Extensive and High Quality Genome Recovery from full-scale WWTPs using Nanopore Sequencing

Singleton, Caitlin Margaret; Andersen, Martin Hjorth; Kirkegaard, Rasmus Hansen; Kristensen, Jannie Munk; Yssing Michaelsen, Thomas; Karst, Søren Michael; Dueholm, Morten Simonsen; Nielsen, Per Halkjær; Albertsen, Mads

Publication date:
2019

Long-read sequencing enables the recovery of >1000 **high-quality** metagenome-assembled genomes from **full-scale** activated sludge WWTPs

Extensive and High Quality Genome Recovery from Full-Scale WWTPs Using Nanopore Sequencing

Caitlin M Singleton, Martin H Andersen, Rasmus H Kirkegaard, Jannie M Kristensen, Thomas Y Michaelsen, Søren M Karst, Morten S Dueholm, Per H Nielsen, Mads Albertsen

Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University

INTRODUCTION

Microorganisms underpin wastewater treatment processes. Although thousands of metagenome-assembled genomes (MAGs) have been recovered from a wide range of environments, including activated sludge, only a handful are true high-quality (HQ), encoding the 16S, 23S, 5S rRNA and 18x tRNA genes. HQ MAGs are required to link structure to potential function in ecosystems.

AIM: To create a foundational HQ genome database for WWTP microbes that will facilitate links between genomes, amplicons and FISH probes.

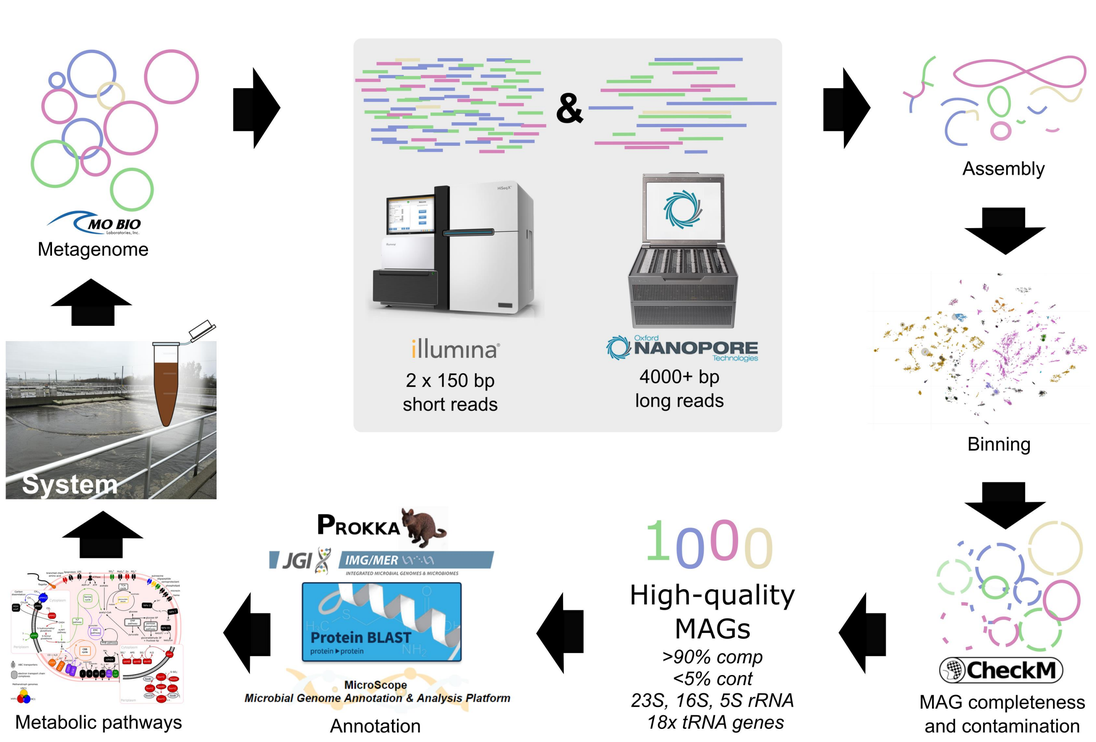

Fig. 1: Schematic of the overall methodology followed for long-read assembly and short read polishing, binning, quality checks, annotation and metabolic inferences.

METHODS

- Metagenome sequencing of 23 WWTPs across Denmark.
- Produced short-read Illumina data (0.9 Tbp) and long-read Oxford Nanopore PromethION data (>1 Tbp).
- After assembly (long-read), assembly polishing (short-read), and three binning iterations, MAGs were assessed for completeness >90%, contamination <5% and presence of full-length rRNA genes (Fig.1).

SIGNIFICANCE

We recovered **1066 true HQ** MAGs from 23 Danish WWTPs. **93%** of the MAGs that met the completeness thresholds (1045 out of 1124) encoded the rRNA genes. This is a vast improvement over traditional short-read MAGs, of which <10% encode even a partial 16S rRNA gene.

We recovered ~**57%** of the microbial community in the metagenomes at the genus level (30% at species). Providing a large primary database for linkage to amplicon studies that can be expanded in the future.

RESULTS

- 1045 MAGs met the completeness, contamination and rRNA requirements of true HQ.
- A further 21 MAGs had circular, ultra-small genomes, resulting in 1066 HQ MAGs belonging to 566 different species (Fig. 2).
- MAGs were recovered from 32 phyla and represent ~30% of the community in the metagenomes (~57% genus level) (Fig. 3). Many are from uncharacterized groups (Fig. 4).

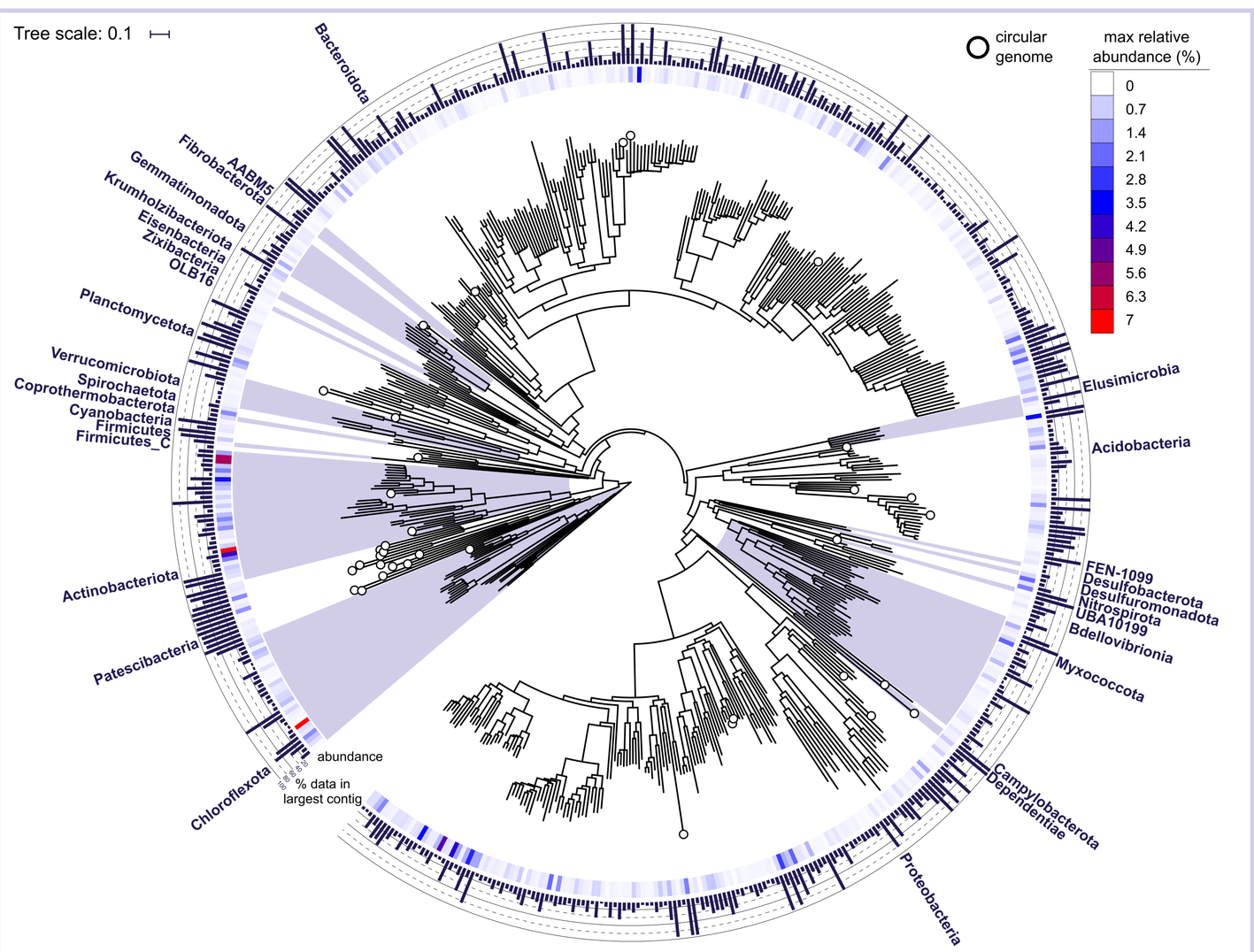

Fig. 2: Phylogenetic genome tree of the 563 bacterial species, created using GTDB-Tk and the concatenated alignment of 120 single copy marker genes. White circles indicate circular MAGs. Heatmap indicates the maximum relative abundance of the species in the metagenomes. Bar chart indicates MAG contiguity.

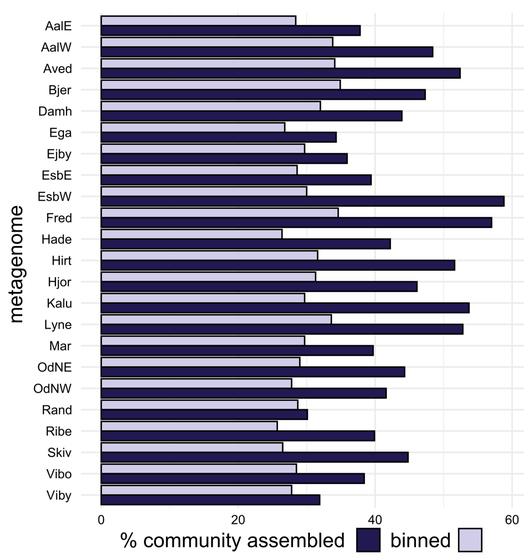

Fig. 3: Fraction of microbial community in the metagenome successfully assembled, or binned as a HQ MAG.

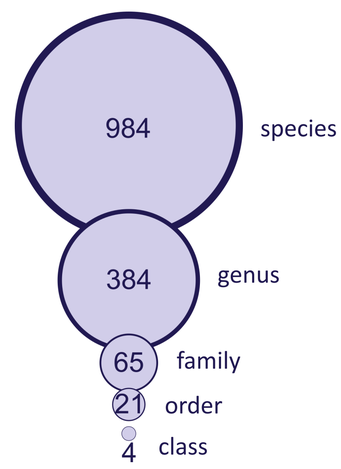

Fig. 4: Number of MAGs belonging to uncharacterised lineages without representatives in GTDB. Only 82 MAGs (7%) have a representative in Refseq.

AALBORG UNIVERSITY
DENMARK

Center for Microbial Communities

VILLUM FONDEN

@PHNLab

@caitysing

Take a picture to download the **poster**