A shotgun metagenomics analysis applied to the ecological engineering of anaerobic microbial communities for the production of hydrogen from Citrus Peel Waste

Cedric Midoux, Franciele Camargo, Olivier Rué, Mahendra Mariadassou, Valentin Loux, Maria Bernadete Amâncio Varesche, Ariane Bize

A shotgun metagenomics analysis applied to the ecological engineering of anaerobic microbial communities for the production of hydrogen from Citrus Peel Waste

Cédric MIDOUX [1, 2, 3], Franciele CAMARGO [4], Olivier RUÉ [2, 3], Mahendra MARIADASSOU [2, 3], Valentin LOUX [2, 3], Maria Bernadete AMÂNCIO VARESCHE [4] and Ariane BIZE [1].

1 Université Paris-Saclay, INRAE, PROSE, 92160, Antony, France
2 Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France
3 Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France
4 Department of Hydraulics and Sanitation, School of Engineering of São Carlos, University of São Paulo, Brazil

Context

Brazil produces annually 88 million tons of Citrus. This production generates waste, composed mainly by peels and bagasse, which accounts for half of the processed mass. More than half of this waste is not yet valorized, although it could be used in biorefineries to produce various invaluable compounds, including biogas.

In this project, we focused on the **production of hydrogen from Citrus Peel Waste by anaerobic microbial communities**. We used a shotgun metagenomics approach to identify key strains and biological functions for this process, by comparing the inoculum (not yet adapted) with the microbial communities sampled after hydrogen production in optimized conditions.

Data

6 samples were sequenced with *Illumina HiSeq* platform (2x150 bp) with a sequencing depth of ≃ 7M reads/sample:

- Inoculum samples (biological triplicates)
- Optimized samples (biological triplicates)

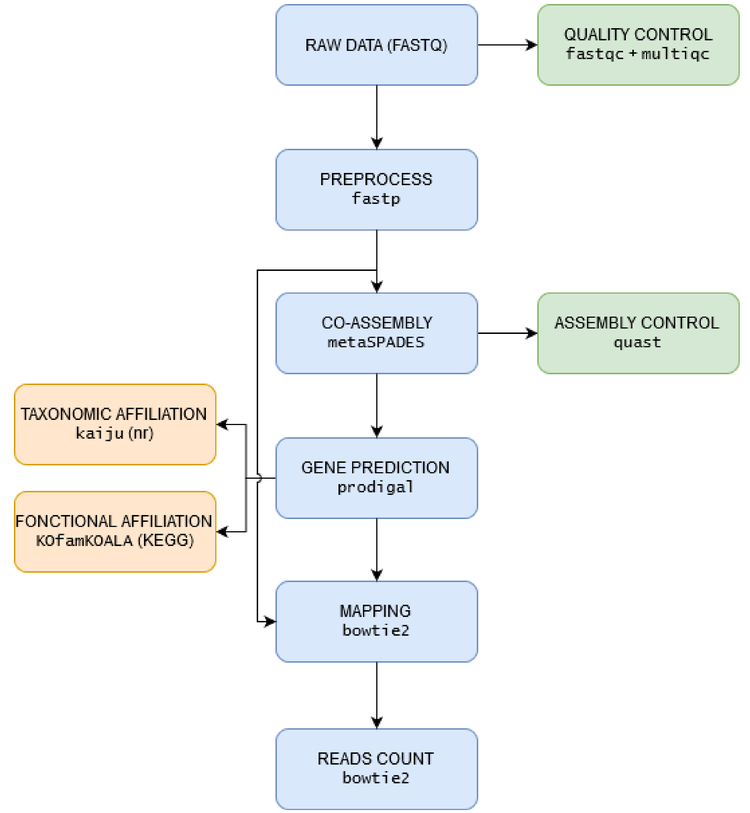
Raw sequences were submitted to NCBI: PRJNA605706

Coassembly with metaSPADES produced 78123 contigs with a total length of 289 Mb and N50 of 6819 nct..

Pipeline

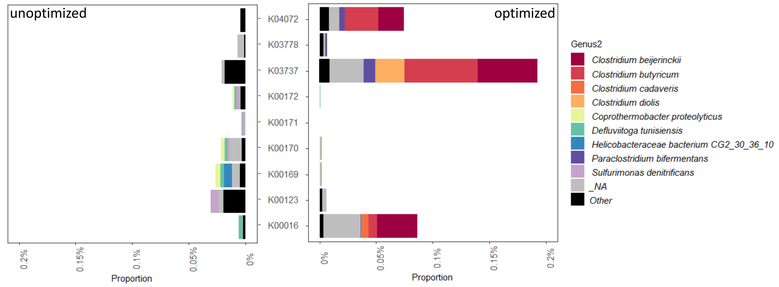
This pipeline was developed with snakemake to favor reproducible and scalable data analysis. It was run on the cluster of the INRAE MIGALE bioinformatics platform.

This snakemake pipeline is scalable simply with a configfile. It is possible to add a normalization step with `khmer` before assembly and removal of rRNA gene sequences with `sortmerna`. `Megahit` can be selected instead of `metaSPADES` if desired. We can use `FragGeneScan` or `prodigal` for gene prediction.
A gene catalog approach is available if desired. In such case, the assembly is performed sample by sample, the genes are clustered after prediction and reads are mapped on this gene catalog.



Biologicals results

Using this pipeline, we were able to observe that reactor optimization led to the strong selection of *Clostridium* genus (its relative abundance increased from 0.4% to 79.7%).
In terms of functional analysis, the gene *pyruvate-ferredoxin / flavodoxin oxidoreductase* (K03737) was the main gene related with H_2 production. This function was largely associated with several strains of *Clostridium*. Moreover, the cellulose degradation pathway was overabundant after optimization.



Conclusions

Using the developed pipe-line, we were able to identify key strains and biological functions selected during this process, especially those related to lignocellulose deconstruction and to hydrogen production.
The pipeline will be reused in future metagenomics projects and will continue to be improved by the addition of new features.

You can fork and contribute to this pipeline on GitLab :
https://gitlab.irstea.fr/cedric.midoux/workflow_metagenomics/