

A comparison of two pipelines for metagenomic 16S rRNA using Ion Torrent (PGM) Sequencing Platform

Suzana Eiko Sato Guima, Daniel Vasconcelos Rissi, Rodrigo Matheus Pereira
FCBA - UFGD (Universidade Federal da Grande Dourados), UFPR (Universidade Federal do Paraná) and FCBA - UFGD

Targeted metagenomics is a powerful tool for inferring the phylogenetic distribution of microbial community in samples from different sites or time series. Several pipelines for metagenomic analysis have been developed, and they present different sets of softwares, which makes it difficult to choose the most appropriate one for metagenomic analysis. Here, the goal is to present a comparison between BMP (Brazilian Microbiome Project) and MICCA (Microbial Community Analysis) pipelines for metagenomic 16S rRNA of Ion Torrent Sequencing Data. Microcosm experiments were performed using the native forest soil collected from the experimental farm of EMBRAPA Agropecuária Oeste with weekly applications of fipronil (final concentration of 200µg/kg of soil) for four weeks. Metagenomic library was prepared using V4 and V5 hypervariable regions of the 16S rRNA gene from the extracted microcosm DNA. Sequencing was carried on Ion Torrent PGM Hi-Q using Ion 318 V2 chip. Primers and adapters of the raw reads were removed using FastX-Toolkit software v.0.0.14. Reads were filtered using Sickle software v.1.33, discarding reads shorter than 150bp and trimming them if its quality drops below Q23. From the filtering step, both pipelines were run separately. Alpha diversity and rarefaction curve were processed using QIIME scripts. BIOM format files from both pipelines were statistically analyzed on STAMP software. From the total (321,561) of raw reads, 94,416 sequences were filtered. Rarefaction curve for BMP achieved the plateau while MICCA did not. This result is related to the number of OTUs (Operational Taxonomic Units) clustered in each pipeline. BMP has a lower number of OTUs because it discards singletons (OTU represented by a unique DNA sequence) while MICCA does not, resulting in a higher number of OTUs. The number of classified genus in MICCA was higher than in BMP. Although the results at the genus level were different, results between both pipelines were statistically similar at the family and order level. Therefore, both pipelines can infer a similar bacterial taxonomic diversity from phylum to family level. It is suggested the use of BMP for the bacterial diversity analysis considering the most common genus, and the use of MICCA for bacterial community analysis considering rare genus and species.