Natural products: what is the biosynthetic potential of unicellular eukaryotes?

Loïc Meunier[1,2], **Philippe Jacques**[2], **Denis Baurain**[1]

[1] InBioS–PhytoSYSTEMS, University of Liège, Belgium;

[2] Microbial Processes and Interactions, Gembloux Agro-Bio Tech, University of Liège, Belgium

lmeunier@uliege.be, denis.baurain@uliege.be

Microbial organisms are a precious source of natural products for medical and agricultural fields. These natural products are produced by different enzymes classes from the secondary metabolism, resulting in a large diversity of molecular structures and bioactivities. Currently, the main known microbial producers of bioactive compounds are bacteria and fungi, which have already been extensively exploited for the pharmaceutical and phytosanitary drugs. However, other unicellular eukaryotes seem to also have potential for the production of bioactive compounds, which present somehow a new continent for the discovery of natural products. A preliminary large-scale *in silico* analysis through unicellular eukaryotic lineages, for which genomic data are available, indicates that a part of these organisms might possess diverse classes of secondary metabolism pathways and thus be potential candidates for the discovery of novel compounds.

Analysis pipeline for the large-scale detection of secondary metabolism pathways in eukaryotes

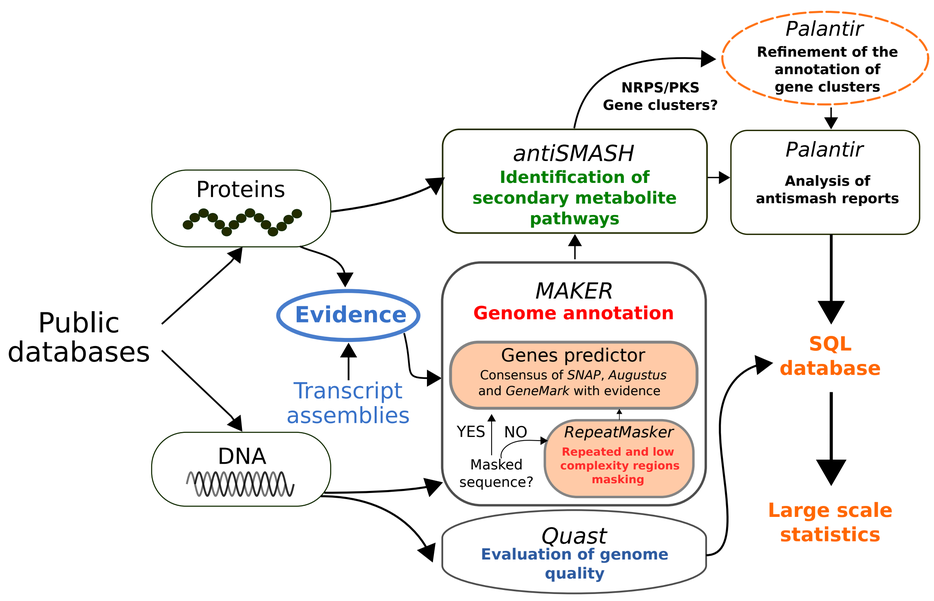

Figure 1: ***Steps involved in the detection of secondary metabolism pathways in eukaryotic organisms.*** *Genomic data for 11 eukaryotic lineages were extracted from public databases (NCBI, Ensembl), while protein sequences (acquired preferentially when available) can directly be used to apply a protein signature search with antiSMASH[1] for detecting biosynthetic genes from secondary metabolism, the nucleotide sequences need to be annotated in protein sequences first. The annotation in proteins was performed with an automated application of MAKER2[2] (a pipeline for the structural annotation of genes) which uses homologous protein and RNA-Seq expression data for supporting in silico gene predictions. Once all genome sequences were processed with antiSMASH, the information from hundreds of secondary metabolism pathways identified was extracted with Palantir (a home-made Perl module for the post-processing of antiSMASH results) and their storage into a SQL database. Furthermore, Palantir also offers an additional step for refining the annotation of the biosynthetic genes of certain classes of secondary metabolites (i.e., NonRibosomal Peptide Synthetases, NRPS, and PolyKetides Synthetases, PKS).*

in silico detection of secondary metabolism classes through unicellular eukaryotic lineages

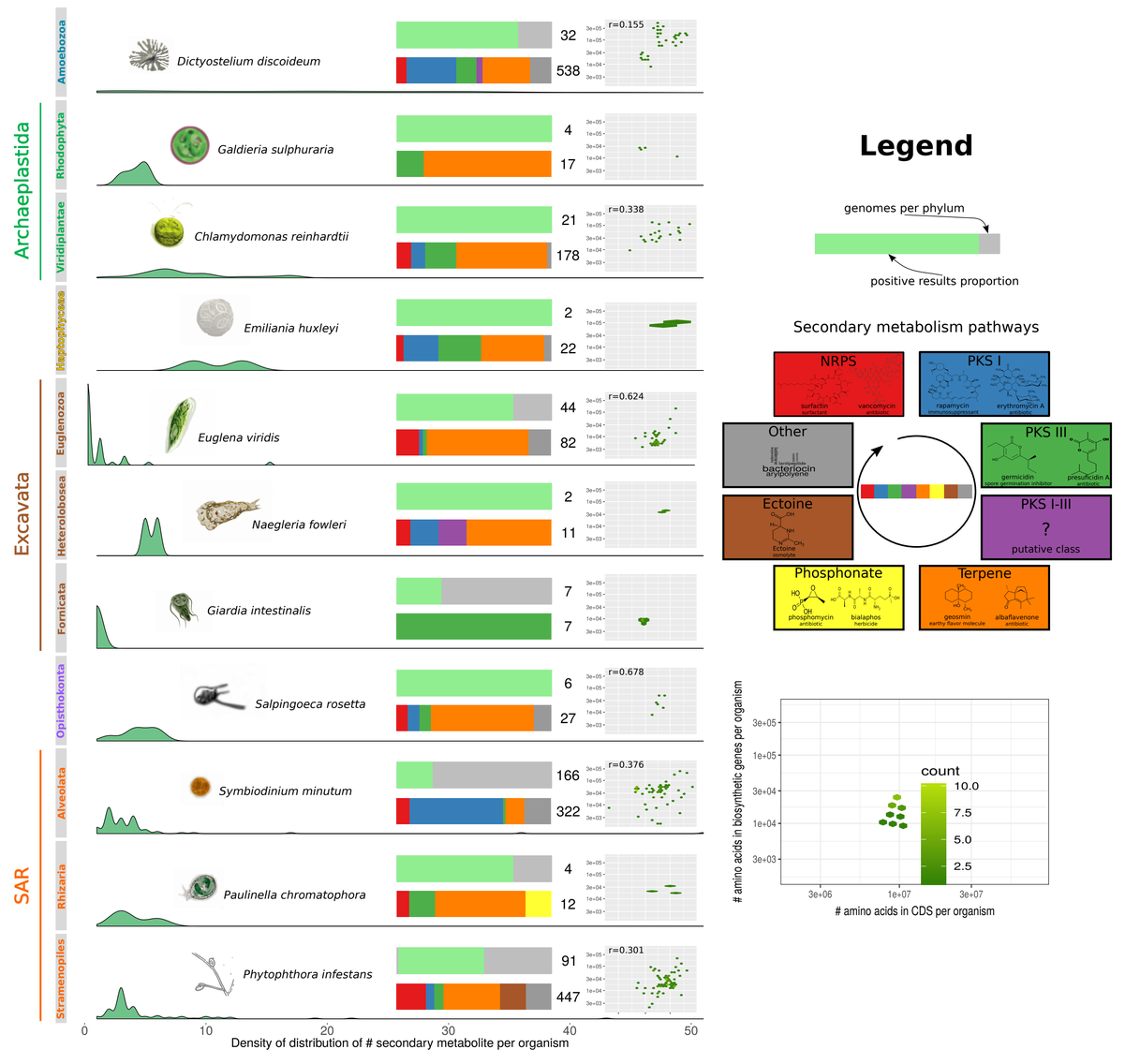

Figure 2: ***Preliminary results of the distribution of secondary metabolism pathways in 11 eukaryotic lineages.*** *First, the main plot represents the distribution of the number of secondary metabolism pathways detected per phylum. Second, the green bar of the top left stacked bar chart represents the propotion of genomes with positive results, the total number being written beside. The bottom left stacked bar chart represents the proportions of the classes of secondary metabolism pathways detected, the colors are described in the legend and the total number of gene pathways detected is written beside. Lastly, the right scatter plot illustrates the number of amino acids contained in the detected gene pathways in fuction of the number of amino acids in the CDS of the organism.*

References

[1] Blin, K., *et al.* (2017). antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Research, (1), 1–6. http://doi.org/10.1093/nar/gkx319.

[2] Holt, C., & Yandell, M. (2011). MAKER2 : an annotation pipeline and genome- database management tool for second- generation genome projects. BMC Bioinformatics, 12(1), 491. http://doi.org/10.1186/1471-2105-12-491.

Visualizing Biological Data (VIZBI 2019), Heidelberg, Germany, March 13th-15th 2019

L. Meunier is a Ph.D. student funded by the FRIA (F.R.S.-FNRS).