INTEGRATIVE META-OMICS IN GALAXY

Francesco Delogu[1], Praveen Kumar[2], Benoit Kunath[1], Subina Mehta[2], James E. Johnson[3], Bjoern Gruning[4], Phil B. Pope[1], Pratik Jagtap[2], Timothy J. Griffin[2], Magnus Ø. Arntzen[1]

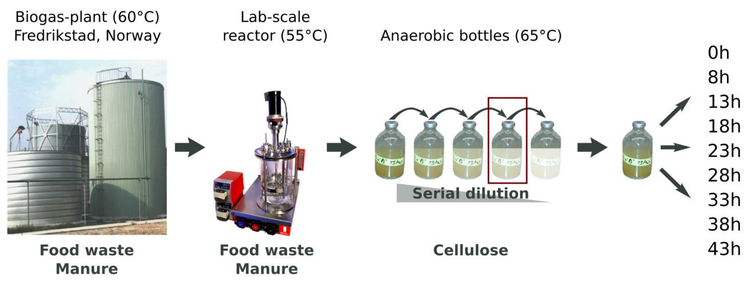
1: Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway; 2: Department of Biochemistry, Biophysics and Molecular Biology, University of Minnesota, Minneapolis, MN 55455; 3: Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455; 4: Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany

Introduction

Individually, the different meta-omics technologies can provide great insight into a microbial community; however, in combination, they can provide a detailed understanding of which organisms occupy specific metabolic niches, how they interact, and how they utilize environmental nutrients.

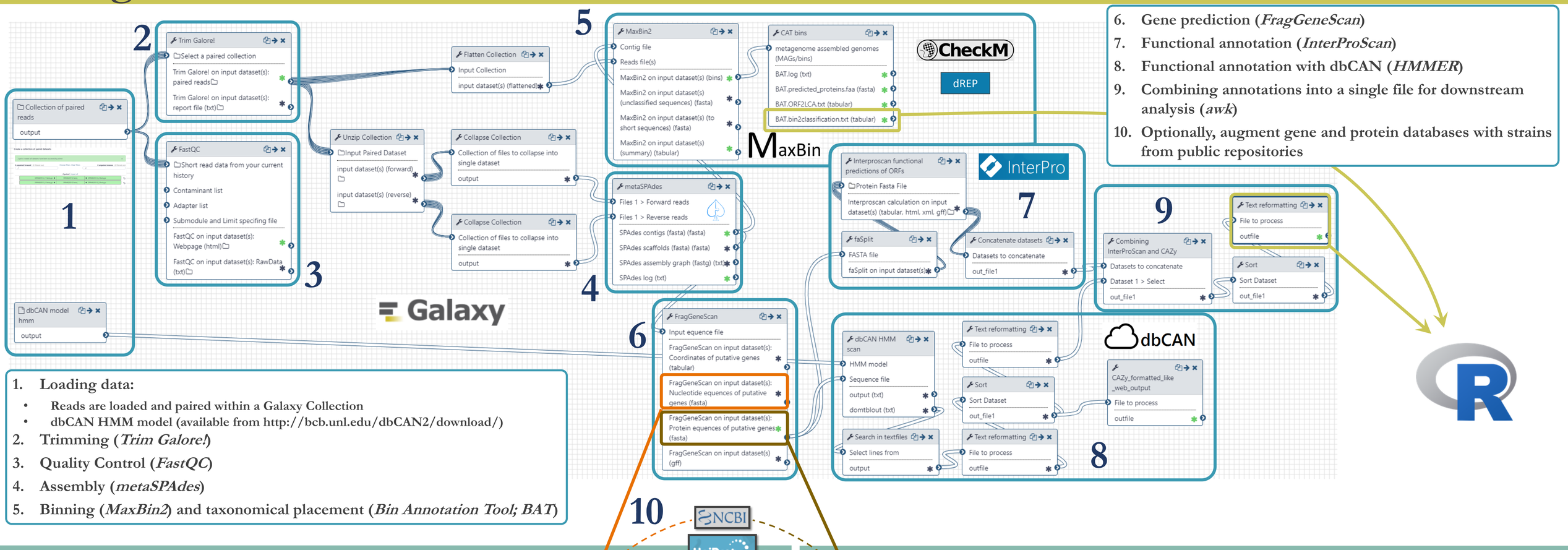
The complexity of informatics approaches required for multi-omics analysis limits their adoption by the wider research community. Here we aimed at implementing a repertoire of commonly used meta-omics tools spanning the three technologies metagenomics, metatranscriptomics and metaproteomics into Galaxy, in order to generate a user-accessible, scalable and robust analytical pipeline for integrated multi-omics analysis.

Dataset



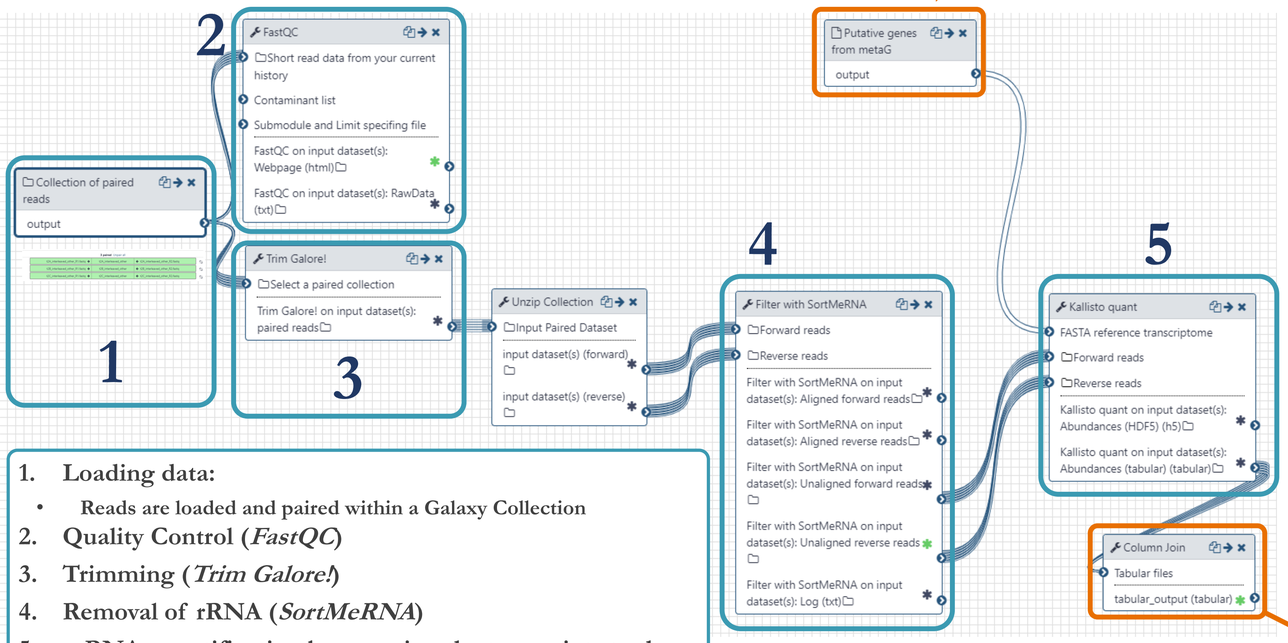
The sample studied in this work originated from a thermophilic biogas-plant operated on muncipal food waste and manure. After a round in a lab-scale reactor, we performed a serial dilution to extrinction experiment to simplify and enrich the community for growth on cellulose (Norwegian Spruce).

Metagenomics



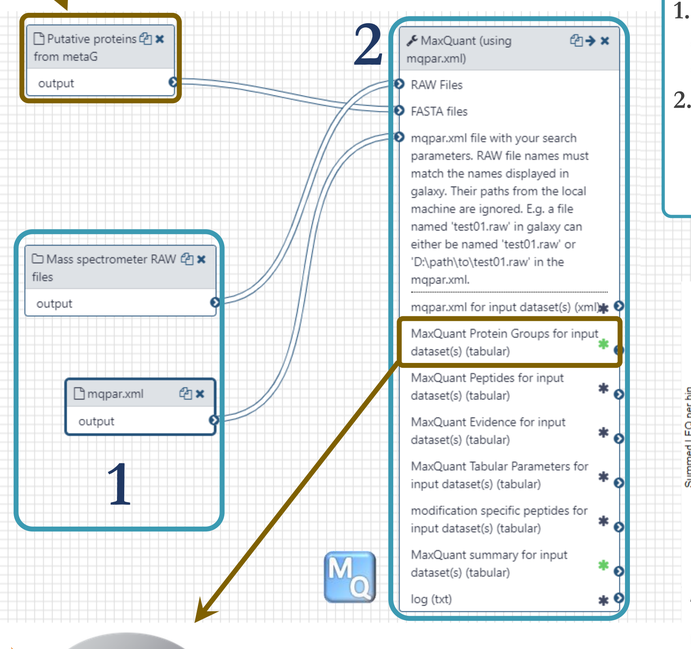
1. Loading data:
   - Reads are loaded and paired within a Galaxy Collection
   - dbCAN HMM model (available from http://bcb.unl.edu/dbCAN2/download/)
2. Trimming (*Trim Galore!*)
3. Quality Control (*FastQC*)
4. Assembly (*metaSPAdes*)
5. Binning (*MaxBin2*) and taxonomical placement (*Bin Annotation Tool; BAT*)
6. Gene prediction (*FragGeneScan*)
7. Functional annotation (*InterProScan*)
8. Functional annotation with dbCAN (*HMMER*)
9. Combining annotations into a single file for downstream analysis (*awk*)
10. Optionally, augment gene and protein databases with strains from public repositories

Metatranscriptomics

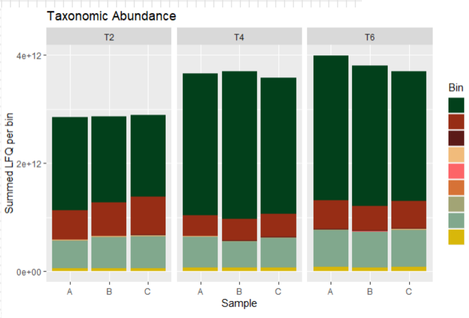


1. Loading data:
   - Reads are loaded and paired within a Galaxy Collection
2. Quality Control (*FastQC*)
3. Trimming (*Trim Galore!*)
4. Removal of rRNA (*SortMeRNA*)
5. mRNA quantification by mapping the transcripts to the putative genes identified from metagenomics (*Kallisto*)

Metaproteomics

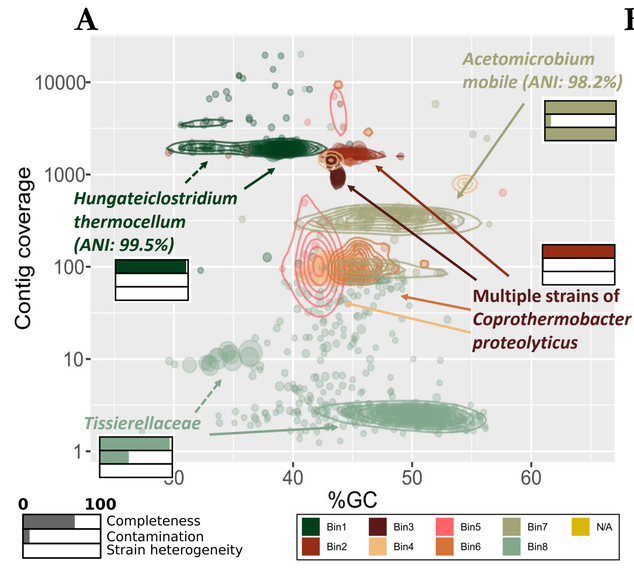


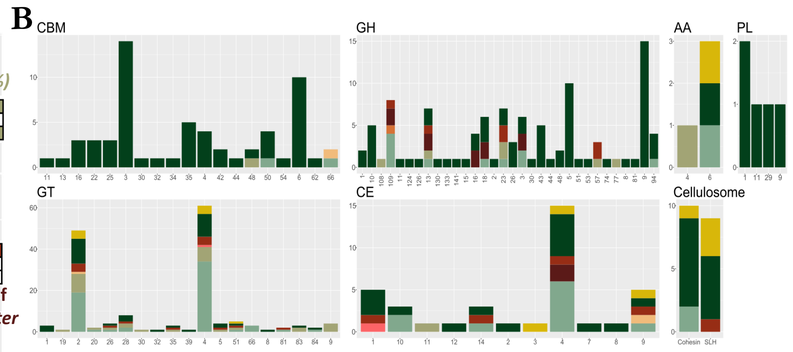
1. Loading data:
   - Mass spectrometer RAW files
   - Parameter set for MaxQuant
2. Protein quantification by mapping MS/MS spectra to the putative proteins identified from metagenomics (*MaxQuant*)



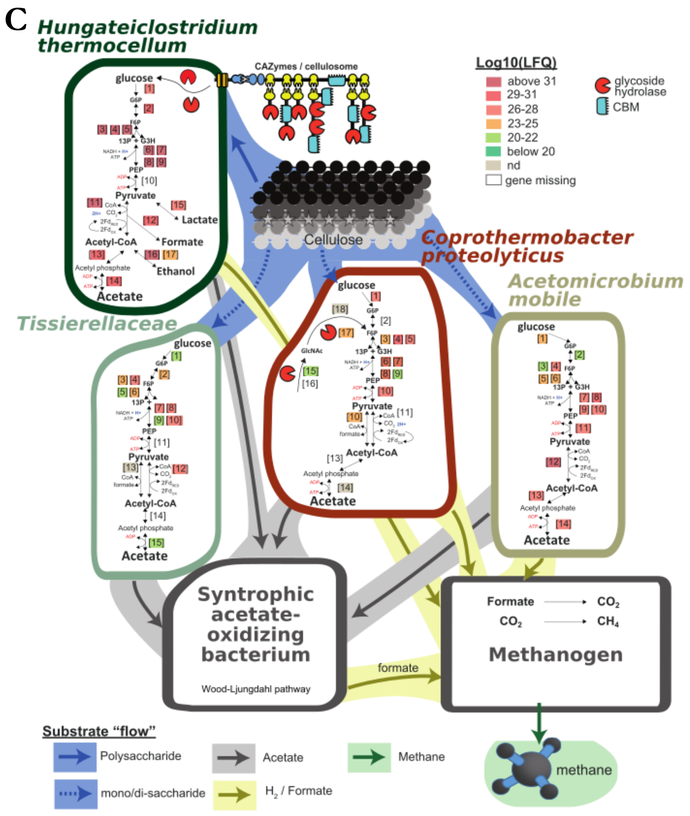
Taxonomic abundances can be calculated using the summed abundances of all proteins per bin, here LFQ-values reported from MaxQuant.

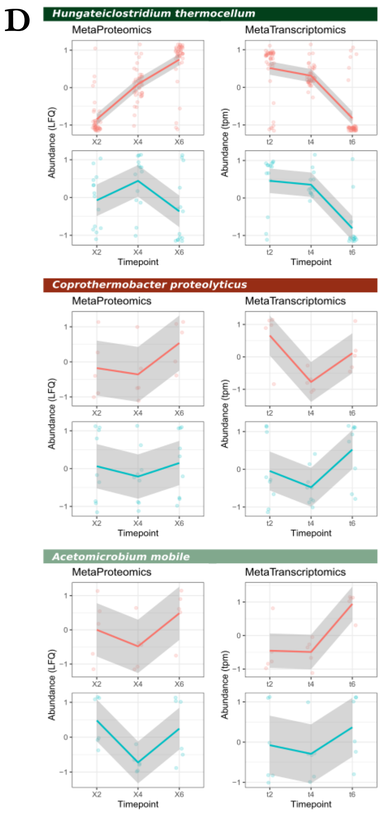
Integration





A) Phylogenetic binning revealed that the microbioal community was co-dominated by *Hungateiclostridium thermocellum* (Bin1) and multiple heterogenic strains affiliated with *Coprothermobacter proteolyticus* (Bin2-6). **B)** An overview of all CAZymes present in the metagenomic data suggests that Bin1 (*Hungateiclostridium thermocellum*) is the main polysaccharide hydrolyzer (see also C). The cellulose monomer, glucose, can act as substrate for the many sugar fermentors.





C) Cellulose is primarily degraded by *Hungateiclostridium thermocellum* while the cellulose monomer, glucose, is fermented to acetate by *Coprothermobacter proteolyticus*, *Acetomicrobium mobile* and *Tissierellaceae*. Only a subset of the original data was used in this Galaxy proof-of-principle analysis, while the full dataset identifies additionally a SAOB and a methanogen converting acetate to formate, H_2, CO_2 and methane. LFQ-values are from the full dataset. **D)** Quantification of selected CAZymes for both omics technologies. Red = glycoside hydrolases, blue = glycosyl transferases.

Contact and references

Contact information:

Magnus Ø. Arntzen, Norwegian University of Life Sciences, magnus.arntzen@nmbu.no
Pratik Jagtap, University of Minnesota, pjagtap@umn.edu
Timothy Griffin , University of Minnesota, tgriffin@umn.edu

Kunath BJ, Delogu F, et. al. From proteins to polysaccharides: lifestyle and genetic evolution of *Coprothermobacter proteolyticus*. ISME J. 2019 Mar; 13(3):603-617

Delogu F, Kunath BJ, et. Al. Integration of absolute multi-omics reveals translational and metabolic interplay in mixed-kingdom microbiomes. bioRxiv. doi: https://doi.org/10.1101/857599