Analyzing microbial diversity using the metagenomics tools of BioNumerics®

J. Dombrecht, K. De Bruyne, H. Pouseele, and K. Janssens

Applied Maths NV, Keistraat 120, B-9830 Sint-Martens-Latem, Belgium, E-mail: info@applied-maths.com - Phone: +32 9 2222 100

Introduction

We recently developed a tool in the BioNumerics® software suite for the analysis, quantification, visualization and comparison of microbial communities, starting from sequence reads from a variety of NGS platforms.
Within the graphical user interface a metagenomics analysis workflow is introduced allowing actions to be added or deleted according to the users own interest: filtering of reads, calculation of OTU's, computation of sample statistics, calculation of diversity indices, …

Methods

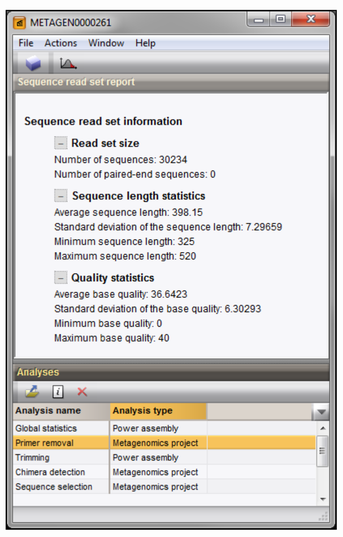
BioNumerics® uses the Mothur[1] project, initiated by Dr. Patrick Schloss and colleagues (Dept., of Microbiology & Immunology, The University of Michigan). The Mothur project filled in the needs of the microbial ecology community by incorporating the functionality of numerous other applications like Dotur, Treeclimber, S-libshuff, Unifrac into a single command line application.

BioNumerics® adds the flexibility of the algorithms implemented in Mothur and elaborates further by creating a **fully interactive reporting service** including a **geographical visualization tool** and various **chart tools** for the interpretation and manipulation of the results.

The integrated follow-up analysis is possible with the same BioNumerics® platform and includes an environment for elaborated **data mining and statistics**.

In this study, we illustrate this novel tool with publicly available genomic data sets.

Import sequence read sets in BioNumerics®



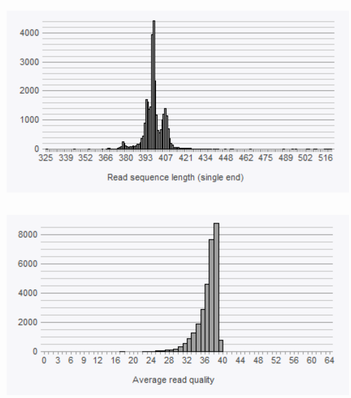


Figure 1. The sequence read set experiment card. Graphs on read length and read quality from the global statistics calculated from the read set.

Quality control of the imported sequence read sets: primer removal, demultiplexing, chimera detection and quality trimming

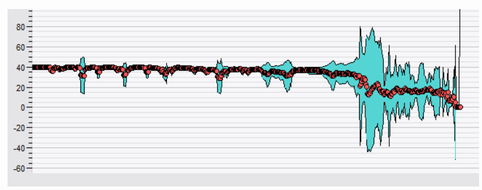


Figure 2. The average base quality distribution.

Data preprocessing includes primer removal, a combination of different trimming operations, chimera detection using ChimeraSlayer (from the Broad Institute), etc…

The analysis of multiple samples, tagged by a multiplex identifier, is supported by a dedicated functionality for de-multiplexing and parallel analysis.

Identification against taxonomic databases

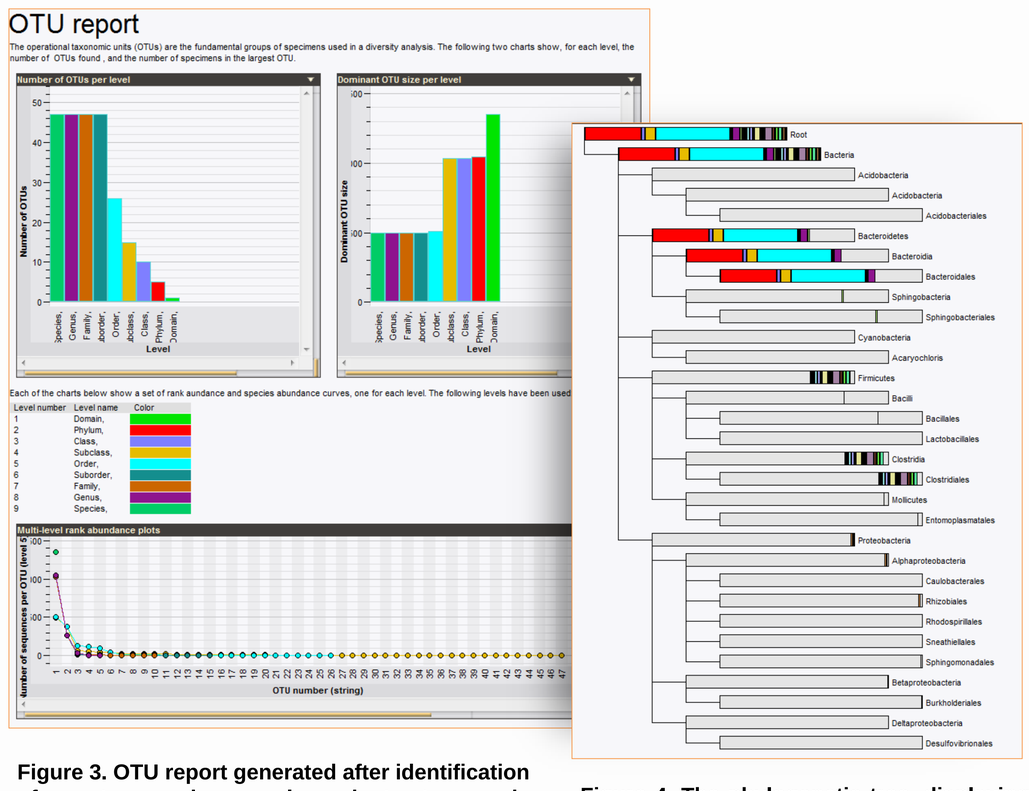


Figure 3. OTU report generated after identification of a metagenomics sample against a taxonomic database.

Figure 4. The phylogenetic tree, displaying the relative OTU abundances present in the sample.

This analysis identifies all metagenomics sample sequences against a taxonomic reference database that can be user-specific or downloaded from public repositories such as the 16S rRNA reference databases from RDP, SILVA or Greengenes. The analysis results in a variety of different visualizations.

Diversity analysis

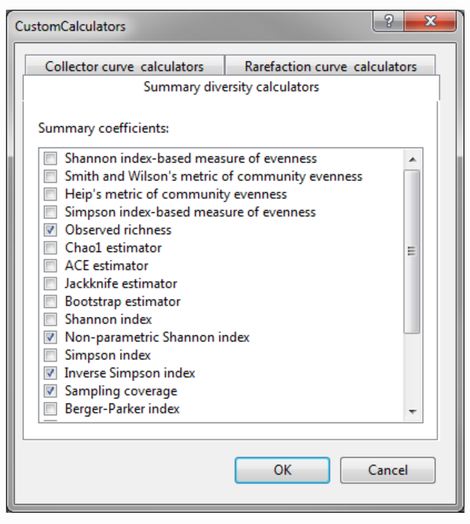


Figure 5. The different diversity calculators.

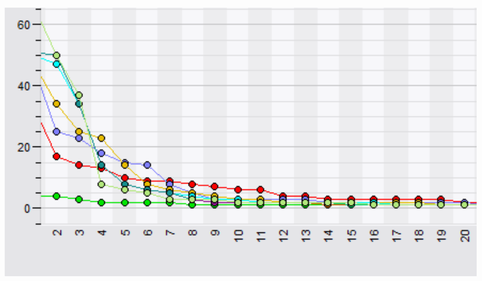


Figure 6. The multi-level (unique up to 91% sim.) rank abundance plots.

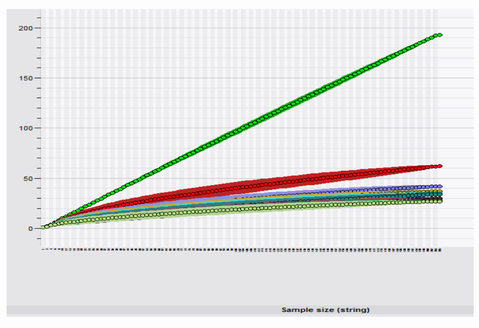
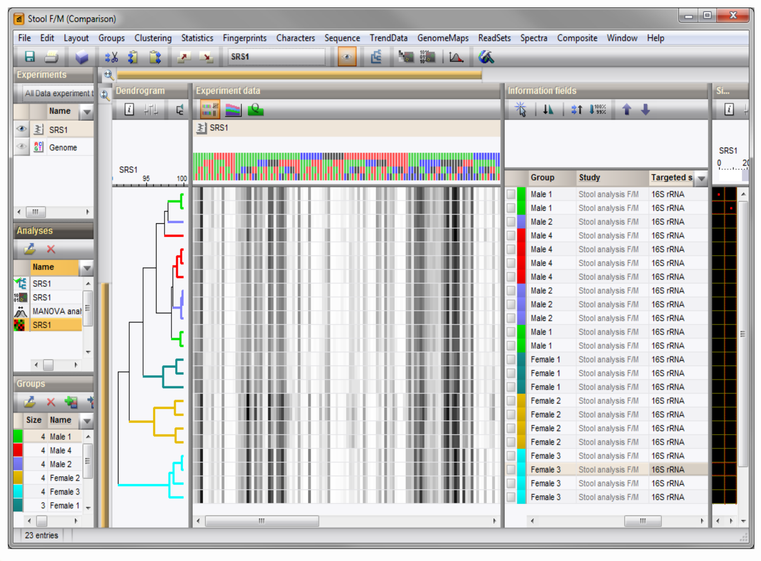
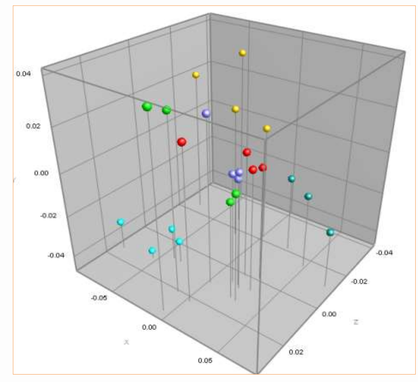


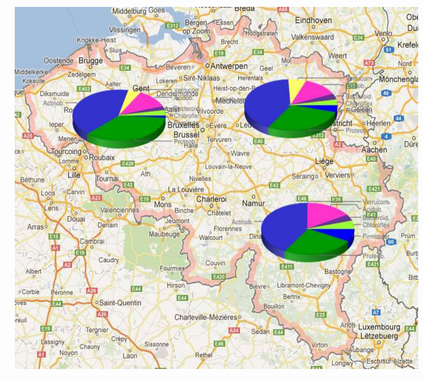
Figure 7. Rarefaction curve for the observed community richness.

In this analysis, the alpha-diversity of a single sample was assessed. First, OTUs were defined either from a similarity cutoff on the sequence clustering results, or from the consensus taxonomy on a specific phylotypic level, defined per cluster. Next, the within-sample diversity, the community evenness, the community richness and the community diversity indices were calculated for the different levels.

Follow up analysis examples: MDS, PCA, Data Mining, Charts, Statistics, Geo plot, …







Conclusion

The BioNumerics® software offers a **graphical user environment** to import **raw read sequences**, perform **trimming, quality control, sequence clustering** and end up with a wide range of **visualization of the OTU abundances** or an **evaluation of the α- and β-diversity** using a plethora of indices. Starting from the BioNumerics® platform, a further, integrated analysis of the metagenomics results can be performed, including a wide range of **data mining**, **clustering**, **identification** and **statistical analysis tools**.

Reference

[1]Schloss PD *et al.* Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009 Dec;75(23):7537-41.