands generally, have also been found in the

brain of worker bees (Kucharski et al. 1998; Kucharski and Maleszka 2002; Garcia et al. 2009; Peixoto et al. 2009). Our findings agreed with the view that the MRJPs can be multifunctional, executing additional roles involved in metabolism, cell differentiation, and morphogenesis besides having a nutritional role (Drapeau et al. 2006; Peixoto et al. 2009). However, the molecular mechanism of MRJPs in the determination of sophisticated behaviors in the honeybee is still obscure. Besides, the western honeybees, *A. mellifera ligustica*, raised in China have a stronger ability to secrete royal jelly than that of *A. cerana cerana* thanks to several years of artificial selection. This is a possible reason why the *mrjps* were expressed at a significantly higher level in *A. mellifera ligustica* than in *A. cerana cerana*.

Compared with *A. mellifera ligustica*, the genes involved in response to stress were upregulated in *A. cerana cerana*. Additionally, it was notable that the gene encoding defensin 2 was expressed at a significantly higher level ($P<10^{-23}$) with 16-fold difference in *A. cerana cerana* than in *A. mellifera ligustica*. However, the gene encoding defensin 1 was expressed at a significantly higher level ($P<10^{-9}$) with four-fold difference in *A. mellifera ligustica* than in *A. cerana cerana*. Defensin 2 and defensin 1, which are antimicrobial peptides in the honeybee, are expressed in heads and thoraces (Klaudiny et al. 2005). The differential expression of defensin suggested that the defense mechanisms against pathogenic microbes in *A. cerana cerana* and *A. mellifera ligustica* are different from each other and may be a reason why there were no reports about the unusual mortality of eastern honeybee (*A. cerana*) colonies in comparison to the drastic loss of western honeybee (*A. mellifera*) that has been occurring worldwide in recent years (Neumann and Carreck 2010). Interestingly, we found that the gene encoding melittin was expressed significantly in *A. cerana cerana*, but not in *A. mellifera ligustica*. This may explain why *A. cerana cerana* is more aggressive than *A. mellifera ligustica*. In addition, 12 genes, encoding different kinds of proteins involved in different pathways, were expressed at significantly higher levels in *A. cerana cerana* than in *A. mellifera ligustica*.

Previous studies have demonstrated the wide existence of antisense transcription and its biological relevance (Carninci et al. 2005; Katayama et al. 2005; Beiter et al. 2009; Hegedùs et al. 2009). In our study, some genes were also detected as antisense tags for the sense transcripts in *A. cerana cerana* and *A. mellifera ligustica*, respectively. For example, antisense genes encoding Period clock proteins (Per) were detected in *A. mellifera ligustica* and *A. cerana cerana*, respectively, after GO analysis of their orthologues. It has been shown that a nonvisual vertebrate-like opsin, which may be involved in linking the circadian clock to daylight, is expressed deep in honeybee brains (Velarde et al. 2005; Rubin et al. 2006). It is of interest to understand whether Period clock proteins have a relationship with circadian rhythms, and further experiments are required to assess the relations between them.

In this study, we found a great many novel genes that were not reported previously in either *A. cerana cerana* or *A. mellifera ligustica* by using Illumina–Solexa's DGE, and these novel genes could guide future studies. At the same time, we also found some genes that were not matched to the honeybee genome at all. The no hit tags existed in both *A. cerana cerana* and *A. mellifera ligustica*, some no hit tags may be genes from microorganisms (bacteria, virus, fungi, etc.) that are parasites of honeybees, and the others in *A. cerana cerana* may represent some special genes that can only be obtained through mapping of the *A. cerana* genome. Given that genomic information for *A. cerana cerana* remains scarce, there exist few effects in analyzing the differential abundance of common genes in *A. mellifera ligustica* and *A. cerana cerana* based on the *A. mellifera* genome. Because there exist recognizable homologues to a considerable degree, many conserved genes also exist in most living cells (The Gene Ontology Consortium, 2000); therefore, genome

information from closely related species can be used as a scaffold to investigate gene expression profiles if there are no reports on genome information for the species. However, only a few genes detected in our study were clearly annotated because of limited EST and cDNA data for *Apis*, which prevented us from making further generalizations on differences in gene expression between *A. cerana cerana* and *A. mellifera ligustica*. A large number of genes reported in other organisms (fruit fly, nematode, etc.) were also found in our study by searching the GenBank data. The results indicated that there are a great many genes in the honeybee that require further characterization.

ACKNOWLEDGEMENTS

The research was supported by earmarked funds for Modern Agro-industry Technology Research System (no. nycytx-43-kxj4), 863 Project of National Science and Technology Ministry of China (no. 2007AA10Z332), and Nature and Science Foundation Commission of Zhejiang Province (R3080306) to SKS. The deep sequencing and bioinformation analysis work were carried out in the Beijing Genome Institute (http://www.genomics.cn/en/index.php). We thank Shuangjin Dai for his beekeeping and Yunjuan Hu and Qin Chen for their help in the research and Dr. Aung Si for his helpful suggestions that improved the manuscript. We thank the two anonymous referees for the constructive comments, which helped improve the manuscript.

Profils transcriptomiques différentiels de la tête des butineuses : comparaison entre *Apis mellifera ligustica* et *Apis cerana cerana*

Comportement d'approvisionnement / séquençage Illumina-Solexa / expression génique différentielle.

Differentielle Transkriptomprofile aus Köpfen von Arbeiterinnen: ein Vergleich zwischen *Apis mellifera ligustica* und *Apis cerana cerana.*

Sammelverhalten / Illumina-Solexa Sequenzierung / differentielle Genexpression

REFERENCES

't Hoen, P.A.C., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H.A.M., de Menezes, R.X., Boer, J.M., van Ommen, G.J.B., den Dunnen, J.T. (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. **36**, e141

Albert, S., Klaudiny, J. (2004) The MRJP/YELLOW protein family of *Apis mellifera*: identification of new members in the EST library. J. Insect Physiol. **50**, 151–159

Audic, S., Claverie, J.M. (1997) The significance of digital gene expression profiles. Genome Res. **7**, 986–995

Beiter, T., Reich, E., Williams, R.W., Simon, P. (2009) Antisense transcription: a critical look in both directions. Cell. Mol. Life Sci. **66**, 94–112

Benjamini, Y., Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. Ann. Stat. **29**, 1165–1188

Ben-Shahar, Y., Robichon, A., Pak, W.L., Sokolowski, M.B., Robinson, G.E. (2002) Influence of gene action across different time scales on behavior. Science **296**, 741–744

Ben-Shahar, Y., Leung, H., Pak, W.L., Sokolowski, M. B., Robinson, G.E. (2003) cGMP-dependent changes in phototaxis: a possible role for the foraging gene in honeybee division of labor. J. Exp. Biol. **206**, 2507–2515

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., et al. (2005) The transcriptional landscape of the mammalian genome. Science **309**, 1559–1563

Chen, S.L. (2001) The Apiculture Science in China. Agricultural Publishing House of China, Beijing

Cloonan, N., Grimmond, S.M. (2008) Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol. **9**, 234

Drapeau, M.D., Albert, S., Kucharski, R., Prusko, C., Maleszka, R. (2006) Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honeybees. Genome Res. **16**, 1385–1394

Fries, I., Wei, H.Z., Shi, W., Chen, S.J. (1996) Grooming behavior and damaged mites (*Varroa jacobsoni*) in *Apis cerana cerana* and *Apis mellifera ligustica*. Apidologie **27**, 3–11

Garcia, L., Garcia, C.H.S., Calábria, L.K., Cruz, G.C.N., Puentes, A.S., Báo, S.N., Fontes, W., Ricart, C.A.O., Espindola, F.S., Sousa, M.V. (2009) Proteomic analysis of honeybee brain upon ontogenetic and behavioral development. J. Proteome Res. **8**, 1464–1473

Hall, N. (2007) Advanced sequencing technologies and their wider impact in microbiology. J. Exp. Biol. **210**, 1518–1525

Hegedùs, Z., Zakrzewska, A., Ágoston, V.C., Ordas, A., Racz, P., Mink, M., Spaink, H.P., Meijer, A.H.

(2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. Mol. Immunol. **46**, 2918–2930

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. Nucl. Acids. Res. **34**, D354–D357

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., et al. (2005) Antisense transcription in the mammalian transcription. Science **309**, 1564–1566

Klaudiny, J., Albert, S., Bachanová, K., Kopernický, J., Šimúth, J. (2005) Two structurally different defensin genes, one of them encoding a novel defensin isoform, are expressed in honeybee *Apis mellifera*, Isect Biochem. Mol. Biol. **35**, 11–22

Kucharski, R., Maleszka, R. (2002) Evaluation of differential gene expression during behavioral development in the honeybee using microarrays and northern blots. Genome Biol. **3**, research0007.1–research0007.9

Kucharski, R., Maleska, R., Hayward, D.C., Ball, E. E. (1998) A royal jelly protein is expressed in a subset of Kenyon cells in the mushroom bodies of the honeybee brain. Naturwissenschaften **85**, 343–346

Li, R., Li, Y., Kristiansen, K., Wang, J. (2008) SOAP: short oligonucleotide alignment program. Bioinformatics **24**, 713–714

Liu, F., Li, W.F., Li, Z.G., Zhang, S.W., Chen, S.L., Su, S.K. (2011) High-abundance mRNAs in *Apis mellifera*: comparison between nurses and foragers. J. Insect Physiol. **57**, 274–279

Mardis, E.R. (2008) The impact of next-generation sequence technology on genetics. Trends Genet. **24**, 133–141

Morozova, O., Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics **92**, 255–264

Morrissy, A.S., Morin, R.D., Delaney, A. (2009) Next-generation tag sequencing for cancer gene expression profiling. Genome Res. **19**, 1825–1835

Nelson, C.M., Ihle, K.E., Fondrk, M.K., Page, R.E., Amdam, G.V. (2007) The gene vitellogenin has multiple coordinating effects on social organization. PLoS Biol. **5**, e62

Neumann, P., Carreck, N.L. (2010) Honey bee colony losses. J. Apic. Res. **49**, 1–6

Patel, A., Fondrk, M.K., Kaftanoglu, O., Emore, C., Hunt, G., Frederick, K., Amdam, G.V. (2007) The making of a queen: TOR pathway is a key player in diphenic caste development. PLoS One **6**, e509

Peixoto, L.G., Calábria, L.K., Garcia, L., Capparelli, F. E., Goulart, L.R., Sousa, M.V., Espindola, F.S. (2009) Identification of major royal jelly proteins in the brain of the honeybee *Apis mellifera*. J. Insect Physiol. **55**, 671–677

Robinson, G.E. (1992) Regulation of division of labor in insect societies. Annu. Rev. Entomol. **37**, 637–665

Robinson, G.E., Fahrbach, S.E., Winston, M.L. (1997) Insect societies and the molecular biology of social behavior. Bioessays **19**, 1099–1108

Rubin, E.B., Shemesh, Y., Cohen, M., Elgavish, S., Robertson, H.M., Bloch, G. (2006) Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honeybee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. Genome Res. **16**, 1352–1365

Schippers, M.P., Dukas, R., Smith, R.W., Wang, W., Smolen, K., McClelland, G.B. (2006) Lifetime performance in foraging honeybees: behaviour and physiology. J. Exp. Biol. **209**, 3828–3836

Seeley, T.D. (1995) The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies. Harvard University Press, Cambridge

Stanfel, M.N., Shamieh, L.S., Kaeberlein, M., Kennedy, B. (2009) The TOR pathway comes of age. BBA – General Subjects **1790**, 1067–1074

Su, S.K., Albert, S., Chen, S.L., Zhong, B.X. (2005) Molecular cloning and analysis of four cDNAs from the heads of *Apis cerana cerana* nurse honeybees coding for major royal jelly proteins. Apidologie **36**, 389–401

Tan, K., Radloff, S.E., Li, J.J., Hepburn, H.R., Yang, M.X., Zhang, L.J., Neumann, P. (2007) Bee-hawking by the wasp, *Vespa velutina*, on the honeybees *Apis cerana* and *A. mellifera*. Naturwissenschaften **94**, 469–472

Toma, D.P., Bloch, G., Moore, D., Robinson, G.E. (2000) Changes in period mRNA levels in the brain and division of labor in honey bee colonies. PNAS **97**, 6914–6919

Velarde, R.A., Sauer, C.D., Walden, K.K., Fahrbach, S.E., Robertson, H.M. (2005) Pteropsin: a vertebrate-like non-visual opsin expressed in the honeybee brain. Insect Biochem. Mol. Biol. **35**, 1367–1377

von Frisch, K. (1967) The Dance Language and Orientation of Bees. Harvard University Press, Cambridge

Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. **10**, 57–63

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., et al. (2006) Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature **443**, 931–949

Wilhelm, B.T., Landry, J.R. (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods **48**, 249–257

Winston, M.L. (1987) The Biology of the Honey Bee. Harvard University Press, Cambridge

Wu, T., Qin, Z.W., Zhou, X.Y., Feng, Z., Du, Y.L. (2010) Transcriptome profile analysis of floral sex determination in cucumber. J. Plant Physiol. **167**, 905–913

Xu, P., Atkinson, R., Jones, D.N., Smith, D.P. (2005) *Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons. Neuron **45**, 193–200

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K., Rotman, G. (2003) Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. **21**, 379–386