Whole Genome Sequencing and Assembly of the Asian Honey Bee *Apis dorsata*

Sara Oppenheim[1,†], Xiaolong Cao[2,†], Olav Rueppel[3], Sasiprapa Krongdang[4], Patcharin Phokasem[4], Rob DeSalle[1], Sara Goodwin[5], Jinchuan Xing[2], Panuwan Chantawannakul[4,*], and Jeffrey A. Rosenfeld[1,6,7,*]

[1]Sackler Institute for Comparative Genomics, American Museum of Natural History

[2]Department of Genetics, Human Genetic Institute of New Jersey, Rutgers, The State University of New Jersey

[3]Biology Department, University of North Carolina at Greensboro

[4]Department of Biology & Environmental Science Research Center (ESRC), Faculty of Science, Chiang Mai University, Thailand

[5]Cold Spring Harbor Laboratory

[6]Rutgers Cancer Institute of New Jersey

[7]Department of Pathology, Robert Wood Johnson Medical School

[†]These authors contributed equally to this work.

*Corresponding authors: E-mails: panuwan@gmail.com; jeffrey.rosenfeld@rutgers.edu.

Accepted: December 13, 2019

Data deposition: This project has been deposited at NCBI BioProject under the accession PRJNA577936.

Abstract

The Asian honey bee (*Apis dorsata*) is distinct from its more widely distributed cousin *Apis mellifera* by a few key characteristics. Most prominently, *A. dorsata,* nest in the open by forming a colony clustered around the honeycomb, whereas *A. mellifera* nest in concealed cavities. Additionally, the worker and reproductive castes are all of the same size in *A. dorsata*. In order to investigate these differences, we performed whole genome sequencing of *A. dorsata* using a hybrid Oxford Nanopore and Illumina approach. The 223 Mb genome has an N50 of 35 kb with the largest scaffold of 302 kb. We have found that there are many genes in the *dorsata* genome that are distinct from other hymenoptera and also large amounts of transposable elements, and we suggest some candidate genes for *A. dorsata's* exceptional level of defensive aggression.

Key words: bees, honey, *Apis*.

Introduction

The Asian honey bee (*Apis dorsata)* is an important pollinator and source of honey throughout its natural range in Southeast and East Asia (Sheppard and Berlocher 1989; Thapa 2006; Chantawannakul et al. 2018). This species is distinct from the European honey bee (*Apis mellifera*) both morphologically and behaviorally. Unlike *A. mellifera*, which is now found throughout the world, *A. dorsata* is found only in Asia (Chantawannakul et al. 2018). *Apis dorsata* is much larger than *A. mellifera*, and *A. dorsata* workers are almost twice as long as *A. mellifera* workers. In addition, *A. dorsata* is distinguished by a lack of body size variation between castes as in other honey bees.

The flavor of *A. dorsata* honey differs from that of *A. mellifera*. Biochemical assays of honey have demonstrated variation in antibacterial activity, protein content, and glucose content (Iftikhar et al. 2011; Pattamayutanon et al. 2015). While much of this variation can be explained by the nectar source that bees use to produce the honey, honey characteristics also differ based on honeybee species.

From an anthropocentric point of view, the most striking difference between *A. dorsata* and other honey bees is in their nesting behavior. *Apis dorsata* do not nest in cavities but instead build exposed nests that hang from tree branches or cliffs. The exposed comb is covered at all times by a blanket of up to 100,000 worker bees. Unlike species that nest in

enclosed spaces like tree trunks and rock crevices, *A. dorsata* colonies are unwilling to live in the traditional Langstroth hives used by commercial beekeepers (Koeniger et al. 2010). This distinct lifestyle, together with an exceptional level of defensive aggression (Breed et al. 2004), has limited the ability of humans to domesticate the species and to commercially produce its honey (Koeniger et al. 2010).

To fully understand the genetic basis of important *Apis* traits requires comparative data from a variety of *Apis* species. In this study, we combine short-read and long-read sequencing technologies to generate a high-quality assembly of the *A. dorsata*.

Materials and Methods

Sample Collection

Multiple *A. dorsata* drones were collected by on the campus of Chiang Mai University (Chiang Mai, Thailand). The drones were collected in the early evening when they visited artificial light sources near the University's gymnasium. Specimens were sacrificed immediately. One exemplar was kept intact; each of the remaining specimens was cut into three sections and placed in RNAlater Stabilization Solution (Invitrogen). Tissues were held in RNAlater at 4C for at least 24 h to allow the solution to saturate the tissue, then shipped to the American Museum of Natural History (AMNH). The exemplar specimen was accessioned into the Ambrose Monell Cryo Collection at AMNH. The remaining specimens were sent to the Cold Spring Harbor Laboratories for DNA extraction and sequencing.

DNA Extraction and Sequencing

Individual heads and thoraces were chopped into small pieces and placed in 1.7 ml centrifuge tubes (~25 mg of tissue per tube). DNA was isolated using the MagAttract HMW DNA kit (Qiagen), following the "Manual Purification of High-Molecular-Weight Genomic DNA from Fresh or Frozen Tissue" protocol. Several tubes were pooled after extraction.

Extracted DNA was sheared to 20 kb using a Covaris G-tube following manufacturer's recommendations. The sheared DNA was prepared for Nanopore sequencing using the Ligation Sequencing 1D kit (SQK-LSK108). Briefly, 1–1.5 µg of fragmented DNA was end-repaired with the NEB FFPE repair kit, followed by end repair and A-tailing with the NEB Ultra II end-prep kit. Following an Ampure clean-up step, prepared fragments were ligated to ONT-specific adapters with the NEB blunt/TA master mix kit. After a final Ampure clean-up, the library was loaded on to a MinION R9.0 SpotON flowcell as per manufacturer's instructions. The flowcell was sequenced with standard parameters for 2 days. The resulting sequence data were processed with the Albacore pipeline (Oxford Nanopore).

Illumina libraries were prepared with the Kapa hyper prep kit, following manufactures guidelines. Prepared libraries pooled and sequenced on an Illumina NextSeq 500. The sequencing was done with PE150 High Output format.

Genome Assembly and Annotation

The Nanopore and Illumina reads were assembled together using the SPAdes software (Antipov et al. 2016). Structural annotation of the assembled genome was performed with MAKER as described in Campbell et al. (2014). Briefly, in the first round, transcript sequences from a previous *A. dorsata* study (NCBI BioBroject PRJNA174631) were provided to MAKER as EST evidence. Protein evidence was obtained using protein sequences from closely related hymenopteran species from NCBI, including *A. dorsata* (PRJNA174631), *A. mellifera* (PRJNA477511), *Apis cerana* (PRJNA324433), *Apis florea* (PRJNA45871), *Polistes dominula* (PRJNA307991), and *Papilio canadensis* PRJNA301748). In addition, the full Swiss-Prot database was included in the analysis (Boutet et al. 2016). Transcript sequences from the non-*dorsata* species were provided to MAKER as alternative EST evidence. Repetitive elements were identified using (Chen 2004) (A.F.A. Smit).

Coding sequences were predicted in the MAKER pipeline with Exonerate (Slater and Birney 2005) and BLAST (Camacho et al. 2009). The output of the first MAKER run was used to train the gene predictors SNAP (Korf 2004) and AUGUSTUS (Keller et al. 2011). The trained models were then provided to MAKER and a new set of training models generated from the result used for another round of running MAKER. Another sets of SNAP and AUGUSTUS models were trained with the new outputs. The assembled contig sequences were used to train GeneMark-ES (Lomsadze et al. 2005). The SNAP, AUGUSTUS, and GeneMark-ES models were used in a final MAKER run, and the resulting predicted coding sequences used in all subsequent analyses. The completeness of the assembly and annotation were analyzed with BUSCO. Simão et al. (2015) using the Insecta data set (1,658 single-copy conserved genes) as a reference. We used the assembly-stats tool (https://github.com/sanger-pathogens/assembly-stats) to determine assembly statistics for the *dorsata* and other *Apis* genomes.

Functional Annotation

Annotation of the predicted genes was undertaken with a variety of tools. First, we conducted BLASTp (BLAST+ 2.8.1) searches against all Hymenoptera entries in the NCBI nr database (Wheeler et al. 2007) (default parameters except that we used the -taxidlist option to restrict the search to species within NCBI Taxonomy ID 7399 [Hymenoptera]; nr database accessed on 09/09/2019). For sequences that had no hit, we did a second BLASTp search against all sequences in the NCBI nr database. We used InterProScan 5 (Jones et al. 2014) to

classify genes into protein families and identify functional domains(Mitchell et al. 2015). The results were imported into Blast2GO (Conesa et al. 2005), where we mapped them to Gene Ontology (GO) terms (Harris et al. 2004).

We also compared the content of the assembled *A. dorsata* genome with that of *A. mellifera*.

Ortholog Identification

We used OrthoPipe, a stand-alone pipeline version of OrthoDB 2.3.1 (Kriventseva et al. 2015), to classify the protein-coding genes from *A. dorsata, A. mellifera*, *A. cerana,* and *A. florea* into orthologous clusters. To evaluate the resulting clusters, we used the reference topology presented by Han et al. (2012).

Identification of Species-Specific *A. dorsata* Genes

Species-specific genes were narrowly defined as sequences that lacked orthologs in any other *Apis* species. A more stringent definition requires a lack of orthologs in any published data set (Tautz and Domazet-Lošo 2011), and we used BLASTp searches against NCBI's nr database to test whether our narrowly defined species-specific genes were legitimate orphans.

Sequence Comparison of *A. dorsata* Genes to Others

To examine how genes from *A. dorsata* have diverged from those of the other examined *Apis* species, we generated multiple sequence alignments with MAFFT (v7.312) (Katoh and Standley 2013). All alignments were generated using "high accuracy" mode (–maxiterate 1000 –localpair). The resulting alignments were used to generate cladograms to visualize the relationship between different genes. We used FastTree (v 2.1.10) (Price et al. 2009) to approximately-maximum-likelihood phylogenetic trees, with the options –nome (to compute minimum-evolution bootstrap support values) and -slow (to conduct exhaustive searches).

Functional Enrichment Tests

Because raw frequency-based comparisons are difficult to interpret in evolutionary terms, we used enrichment tests to examine whether any gene functions were overrepresented in the genome of *A. dorsata* as compared with *A. mellifera*. We conducted Fisher's exact tests (q-value < 0.01) to identify GO terms or InterPro signatures enriched in a test set of annotated genes relative to a reference set. Since GO annotations, which are assigned transitively via BLAST hits, can sometimes lead to incorrect functional annotation (Yon Rhee et al. 2008), we used InterPro signatures as the primary source of functional annotation and comparison.

Results and Discussion

Genome Assembly

Our assembly results were consistent with previous Apis genome sequencing projects (fig. 1): The total assembly size was 224 Mb (vs 225–230 Mb in other apid bees, Consortium 2006; Wang et al. 2013; Park et al. 2015), with a contig N50 of 28 kb (largest contig: 302 kb). This assembly puts the *A. dorsata* on the same scale as the *A. florea* and *A. cerana* genomes, which have contig N50s of 20–30 kb and largest contigs of over 300 kb. In contrast to the previous *A. dorsata* assembly from 2013 (PRJNA174631), we have an improved assembly from a lower level of sequencing coverage. None of these genomes compared with *A. mellifera*, an intensely studied model organism with a very high-quality contiguous genome assembly. BUSCO analysis of our assembly showed that the assembly was 98.9% complete. We identified a wide variety of repetitive elements (supplementary table 1, Supplementary Material online), and these comprised ~7% of the genome. Using MAKEr, we identified 13,517 protein coding genes (vs 12,145–12,940 genes in other apids, Consortium 2006; Wang et al. 2013; Park et al. 2015).

Functional Annotation

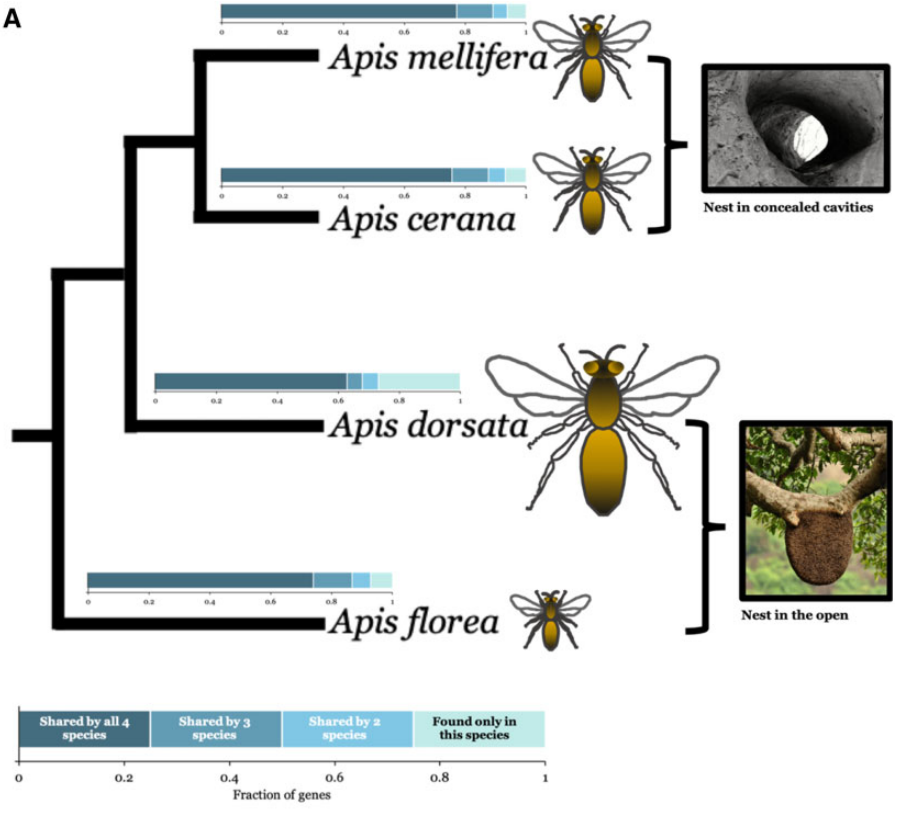
More than 80% (11,264) of the predicted *A. dorsata* genes had matches to other Hymenopteran sequences. The remaining 2,253 genes had no hits to other hymenoptera (or only had hits to *A. dorsata* sequences). When we blasted these genes against the entire nr database, 182 of them had hits but 2,071 still had no hit. The results of InterProScan analyses were similar: 1,956 genes had no InterPro signature. The overlap between these annotations left us with 1,251 genes that lacked functional annotation.

Ortholog Identification

We identified 3,451 single copy universal ortholog clusters that contained one gene from each of the four species included in ortholog analysis. Consistent with the idea that single copy universal orthologs represent highly conserved genes, every *A. dorsata* single copy universal ortholog had a hit to other Hymenoptera in the nr database. Seventy-three percent of the *A. dorsata* genes (9,900) had orthologs in at least one other species, versus 93% of the *A. mellifera* genes. In the specific comparison between *A. mellifera* and *A. dorsata*, 22% (3,034) of the *A. dorsata* genes had no ortholog in *A. mellifera*, and 2,177 of these had no ortholog in any of the included *Apis* species and no hit to any other hymenopteran in our BLAST searches (fig. 1).

Species-Specific *A. dorsata* Genes

Some 3,617 *A. dorsata* genes did not have orthologs in any of the included *Apis* species. We used BLAST searches against all



B

Organism	Assembly	Genome size (Mb)	Percent GC	Contig N50	Largest contig	Predicted genes	Busco % missing	Coverage
Apis florea	PRJNA45871	230.49	34	24,306	306,154	12,145	0.4	20.5x
Apis cerana	PRJNA324433	228.32	34.9	30,181	334,921	12,940	0.8	152x
Apis mellifera	PRJNA477511	225.25	32.6	5,382,476	13,404,451	12,337	0.5	192x
Apis dorsata (2013)	PRJNA174631	219.17	31.9	8,020	102,139	11,378	1.4	60x
Apis dorsata	PRJNA577936	223.57	31.9	28,586	302,553	13,776	1.1	38x

Fig. 1.—(*A*) Phylogenetic relationship between the four *Apis* species used for ortholog identification, relative sizes of drones are shown. Inset bar charts indicate the fraction of genes from each species that were orthologous at different levels: Universal orthologs (present in all four species), species-specific genes (those found only in one species), and others. (*B*) Assembly statistics for *Apis* species.

hymenopteran sequences in the nr database to determine whether these were species-specific genes. We found that 1,424 of these genes had a BLAST hit to another hymenopteran species, and an additional 75 had a hit to a nonhymenopteran species. This left 2,118 genes that appear to be either pseudogenes or species-specific *A. dorsata* genes because they had no hit to any sequence in the nr database.

Divergence in 5-Hydroxytryptamine Receptor 2A (5-HT2a)

We were particularly interested in identifying genes that might contribute to *A. dorsata*'s behavioral divergence from other *Apis* species. We focused on receptors for the neurotransmitter serotonin (5-hydroxytryptamine, 5-HT) because of serotonin's role as a neuromodulator (Turlejski 1996). The effects of serotonin depend on receptor-specific binding, and in *Drosophila* and other insects 5-HT receptors have been associated with variation in locomotion, feeding behavior, learning and memory, and aggression (Majeed et al. 2016; Huser et al. 2017).

We identified four 5-HT2A receptors in *A. dorsata* (fig. 2*A* and *B*). Although three of these had orthologs in other *Apis* species, one was sufficiently divergent that it did not cluster with any other *Apis* genes. Because serotonin has a stimulatory effect on aggression in some insects (Bubak et al. 2019), this gene is an attractive candidate for future studies on the genetic basis of aggression in *A. dorsata*.

Functional Enrichment

Single Copy Universal Orthologs versus All Other Genes

These highly conserved genes were enriched for functions related to development, including a growth factor responsible

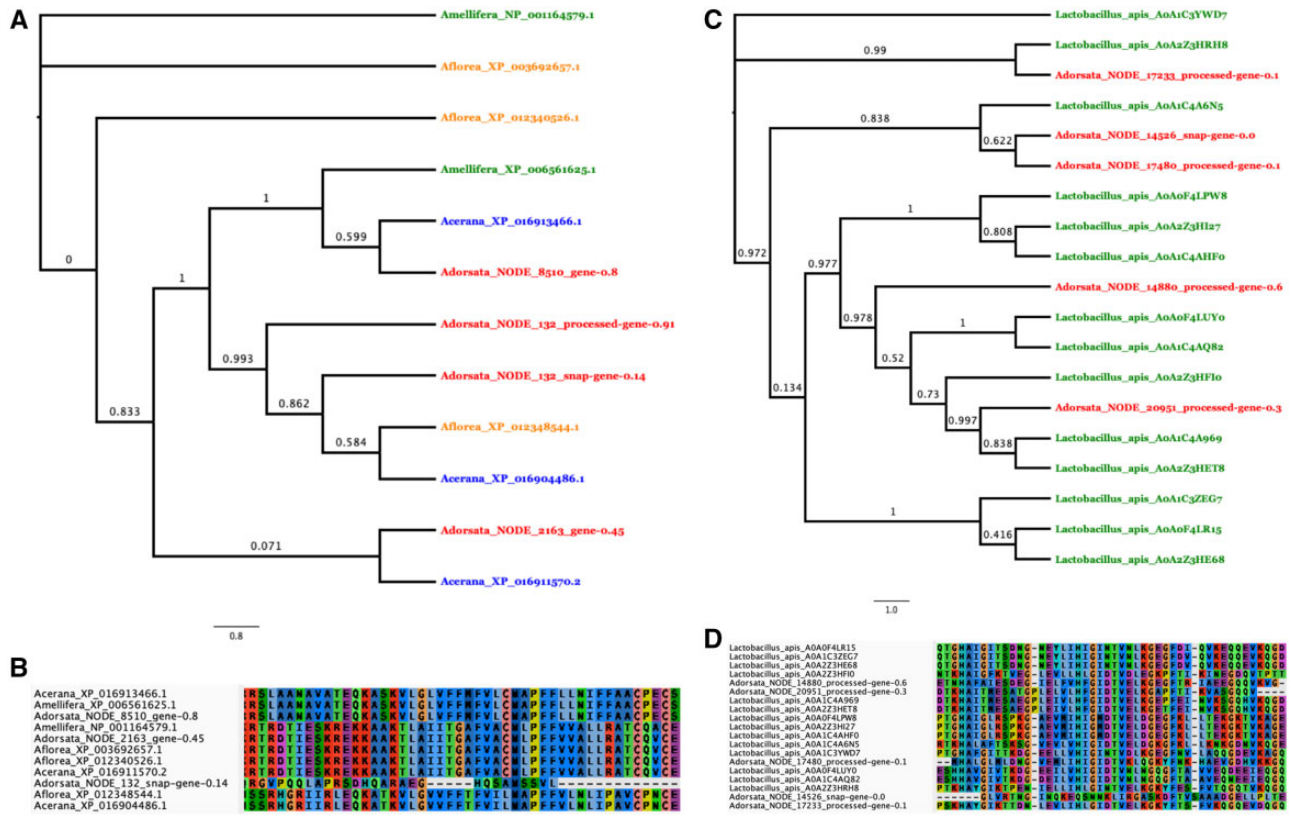


Fig. 2.—(*A*) Cladogram with bootstrap support values and (*B*) sequence alignment of putative 5-hydroxytryptamine (serotonin) receptor 2A (5-HT2A) genes from *Apis dorsata* and other *Apis* species. (*C*) Cladogram with bootstrap support values and (*D*) sequence alignment of *A. dorsata* species-specific genes with InterPro signature IPR001127 (Lactobacillus PTS carbohydrate transport). *Lactobacillus apis* genes are from the gut of *Apis mellifera*. Both cladograms (*A* and *C*) are based on full-length gene alignments; in the interests of space, only a representative section of the full alignment is shown (*B* and *D*).

for cell proliferation and differentiation (IPR001111), and juvenile hormone binding proteins (JHBP; IPR010562), which regulates embryogenesis, larval development, and reproductive maturation in the adult forms (supplementary table 2, Supplementary Material online).

Species-Specific A. mellifera Genes versus All Other Genes

The *A. mellifera* genes that lacked *A. dorsata* homologs were enriched for signal transduction functions. Several enriched signatures related to small GTPases, which are critical components in cellular signal transduction pathways,

Species-Specific A. dorsata Genes versus All Other Genes

The InterPro terms that were enriched in the *A. dorsata* genes that lacked *A. mellifera* homologs fell into two main functional categories. First, there were 28 genes with Ty1/copia-like retrotransposon signatures. The "domestication" of transposable elements can lead to the emergence of species-specific genes by providing a source of biochemically active elements such as transcription factor-binding sites, and by generating genomic rearrangements (Toll-Riera et al. 2008; Kaessmann 2010; Lisch 2013; Göke and Ng 2016), and the overrepresentation of transposable element-related signatures in the genome of *A. dorsata*, suggests that the evolution of "new" genes may have played a role in the behavioral diversification of *A. dorsata* from other honeybees.

Second, there were 23 sequences that appear to be part of *A. dorsata*'s microbiome assemblage. These all had high-scoring hits to bacterial species in the nr database, and when aligned with bacterial sequences known to occur in Hymenopteran microbiomes showed high levels of similarity. The enriched InterPro signature IPR001127, a domain that is associated with the carbohydrate transport system in *Lactobacillus* bacteria, was found in six of the species-specific *A. dorsata* genes. We compared the *A. dorsata* sequences to *Lactobacillus apis* genes from the microbiome of *A. mellifera* (fig. 2*C* and *D*).

In honeybees, the microbiome of worker bees is of especial importance. All worker gut microbiomes studied to date contain abundant carbohydrate-processing genes, but the exact components of the microbiome community show variation between honeybee strains and species (Engel et al. 2012). This variation suggests that some of the dietary differences

between honeybees in different localities (in their use of pollen/nectar from different plant species) might be controlled by microbiome components (Disayathanoowat et al. 2012; Saraithong et al. 2015, 2017). Previous research has shown that *Lactobacillus* species, which were the most common top hits for our *A. dorsata*-only sequences, play a major role in carbohydrate digestion (Lee et al. 2015; Moran 2015). Thus, the specific microbiome components present in different colonies or species might determine their nutritional ecology and/or their ability to cope with dietary toxins and could explain the geographic distribution of *A. dorsata* as a case of coevolution between bees, bacteria, and flowering plants.

The microbiome of *A. dorsata* may also be involved in the distinctive flavor of the honey they produce. *Lactobacillus* species are thought to prevent the fermentation of stored honey by retarding the growth of fermenting yeasts (Olofsson and Vásquez 2008). Variation in the activity of *Lactobacillus* strains in honeybee microbiomes can affect the flavor of honey since flavor is partially dependent on lactic acid bacterial metabolites (Olofsson et al. 2016). Thus, the "terroir" of different honeys (Marchese and Flottum 2013) may result as much from the honeybee microbiome (Tajabadi et al. 2011) as from the particular locale where the honeybees forage.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This research work was partially supported by Chiang Mai University. This research used resources from the Rutgers Discovery Informatics Institute, which are supported by Rutgers and the State of New Jersey. Some of this work was conducted on the BIOMIX compute cluster, made possible through funding from Delaware INBRE (NIGMS P20GM103446), the State of Delaware, and the Delaware Biotechnology Institute. Dr Jeffrey Rosenfeld is supported by Cancer Center Support grant (2P30-CA072720-20) and the Systems Biology program at the Cancer Institute of New Jersey. We would like to thank Corey Smith at the American Museum of Natural History for Assistance in navigating the government paperwork for the import of the specimens.

Literature Cited

Antipov D, Korobeynikov A, McLean JS, Pevzner PA. 2016. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32(7):1009–1015.

Boutet E, et al. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. In Plant Bioinformatics, New York, NY: Humana Press. pp. 23–54.

Breed MD, Guzmán-Novoa E, Hunt GJ. 2004. Defensive behavior of honey bees: organization, genetics, and comparisons with other bees. Annu Rev Entomol. 49(1):271–298.

Bubak AN, et al. 2019. Sex differences in aggression: differential roles of 5-HT2, neuropeptide F and tachykinin. PLoS One 14(1):e0203980.

Camacho C, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10(1):421.

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. Curr Protoc Bioinformatics. 48:4.11.1–39.

Chantawannakul P, Williams G, Neumann P. Editors. 2018. Asian beekeeping in the 21st century. Singapore: Springer.

Chen N. 2004. Using Repeat Masker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 5(1):4–10.

Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676.

Consortium HGS. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443:931.

Disayathanoowat T, Young JPW, Helgason T, Chantawannakul P. 2012. T-RFLP analysis of bacterial communities in the midguts of *Apis mellifera* and *Apis cerana* honey bees in Thailand. FEMS Microbiol Ecol. 79(2):273–281.

Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. Proc Natl Acad Sci U S A. 109(27):11002–11007.

Göke J, Ng HH. 2016. CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. EMBO Rep. 17(8):1131–1144.

Han F, Wallberg A, Webster MT. 2012. From where did the Western honeybee (*Apis mellifera*) originate? Ecol Evol. 2(8):1949–1957.

Harris MA, et al.; Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32(Database issue):D258–261.

Huser A, et al. 2017. Anatomy and behavioral function of serotonin receptors in *Drosophila melanogaster* larvae. PLoS One 12(8): e0181865.

Iftikhar F, Masood MA, Waghchoure ES. 2011. Comparison of *Apis cerana, Apis dorsata, Apis florea* and *Apis mellifera*: honey from different areas of Pakistan. Asian J Exp Biol Sci. 2:399–403.

Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res. 20(10):1313–1326.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27(6):757–763.

Koeniger N, Koeniger G, Tingek S. 2010. Honey bees of Borneo: exploring the centre of Apis diversity. Borneo (Asia): Natural History Publications.

Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5(1):59.

Kriventseva EV, et al. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 43(D1):D250–D256.

Lee FJ, Rusch DB, Stewart FJ, Mattila HR, Newton IL. 2015. Saccharide breakdown and fermentation by the honey bee gut microbiome. Environ Microbiol. 17(3):796–815.

Lisch D. 2013. How important are transposons for plant evolution? Nat Rev Genet. 14(1):49–61.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33(20):6494–6506.

Majeed ZR, et al. 2016. Modulatory action by the serotonergic system: behavior and neurophysiology in *Drosophila melanogaster*. Neural Plast. 2016:1–23.
Marchese CM, Flottum K. 2013. Honey Connoisseur: Selecting, Tasting, and Pairing Honey, With a Guide to More Than 30 Varietals. Hachette UK: Black Dog & Leventhal Publishers.
Mitchell A, et al. 2015. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43(D1):D213–D221.
Moran NA. 2015. Genomics of the honey bee microbiome. Curr Opin Insect Sci. 10:22–28.
Olofsson TC, et al. 2016. Lactic acid bacterial symbionts in honeybees—an unknown key to honey's antimicrobial and therapeutic activities. Int Wound J. 13(5):668–679.
Olofsson TC, Vásquez A. 2008. Detection and identification of a novel lactic acid bacterial flora within the honey stomach of the honeybee *Apis mellifera*. Curr Microbiol. 57(4):356–363.
Park D, et al. 2015. Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. BMC Genomics 16:1.
Pattamayutanon P, et al. 2015. Biomedical activity and related volatile compounds of Thai honeys from 3 different honeybee species. J Food Sci. 80(10):M2228–M2240.
Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26(7):1641–1650.
Saraithong P, Li Y, Saenphet K, Chen Z, Chantawannakul P. 2015. Bacterial community structure in *Apis florea* larvae analyzed by denaturing gradient gel electrophoresis and 16S rRNA gene sequencing. Insect Sci. 22(5):606–618.
Saraithong P, Li Y, Saenphet K, Chen Z, Chantawannakul P. 2017. Midgut bacterial communities in the giant Asian honeybee (*Apis dorsata*) across 4 developmental stages: a comparative study. Insect Sci. 24(1):81–92.
Sheppard W, Berlocher S. 1989. Allozyme variation and differentiation among four Apis species. Apidologie 20(5):419–431.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212.
Slater GS, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6(1):31.
Tajabadi N, et al. 2011. Detection and identification of Lactobacillus bacteria found in the honey stomach of the giant honeybee *Apis dorsata*. Apidologie 42(5):642–649.
Tautz D, Domazet-Lošo T. 2011. The evolutionary origin of orphan genes. Nat Rev Genet. 12(10):692–702.
Thapa R. 2006. Honeybees and other insect pollinators of cultivated plants: a review. J Inst Agric Anim Sci. 27:1–23.
Toll-Riera M. et al. 2008. Origin of primate orphan genes: a comparative genomics approach. Mol Biol Evol 26(3):603–612.
Turlejski K. 1996. Evolutionary ancient roles of serotonin: long-lasting regulation of activity and development. Acta Neurobiol Exp. 56(2): 619–636.
Wang AR, et al. 2013. Complete mitochondrial genome of the dwarf honeybee, *Apis florea* (Hymenoptera: Apidae). Mitochondrial DNA 24(3):208–210.
Wheeler DL, et al. 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 36(Database):D13–D21.
Yon Rhee S, Wood V, Dolinski K, Draghici S. 2008. Use and misuse of the gene ontology annotations. Nat Rev Genet. 9(7):509–515.

Associate editor: Nancy Moran