PP-378 [Bee Biology]
Construction and annotation of the Apis cerana full-length transcriptome

Kunze Li[1], Jianfeng Qiu[1], Jianfeng Qiu[2], Jianfeng Qiu[3], Dafu Chen[1], Dafu Chen[2], Dafu Chen[3], Rui Guo[1], Rui Guo[2], Rui Guo[3]
[1]College of Bee Science and Biomedicine, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]National & Local United Engineering Laboratory of Natural Biotoxin, Fuzhou 350002, China
[3]Apitherapy Research Institute of Fujian Agriculture and Forestry University, Fuzhou 350002, China

Apis cerana is a major bee species used in beekeeping practice in many Asian countries including. Third-generation sequencing technology, such as Nanopore long-read sequencing, has been applied for construction of chromosome-scale genome and full-length transcriptome in an array of animals and plants. However, Nanopore sequencing-based study on honeybees is currently limited. In the present study, the gut tissues of A. cerana worker 4-, 5-, and 6-day-old larvae were prepared and subjected to Nanopore sequencing, followed by identification of full-length transcripts as well as construction and annotation of the full-length transcriptome. Here, from tehe 4-, 5-, and 6-day-old larval guts, 14474634, 10461827, and 11890978 raw reads were obtained, including 27815, 21781, and 20004 non-redundant full-length transcripts, with the N50 lengths of 1900, 1961, and 2294 bp. A total of 40562 non-redundant full-length transcripts were identified, and 35415, 24646, 34054, and 23053 transcripts could be annotated to the Nr, KOG, eggNOG, and GO databases, respectively. The species with the highest number and proportion of annotated full-length transcripts was A. cerana (20310 transcripts, accounting for 57.35%), followed by A. mellifera (4686 transcripts, accounting for13.23%), A. dorsata (2536 transcripts, 7.16%) and A. florea (2079, accounting for 5.87%) in the Nr database. The non-redundant full-length transcripts were annotated to 25 functional categories such as unknown function, and post-translational modification, protein turnover and molecular chaperones in the eggNOG database, 25 functional categories such as general function prediction only and signal transduction mechanisms in the KOG database, 50 functional terms relevant to three major categories of biological process, cellular component and molecular function in the GO database, and 196 pathways such as ribosome and RNA transduction in the KEGG database. A total of 2301 lncRNAs with high confidence were identified, involving four types including sense lncRNA, anti-sense lncRNA, intronic lncRNA and intergenic lncRNA. In conclusion, the constructed full-length transcriptome provides high-quality reference background and key foundation for study on molecular biology and omics of A. cerana.