

Transcriptome analysis of the Formosan subterranean termite, *Coptotermes formosanus* Shiraki (Isoptera: Rhinotermitidae)

by

Wenjing Wu, Zhiqiang Li, Shijun Zhang, and Yunling Ke

Guangdong Institute of Applied Biological Resources, Guangzhou 510260, China

Abstract

The Formosan subterranean termite, *Coptotermes formosanus* Shiraki, is one of the most economically important termite pests in the world. It has been studied using the transcriptome-based conventional Sanger sequencing method. Next generation sequencing methods, which could generate over one billion bases of high-quality DNA sequence have yet to be applied to *C. formosanus* transcriptome research.

To facilitate future research on this subterranean termite, a cDNA library of pooled castes - worker, soldier and reproductives was constructed and used for transcriptome sequencing. A total of 11.02 GB of 125-bp paired-end clean reads were generated using the Illumina HiSeqTM 2500 platform. The obtained reads were assembled into 189,421 unigenes with mean length and N50 length of 629 bp and 974 bp, respectively. In total 61,407 unigenes were annotated in the NCBI non-redundant protein (NR) database. Species distribution showed that 35% NR annotated unigenes were homologous to genes in *Trichomonas vaginalis*, which may be an important symbiont of *C. formosanus*. Moreover, 16,552 unigenes were assigned to 3,995 gene ontology terms, with 2,208 terms in biological processes, 511 in cellular components, and 1,276 in molecular functions. Using the Kyoto Encyclopedia of Genes and Genomes pathway database, 16,444 unigenes were mapped onto 221 pathways. A total of 9,054 simple sequence repeat (SSR) markers were identified. Mono-nucleotide repeats were the most common repeat motif, accounting for 42.42%, followed by di-nucleotide repeats at 29.20%.

Our data provides the most comprehensive sequence resource available for a subterranean termite and demonstrates that the Illumina sequencing allows *de novo* transcriptome assembly and gene expression analysis in a species lacking genome information. We anticipate that next generation sequencing technologies hold great potential for the study of the transcriptome of other non-model organisms.

Key words: *Coptotermes formosanus*, *de novo* transcriptome assembly, functional annotation, SSR markers