

Mengyi Wang¹, Xiaoxue Fan¹, He Zang¹, Dafu Chen², Rui Guo³

¹College of Bee Science and Biomedicine, Fujian Agriculture and Forestry University, Fuzhou 350002, China

²National & Local United Engineering Laboratory of Natural Biotoxin, Fuzhou 350002, China

³Apitherapy Research Institute of Fujian Agriculture and Forestry University, Fuzhou 350002, China

Nosema ceranae is a unicellular fungal parasite for bees and causes bee nosemosis, severely damaging colony health and apicultural industry. Currently, the chromosome-scale reference genome of *N. ceranae* is still lacking. Here, the clean *N. ceranae* spores were subjected to cDNA library construction and PacBio single molecule real time sequencing, followed by assembly and annotation of the chromosome-level genome of *N. ceranae*. Following construction and annotation of the full-length transcriptome, and annotations of the current reference genome of *N. ceranae* was than improved with the generated long reads. The genome sequencing produced approximately 1.73 Gb data, including 36 contigs with a N50 of 457301 bp. The assembled genome included 19 scaffolds (N50=728543 bp) and 1685624 bp of repeated sequences and 2047 transposons. Additionally, 3112 protein-coding genes were predicted, among which 1076, 872, 1310, 1349, 1154, 2709, and 2696 could be annotated to the GO, KEGG, KOG, Pfam, Swiss-Prot, TrEMBL and Nr databases. There were 1598 (1885) homologous genes between *N. ceranae* and *N. apis* (*Nosema bombycis*), and 1838 genes were found to be homologous between *N. ceranae* and (*Encephalitozoon cuniculi*). Intriguingly, the ABC transporter gene family were highly conservative among different microsporidian species. The full-length transcriptome of *N. ceranae* was constructed, including 4387 non-redundant full-length transcripts. Among these, 778, 2083, 1202, 1559, 1457, 1232, 1702, and 3896 full-length transcripts could be annotated to COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG and Nr databases, respectively. A total of 362 (699) new genes (transcripts) unannotated on the reference genome were identified, among which 359 (695), 202 (329), 133 (153), 65 (108), 145 (251), 5 (351), 103 (420) and 166 (310) could be annotated to the aforementioned databases. The structures of 225 annotated genes in the reference genome were optimized. Moreover, 172 SSRs, 472 complete ORFs and 29 lncRNAs were discovered. Totally, 11 alternative splicing (AS) events occurred in six genes were identified, including 10 intron retained and one alternative 5' splice site. Together, these results lay a solid foundation for continuous and further study and enrich the information about genes and transcripts in *N. ceranae*.