

## RESEARCH ARTICLE

## Open Access

# Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing

Doori Park<sup>1</sup>, Je Won Jung<sup>1</sup>, Beom-Soon Choi<sup>2</sup>, Murukarthick Jayakodi<sup>3</sup>, Jeongsoo Lee<sup>2</sup>, Jongsung Lim<sup>2</sup>, Yeisoo Yu<sup>4</sup>, Yong-Soo Choi<sup>5</sup>, Myeong-Lyeol Lee<sup>5</sup>, Yoonseong Park<sup>6</sup>, Ik-Young Choi<sup>2</sup>, Tae-Jin Yang<sup>3</sup>, Owain R Edwards<sup>7</sup>, Gyoungju Nah<sup>3\*</sup> and Hyung Wook Kwon<sup>1\*</sup>

## Abstract

**Background:** The honey bee is an important model system for increasing understanding of molecular and neural mechanisms underlying social behaviors relevant to the agricultural industry and basic science. The western honey bee, *Apis mellifera*, has served as a model species, and its genome sequence has been published. In contrast, the genome of the Asian honey bee, *Apis cerana*, has not yet been sequenced. *A. cerana* has been raised in Asian countries for thousands of years and has brought considerable economic benefits to the apicultural industry. *A. cerana* has divergent biological traits compared to *A. mellifera* and it has played a key role in maintaining biodiversity in eastern and southern Asia. Here we report the first whole genome sequence of *A. cerana*.

**Results:** Using *de novo* assembly methods, we produced a 238 Mbp draft of the *A. cerana* genome and generated 10,651 genes. *A. cerana*-specific genes were analyzed to better understand the novel characteristics of this honey bee species. Seventy-two percent of the *A. cerana*-specific genes had more than one GO term, and 1,696 enzymes were categorized into 125 pathways. Genes involved in chemoreception and immunity were carefully identified and compared to those from other sequenced insect models. These included 10 gustatory receptors, 119 odorant receptors, 10 ionotropic receptors, and 160 immune-related genes.

**Conclusions:** This first report of the whole genome sequence of *A. cerana* provides resources for comparative sociogenomics, especially in the field of social insect communication. These important tools will contribute to a better understanding of the complex behaviors and natural biology of the Asian honey bee and to anticipate its future evolutionary trajectory.

**Keywords:** *Apis cerana*, Asian honey bee, Genome, Social insect, Chemosensory receptors, Honey bee immunity

## Background

The honey bee, a social insect, has received considerable attention as a model for studying neurobiology [1], development [2], social behavior [3] and, most recently, epigenomics [4]. To facilitate studies of this important biological model, the genomic sequence of the western honey bee, *Apis mellifera*, was published in 2006,

providing a wealth of information for understanding the molecular basis of social behavior and eusocial evolution [5].

Honey bees are classified in the order Hymenoptera, which also includes ants, bees, sawflies, and wasps [6]. The genus *Apis* consists of eight Asian species and one western species [7]. Ancient bees first emerged 120–130 million years ago (mya) coincident with the emergence of early angiosperms [8]. The ancestors of modern bees were geographically isolated in the Middle East during the late Pleistocene approximately 1–2 mya. It was from this population that the ancestral lineage of *A. mellifera* spread to Africa, and the ancestral lineages of *A. cerana*

\* Correspondence: gjnah@snu.ac.kr; biomodeling@snu.ac.kr

<sup>3</sup>Department of Plant Science, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea

<sup>1</sup>Biomodulation Major, Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea  
Full list of author information is available at the end of the article

expanded throughout Europe and Asia [7,9,10]. As a result of allopatric speciation, these two honey bee species may have evolved in different ecological environments, which gave rise to different behavioral and physiological characteristics observed in the present day [7].

Compared to *A. mellifera*, the Asian honey bee, *A. cerana*, has several distinguishing behavioral traits (Table 1). It is able to adapt to extreme weather conditions [11] and has long flight duration [7], effective grooming and hygienic behaviors [12], and cooperative group-level defenses [13]. A well-known behavior of *A. cerana* is aggregation when a colony is exposed to dangers, such as predators or intruders. In these situations, guard bees produce alarm pheromones that dictate group behavior [13-15]. In addition, *A. cerana* provides considerable economic benefits to the apicultural industry through its high quality bi-products, perhaps even more so than *A. mellifera* [16].

In recent years, population decreases similar to those documented in the western honey bee have also been seen in *A. cerana* colonies [25]. In Korea, more than ninety percent of *A. cerana* colonies were destroyed by sacbrood virus (SBV) infection [26]. However, few studies have been

conducted on the underlying molecular mechanisms and immune responses to this virus [11].

This study reports the *A. cerana* genome produced with deep sequence coverage using next-generation technologies. We generated gene sets of *A. cerana* using transcriptome data from seven tissues. Then, we focused on the characterization of important genes related to chemosensory reception and immunity. This genome sequence will provide invaluable information on the novel characteristics of the honey bee species indigenous to eastern and Southern Asia. The data will also provide a resource for comparative sociogenomic studies with the seven ants and western honey bee species for which genomes are already available [5,27-30].

## Results and discussion

### Genomic features of *A. cerana*

#### Sequencing and assembly

We performed whole genome sequencing of the Asian honey bee using seven drones derived from a single colony. Because the honey bee has a haplodiploid mating system, males (drones) are haploid and females (workers and

**Table 1 Comparison of biological differences between *A. cerana* and *A. mellifera***

	<i>A. cerana</i>	<i>A. mellifera</i>	Reference
Aria of origin	Southeast Asia	Africa or Western Asia	
Body mass	43.8 mm	77.2 mm	[17]
Wing length (worker)	7.54 ± 1.14 mm	9.32 ± 0.16 mm	[18]
Wing beat frequencies (worker)	306 beats/s	235 beats/s	
Nest cavity volume	10-15 liters in general	35 liters,	[19]
Comb structure	Uneven round	Square	
Dismantling of old comb	Yes	No	[4]
Pore in the drone cell	Yes	No	[20]
Flight pattern	Rapid, hasty, and unpredictably zig-zag	Steady and clumsy	[4]
Homing speed (~50 m)	192 sec	295 sec	[21]
Foraging range	Maximum 1500 m to 2500 m	Average 1650 m, maximum 6 km	[22]
Active foraging time and temperature range	9 am and 11.30 am/15.5 - 21°C	11 am and 1 pm/21-25°C	[23]
Flower preference	Wide range of flowers, including wild plants	Mainly on <i>Trifolium</i> and <i>Brassica</i>	[24]
Pollination	Larger pollen collector	Smaller pollen collector (compared to <i>A. cerana</i> )	[10]
Robbing defense	Weak	Strong	[10]
Frequencies of absconding	Often (heavily depends on condition)	Rarely (heavily depends on condition)	[8]
Collecting propolis	No	Yes	[4]
Defense to wasps	Forming a ball with worker	Individual	[11,12]
<i>Varroa</i> mite resistance	Yes	No	[13]
Ventilation direction	Head toward outside	Head toward entrance	[7,12]
Royal jelly production rate	3.21 ± 0.43 g	80.5 ± 7.83 g	[23]
Rate of stinging	Low	High	[4]

queens) are diploid. To minimize possible contamination from foreign genomes such as bacteria and viruses, we eliminated mid-gut tissues from the individual drone bees prior to sequencing. Genomic sequence libraries were constructed with a combination of short reads (500 bp) and two longer insert libraries (3 and 10 Kb), using Illumina sequencing technology (152-fold coverage) (Table 2). The assembly consisted of 2,430 scaffolds with a total length of 228 Mb, which covered 96% of the estimated genome size (238 Mb) [31]. General information on the genome assembly is presented in Table 3. The N50 scaffold size was 1,421 kb (Table 3), much longer than the N50 scaffold size found in the initial and recently improved assemblies of *A. mellifera* (359 kb and 997 kb, Amel\_4.0 and Amel\_4.5, respectively; Additional file 1: Table S1) [32]. To assess the accuracy of the scaffolds, we compared the genome of *A. mellifera* and *A. cerana* to identify genomic synteny (Additional file 1: Figure S1). Results revealed several scaffolds of *A. cerana* and chromosome 3 of *A. mellifera* that showed syntenic relationships with no large-scale rearrangement. Additionally, we found that the mitochondrial genome of *A. mellifera* (NCBI GQ162109, [33]) and one contig of *A. cerana* had high sequence similarity, ~99% (Additional file 1: Figure S2). This contig, covering the entire *A. cerana* mitochondrial genome, is 15,915 bp and includes 13 protein-coding genes (Additional file 1: Figure S3). All sequence information was submitted in the NCBI [Submission number: SUB582139].

#### Guanine plus cytosine (GC) content

The *A. cerana* assembly contains 30% GC (Table 3), similar to the mean GC content of *A. mellifera* (33%). In addition, six ant species (*Linepithema humile*, *Camponotus floridanus*, *Pogonomyrmex barbatus*, *Solenopsis invicta*, *Atta cephalotes*, and *Acromyrmex echinatior*) have similar GC contents ranging from 33% to 38% [5,27-30,34,35]. In contrast, *Drosophila melanogaster* (42%), *Nasonia vitripennis* (42%), and *Harpegnathos saltator* (45%) have higher GC contents compared to *A. cerana* [34,36]. According to comparative studies of two ant species, *C. floridanus* and *H. saltator*, organisms with more complex social traits may have AT-biased genomes [30,34].

Relative AT bias correlates with DNA methylation, as DNA methyltransferases (*Dmmts*) are almost entirely targeted to cytosine residues followed by guanines in the 5' to 3' orientation (CpG dinucleotides). Methylcytosine tends to mutate to thymine (T), thus the gradual accumulation of mutation that convert CpG dinucleotides to TpG dinucleotides lead to AT-rich genomes [37-39]. In particular, normalized CpG observed/expected (CpG o/e) values have a negative relationship with levels of DNA methylation [40]. DNA methylation is one of the major part of epigenetic regulation and has functional roles in gene expression regulation in vertebrates and insects [40]. In contrast to vertebrate genomes, which are depleted of CpG dinucleotides [41], most hymenopteran insects, including *A. cerana* (1.61), *A. mellifera* (1.65), *C. floridanus* (1.58), *H. saltator* (1.49), and *N. vitripennis* (1.35), exhibit high levels of CpG o/e in their genomes [34]. Another intriguing discovery is that normalized CpG o/e value within protein coding sequences of *A. cerana* showed bimodal distribution, similar to *A. mellifera* (Figure 1, Additional file 1: Figure S4) and the pea aphid *Acyrtosiphon pisum* [5,42]. Interestingly, two distinct classes of genes are documented to perform different functions [43], which low-CpG genes are mainly involved in housekeeping function and high-CpG genes are involved in development. Indeed, we found that genes that are represented low-CpG classes are categorized with metabolic process, and transcriptional and translational regulation. In contrast, high-CpG genes represented GO categories that specific to biological functions.

The genome of *A. cerana*, *A. mellifera*, and *A. pisum* encode full complements of DNA methylation proteins (*Dmmts*) [5,42], but, according to a recent discovery, several insects possess a full set of *Dmmts* without any striking depletion pattern of coding exons [27,36,44]. Thus, this genomic feature may not be species specific, but mechanisms of epigenetic regulation may be conserved in both honey bee species.

#### Repetitive elements

The *A. cerana* assembly comprised 6.48% (14.79 Mb) repetitive elements, consisting of 3.58% (8.16 Mb) simple repeats and 1.95% (4.44 Mb) interspersed repeats (Additional file 1: Table S2). Seventy-five *A. cerana*-

**Table 2 Sequencing raw data summary**

Library	Raw data		Filtered data		Sequence coverage (X)
	Number of reads	Length (bp)	Number of reads	Length (bp)	
PE <sup>a</sup> 500 bp	290,571,627	58,695,468,654	101,647,459	20,532,786,718	86
MP <sup>b</sup> 3 kb	214,359,129	21,435,912,900	87,974,983	8,797,498,300	37
MP 10 kb	229,499,284	22,949,928,400	68,408,294	6,840,829,400	29
Total	734,430,040	103,081,309,954	258,030,736	36,171,114,418	152

The raw data were filtered by high stringency and more detailed information described in Method. <sup>a</sup>Paired end sequencing, <sup>b</sup>Mate-paired sequencing.

**Table 3 Genome assembly summary**

Summary	Number
Estimated genome size (bp)	238,934,385
Total assembly length (bp)	228,315,917
Total assembly gap length (bp)	22,358,847
Number of contigs	18,160
Contig N50 length (bp)	21,729
Number of scaffolds	2,430
Scaffold N50 length (bp)	1,421,626
Largest scaffold length (bp)	6,352,280
Average scaffold length (bp)	93,957
Number of (A + T)s (%)	60.19
Number of (G + C)s (%)	30.02
Number of Ns (%)	9.79
Repeats length (bp)	14,794,603
Interspersed repeats (bp)	4,448,613
Simple repeats (bp)	8,167,274

Size of estimated genome and statistics of assembled scaffolds. The N50 scaffold size indicated that 50% of nucleotides in the assembly occur in scaffolds of length more than or equal to the N50 size.

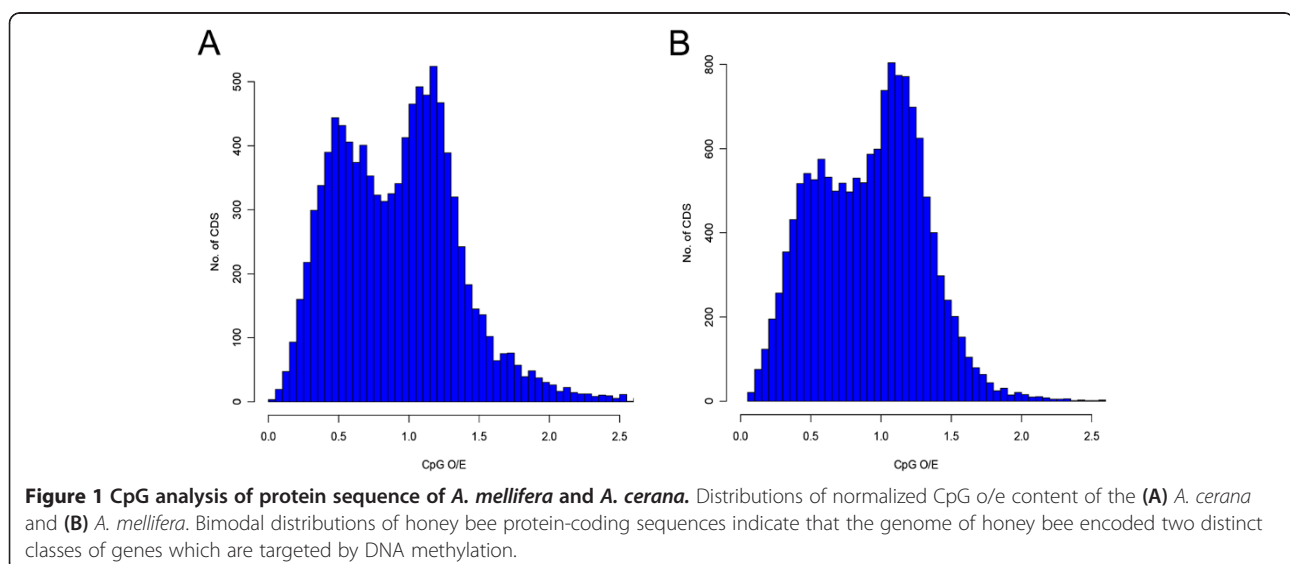
specific repeat elements were found using the *de novo* repeat finding program, RepeatModeler (version 1.0.7). Since the *A. cerana* genome assembly contained 9.79% N's, we assumed that repetitive sequences in the current assembly may be underestimated. In comparison to *A. mellifera*, only long terminal repeat elements were over-represented in *A. cerana*, which accounted for 0.1% (218 kb) of the genome, compared to 0.02% (49.6 kb) in *A. mellifera*. In contrast, long interspersed elements and short interspersed elements were not detected in the genome of *A. cerana*. Both were found in the *A. mellifera* genome at frequencies of 0.04% (83.1 kb) and 0.03% (70 kb),

respectively. DNA transposons constitute 0.11% (247 kb) of the *A. cerana* genome and 0.57% (1.34 Mb) of the *A. mellifera* genome. *Mariner* transposable elements, first discovered in the fruit fly, have been found across honey bee species [45,46]. The genome of the western honey bee, *A. mellifera*, contained multiple copies of *mariner* transposons, ranging from *AmMar1* to *AmMar6* [5]. In contrast, the *A. cerana* genome contained orthologs of *AmMar1* and *AmMar3-6*, but *AmMar2* ortholog was not found. This is consistent with the speculation that *AmMar1* and *AmMar2* were transferred to the *A. mellifera* genome relatively recently [45,46].

Although genome-wide repeat analyses need to be investigated further, the results of this study showed a striking reduction of transposable elements (TEs) and retrotransposons in the *A. cerana* genome compared to *A. mellifera*. Lack of TEs is one of the major features of the honey bee genome, compared to other sequenced Hymenoptera [30,32]. Some evidence suggests that grooming and hygienic behaviors in eusocial organisms reduced insertion of TEs from foreign genomes [30]. However, both social and non-social hymenopteran insect genomes, including seven ants and the parasitoid *Nasonia*, have been sequenced, and they include significantly different quantities of repetitive elements comprising 11% to 28% of genomes [30,36]. Therefore, hygienic behavior is not the only factor influencing the accumulation of repetitive elements in genomes.

#### Analysis of the *A. cerana* gene set

Due to limited data on expressed sequence tags (ESTs) and complementary DNAs (cDNAs) available for *A. cerana*, we established a gene annotation pipeline using multiple evidence data (Table 4). First, we generated 213,327 transcripts covering 515,809,639 bp using a *de novo* assembly from 68 Gb of *A. cerana* RNA-seq reads.



**Table 4 General statistics for gene modeling**

	Number of gene models	Length (bp)
<b>Initial evaluation (Evidence data)</b>		
RNA-seq, <i>de novo</i>	213,327	515,809,639
Genome + RNA-seq	31,027	96,495,948
Genome sequence, <i>de novo</i>	24,579	18,397,306
<i>A. mellifera</i> RefSeq <sup>a</sup>	18,215	23,654,031
<b>Primary gene set</b>		
Integrated evidence data	11,458	32,199,638
<b>Final gene set (OGS v1.0)</b>		
Manual annotation	10,651	15,655,993

<sup>a</sup>NCBI RefSeq from [http://www.ncbi.nlm.nih.gov/genome/48?project\\_id=13343](http://www.ncbi.nlm.nih.gov/genome/48?project_id=13343).

Second, RNA-seq data was aligned to scaffold sequences, which resulted in 31,027 gene models representing 96,495,948 bp. Third, we performed computational gene prediction based on the scaffold sequence information, which generated 24,579 genes covering 18,397,306 bp. We also employed *A. mellifera* gene sequences collected from the National Center for Biotechnology Information Reference Sequence Database (NCBI RefSeq, [47]) as a model to obtain homology-based gene annotation. Subsequently, we merged all predicted gene models using the MAKER [48] program to generate a primary gene set. All genes were queried with the NCBI non-redundant database using BLASTX. Lastly, we manually checked for missing genes, partial genes, or separated genes. Chemo-receptor genes, including gustatory receptors (Grs), odorant receptors (Ors), and ionotropic receptors (Irs), were investigated more carefully using analyses of functional sequence domains. Finally, 10,651 genes were annotated as the official gene set (OGS) of *A. cerana*, OGS version 1.0 (Table 4), of which approximately 84% of genes were annotated with NCBI non-redundant database and 70% were annotated in the Uniprot database [49]. Overall, the total number of genes in the *A. cerana* OGS v1.0 was comparable to the number in the *A. mellifera* OGS v1.0 (10,157 genes). However, the number is less than the current release of the *A. mellifera* genome, OGS v3.2 (15,314 genes; Table 5) [5,32].

We classified genes by function using the gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases [50,51]; 6,338 genes (60%) had more than one GO term and 1,696 enzymes were categorized into 125 pathways (Additional files 2 and 3). Here, several interesting molecular pathways that could represent honey bee-specific molecular mechanisms were revealed. For example, the fatty acid biosynthesis, glutathione metabolism, and cytochrome P450 pathways may be involved in nest-mate recognition and detoxification of pesticides (Additional file 1: Figure S5) [52,53]. The surface of the

**Table 5 Comparison of the official gene set of *A. cerana* and *A. mellifera***

	<i>A. cerana</i> OGSv1.0 (2014)	<i>A. mellifera</i> OGSv1.0 (2006)	<i>A. mellifera</i> OGSv3.2 (2014)
Number of genes	10,651	10,157	15,314
Total bases (bp)	15,655,993	16,484,776	19,342,383
Minimum length (bp)	33	24	75
Maximum length (bp)	37,824	53,649	70,263
Average length (bp)	1,469	1,623	1,263
Median length (bp)	1,071	1,197	840
(A + T)s (%)	61.88	60.41	60.42
(G + C)s (%)	38.12	39.43	39.47
Ns (%)	0.00	0.16	0.10

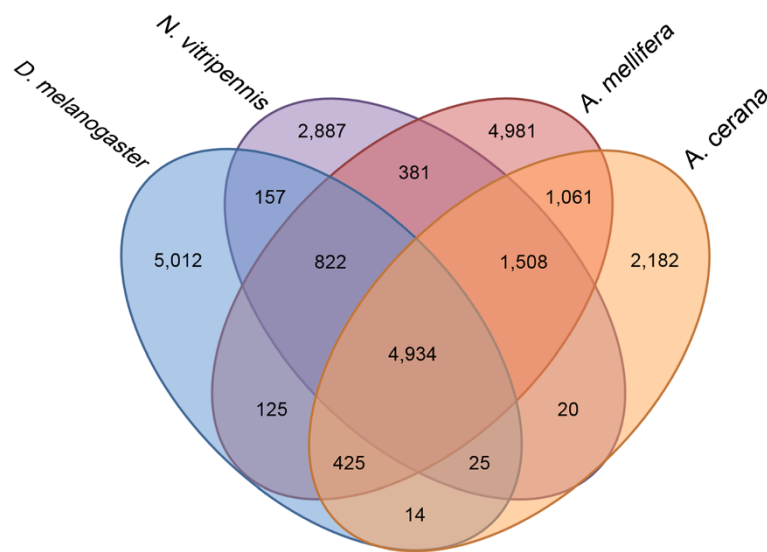
honey bee is composed of fatty acids and hydrocarbons, which reflect identity, and guard bees recognize these compounds to discriminate colony members from intruders [52]. KEGG analyses revealed that classes of enzymes involved in fatty acid biosynthesis are shared between *A. cerana* and *A. mellifera*, and *A. cerana* has fewer detoxification enzymes compared to fly and mosquito but similar numbers to *A. mellifera* [53]. The contribution of pesticides to global colony losses of *A. mellifera* is still a controversial issue, but data indicate that *A. mellifera* is unusually sensitive to various insecticides [54,55]. Interestingly, colonies of *A. cerana* have not shown similar levels of collapse to *A. mellifera*, but this could be explained by other differences that may reduce exposure to pesticides, such as frequent absconding behavior, small nest architecture, and foraging in high altitude regions [7,56].

#### Genes unique to *A. cerana* and orthologous to honey bee

To investigate whether non-orthologous genes are associated with features of *A. cerana* biology, we compared three hymenopteran insects, *Apis mellifera* (social), *Apis cerana* (social), *Nasonia vitripennis* (non-social), and one dipteran insect, *Drosophila melanogaster* (non-social) by orthology based clustering. Amongst 2,182 unique genes in *A. cerana* (Figure 2), most of the significantly enriched GO-terms were involved in neuromuscular junction, neuromuscular process, regulation of muscle organ development, muscle cell differentiation, and muscle tissue development ( $p < 0.05$ , Additional file 4). *A. cerana* has a higher wing beat frequency (306 beats/s) compared to *A. mellifera* (235 beats/s) and quick, impetuous, and unpredictable flight patterns, therefore some of the enriched proteins involved in muscle movement might account for *A. cerana*-specific flight patterns [56]. Future studies should be performed to dissect this relationship.

Notably, neural signaling-related GO-terms, including neuron recognition, signaling receptor activity,





**Figure 2 Comparative analysis of orthologous protein groups among the four insect genomes.** Orthology analysis of the proteins of *A. cerana* (orange oval) with three well known insect models, *D. melanogaster* (blue oval), *N. vitripennis* (purple oval), and *A. mellifera* (red oval). Both *D. melanogaster* and *N. vitripennis* are non-social and *A. mellifera* and *A. cerana* are social insect species. \*indicates *A. cerana*-specific proteins.

transmembrane receptor signaling pathway, ionotropic glutamate receptor signaling pathway, and active transmembrane transporter activity, which are closely related with chemosensory reception and chemical signaling, were also enriched ( $p < 0.05$ ) in the *A. cerana* unique gene set (Additional file 4). Genes involved in chemical signaling have evolved rapidly, especially in eusocial organisms [57,58]. Neural signaling processes play major roles mediating social communication in honey bee society. *A. cerana* shows a number of group-level behaviors distinct from *A. mellifera* [12], including a unique defense behaviors against a hornets. *A. cerana* guard bees raise their abdomens and shake or flutter, producing alarm pheromones, when hornets approach the hive [7,56]. Additional research is required to determine whether molecular regulation mechanisms found uniquely in *A. cerana* may be responsible for these unique social defense behaviors.

Since *A. mellifera* and *A. cerana* diverged recently, we hypothesized there would be protein orthologs conserved in both honey bee species that explain shared honey bee characteristics. A total of 1,061 *A. cerana* proteins were identified with orthologs in *A. mellifera* but no other non-social species. These orthologs were categorized with GO-terms “sensory perception of smell” ( $p < 1.75E^{-04}$ ) and “sensory perception of chemical stimulus” ( $p < 7.55E^{-04}$ ), which are crucial features for social communication and physical interaction [27]. In addition, the GO-term “carbohydrate transporter activity” ( $p < 1.87E^{-02}$ ), which describes cuticular hydrocarbon detection [27], and “regulation of short-term neuronal synaptic plasticity” ( $p < 2.21E^{-02}$ ) and

“transmembrane signaling receptor activity” ( $p < 3.04E^{-02}$ ), which are involved in neuronal signaling during social interaction, were also enriched in orthologs shared by the two honey bee species (Additional file 1: Table S3).

#### Chemoreceptor gene family

Chemoreceptors play important roles in communication and social behaviors, in part by mediating detection of chemical signals from nest-mates [59]. Major groups of chemoreceptor genes include gustatory receptors (*Grs*), odorant receptors (*Ors*), and ionotropic receptors (*Irs*) [59-63]. In social insects, such as ants and honey bees, chemical communication is crucial for colony maintenance and cooperation [64-66]. Here, we have characterized 10 new *Grs*, 119 new *Ors*, and 10 new *Irs* in the *A. cerana* genome. Gene expression patterns, examined using RNA-seq data, revealed that annotated chemoreceptor genes were well organized and comparable to those of *A. mellifera* [59] and *N. vitripennis* [62], although they were slightly underrepresented compared to the *A. mellifera* genome [59].

#### The gustatory receptor family

The gustatory receptor family plays an important role in taste and is used to collect nectar and pollen for energy and brood care [67]. In honey bee society, colony members have division labor and perform different tasks. Nurse bees take care of the brood and the queen, and they clean inside the nest. Forager bees collect food or

resin from outside and bring it to the hive [68]. Peripheral and internal regulation of *Gr* gene expression is involved in this behavioral transition [69].

According to Robertson and Wanner [59], the western honey bee, *A. mellifera* has 13 *Grs* (H. M. Robertson, personal communication), a small number compared to the fruit fly *D. melanogaster* (68 *Grs*, [60]), the mosquito *Aedes aegypti* (79 *Grs*, [70]), the parasitoid wasp *N. vitripennis* (58 *Grs*, [62]), and the ant *Linepithema humile* (116 *Grs*, [27]). Similar to *A. mellifera*, 10 *Gr* genes were identified in the *A. cerana* genome. They were named based on their orthology to *A. mellifera* *Grs* (*AmGrs*). All identified *Grs* in *A. cerana* showed simple orthologous relationships with *Grs* in *A. mellifera*, and *AcGr1*, 2, 3, 6, 7, 9, and 10 also had orthologs in *N. vitripennis* (Additional file 1: Figure S6). These data indicated that *Gr* genes are highly conserved among hymenopteran species. Similar to the *A. mellifera* *Gr* repertoire, *AcGr1* and *AcGr2* were positioned in expanded lineages to sugar receptors of *D. melanogaster*, including *DmGr5a*, *DmGr61a* and *DmGr64a/f* (Figure 3A) [71-73]. In addition, *AcGr3* shared a clade with *DmGr43a*, which functions as a fructose receptor in the periphery and a nutrient sensor in the brain of *Drosophila* (Figure 3A) [74]. In contrast, *AcGr6*, 7, 9, 10, and *X* lineages showed no apparent relationships with *DmGrs*, implying that they may be unique to the honey bee. Bitter taste receptors also seem to be lost in the *A. cerana* genome, which may be related to the evolution of flower preference in the honey bee [59] compared to other social insects such as the ant in which bitter receptors are preserved [27,28]. Additionally, orthologs to *Drosophila* carbon dioxide (CO<sub>2</sub>) receptors, *Gr21a* and *Gr63a* [28,75], were not present in the *A. cerana* genome, similar to *A. mellifera* [59]. However, honey bees are known to detect CO<sub>2</sub> [76], indicating that they may have evolved novel molecular mechanisms similar to the acid sensing mechanism found in *Drosophila* for detecting high CO<sub>2</sub> concentrations [77]. Partial sequences of *A. cerana* *Gr4* and *Gr5* orthologs were located using TBLASTN searches. A *Gr8* ortholog could not be found in the *A. cerana* genome.

Expression patterns of *Gr* orthologs in *A. cerana* and *A. mellifera* were determined by relative gene expression analyses (Figure 3B). Surprisingly, expression patterns of *Gr* orthologs between the two honey bee species were distinct. Candidate sugar receptors, *Gr1* and *Gr2*, were expressed higher in *A. cerana* compared to *A. mellifera* (Figure 3B), suggesting that *A. cerana* may have a greater ability to sense sugars. Similarly, *Gr5* and *Gr7* were highly expressed in *A. cerana* compared to *A. mellifera*. In contrast, *Gr3*, 6, 9, and 10 were more highly expressed in *A. mellifera* compared to *A. cerana*. *Gr4* and *GrX* were not detected in antennal transcriptome of *A. cerana* (data not shown), implying that *Gr4* and *GrX* might be expressed at undetectable levels or in

other tissues, such as the tongue or legs. Future functional studies on *Grs* may reveal taste sensing and internal regulation differences between the species.

### The odorant receptor family

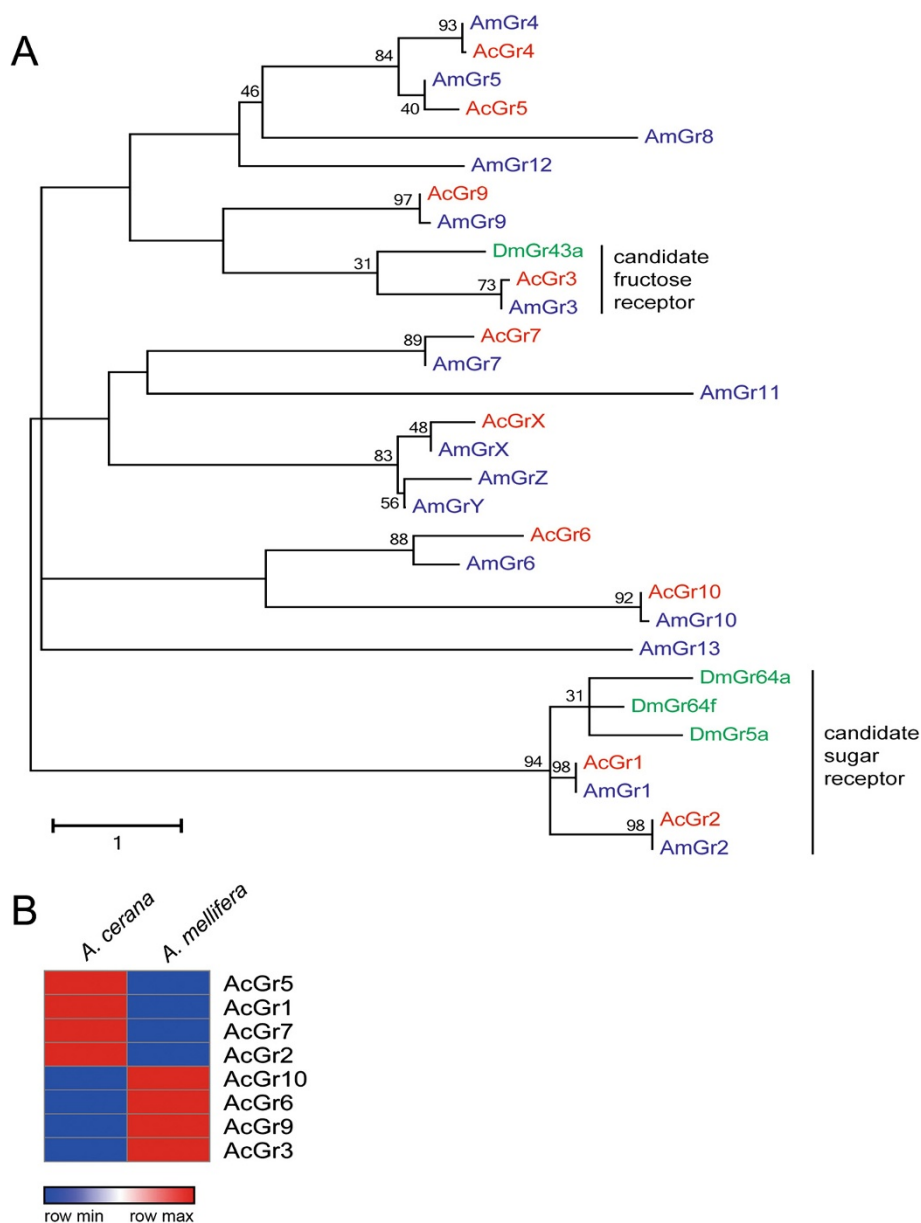
Insect odorant receptors play important roles in environmental signal recognition and inter- and intra-species communication [59]. Honey bees use odorant receptors in contexts, including kin recognition, food navigation, and pheromone detection [59,65,78]. However, despite the importance of odorant receptors, the functional identification of *Ors* in honey bees is lacking compared to other model insects, including fly and mosquito species [79].

In the *A. cerana* genome, 119 *AcOrs*, including a few 5'- or 3'- partial sequences containing the odorant receptor domain, were identified. We named *A. cerana* *Ors* by sequence positions in scaffolds. Most *AcOrs* were not spread evenly across scaffolds, but were clustered at a few locations in the genome. For example, clusters of 37 *Ors*, 15 *Ors*, and 17 *Ors* were arrayed on scaffold 3, 103, and 139, respectively (Additional file 1: Figure S7). In *A. mellifera*, the largest tandem array of 60 *Ors* was found on chromosome 2 [59]. This expansion of *Ors* implies unequal crossing-over by neighboring genes occurred. The large number of *Or* paralogs indicate diverse roles for odorant recognition in honey bee society, such as pheromone blends, cuticular hydrocarbons, and floral odor cocktails [5]. Since *A. mellifera* and *A. cerana* diverged recently [7], it was hypothesized that there may be synteny between *Or* clusters. Regions of *A. mellifera* chromosome 2 with microsynteny conservation were identified by comparing the *Or* gene arrangement in the *A. cerana* genome to the *A. mellifera* genome. Consistent with the hypothesis, conserved microsynteny and clear orthologs of *A. cerana* *Ors* to *A. mellifera* *Ors* were found (Figure 4C, Additional file 5), suggesting that honey bee *Or* paralogs are clustered in conserved genomic regions [80,81].

Insects have a number of variable *Ors*, which form a chaperone with *olfactory receptor co-receptor (Orco)* *in vivo* [82-85]. In the present study, *A. cerana* *Or5* shared orthology with insect *Orcos* including *D. melanogaster* *Or83b*, *N. vitripennis* *Or1*, and *A. mellifera* *Or2* (Figure 4A). Overall, identified *AcOrs* showed simple orthologous relationships with *AmOrs*, such as 1:1, 1:2, and 1:3 (*AcOrs* : *AmOrs*).

Among 177 *A. mellifera* *Ors*, *AmOr11* was functionally characterized as a queen pheromone receptor responding to 9-oxo-2-decenoic acid (9-ODA) [78]. In our study, *AcOr30* showed 1:1 orthology to *AmOr11* with 98.7% identity (Additional file 1: Figure S7), implying that queen pheromone components may be conserved between *A. mellifera* and *A. cerana*.

Transcriptome data revealed that *Or* homologs are differentially expressed between *A. cerana* and *A. mellifera* (Figure 4B). Forty-four *Or* homologs were more highly



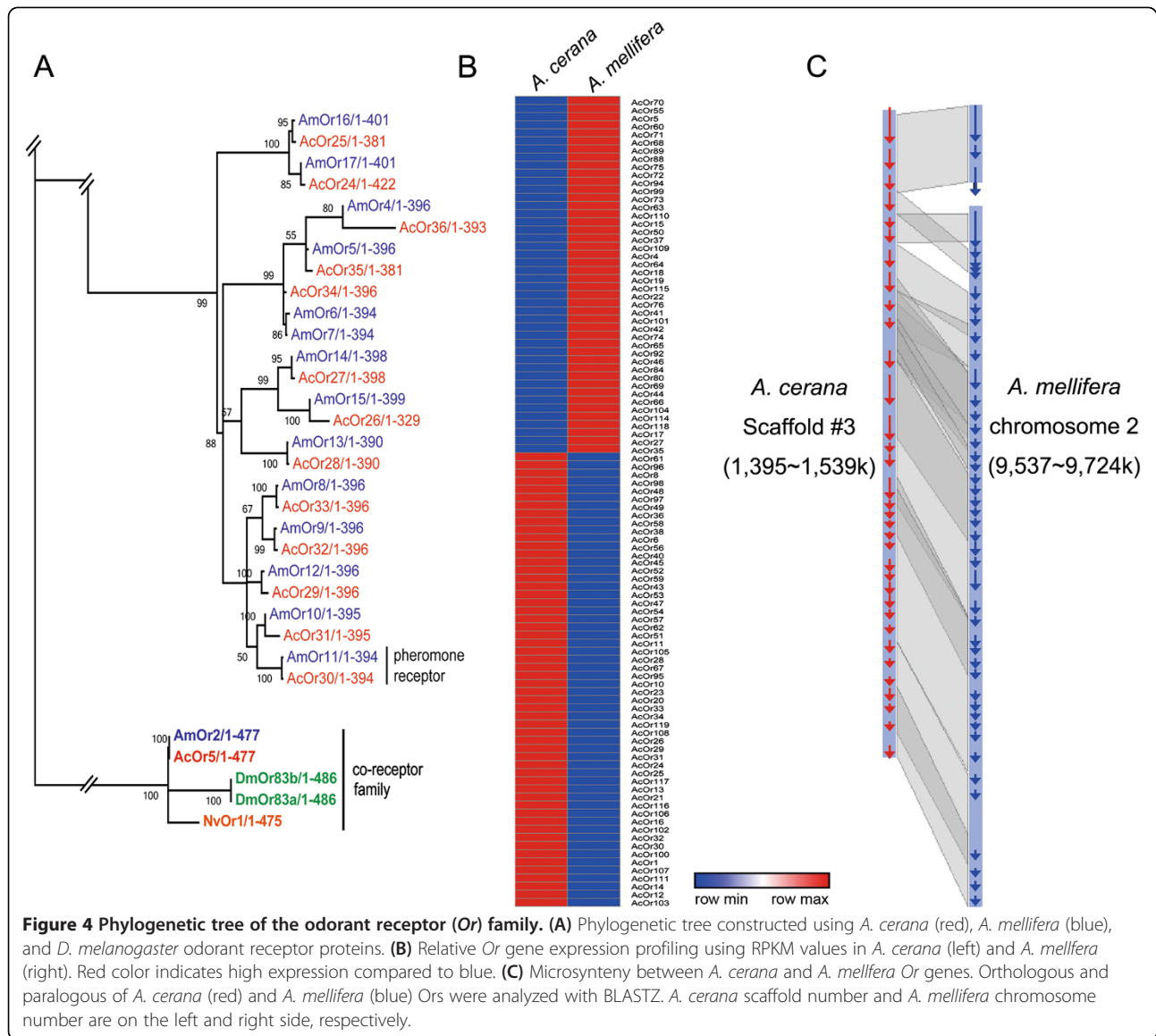
**Figure 3 Phylogenetic tree of the gustatory receptor (Gr) family.** (A) Phylogenetic tree constructed using *A. cerana* (red), *A. mellifera* (blue), and *D. melanogaster* gustatory receptor proteins (B) Relative Gr gene expression profiling using RPKM values in *A. cerana* (left) and *A. mellifera* (right). Red color indicates high expression compared to blue.

expressed in *A. mellifera*, and 56 *Or* homologs were more highly expressed in *A. cerana*. Different expression patterns support the idea that coding sequences are well conserved among *Or* homologs, but their promoter sequences have diverse regulatory motifs. These data imply that the two honey bee species express different odorant spectra. Specifically, seven *AcOrs* (*AcOr21*, 38, 40, 45, 56, 58, and 116) were expressed only in *A. cerana*, indicating functions specific to *A. cerana*. Functional studies using heterologous expression systems are needed to better understand the various functions of *Ors* in honey bees.

### The ionotropic receptor family

Recently, a new family of chemosensory receptors, the ionotropic receptor (*Ir*) family, was identified in *D. melanogaster* [61]. *Irs* in *D. melanogaster* constitute distinct and divergent subfamilies of ionotropic glutamate receptors (*iGluRs*) [63]. Sixty-six *Ir* homologs have been identified in *D. melanogaster*, and 16 were expressed specifically in antennae [61,63]. This suggested that *Irs* belong to two subgroups: conserved antennal *Irs* and species-specific divergent *Irs*. These subgroups represent classes of *Ors* and *Grs*, respectively [63]. In contrast to

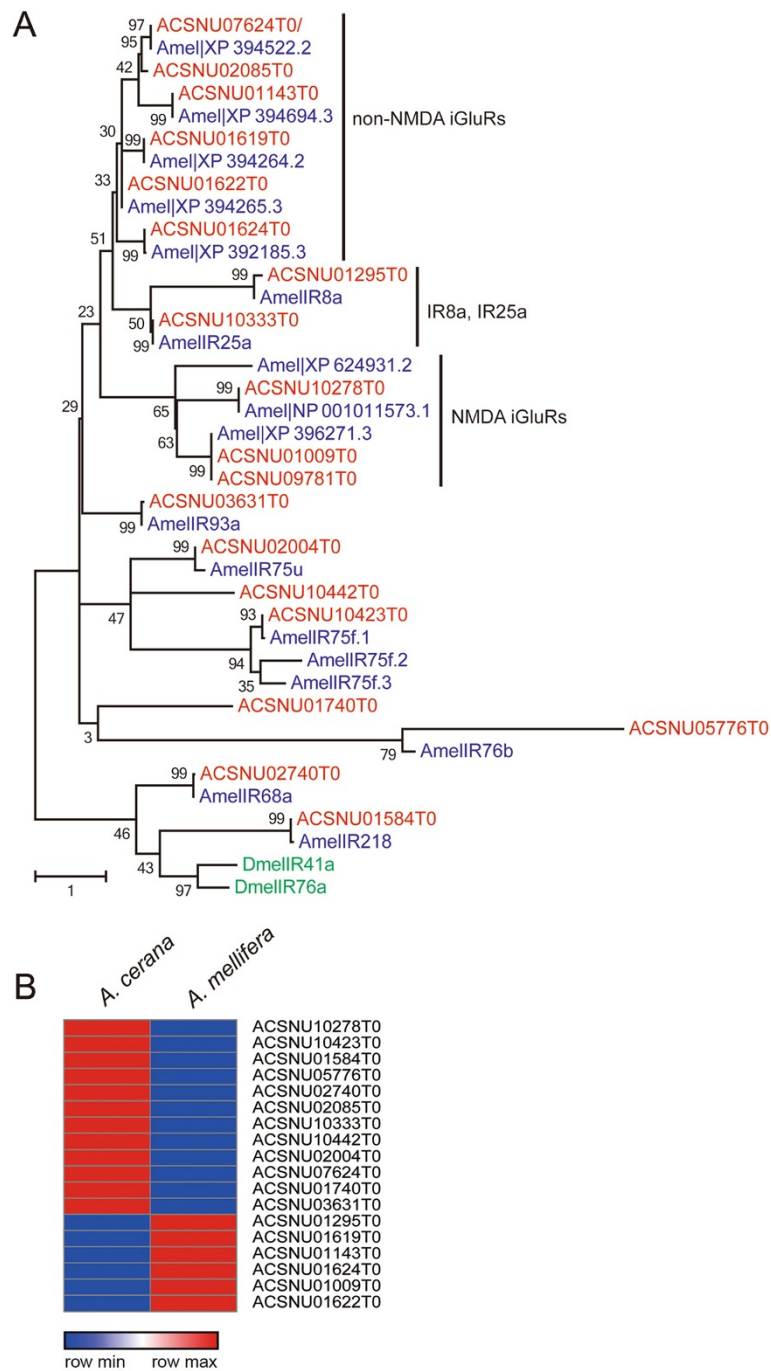




Ors, which respond broadly to alcohols, ketones, and esters, Irs primarily respond to acids, amines, and carbon dioxide, which can be physiologically important in many insect species [77,86-88]. Although the functions of these receptors are not yet known, Irs may have more general function in the detection of environmental chemicals including odorants and tastes [89].

The number of identified *Irs* in insects is increasing [63,90,91], and a large complement of *Irs* has been described in the complete genomes of four hymenopteran species: *A. mellifera* (10 *Irs*), *N. vitripennis* (10 *Irs*), *L. humile* (32 *Irs*), and *P. barbatus* (24 *Irs*) [5,27,28,36]. In this study, 10 *Ir* homologs were found in the *A. cerana* genome (Figure 5A). Sequence comparison and phylogenetic analyses of *Irs* with *D. melanogaster* and *A. mellifera* identified

putative orthologs of conserved *Irs* in the *A. cerana* genome: *Ir8a*, *Ir25a*, *Ir68a*, *Ir75a*, *Ir76a*, and *Ir93a* [63]. As expected, highly conserved orthologs of antennal *Irs* were identified in the *A. cerana* genome. These results support the hypothesis that antennal expression of *Ir* orthologs has been conserved for over 350 mya since dipteran and hymenopteran insects diverged [92]. Other *Irs* in *A. cerana* with low similarity to orthologs of other insect receptors appear to be honey bee-specific. These *Irs* may be used for species-specific recognition, including candidates for cuticular hydrocarbon receptors and brood pheromone receptors. However, expression patterns for the vast majority of *Irs* are unknown and no ligands for honey bee *Irs* have been identified. In this study, *Aclr* expression profiles were different in *A. mellifera* and *A.*



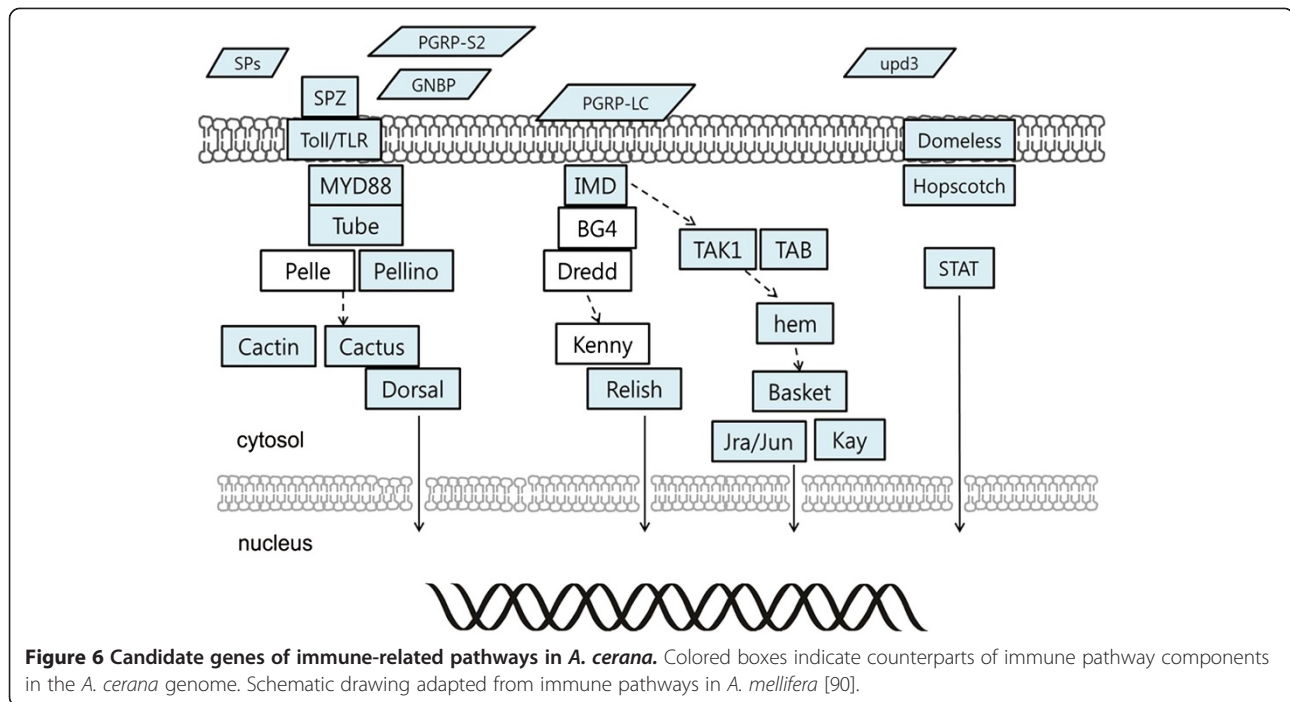
**Figure 5 Phylogenetic tree of the ionotropic receptor (*Ir*) family. (A)** Phylogenetic tree constructed using *A. cerana* (red), *A. mellifera* (blue), and *D. melanogaster* ionotropic receptor proteins. **(B)** Relative *Ir* gene expression profiling using RPKM values in *A. cerana* (left) and *A. mellifera* (right). Red color indicates high expression compared to blue.

*cerana* (Figure 5B). Their functions and evolutionary basis for diversity remain to be investigated.

#### Immune-related genes

Honey bees are invaluable models for studying social defense dynamics and individual molecular and behavioral

defense mechanisms [93]. In contrast to *A. mellifera*, *A. cerana* is not susceptible to the ectoparasitic mite, *Varroa destructor*, one of the major vectors of bee pathogens. In contrast, *A. cerana* has suffered greatly from viral and bacterial diseases in recent years [94]. A recent report indicated that more than 90% of Asian honey bee colonies



collapsed due to sacbrood virus (SBV) infection in Korea [95]. Many Asian countries also suffered declines in *A. cerana* colonies for several reasons. However, the molecular defense mechanisms of *A. cerana* are still unknown. Therefore, we investigated immune genes present in the *A. cerana* genome by comparing genomic information to other sequenced insect genomes [96].

Using multiple TBLASTN searches, 160 immune gene orthologs were identified in *A. cerana* and 11 additional genes were detected by manual annotation. All major pathways were identified in *A. cerana*, including components of the Toll, Imd, Jak/Stat, and JNK pathways [96]. Notably, the *FADD*, *Dredd*, and *Kenny*, components of the Imd pathway and *Pelle* of the Toll pathway were not detected in the *A. cerana* genome (Figure 6). The total number of innate immune genes in *A. cerana* is similar to other social Hymenoptera [28] (Additional file 1: Table S4), and most immune genes in *A. cerana* shared higher sequence similarity with *A. mellifera* compared to other sequenced insect species. This may be explained by conservation of the innate immune system between *A. cerana* and *A. mellifera*. Eusocial insects have additional social immune systems, such as cleaning behaviors (hygienic behavior, grooming, and undertaking), thermal defenses (*A. mellifera* lacks this behavior), and antibiotic nest architecture (resin collection), which may contribute to reducing exposure to pathogens [93].

Previous studies indicate more antimicrobial proteins are encoded in the *A. cerana* genome compared to *A. mellifera* [11]. Indeed, defensive peptides, including venomous peptides, in *A. cerana* are more strongly

expressed than those in *A. mellifera* [97]. In addition, some reports show the uniqueness of strong behavioral defenses of *A. cerana*, such as hygienic and grooming behaviors [11,12,94]. Together, these data indicate that *A. cerana*, through a combination of elaborate molecular and behavioral mechanisms, may have a more effective social defense system compared to *A. mellifera*. Functional studies of immune genes will inform knowledge of *A. cerana*-specific disease control methods and provide a valuable model for comparative studies of social insect immune systems.

## Conclusion

The Asian honey bee, *A. cerana*, a close relative of the western honey bee, *A. mellifera*, provides considerable economic benefits to the apicultural industry and serves as a new model for biodiversity in eastern and southern Asia. Despite similarities in Asian and western honey bees, differences in biological traits led us to explore the genome and transcriptome of *A. cerana*. This is the first report to elucidate the whole genome sequence of the Asian honey bee, *A. cerana*, by employing *de novo* assembly of the *A. cerana* genome and computational gene prediction followed by manual annotation. *A. cerana* has unique features for muscle movement and neural signaling, reflecting the wild nature of *A. cerana*. In addition, chemosensory receptors and immune-related genes, which might be responsible for sophisticated and well-organized life styles in honey bees, were identified. Comparative studies of chemoreceptor gene families showed similarities in receptor family expansions relative to *A.*

*mellifera*. However, significant differences in gene expression were also identified, potentially reflecting different capabilities for odor perception. In addition, immune-related genes in *A. cerana* exhibited expression patterns that may reflect an advanced social immune system, advancing understanding of the molecular basis of social immunity. This genome analysis will provide invaluable information on the novel characteristics of *A. cerana*, and contribute to the understanding of comparative sociogenomics with other sequenced social insect genomes.

## Methods

### Sample collection and DNA and RNA extraction

Adult drone bees of *A. cerana* ( $n = 4$ ) were collected from a single colony at the College of Agriculture and Life Sciences, Seoul National University (SNU), Seoul, Korea during summer 2012. Mid-gut tissues were removed to reduce contamination with genomes of intestinal microorganisms. Genomic DNA was extracted from individual drone bees using the Wizard Genomic DNA Purification kit (Promega, MI, US) according to the manufacturer's protocol. For RNA sequencing, *A. mellifera* and *A. cerana* worker bees were collected from 3 different colonies at SNU and stored immediately in liquid nitrogen. After flash freezing, samples were stored at  $-80^{\circ}\text{C}$  until dissection. Tissues from *A. cerana* (brain, antenna, hypopharyngeal gland, gut, fat body, and venom gland, and larvae) and from *A. mellifera* (antenna) were dissected in cold RNase free PBS (pH = 7.4). Total RNA was isolated using a QIAGEN RNeasy Mini kit (Qiagen, CA, US) according to the manufacturer's protocol.

### DNA and cDNA library construction

Genomic DNA for shotgun sequencing was sheared with average fragment sizes of 500 bp, 3 kb, and 10 kb. Three DNA libraries were constructed using the Illumina sequencing platform according to the manufacturer's protocol (Illumina CA, US). Briefly, 5, 10, and 20  $\mu\text{g}$  of genomic DNA were used for whole genome random shearing to construct 500 bp, 3 kb, and 10 kb libraries, respectively. Fragments were ligated with 3' and 5' adapters to produce libraries for PCR amplification.

For mRNA sequencing, cDNA libraries were prepared for each tissue using the Illumina mRNA sequencing kit (Illumina, CA, US) and the Clontech SMART cDNA Library Construction Kit (Invitrogen). Libraries were sequenced using the Illumina HiSeq2000 and Roche 454 Genome Sequencer FLX instrument (Roche, UK).

### De novo genome assembly and sequence analyses

Raw sequencing reads with phred scores  $\leq 20$  were filtered out using the CLC\_quality\_trim (CLC 3.22.55705). Duplicate sequences were removed with the remove\_duplicate program (CLC-bio) using the default options. After

filtration, genome libraries with inserts of 500 bp, 3 kb, and 10 kb were assembled using the AllPaths-LG (version 42411, [31]) algorithm with default parameters. The *A. cerana* genome sequence is available from the NCBI with project accession PRJNA235974. Repeat elements in the *A. cerana* genome were identified using RepeatModeler (version 1.0.7, [98]) with default options. Subsequently, RepeatMasker (version 4.03, [99]) was used to screen DNA sequences against RepBase (update 20130422, [100]), the repeat database, and mask all regions that matched known repetitive elements. Comparison of experimental mitochondrial DNA to published mitochondrial DNA (NCBI accession GQ162109) was performed using the CGView Server with the default options [101]. The percent identity shared between the *A. cerana* mitochondrial genome assembly and NCBI GQ162109 was determined by BLAST2 [102]. To examine the distribution of observed to expected (o/e) CpG ratios in protein coding sequences of *A. cerana*, we used in-house perl scripts to calculate normalized CpG o/e values [43]. Normalized CpG was calculated using the formula:

$$\text{freq}(\text{CpG})/\text{freq}(\text{C}) \times \text{freq}(\text{G}),$$

where  $\text{freq}(\text{CpG})$  is the frequency of CpG,  $\text{freq}(\text{C})$  is the frequency of C and  $\text{freq}(\text{G})$  is the frequency of G observed in a CDS sequence.

### Evidence-based gene model prediction

Assembly of RNAseq data was performed using *de novo* assembler software, Trinity (version r2013-02-25, [98]). Alignment of RNAseq reads against genome assemblies was performed using Tophat and transcript assemblies were determined using Cufflinks (version 2.1.1, [103]). Gene set predictions were generated using GeneMark.hmm (version 2.5f, [99]). Homolog alignments were made using NCBI RefSeq and *A. mellifera* as a reference gene set (Amel\_4.5). A final gene set was created synthetically by integrating evidence-based data using the gene modeling program, MAKER (version 2.26-beta), including the exonerate pipeline with default options [48,104]. Subsequently, we performed blast searches with the NCBI non-redundant dataset to annotate combined gene models. All gene predictions were provided as input to the Apollo genome annotation editor (version 1.9.3, [105]), and genes included in phylogenetic analyses were manually checked against transcript information generated by Cufflinks to correct for 1) missing genes, 2) partial genes, and 3) separated genes.

### Gene orthology and ontology analysis

The protein sets of four insect species were obtained from *A. cerana* OGS v1.0, *A. mellifera* OGS v3.2 [32], *N. vitripennis* OGS v1.2 [36], and *D. melanogaster* r5.54



[106]. We used OrthoMCL v 2.0 [107] to perform ortholog analysis with default parameter for all steps in the program. GO annotation proceeded in Blast2GO (version 2.7) with default Blast2GO parameters. Enrichment analysis for statistical significance of GO annotation between two groups of annotated sequences was performed using Fisher's Exact Test with default parameters.

#### Gene family identification and phylogenetic analysis

Total 10,651 sequences of OGS v1.0 were classified with Gene Ontology (GO) and KEGG database using blast2GO (version 2.7) with MySQL DBMS (version 5.0.77). To search the sequence of *A. cerana* odorant receptors (Ors), gustatory receptors (Grs), and ionotropic receptors (Irs), we prepared three sets of query protein sequences: 1) first set includes Or and Gr protein sequences from *A. mellifera* (provided by Dr. Robertson H. M. at the University of Illinois, USA), 2) second set includes Or, Gr, and Ir protein sequences of previously known insects from NCBI Refseq [108], 3) third set includes functional domain of chemoreceptor from Pfam (PF02949, PF08395, PF00600) [109]. The TBLASTN of these three sets of receptor proteins was performed against *A. cerana* genome. Candidate chemoreceptor sequences from the results of TBLASTN were compared with *ab initio* gene predictions (see Gene annotation section) and verified its functional domain using the MOTIF search program [110]. Annotated Or, Gr, and Ir proteins were aligned with ClustalX [111] to corresponding proteins of *A. mellifera* and were manually corrected. Alignments were performed iteratively and each sequence was refined based on alignments to make complete Or, Gr, and Ir sequences for *A. cerana*. Sequences were aligned with ClustalX [111], and a tree was built with MEGA5 [112] using the maximum likelihood method. Bootstrap analysis was performed using 1000 replicates.

#### Tissue expression analysis

To compare tissue transcriptomes of *A. cerana* and *A. mellifera*, RNAseq reads from antenna tissue of each species were mapped against annotated *Gr*, *Or*, and *Ir* gene sequences using RSEM [113]. TMM (Trimmed Mean of M-values)-normalized FPKM (fragments per kilobase per million) expression values were calculated by running perl scripts included in Trinity assembler (version r2013-02-25) and were used to draw a heat map [98].

#### Analysis of microsynteny between AmOrs and AcOrs

Microsynteny analyses of orthologous *Or* genes between *A. cerana* and *A. mellifera* were carried out based on a phylogenetic tree. In addition, we performed reciprocal BLASTZ searches [114] between two species in syntenic regions and identified clear orthologs based on E-values (cut-off:  $1e^{-40}$ ) and phylogenetic analyses.

#### Immune related gene family

Holometabola immune gene orthologs obtained from Immunodb [96] were used as queries against the *A. cerana* genome using TBLASTN. Genomic sequences encoding immune gene candidates were compared with *ab initio* gene predictions to confirm completeness of the gene models.

#### Availability of supporting data

The data sets supporting the results of this article are available through a BioProject (accession number PRJNA235974) at the NCBI.

#### Data availability

The Illumina transcript data generated for *A. cerana* are available in the NCBI SRA [115] with accessions SRR1380970 (brain), SRR1380976 (antenna), SRR1380979 (hypopharyngeal gland), SRR1380984 (gut), SRR1388774 (fat body), SRR1406762 (venom gland), and the accession for 454 transcript data is SRR1408614 (head). The Illumina raw data from *A. mellifera* are available at the NCBI SRR1386316 (brain), SRR1407793 (hypopharyngeal gland), and SRR1408090 (antenna). Whole genome sequences of *A. cerana* are submitted in the NCBI Whole Genome Shotgun database (SUB582139). The genome and gene information is also available at *A. cerana* database ([http://mnblab.snu.ac.kr/genome\\_seq.php](http://mnblab.snu.ac.kr/genome_seq.php)).

#### Additional files

**Additional file 1: Table S1.** Statistics of the honey bee genome assembly, *Apis mellifera* and *Apis cerana*. **Table S2.** Summary of repetitive elements in the *A. cerana* genome. **Table S3.** GO terms enriched in honey bee shared orthologs. **Table S4.** Immune related gene set counts in social and non-social insect, *Apis cerana*, *Apis mellifera*, *Nasonia vitripennis*, *Linepithema humile*, and *Drosophila melanogaster*. **Figure S1.** Syntenic view of the *A. mellifera* chromosome 3 and *A. cerana* scaffolds. **Figure S2.** Syntenic view of the mitochondrial genome of *A. mellifera* and *A. cerana*. **Figure S3.** The mitochondrial genome of *A. cerana*. **Figure S4.** Normalized CpG dinucleotide in protein-coding sequences in *A. cerana*. **Figure S5.** KEGG pathways in *A. cerana*. **Figure S6.** Phylogenetic tree of the gustatory receptor family. **Figure S7.** Phylogenetic tree of the odorant receptor family. **Figure S8.** Amino acid alignment between *AmOr11* and *AcOr30*.

**Additional file 2:** Is a text file containing results of GO annotations for *A. cerana* OGS v 1.0.

**Additional file 3:** Is a spreadsheet showing results of KEGG analyses of *A. cerana* OGS v 1.0.

**Additional file 4:** Is a spreadsheet showing results of GO enrichment analyses in *A. cerana* specific orthologs.

**Additional file 5:** Is a spreadsheet showing results of reciprocal BLAST between *A. cerana* and *A. mellifera* *Or* genes.

#### Abbreviations

Mya: Million years ago; SBV: Sacbrood virus; PE: Paired end; MP: Mate paired end; GC: Guanine plus cytosine; CpG: Cytosine followed by guanine in 5' to 3' orientation; TpG: Thymine followed by guanine in 5' to 3' orientation; o/e: observed/expected; DNMT: DNA methylation protein; TE: Transposable element; EST: Expressed sequence tag; RNA-seq: RNA sequencing;



Refseq: Reference sequences; NCBI: National Center for Biotechnology Information; OGS: Official gene set; GO: Gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; Gr: Gustatory receptor; Or: Odorant receptor; Ir: Ionotropic receptor; 9-ODA: 9-oxo-2-decenoic acid; iGluR: Ionotropic glutamate receptors; SBV: Sacbrood virus.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

BC, MJ, JLim, IC, YP, and TY carried out the sequence analysis and bioinformatics work from genome sequence. Specifically, BC conducted genome analysis and DP and MJ analysed RNAseq data. JLee participated in phylogenetic analysis and YY performed CpG analysis. DP, JJ, and YC, and ML collected honey bees and isolated RNA and DNA from the samples and participated in the design of the study. DP, ORE, GN, and HK conceived of the study, and participated in its design and coordination. DP, GN, and HK wrote the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

We appreciate RDA (Rural Development Administration) of Korea, Korean honey bee association, and Korea Beekeeping Association for helpful advice on beekeeping. We also thank Dr. Hugh Robertson at the University of Illinois for helpful comments on honey bee chemosensory receptors. We also appreciate NICEM of Seoul National University and two bioinformatics companies, MACROGEN and PHYZEN, for fast and accurate sequencing service and helpful discussion. Many thanks go to the members of the Kwon Lab at Seoul National University for assisting bee hive maintenance and collection of bees. This work was supported by Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ010487) from the RDA of the Republic of Korea to HK.

#### Author details

<sup>1</sup>Biomodulation Major, Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea. <sup>2</sup>National Instrumentation Center for Environmental Management, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea. <sup>3</sup>Department of Plant Science, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea. <sup>4</sup>Arizona Genomics Institute, University of Arizona, Tucson, Arizona 85721, USA. <sup>5</sup>National Institute of Agricultural Biotechnology, Rural development Administration, Suwon 441-707, Republic of Korea. <sup>6</sup>Department of Entomology, Kansas State University, Manhattan, Kansas, USA. <sup>7</sup>CSIRO Ecosystem Sciences, Centre for Environment and Life Sciences, Underwood Avenue, Floreat, WA 6014, Australia.

Received: 1 August 2014 Accepted: 2 December 2014

Published: 2 January 2015

#### References

- Galizia CG, Eisenhardt D, Giurfa M, Menzel R: *Honeybee Neurobiology and Behavior: A Tribute to Randolph Menzel*. Dordrecht Netherlands: New York: Springer; 2012.
- Begna DHB, Feng M, Li J: **Differential expression of nuclear proteomes between honeybee (*Apis mellifera* L.) queen and worker larvae: a deep insight into caste pathway decisions.** *J Proteome Res* 2012, **11**:1317–1329.
- Zayed ARG: **Understanding the relationship between brain gene expression and social behavior: lessons from the honey Bee.** *Annu Rev Genet* 2012, **46**:591–615.
- Foret S, Kucharski R, Pellegrini M, Feng S, Jacobsen SE, Robinson GE, Maleszka R: **DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees.** *Proc Natl Acad Sci U S A* 2012, **109**:4968–4973.
- Honeybee: **Insights into social insects from the genome of the honeybee *Apis mellifera*.** *Nature* 2006, **443**:931–949.
- Monica C, Munoz-Torres JTR, Christopher P, Childers Anna K, Bennett Jaideep P, Sundaram Kevin L, Childs Juan M, Anzola Natalia M, Elsik CG: **Hymenoptera genome database: integrated community resources for insect species of the order hymenoptera.** *Nucleic Acids Res* 2011, **39**:D658–D662.
- Oldroyd BP, Wongsiri S: *Asian Honey Bees: Biology, Conservation, and Human Interactions*. Cambridge, Mass: Harvard University Press; 2006.
- Engel MS: **A monograph of the Baltic amber bees and evolution of the Apoidea (Hymenoptera).** *Bull Am Mus Nat Hist* 2001, **259**:5–192.
- Arias MC, Sheppard WS: **Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence.** *Mol Phylogenet Evol* 1996, **5**:557–566.
- Arias MC, Sheppard WS: **Phylogenetic relationships of honey bees (Hymenoptera:Apinae:Apini) inferred from nuclear and mitochondrial DNA sequence data.** *Mol Phylogenet Evol* 2005, **37**:25–35.
- Xu P, Shi M, Chen XX: **Antimicrobial peptide evolution in the Asiatic honey bee *Apis cerana*.** *PLoS one* 2009, **4**:e4239.
- YS Peng YF, Xu S, Ge S: **The resistance mechanism of the Asian honey Bee, *apis cerana* fabr., to an ectoparasitic mite, *varroa jacobsoni* oudemans.** *J Invertebr Pathol* 1987, **49**:54–60.
- Ono M, Okada I, Sasaki M: **Heat protection by balling in the Japanese honeybee *Apis cerana japonica* as a defensive behavior against the hornet, *Vespa similima xanthoptera* (Hymenoptera: Vespidae).** *Experientia* 1987, **43**:1031–1032.
- Morse RA, Shearer DA, Bosh SR, Benton AW: **Observation on alarm substances in the genus *Apis*.** *J Api Res* 1967, **6**:113–118.
- Ono M, Garashi T, Ohno E, Sasaki M: **Unusual thermal defence by a honeybee against mass attack by hornets.** *Nature* 1995, **377**:334–336.
- Wang Zi L, Liu TT, Huang Zachary Y, Wu Xiao B, Wei Yu Y, Zhi Jiang Z: **Transcriptome analysis of the Asian honey Bee *apis cerana cerana*.** *PLoS one* 2012, **7**:e47954.
- Dyer FC, Seeley TD: **Interspecific comparisons of endothermy in honeybees (*Apis*): deviations from the expected size-related patterns.** *J Exp Biol* 1987, **122**:1–26.
- Seeley TD, Seeley RH: **Colony defense strategies of the honeybee in Thailand.** *Ecol Monogr* 1982, **52**:43–63.
- Seeley TD, Morse RA: **Dispersal behavior of honey Bee swarms.** *Psyche* 1977, **84**:199–209.
- Hadisoelilo S, Otis G: **Differences in drone cappings of *Apis cerana* and *Apis nigrocincta*.** *J Api Res* 1998, **37**:11–15.
- Atwal AS, Dhaliwal GS: **Some behavioural characteristics of *Apis indica* F and *Apis mellifera* L.** *Indian Bee J* 1969, **31**:83–90.
- Koetz AH: **Ecology, behaviour and control of *apis cerana* with a focus on relevance to the Australian incursion.** *Insects* 2013, **4**:558–592.
- Li JKFM, Zhang L, Zhang ZH, Pan YH: **Proteomics analysis of major royal jelly protein changes under different storage conditions.** *J Proteome Res* 2008, **7**:3339–3353.
- Miyamoto S: **Biological studies on Japanese Bees. X Differences in flower relationship between a Japanese and a European honeybee.** *Sci Rep Hyogo Univ Agric Ser agric Biol* 1958, **3**:99–101.
- Hongxia Ai XY, Richou H: **Occurrence and prevalence of seven bee viruses in *Apis mellifera* and *Apis cerana* apiaries in China.** *J Invertebr Pathol* 2012, **109**:160–164.
- Choe SE, Nguyen TT, Hyun BH, Noh JH, Lee HS, Lee CH, Kang SW: **Genetic and phylogenetic analysis of South Korean sacbrood virus isolates from infected honey bees (*Apis cerana*).** *Vet Microbiol* 2012, **157**:32–40.
- Smith CD, Zimin A, Holt C, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Fave MJ, Fernandes V, Gadau J, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Helmkamp M, Holley JA, Hu H, Vinięra AS, Johnson BR, Johnson RM, Khila A, Kim JW, Laird J, Mathis KA, Moeller JA, Muñoz-Torres MC, Murphy MC, et al: **Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*).** *Proc Natl Acad Sci U S A* 2011, **108**:5673–5678.
- Smith CR, Smith CD, Robertson HM, Helmkamp M, Zimin A, Yandell M, Holt C, Hu H, Abouheif E, Benton R, Cash E, Croset V, Currie CR, Elhaik E, Elsik CG, Favé MJ, Fernandes V, Gibson JD, Graur D, Gronenberg W, Grubbs KJ, Hagen DE, Vinięra AS, Johnson BR, Johnson RM, Khila A, Kim JW, Mathis KA, Munoz-Torres MC, Murphy MC, et al: **Draft genome of the red harvester ant *Pogonomyrmex barbatus*.** *Proc Natl Acad Sci U S A* 2011, **108**:5667–5672.
- Wurm Y, Wang J, Riba-Grognuz O, Corona M, Nygaard S, Hunt BG, Ingram KK, Falquet L, Nipitwattanaphon M, Gotzek D, Dijkstra MB, Oettler J, Comtesse F, Shih CJ, Wu WJ, Yang CC, Thomas J, Beaudoin E, Pradervand S, Flegel V, Cook ED, Fabbretti R, Stockinger H, Long L, Farmerie WG, Oakey J, Boomsma JJ, Pamilo P, Yi SV, Heinze J, et al: **The genome of the fire ant *Solenopsis invicta*.** *Proc Natl Acad Sci U S A* 2011, **108**:5679–5684.

30. Gadau J, Helmkamp M, Nygaard S, Roux J, Simola DF, Smith CR, Suen G, Wurm Y, Smith CD: **The genomic impact of 100 million years of social evolution in seven ant species.** *Trends Genet* 2012, **28**:14–21.
31. Gnerre S, MacCallum I, Przybylski D, Ribeiro F, Burton J, Walker B, Sharpe T, Hall G, Shea T, Sykes S, Berlin A, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB: **High-quality draft assemblies of mammalian genomes from massively parallel sequence data.** *Proc Natl Acad Sci U S A* 2010, **108**:1513–1518.
32. Elsik CG, Worley KC, Bennett AK, Beyre M, Camara F, Childers CP, de Graaf DC, Debysers G, Deng J, Devreese B, Elhaik E, Evans JD, Foster LJ, Graur D, Guigo R, HGSC production teams, Hoff KJ, Holder ME, Hudson ME, Hunt GJ, Jiang H, Joshi V, Khetani RS, Kosarev P, Kovar CL, Ma J, Maleszka R, Moritz RF, Muñoz-Torres MC, Murphy TD, et al: **Finding the missing honey bee genes: lessons learned from a genome upgrade.** *BMC Genomics* 2014, **15**:86.
33. Tan HW, Liu GH, Dong X, Lin RQ, Song HQ, Huang SY, Yuan ZG, Zhao GH, Zhu XQ: **The complete mitochondrial genome of the Asiatic cavity-nesting honeybee *apis cerana* (hymenoptera: apidae).** *PLoS one* 2011, **6**:e23008.
34. Bonasio R, Zhang G, Ye C, Mutti NS, Fang X, Qin N, Donahue G, Yang P, Li Q, Li C, Zhang P, Huang Z, Berger SL, Reinberg D, Wang J, Liebig J: **Genomic comparison of the ants *Camponotus floridanus* and *Harpegnathos saltator*.** *Science* 2010, **329**:1068–1071.
35. Suen G, Teiling C, Li L, Holt C, Abouheif E, Bombard-Bauer E, Bouffard P, Caldera EJ, Cash E, Cavanaugh A, Denas O, Elhaik E, Favé MJ, Gadau J, Gibson JD, Graur D, Grubbs KJ, Hagen DE, Harkins TT, Helmkamp M, Hu H, Johnson BR, Kim J, Marsh SE, Moeller JA, Muñoz-Torres MC, Murphy MC, Naughton MC, Nigam S, Overson R, et al: **The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle.** *PLoS Genet* 2011, **7**:e1002007.
36. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Reichsteiner A, Reese JT, Reid JG, Riddle M, Robertson HM, Romero-Severson J, Rosenberg M, Sackton TB, Sattelle DB, Schlüns H, Schmitt T, Schneider M, Schüler A, Schurko AM, Shuker DM, Simões ZL, Sinha S, Smith Z, Solovoyev V, Souvorov A, Springauf A, Stafflinger E, Stage DE, Stanke M, et al: **Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species.** *Science* 2010, **327**:343–348.
37. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W: **Molecular basis of base substitution hotspots in *Escherichia coli*.** *Nature* 1978, **274**:775–780.
38. Yi SV, Goodisman MA: **Computational approaches for understanding the evolution of DNA methylation in animals.** *Epigenetics Off J DNA Methylation Soc* 2009, **4**:551–556.
39. Bird AP: **DNA methylation and the frequency of CpG in animal DNA.** *Nucleic Acids Res* 1980, **8**:1499–1504.
40. Glastad KM, Hunt BG, Yi SV, Goodisman MA: **DNA methylation in insects: on the brink of the epigenomic era.** *Insect Mol Biol* 2011, **20**:553–565.
41. Okamura K, Matsumoto KA, Nakai K: **Gradual transition from mosaic to global DNA methylation patterns during deuterostome evolution.** *BMC Bioinform* 2010, **7**(11 Suppl):S2.
42. International Aphid Genomics Consortium: **Genome sequence of the pea aphid *Acyrtosiphon pisum*.** *PLoS Biol* 2010, **8**:e1000313.
43. Simola DF, Wissler L, Donahue G, Waterhouse RM, Helmkamp M, Roux J, Nygaard S, Glastad KM, Hagen DE, Viljakainen L, Reese JT, Hunt BG, Graur D, Elhaik E, Kriventseva EV, Wen J, Parker BJ, Cash E, Privman E, Childers CP, Muñoz-Torres MC, Boomsma JJ, Bombard-Bauer E, Currie CR, Elsik CG, Suen G, Goodisman MA, Keller L, Liebig J, Rawls A, et al: **Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality.** *Genome Res* 2013, **23**:1235–1247.
44. International Silkworm Genome C: **The genome of a lepidopteran model insect, the silkworm *Bombyx mori*.** *Insect Biochem Mol Biol* 2008, **38**:1036–1045.
45. Robertson HM: **The mariner transposable element is widespread in insects.** *Nature* 1993, **362**:241–245.
46. Lampe DJ, Witherspoon DJ, Soto-Adames FN, Robertson HM: **Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer.** *Mol Biol Evol* 2003, **20**:554–562.
47. Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, Farrell CM, Hart J, Landrum MJ, McGarvey KM, Murphy MR, O'Leary NA, Pujar S, Rajput B, Rangwala SH, Riddick LD, Shkeda A, Sun H, Tamez P, Tully RE, Wallin C, Webb D, Weber J, Wu W, DiCuccio M, Kitts P, Maglott DR, Murphy TD, Ostell JM: **RefSeq: an update on mammalian reference sequences.** *Nucleic Acids Res* 2014, **42**:D756–D763.
48. Brandi L, Cantarel IK, Sofia MC, Robb G, Parra Eric R, Barry M, Carson H, Alejandro Sánchez A, Mark Y: **MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes.** *Genome Res* 2008, **18**:188–196.
49. Consortium UP: **Ongoing and future developments at the Universal Protein Resource.** *Nucleic Acids Res* 2011, **39**:D214–D219.
50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology.** *The Gene Ontology Consortium. Nat Genet* 2000, **25**:25–29.
51. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28**:27–30.
52. Breed MD, Seth Perry P, Bjostad LB: **Testing the blank slate hypothesis: why honey bee colonies accept young bees.** *Insect Sociol* 2004, **51**:12–16.
53. Claudianos C, Ranson H, Johnson RM, Biswas S, Schuler MA, Berenbaum MR, Feyerisen R, Oakeshott JG: **A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee.** *Insect Mol Biol* 2006, **15**:615–636.
54. Devillers J, Pham-Delegue MH, Decourtye A, Budzinski H, Cluzeau S, Maurin G: **Structure-toxicity modeling of pesticides to honey bees.** *SAR QSAR Environ Res* 2002, **13**:641–648.
55. Stefanidou M, Athanasiou S, Koutselinis A: **The toxicology of honey bee poisoning.** *Vet Hum Toxicol* 2003, **45**:261–265.
56. Ruttner F: *Biogeography and Taxonomy of Honeybees.* Berlin New York: Springer-Verlag; 1988.
57. Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, Clark AG, Robinson GE: **Genes involved in convergent evolution of eusociality in bees.** *Proc Natl Acad Sci U S A* 2011, **108**:7472–7477.
58. Fischman BJ, Woodard SH, Robinson GE: **Molecular evolutionary analyses of insect societies.** *Proc Natl Acad Sci U S A* 2011, **108**(Suppl 2):10847–10854.
59. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family.** *Genome Res* 2006, **16**:1395–1403.
60. Robertson HM, Warr CG, Carlson JR: **Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*.** *Proc Natl Acad Sci U S A* 2003, **100**:14537–14542.
61. Benton R, Vannice KS, Gomez-Diaz C, Vossell LB: **Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*.** *Cell* 2009, **136**:149–162.
62. Robertson HM, Gadau J, Wanner KW: **The insect chemoreceptor superfamily of the parasitoid jewel wasp *Nasonia vitripennis*.** *Insect Mol Biol* 2010, **19**:121–136.
63. Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R: **Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction.** *PLoS Genet* 2010, **6**:e1001064.
64. Morel RKVMA: *NESTMATE RECOGNITION IN ANTS.* 1998.
65. Dani FR, Jones GR, Corsi S, Beard R, Pradella D, Turillazzi S: **Nestmate recognition cues in the honey Bee: differential importance of cuticular alkanes and alkenes.** *Chem Senses* 2005, **30**:477–489.
66. Châline N, Sandoz JC, Martin SJ, Ratnieks FL, Jones GR: **Learning and discrimination of individual cuticular hydrocarbons by honeybees (*apis mellifera*).** *Chem Senses* 2005, **30**:327–335.
67. de Brito Sanchez MG: **Taste perception in honey bees.** *Chem Senses* 2011, **36**:675–692.
68. Robinson GE: **Genomics and integrative analyses of division of labor in honeybee colonies.** *Am Nat* 2002, **160**(Suppl 6):S160–S172.
69. Wang Y, Brent CS, Fennern E, Amdam GV: **Gustatory perception and fat body energy metabolism are jointly affected by vitellogenin and juvenile hormone in honey bees.** *PLoS Genet* 2012, **8**:e1002779.
70. Kent LB, Walden KK, Robertson HM: **The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*.** *Chem Senses* 2008, **33**:79–93.
71. Peter J, Clyne CGW, John R, Carlson: **Candidate taste receptors in *Drosophila*.** *Science* 2000, **287**:1830–1834.
72. Montell C: **A taste of the *Drosophila* gustatory receptors.** *Curr Opin Neurobiol* 2009, **19**:345–353.
73. Tetsuya Miyamoto YC, Jesse S, Hubert A: **Identification of a *Drosophila* glucose receptor using Ca<sup>2+</sup> imaging of single chemosensory neurons.** *PLoS one* 2013, **8**:e56304.
74. Miyamoto T, Slone J, Song X, Amrein H: **A fructose receptor functions as a nutrient sensor in the *Drosophila* brain.** *Cell* 2012, **151**:1113–1125.

75. Suh GS, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ: **A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*.** *Nature* 2004, **431**:854–859.
76. Stange G: **The response of the honeybee antennal CO<sub>2</sub>-receptors to N<sub>2</sub>O and Xe.** *J Comp Physiol* 1973, **86**:139–158.
77. Ai M, Min S, Grosjean Y, Leblanc C, Bell R, Benton R, Suh GS: **Acid sensing by the *Drosophila* olfactory system.** *Nature* 2010, **468**:691–695.
78. Wanner KW, Nichols AS, Walden KK, Brockmann A, Luetje CW, Robertson HM: **A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid.** *Proc Natl Acad Sci U S A* 2007, **104**:14383–14388.
79. Hallem EA, Dahanukar A, Carlson JR: **Insect odor and taste receptors.** *Annu Rev Entomol* 2006, **51**:113–135.
80. SCHMIDT J: **Toxinology of venoms from the honeybee genus *Apis*.** *Toxicon* 1995, **33**:917–927.
81. Raffiudin R, Crozier RH: **Phylogenetic analysis of honey bee behavioral evolution.** *Mol Phylogenet Evol* 2007, **43**:543–552.
82. Krieger J, Klink O, Mohl C, Raming K, Breer H: **A candidate olfactory receptor subtype highly conserved across different insect orders.** *J Comp Physiol A* 2003, **189**:519–526.
83. Larsson MC, Domingos AI, Jones WD, Chiappe ME, Amrein H, Vosshall LB: **Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction.** *Neuron* 2004, **43**:703–714.
84. Jones WD, Nguyen TA, Kloss B, Lee KJ, Vosshall LB: **Functional conservation of an insect odorant receptor gene across 250 million years of evolution.** *Curr Biol* 2005, **15**:R119–R121.
85. Benton R, Sachse S, Michnick SW, Vosshall LB: **Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo.** *PLoS Biol* 2006, **4**:e20.
86. Yao CA, Ignell R, Carlson JR: **Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna.** *J Neurosci* 2005, **25**:8359–8367.
87. Hallem EA, Carlson JR: **Coding of odors by a receptor repertoire.** *Cell* 2006, **125**:143–160.
88. Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GS, Benton R: **Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems.** *J Neurosci* 2011, **31**:13357–13375.
89. Rytz R, Croset V, Benton R: **Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in *Drosophila* and beyond.** *Insect Biochem Mol Biol* 2013, **43**:888–897.
90. Chao Liu RJP, Jonathan D, Bohbot JD, Jones Patrick L, Wang G, Zwiebel LJ: **Distinct olfactory signaling mechanisms in the malaria vector mosquito *Anopheles gambiae*.** *PLoS Biol* 2010, **8**:e1000467.
91. Olivier V, Monsempe C, François MC, Poivet E, Jacquin-Joly E: **Candidate chemosensory ionotropic receptors in a Lepidoptera.** *Insect Mol Biol* 2010, **20**:189–199.
92. Wiegmann BM, Trautwein MD, Kim JW, Cassel BK, Bertone MA, Winterton SL, Yeates DK: **Single-copy nuclear genes resolve the phylogeny of the holometabolous insects.** *BMC Biol* 2009, **7**:34.
93. Evans JD, Spivak M: **Socialized medicine: Individual and communal disease barriers in honey bees.** *J Invertebr Pathol* 2010, **103**:S62–S72.
94. Li J, Qin H, Wu J, Sadd BM, Wang X, Evans JD, Peng W, Chen Y: **The prevalence of parasites and pathogens in Asian honeybees *Apis Cerana* in china.** *Plos one* 2012, **7**:e47955.
95. Yoo MS, Yoon BS: **Incidence of honey bee disease in Korea 2009.** *Korean J Apic* 2009, **24**:273–278.
96. Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, Kanost M, Thompson GJ, Zou Z, Hultmark D: **Immune pathways and defence mechanisms in honey bees *Apis mellifera*.** *Insect Mol Biol* 2006, **15**:645–656.
97. Park D, Jung JW, Lee MO, Lee SY, Kim B, Jin HJ, Kim J, Ahn YJ, Lee KW, Song YS, Hong S, Womack JE, Kwon HW: **Functional characterization of naturally occurring melittin peptide isoforms in two honey bee species, *Apis mellifera* and *Apis cerana*.** *Peptides* 2014, **53**:185–193.
98. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, Williams T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nat Protoc* 2013, **8**:1494–1512.
99. Lukashin AV, Borodovsky M: **GeneMark.hmm: new solutions for gene finding.** *Nucleic Acids Res* 1998, **26**:1107–1115.
100. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase Update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110**:462–467.
101. Grant JR, Stothard P: **The CGView Server: a comparative genomics tool for circular genomes.** *Nucleic Acids Res* 2008, **36**:W181–W184.
102. Altschul SFGW, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
103. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.** *Nat Protoc* 2012, **7**:562–578.
104. Slater GS, Birney E: **Automated generation of heuristics for biological sequence comparison.** *BMC Bioinformatics* 2005, **6**:31.
105. Lee E, Harris N, Gibson M, Chetty R, Lewis S: **Apollo: a community resource for genome annotation editing.** *Bioinformatics* 2009, **25**:1836–1837.
106. St Pierre SE, Ponting L, Stefancsik R, McQuilton P, FlyBase Consortium: **FlyBase 102—advanced approaches to interrogating FlyBase.** *Nucleic Acids Res* 2014, **42**:D780–D788.
107. Fischer S, Brunk BP, Chen F, Gao X, Harb OS, Iodice JB, Shanmugam D, Roos DS, Stoeckert CJ Jr: **Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes into new Ortholog Groups.** *Current Protocols in Bioinformatics / Editorial Board, Andreas D Baxevanis* 2011, Chapter 6:Unit 6 12 11–19.
108. Stephen F, Altschul TLM, Alejandro A, Schäffer Jinghui Z, Zheng Zhang WM, David J, Lipman: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389–3402.
109. Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic Acids Res* 2012, **40**:290–301.
110. MOTIF. In <http://www.genome.jp/tools/motif/>.
111. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947–2948.
112. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**:2731–2739.
113. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12**:323.
114. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W: **PipMaker—a web server for aligning two genomic DNA sequences.** *Genome Res* 2000, **10**:577–586.
115. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database C: **The sequence read archive: explosive growth of sequencing data.** *Nucleic Acids Res* 2012, **40**:D54–D56.

doi:10.1186/1471-2164-16-1

Cite this article as: Park et al.: Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genomics* 2015 **16**:1.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

