IJIE Int. J. Indust. Entomol. 27(2) 271-276 (2013)
ISSN 1598-3579, http://dx.doi.org/10.7852/ijie.2013.27.2.271

Comparative Transcriptome Analysis of Queen, Worker, and Larva of Asian Honeybee, *Apis cerana*

Woo Jin Kim[1], Seok Hee Lee[1], Saes Byeol An[1], Song Eun Kim[1], Qin Liu[1], Jae Young Choi[2], and Yeon Ho Je[12]*

[1]*Department of Agricultural Biotechnology, College of Agriculture and Life Science, Seoul National University, Seoul, Korea*
[2]*Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Korea*

Abstract

The Asian honeybee, *Apis cerana*, is a native honeybee species in Korea which is important in agriculture for pollination and honey production. For better understanding of the physiology of *A. cerana*, high-throughput Illumina transcriptome sequencing was performed to analyze the gene expression profiles of queen, worker, and larva. A total of 219,799,682 clean reads corresponding to 22.2 Gb of nucleotide sequences was obtained from the whole body total RNA samples. The *Apis mellifera* reference mRNA sequence database was used to measure the gene expression level with Bowtie2 and eXpress software, and the Illumina short reads were then mapped to 11,459 out of 11,736 *A. mellifera* reference genes. Total of 9,221 genes with FPKM value greater than 5 of each sample group were subjected to eggNOG with BLASTX for gene ontology analysis. The differential gene expression between queen and worker, and worker and larva were analyzed to screen the overexpressed genes in each sample group. In the queen and worker sample group, total of 1,766 genes were differentially expressed with 887 and 879 genes overexpressed over two folds in queen and worker, respectively. In the worker and larva sample group, total of 1,410 genes were differentially expressed with 1,009 and 401 genes overexpressed over two folds in worker and larva, respectively.

Int. J. Indust. Entomol. 27(2), 271-276 (2013)

Received : 5 Dec 2013
Accepted : 20 Dec 2013

Keywords:
Apis cerana,
Transcriptome,
Illumina,
Developmental stage,
Caste

Introduction

Apis cerana, the Asian honeybee is a native bee species to eastern and southeastern Asian countries such as Korea, China, and Japan. In Korea, *A. cerana* is one of the most important honeybee species along with western honeybee, *Apis mellifera* because not only of the economic importance of honey production, but also the importance as one of the pollinator species in agriculture. Also *A. cerana* had been concerned of its strong resistance to the ectoparasitic mites (Peng *et al*., 1987). However, the recent outbreak of sacbrood virus (SBV) belonging to the genus Iflavirus which infects *A. cerana* (Choi *et al*., 2010) caused a devastating colony loss of Korean honeybee industry.

SBV infected larva fails to pupate, and accumulates virus enriched ecdysal fluid beneath its unshed skin (Bailey *et al*., 1964). SBV also infects adult bees, however, the mortality

***Corresponding author.**
Yeon Ho Je
Department of Agricultural Biotechnology, College of Agriculture and Life Science, Seoul National University, Seoul, 151-921 Korea
Tel: +82-2-880-4706 / FAX: +82-2-873-2319
E-mail: btrus@snu.ac.kr

of SBV infected adults is low while SBV results significant larval mortality (Bailey and Fernando, 1972) which suggests that the responses of the hosts against SBV might be different to each other according to their developmental stage and/or caste. Recently, the SBV isolated from *A. cerana* in Korea was sequenced, and characterized by molecular genomics and phylogenetic approach to reveal that the SBV found in Korea is different from other strains previously published (Choe *et al*., 2012a, 2012b), however, the molecular biology and nucleotide sequence information of *A. cerana*, which is very important to study the virus and host interaction, is still limited. Currently, *A .cerana* genome sequence is not available yet, and there are just 115 ESTs and 7,009 amino acid sequences of *A. cerana* are available in the NCBI database which are only 0.2% and 17% of *A. mellifera* based on the number of sequence entries. For comprehensive understanding of the virus and host interaction, accumulation of nucleotide sequence data, and study of functional genomics of *A. cerana* is indispensable.

Illumina sequencing is a recently developed technology which can produce hundreds of millions of short reads from DNA or cDNA sample (Bentley *et al*., 2008). The high sequencing capacity of Illumina sequencing makes possible tagging of the gene or genome sequences by aligning the shorts reads of transcriptome to reference sequences to measure the gene expression level by counting the number of tags produced from each gene (Mardis, 2008; Velculescu *et al*., 1995). In this study, we employed the Illumina sequencing technology to sequence cDNA samples of each developmental stages of *A. cerana*, and mapped the short reads to the closely related species, *A. mellifera* reference genes to search and measure stage and caste specific gene transcripts and their transcription levels for studying the candidate genes responsible to their physiological differences.

Materials and Methods

Insect

The *Apis cerana* samples were collected from the beehive of honeybee farm at Cheongwon-gun, Chungcheongbuk-do, Korea. The honeybee samples were collected in mid-March 2013, therefore the workers were virtually all winter nurse bees, no forager. Newly laid eggs were found in the bee hive which indicate that the queen was actively laying eggs. One queen, five of workers and approximately seven days old larvae were sampled per group for RNA isolation, respectively. The compound eyes, wings, and legs of the queen and worker samples were removed beforehand, and then the total RNA samples were extracted from whole body of each sample for sequencing.

cDNA library preparation and Illumina transcriptome sequencing

To construct a cDNA library, the total RNA was isolated using Qiazol lysis reagent (Qiagen, Germany) according to the manufacture's protocol. Five of the total RNA samples from the workers and larvae groups were pooled equal amount as one sample of 10 μg for transcriptome sequencing, respectively. The integrity of total RNA was examined by using 2100 Bioanalyzer (Agilent, USA) with RNA 6000 nano chip. The mRNA in total RNA was converted into a library of template molecules suitable for subsequent cluster generation using the reagents provided in the Illumina TruSeq RNA Sample Preparation Kit. The first step in the workflow involves purifying the poly-A containing mRNA molecules using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers. This was followed by second strand cDNA synthesis using DNA polymerase I and RNase H. These cDNA fragments then went through an end repair process, the addition of a single 'A' base, and ligation of the adapters. The products were then purified and enriched with PCR to create the final cDNA library. Finally the cDNA library of 250~300 bp insert was subjected to Illumina HiSeq 2000 sequencer to sequence the 101 bases of 5' end.

Analysis of transcriptome sequencing results

The raw Illumina sequence reads were filtered to remove low quality sequences by using NGS QC Toolkit v2.3 (Patel and Jain, 2012) before bioinformatical analysis. The sequence reads of base called with error rate higher than 0.1% (quality score<30) were eliminated, and the remaining clean reads were used for further analysis. The filtered Illumina short reads were then mapped to *A. mellifera* reference mRNA sequences (ftp://ftp.ncbi.nih.gov/

Table 1. Summary of the *A. cerana* transcriptome sequencing

	Queen	Worker	Larva
Total sequence reads	82,069,132	79,919,960	80,424,924
Total clean reads	74,905,753	71,548,307	73,345,622
Total nucleotides of clean reads	7,565,481,053	7,226,379,007	7,407,907,822
Q30 percentage	92.37%	91.03%	92.20%
GC percentage	34.67%	39.10%	37.33%

genomes/Apis_mellifera/RNA) from the honey bee assembly v4.5 of BeeBase (Munoz-Torres *et al*., 2011) using Bowtie2 software (Langmead and Salzberg, 2012) with default parameters. The mapping results of each sample to reference mRNA sequences obtained by Bowtie2 were then quantified by eXpress software (Roberts and Pachter, 2013). The gene expression profiles of each sample obtained by eXpress were calculated as fragments per kilobase of exon per million reads (FPKM) value to compare the gene expression levels between samples.

Gene ontology analysis

The reference mRNA sequences which were mapped with the Illumina short reads by Bowtie2 software were subjected to eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) (Jensen *et al*., 2008) by BLASTX with a cut-off E-value of 10^{-5} for gene ontology (GO) anlaysis. The returned mRNA sequences above cut-off score were annotated and categorized for further analysis.

Results and Discussion

Transcriptome sequencing

The Illumina sequencing generated a total of 242,414,016 single-end raw reads, and a total of 219,799,682 reads which was 90.67% of the raw reads was obtained after Q30 (sequencing error rate, 0.1%) filtration for further analysis. The accumulated length of the total filtered reads was 22,199,767,882 bases with GC percentage of 37.0%. The detailed transcriptome sequencing results of each sample group are summarized in Table 1. The raw Illumina sequencing results were submitted to NABIC (National Agrucultural Biotechnology Information Center, Rural Development Administration, Korea) NGS SRA database. The accession numbers for the transcriptome of Queen, workers, and larvae are NN0646, NN0650, and NN0648, respectively.

Table 2. Detailed statistics of the gene expression levels of each sample

	Queen	Worker	Larva
Max FPKM	328,145.853	324,600.660	1,393,940.143
Min FPKM	0.0095	0.0098	0.0023
Avg FPKM	226.5836	248.9488	339.8804
Median FPKM	21.6844	20.1978	14.8157
No. of mapped genes	10,197	10,129	10,090

Analysis of gene expression level

The Illumina short reads of each *A. cerana* sample were mapped to *A. mellifera* reference mRNA sequences by Bowtie2 software. Total of 11,459 out of 11,736 *A. mellifera* reference sequences were mapped with the short reads of *A. cerana* transcriptome, and the expression levels of each gene were converted to FPKM value by eXpress software. Detailed statistics of the gene expression levels of each sample group are summarized in Table 2. The larva sample includes genes with remarkably higher FPKM value than those of queen and worker samples, and those genes were ribosomal protein genes which are part of the essential cellular machinery for protein synthesis, development, and growth.

Gene ontology analysis of differentially expressed genes

To compare the gene expression profiles, the FPKM values

of each gene were then arithmetically normalized based on the ratio of the total numbers of Q30-filtered short reads between sample groups. The genes which were overexpressed over two-fold between the two sample groups were isolated, and the sum of FPKM value of each gene was calculated. A total of 9,221 genes with sum of FPKM value greater than 5 were subjected to eggNOG analysis. The eggNOG functional groups and their abbreviations are as follows; [Q] Secondary metabolites biosynthesis, transport and catabolism, [P] Inorganic ion transport and metabolism, [I] Lipid transport and metabolism, [H] Coenzyme transport and metabolism, [F] Nucleotide transport and metabolism, [E] Amino acid transport and metabolism, [G] Carbohydrate transport and metabolism, [C] Energy production and conversion, [O] Posttranslational modification, protein turnover, chaperones, [U] Intracellular trafficking, secretion, and vesicular transport, [W] Extracellular structures, [Z] Cytoskeleton, [N] Cell motility, [M] Cell wall/membrane/ envelope biogenesis, [T] Signal transduction mechanisms, [V] Defense mechanisms, [Y] Nuclear structure, [D] Cell cycle

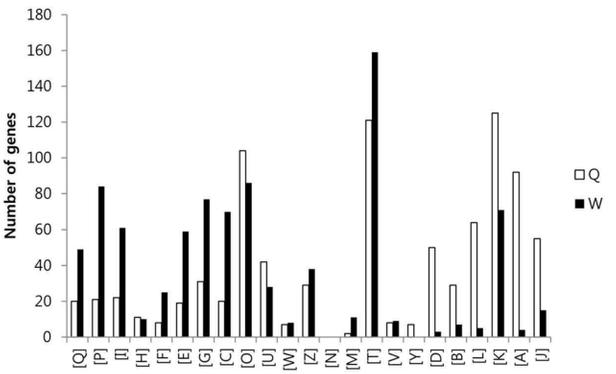

Fig. 1. Graph of eggNOG analysis of the genes which were overexpressed over two-fold in queen and worker. The number of overexpressed genes in the queen and worker of each GO group are indicated as white and black, respectively.

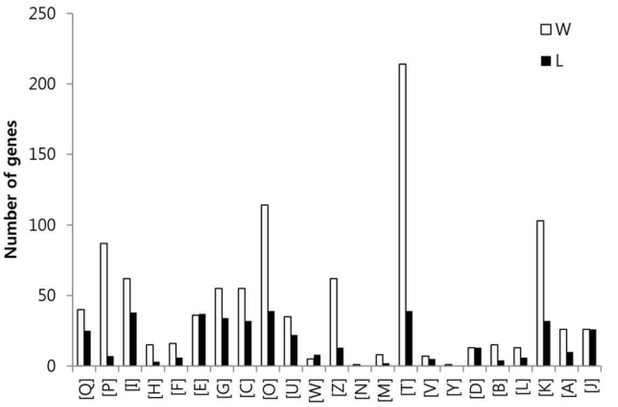

Fig. 2. Graph of eggNOG anlaysis of the genes which were overexpressed over two-fold in worker and larva. The number of overexpressed genes in the worker and larva of each GO group are indicated as white and black, respectively.

Table 3. The differentially express genes over 500 folds in queen and worker.

A. mellifera Ref gene ID	UniprotKB accession no.	Gene function	eggNOG category	FPKM (Queen)	FPKM (Worker)	Q/W ratio
CL10988Contig1	XP_394161	Orthodenticle-1 protein	K	183.198	0.069	2659.336
CL8765Contig1	XP_394847	Gooseberry protein	K	18.735	0.011	1766.998
CL52367Contig1	XP_001119912	Transcription factor	K	66.151	0.041	1625.326
CL8225Contig1	XP_001122398	Protein involved in genital disc development	K	371.386	0.425	873.237
CL47583Contig1	XP_001601536	Protein involved in regulation of transcription from RNApolymerase II promoter	K	474.849	0.681	697.110
CL46357Contig1	XP_624026	Cytochrome P450 protein	Q	85.404	0.123	692.136
CL52111Contig1	Q17061	Major royal jelly protein	S	2.074	4237.383	0.001
CL23420Contig1	XP_394370	Serine-Type enodpeptidase	O	4.599	6503.741	0.000
CL48693Contig1	XP_001121120	Cysteine-Rich venom protein	S	0.520	683.745	0.001
CL49002Contig1	XP_001121120	Cysteine-Rich venom protein	S	0.640	400.828	0.002
CL778Contig2	Q9U8X6	Glucose oxidase	C	0.194	119.962	0.002

Table 4. The differentially express genes over 500 folds in worker and larva.

A. mellifera Ref gene ID	UniprotKB accession no.	Gene function	eggNOG category	FPKM (Worker)	FPKM (Larva)	W/L ratio
CL778Contig2	Q9U8X6	Glucose oxidase	C	119.962	0.010	11811.665
CL21074Contig1	O97432	Major royal jelly protein	S	11022.810	1.227	8980.947
CL9379Contig1	Q25BT8	Alpha-Glucosidase	G	3070.006	0.616	4980.110
CL8792Contig1	XP_397087	Serine-Type endopeptidase	O	555.582	0.128	4330.956
CL14122Contig1	Q17053	Rhodopsin protein	T	259.021	0.209	1239.556
CL21265Contig1	Q17060	Major royal jelly protein	S	257.226	0.263	978.909
CL36969Contig1	NP_001011651	Troponin C protein	T	264.119	0.282	937.569
CL50642Contig1	Q1W647	Odorant binding protein 3	P	400.704	0.456	879.473
CL29065Contig1	O77061	Major royal jelly protein	S	1169.244	1.339	873.426
CL12073Contig1	XP_001122057	Leucine-Rich repeats	S	403.929	0.578	699.182
CL42234Contig1	XP_395084	Neuropeptide Y receptor	T	0.191	1385.589	0.000
CL5768Contig1	A6YLP8	Hexamerin protein	S	1.329	2487.960	0.001
CL46982Contig1	XP_001121128	Structural constituent of cuticle protein	S	5.523	6784.944	0.001
CL11903Contig1	XP_397272	Cytochrome P450 protein	I	0.036	37.487	0.001
CL14883Contig1	Q6J4Q1	Hexamerin protein	S	7.903	6855.978	0.001
CL38195Contig1	Q1W641	Odorant-Binding protein	S	0.936	491.532	0.002

control, cell division, chromosome partitioning, [B] Chromatin structure and dynamics, [L] Replication, recombination and repair, [K] Transcription, [A] RNA processing and modification, and [J] Translation, ribosomal structure and biogenesis.

In the comparison of queen and worker, total of 3,014 and 3,009 eggNOG tagged genes from queen and worker were compared their expression levels, and identified total of 887 (22.7%) and 879 (22.5%) genes which were overexpressed over two folds, respectively. In the comparison of worker and larva, total of 3,492 and 2,884 eggNOG tagged genes from worker and larva were compared, respectively, and identified total of 1,009 (25.9%) and 401 (10.3%) genes which were overexpressed over two folds, respectively.

The GO distributions demonstrated that the genes related to nutritional metabolism were overexpressed in the worker while the genes related to cell division and nucleic acid processing were overexpressed in the queen. These results reflect the role of each caste which are honey and royal jelly production, and reproduction of worker and queen, respectively (Fig. 1). On the other hand, the genes related to transcription, translation, and transduction were overexpressed in the larva which reflect the highly active genes of developing larvae (Fig. 2). The genes significantly overexpressed over 500-fold are summarized in Table 3 and 4.

Acknowledgment

This work was supported by a grant from the Next-Generation BioGreen 21 Program (no.PJ009031), Rural Development Administration, Republic of Korea. Seok Hee Lee, Saes Byeol An, Qin Liu, and Song Eun Kim were supported by the Brain Korea 21 project.

Reference

Bailey L, Fernando EFW (1972) Effects of sacbrood virus on adult

honey-bees. Ann. Appl. Biol. 72, 27-35.

Bailey L, Gibbs AJ, Woods RD (1964) Sacbrood virus of the larval honey bee (*Apis mellifera* linnaeus). Virology 23, 425-429.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, *et al*. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53-59.

Choe SE, Nguyen LTK, Noh JH, Kweon CH, Reddy KE, Koh HB, Chang KY, Kang SW (2012a) Analysis of the complete genome sequence of two Korean sacbrood viruses in the Honey bee, *Apis mellifera*. Virology 432, 155-161.

Choe S-E, Nguyen TT-D, Hyun B-H, Noh J-H, Lee H-S, Lee C-H, Kang S-W (2012b) Genetic and phylogenetic analysis of South Korean sacbrood virus isolates from infected honey bees (*Apis cerana*). Veterinary Microbiology 157, 32-40.

Choi YS, Lee MY, Hong IP, Kim NS, Kim HK, Lee KG, Lee ML (2010) Occurrence of sacbrood virus in Korean apiaries from *Apis cerana* (Hymenoptera: Apidae). Korean Journal of Apiculture 25, 187-191.

Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork, P. (2008) eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res 36, D250-D254.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Meth 9, 357-359.

Mardis ER (2008) Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics 9, 387-402.

Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, Childs KL, Anzola JM, Milshina N, Elsik CG (2011) Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucl. Acids Res. 39, D658-D662.

Patel RK, Jain M (2012) NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. PLoS ONE 7, e30619.

Peng Y-S, Fang Y, Xu S, Ge L (1987) The resistance mechanism of the Asian honey bee, *Apis cerana* Fabr., to an ectoparasitic mite, Varroa jacobsoni Oudemans. Journal of Invertebrate Pathology 49, 54-60.

Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Meth 10, 71-73.

Velculescu V.E, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270, 484-487.