SEQUENTIA

WE TRANSFORM DATA INTO KNOWLEDGE

GAIA: an integrated metagenomics suite

Paytuví A.[1], Battista E.[1], Scippacercola F.[1], Sundesha V[1]., Aiese Cigliano R.[1], Sanseverino W.[1]

[1] Sequentia Biotech SL, Calle Comte d'Urgell 240, 08036 Barcelona, Spain

Introduction

Identifying the biological diversity of a microbial population is of fundamental importance due to its implications in industrial processes, environmental studies and clinical applications. In plants specifically, microbial communities affect the growth and health as well as the productivity of crops. Today, there is still an outstanding need to develop new, easy-to-use bioinformatics tools to analyze both shotgun and targeted metagenomics with the highest accuracy and the lowest running time. With the aim of overcoming this need, we introduce you to GAIA, an online SaaS (Software as a Service) solution that has been designed to give you the maximum information on your sample whatever you perform: 16/18S, virome or shotgun analysis.

Methods

The GAIA pipeline uses BWA to map the reads/pairs from any platform against one of the three custom-made databases created using NCBI as the source (16S rRNA or 18S rRNA for targeted metagenomics and genomic sequences for shotgun metagenomics). Then, reads/pairs are classified into the most specific taxonomic level using an in-house Lowest Common Ancestor (LCA) algorithm. Identity thresholds are applied to classify reads into species, genus, family, phylum and domain levels. Alpha and beta diversities are calculated using phyloseq. Besides, in case the input datasets come from different conditions, GAIA is able to perform a differential abundance analysis using DESeq2 (Figure 1).

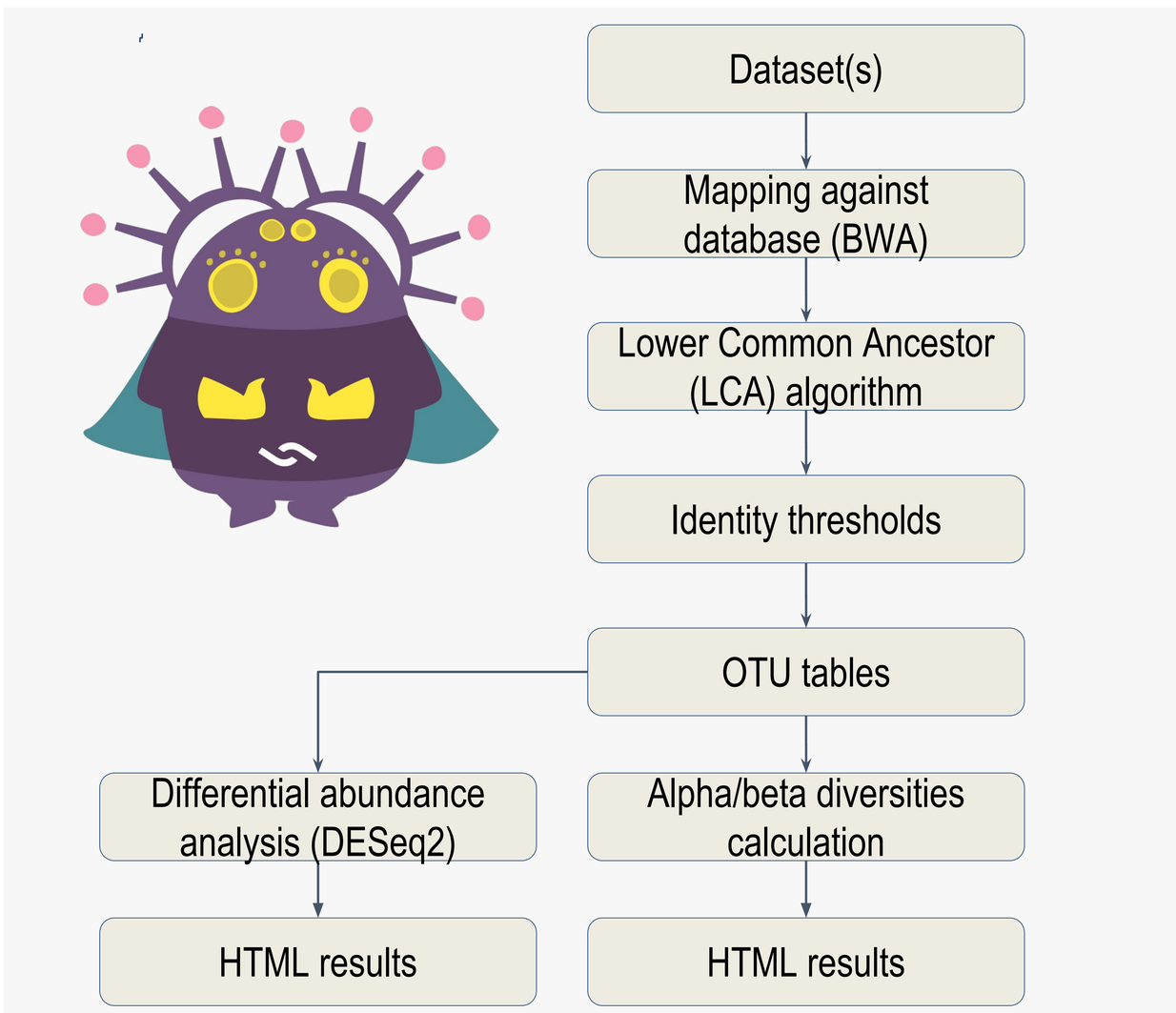

Figure 1. Overview of GAIA, the bioinformatics pipeline to analyze metagenomics data.

Results

Using recently published benchmark datasets from shotgun and 16S experiments (Siegwald, *et al*. 2017; McIntyre, *et al*. 2017), we compared Gaia against several available pipelines. Our results show that for shotgun metagenomics, Gaia obtained the highest F-measures at species level above all tested pipelines (CLARK, Kraken, LMAT, BlastMegan, DiamondMegan and NBC)(Figure 2). For 16S metagenomics, Gaia also obtained excellent F-measures comparable to QIIME at family level (Figure 3).

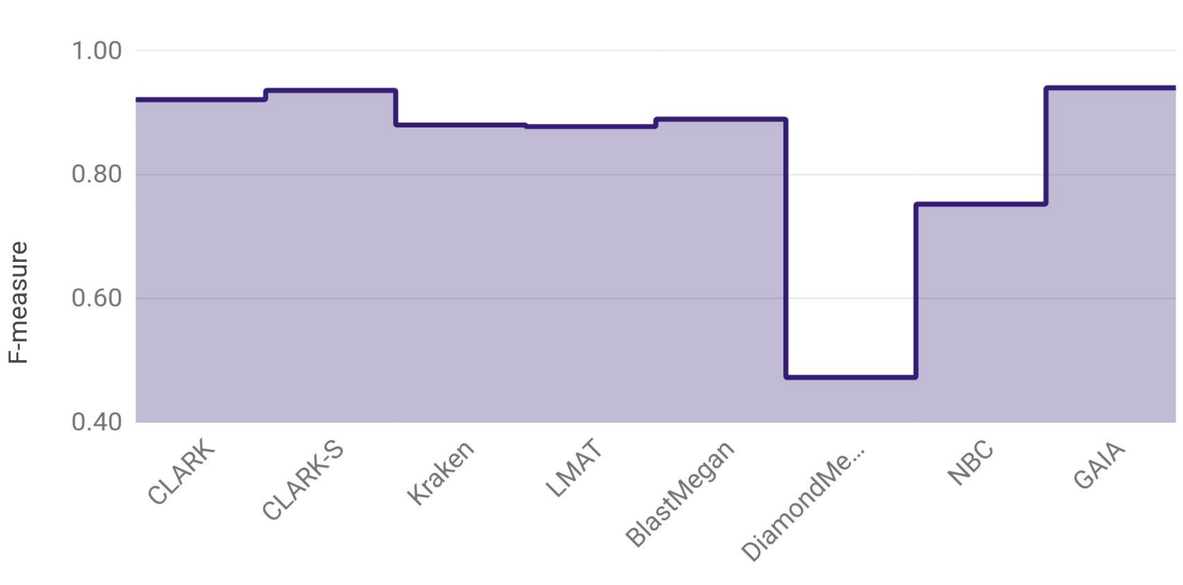

Figure 2. Benchmark results (shotgun) using McIntyre, *et al*. datasets at species level.

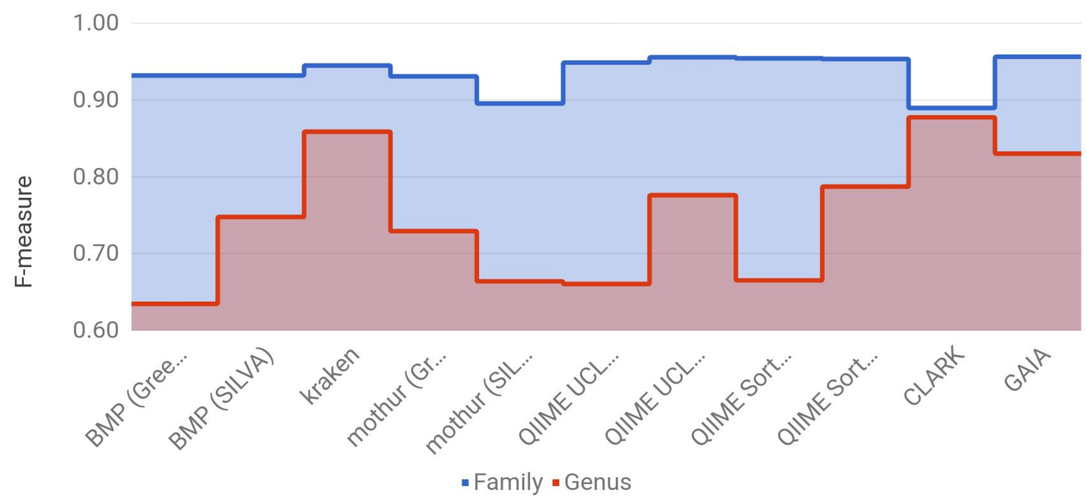

Figure 3. Benchmark results (16S) using Siegwald, *et al*. datasets at family (blue) and genus (red) level.

Thinking of user experience, the pipeline is integrated into an online SaaS solution, which delivers the software in a way it can be accessed from any device with an Internet connection and a web browser. This way, end users of this platform do not need bioinformatics skills to perform an analysis, since the required steps for it are done interactively online. The accurate results include dynamic charts and tables using Google Charts and DataTables (JavaScript-based)(Figure 4).

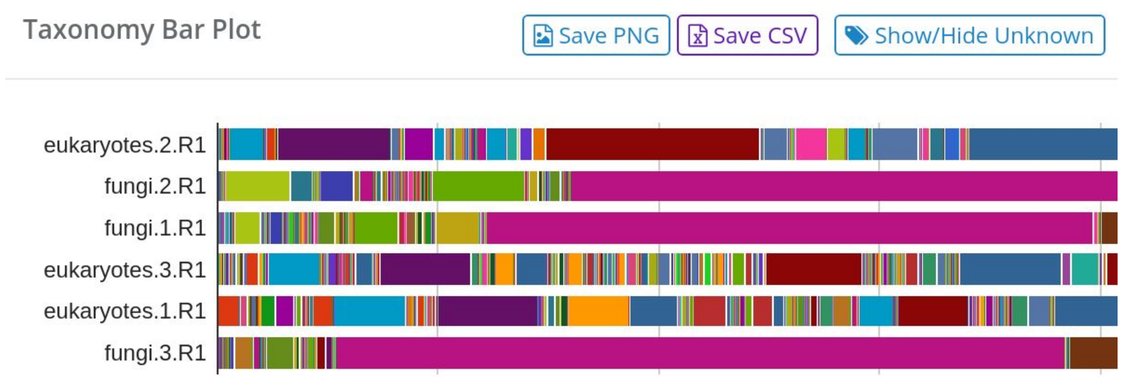

Figure 4. Interactive taxonomy bar plot displayed in the results page after the analysis.

Conclusions

GAIA is able to obtain a comprehensive and detailed overview at any taxonomic level of microbiomes of different origins: human (e.g. stomach or skin), agricultural and environmental (e.g. land, water or organic waste) in an accurate, fast and easy way. The overall objective of GAIA is to provide academia and industries with an integrated metagenomics suite that will allow to perform metagenomics data analysis easily, quickly and affordably. It will be available soon at gaia.sequentiabiotech.com.

References

Siegwald, L., *et al*. (2017). PLoS ONE, 12(1), e0169563. http://doi.org/10.1371/journal.pone.0169563

McIntyre, *et al*. (2017). Genome Biology 18:182. https://doi.org/10.1186/s13059-017-1299-7

CONTACT US

Campus UAB, Edifici Eureka
Av. de Can Domènech s/n
08193 Bellaterra (Cerdanyola del Vallès)

info@sequentiabiotech.com
+34 93 010 73 68
www.sequentiabiotech.com