

METEOR - a platform for quantitative metagenomic profiling of complex ecosystems

Nicolas PONS¹, Jean-Michel BATTO, Sean KENNEDY, Mathieu ALMEIDA, Fouad BOUMEZBEUR, Bouziane MOUMEN, Pierre LEONARD, Emmanuelle LE CHATELIER, S. Dusko EHRLICH and Pierre RENAULT

¹ INSTITUT MICALIS, UMR1319 INRA, Domaine de Vilvert, 78352, Jouy-en-Josas, Cedex, France
nicolas.pons@jouy.inra.fr

The study of complex microbial ecosystems by a quantitative metagenomic approach has been made possible by advancements in high-throughput sequencing technologies. Quantitative metagenomics relies on deep sequencing to construct an ecosystem profile using gene and genome counts. Next generation sequencing (NGS) technologies such as SOLiD or Illumina produce millions of short sequences (35 to 75bp) which can be used as tags to establish gene profiles. This approach requires the use of a specific reference catalog which should be composed of genes present in the ecosystem of interest. The use of classical bioinformatic methods for the analysis of such large amounts of data is not feasible as we overpass the expected dataset size of common tools. We have therefore developed an integrated metagenomic analysis pipeline, METEOR, which includes the indexing of short reads to genomic objects. Data are indexed in an embedded database around the iMOMi framework [1] and organized in a dedicated file system. This optimization facilitates secondary analysis including gene/species abundance evaluation, cross-sample comparison, ecosystem metabolism reconstruction or gene/species diversity analysis. The METEOR pipeline has been implemented for several metagenomic projects such as MicroObes for characterizing the human intestinal microbiome of obese individuals following a restrictive diet or FoodMicrobiomes for studying the ecosystem of fermented food like French traditional cheeses.

In MicroObes, we investigate the changes of gut microbiota in a human model of weight loss induced by restrictive diet in moderately obese subjects. DNA isolated from 195 faecal samples of 49 obese subjects collected at different date (start of the study, 6 weeks after a restrictive diet and 12 weeks) have been sequenced with SOLiD technology yielding about 300 gigabases. Short reads have been indexed against the 3.3 millions

genes of the human gut microbial gene catalog (MetaHIT consortium [2]). Statistical analysis of the gene profiles generated indicate significant variations in gene and genome frequencies during the first 6 weeks of dieting and a subsequent stabilization after 12 weeks according to the observed success of patients dietary.

References

- [1] N. Pons, J.M. Batto, S.D. Ehrlich and P. Renault, Development of software facilities to characterize regulatory binding motifs and application to streptococcaceae. *J Mol Microbiol Biotechnol*, 14(1-3): 67-73, 2008.
- [2] J. Qin, ..., N. Pons, ..., J.M. Batto, ..., P. Renault, ..., S.D. Ehrlich, J. Wang. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59-65, 2010.