Combining metagenomics and metatranscriptomics approaches for prospection of CAZymes of the lower termite Coptotermes gestroi

Luciana Souto Mofatto[1], João Paulo Lourenço Franco Cairo[2], Melline Fontes Noronha[3], Ana Maria Costa Leonardo[4], Fabio Marcio Squina[5], Gonçalo Amarante Guimarães Pereira[6], Marcelo Falsarella Carazzolle[7]

1 UNICAMP
2 LABORATÓRIO NACIONAL DE CIÊNCIA E TECNOLOGIA DO BIOETANOL, CENTRO NACIONAL DE PESQUISA EM ENERGIA E MATERIAIS, CAMPINAS - SP
3 CENTRO PLURIDISCIPLINAR DE PESQUISAS QUÍMICAS BIOLÓGICAS E AGRÍCOLAS, UNICAMP
4 DEPARTAMENTO DE CIÊNCIAS BIOLÓGICAS, INSTITUTO DE BIOCIÊNCIAS, UNESP
5 UNIVERSIDADE DE SOROCABA, PROGRAMA DE PROCESSOS TECNOLÓGICOS E AMBIENTAIS
6 BRAZILIAN BIOETHANOL SCIENCE AND TECHNOLOGY LABORATORY, BRAZILIAN CENTER FOR RESEARCH IN ENERGY AND MATERIALS, BIOLOGY INSTITUTE - UNICAMP
7 BIOLOGY INSTITUTE - UNICAMP, NATIONAL CENTER FOR HIGH PERFORMANCE COMPUTING/UNICAMP

Abstract

Termites are interesting insects to mining new efficient enzymes for biomass degradation (CAZymes - Carbohydrate activity enzymes). Because they live in symbiosis with bacteria, protozoa and fungus inside their guts, termites have the ability to degrade approximately 90% of plant-dry matter in tropical forest, converting lignocellulosic materials into fermentable sugars. In a previous study of our group, we performed bioinformatics analysis of Coptotermes gestroi genomic and transcriptomic data for prospecting CAZymes in this termite, including symbiont genes. The results identified few CAZymes from symbiont species, probably due to the low representation of these genomes in comparison with the genome of C. gestroi. In order to prospect more specific CAZymes from symbionts, a new approach was performed using the combination of metagenomic and metatranscriptomic analysis from C. gestroi gut. The main aim of this study is to compare metatranscriptomic data from insects submitted to different diets using metagenomic assembly as reference, mainly composed by symbiont sequences. For this purpose, we obtained RNA-seq data from five conditions: (1) sugarcane bagasse in natura; (2) sugarcane bagasse treated with phosphoric acid; (3) filter paper (cellulose); (4) filter paper (cellulose) and iron; (5) sugarcane bagasse treated with sodium chlorite and hydrochloric acid.

As results, Coptotermes gestroi metagenome was assembled using MetaSpades, followed by binning pipeline using CONCOCT and MetaGeneMark software for gene prediction in bacterial genomes. For the gene prediction in non-bacterial genomes (fungi and protozoa), the RNA-seq reads were aligned against the assembled metagenome for transcriptomic reconstruction and quantification using STAR, followed by StringTie and Kallisto. The R/Bioconductor package DESeq2 was used to determine the differentially expressed genes between the conditions. As conclusion, we found a strong relationship between termite's diet and gene expression of the symbiont organisms (bacteria, fungus and protozoa) involved on lignocellulose digestion process.