Possibilities and obstacles in recovery of genomes from elusive microbes in complex metagenomes

Karst, Søren Michael; Albertsen, Mads; Nielsen, Jeppe Lund; Nielsen, Kåre Lehmann; Nielsen, Per Halkjær

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Possibilities and obstacles in recovery of genomes from elusive microbes in complex metagenomes

Søren M. Karst, Mads Albertsen, Jeppe L. Nielsen, Kåre L. Nielsen and Per H. Nielsen

Center for Microbial Communities, Aalborg University, Aalborg, Denmark

Introduction

Representative genomes provide an entry point for understanding a given microbial ecosystem. The genomes give insights into the metabolic potential and possible roles of the bacteria, as well as being essential when applying other -omics based techniques. For elusive, unculturable bacteria, metagenomics can be a useful tool to extract their genomes directly from their environments. However the approach has limitations when the community complexity is high and microdiversity is present.

In this study, we explored the potentials and obstacles faced when assembling genomes from complex metagenomes using activated sludge as a model system.

Aims

- To investigate the impact of microdiversity and community complexity on metagenomic assembly, binning and genome extraction.
- To investigate whether short-term enrichment can mitigate the adverse effects of the above

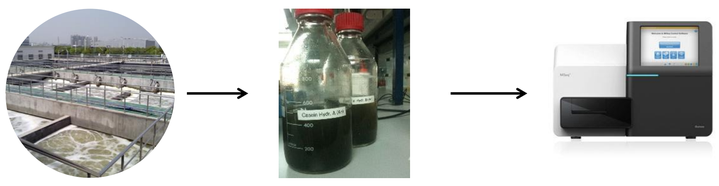
Methods

Short-term Enrichment

Seed sample
- Fullscale EBPR WWTP

Batch Incubations
- Casein, Glucose or Butyrate
- Aerobic or Anaerobic
- All incubations at 20 °C
- 1 sample/d for 7 days

Screen incubations
- V4 16S rRNA amplicons
- 14 of 50+ samples chosen for metagenome sequencing



Metagenome Binning

Sequencing and Assembly
- Illumina PE/MP 2x150bp (300 Gbp)
- *De novo* assembly (kmer: 64) and read mapping using CLC genomics workbench

Coverage Binning
- Read coverage profiles from 14 samples
- Taxonomic assignments
- Essential genes
- Paired-end read connections
- Tetranucleotide frequencies

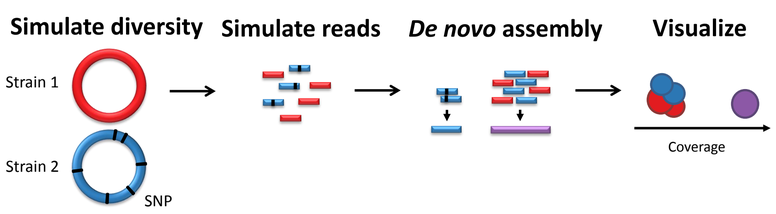
mmgenome

An R toolbox for reproducible genome extraction from metagenomes

Re-Assembly and Finishing
- Bin specific reads *de novo* re-assembled in SPAdes
- Manual finishing using SSPACE, Circos and R-biostrings

GitHub
MadsAlbertsen/mmgenome

Microdiversity Simulation



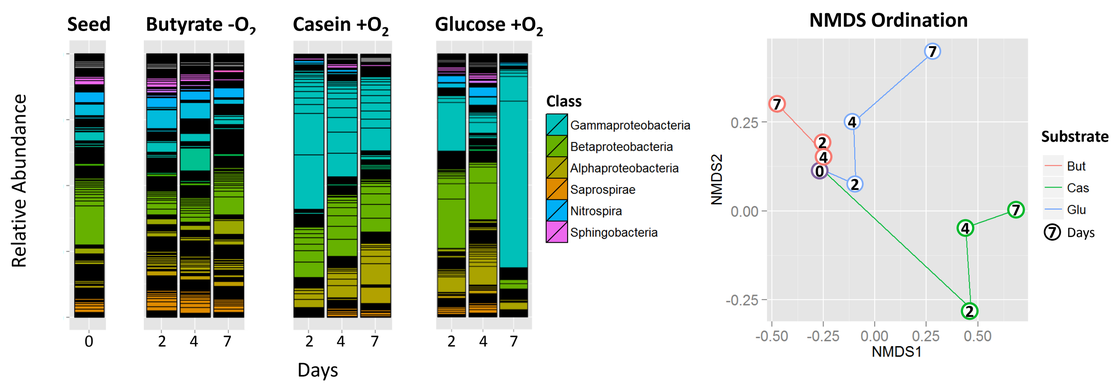
Simulate different degrees of microdiversity using artificial genomes (no repeats)

Some parts of the genomes assembles separately (low coverage), while others co-assemble (high-coverage).

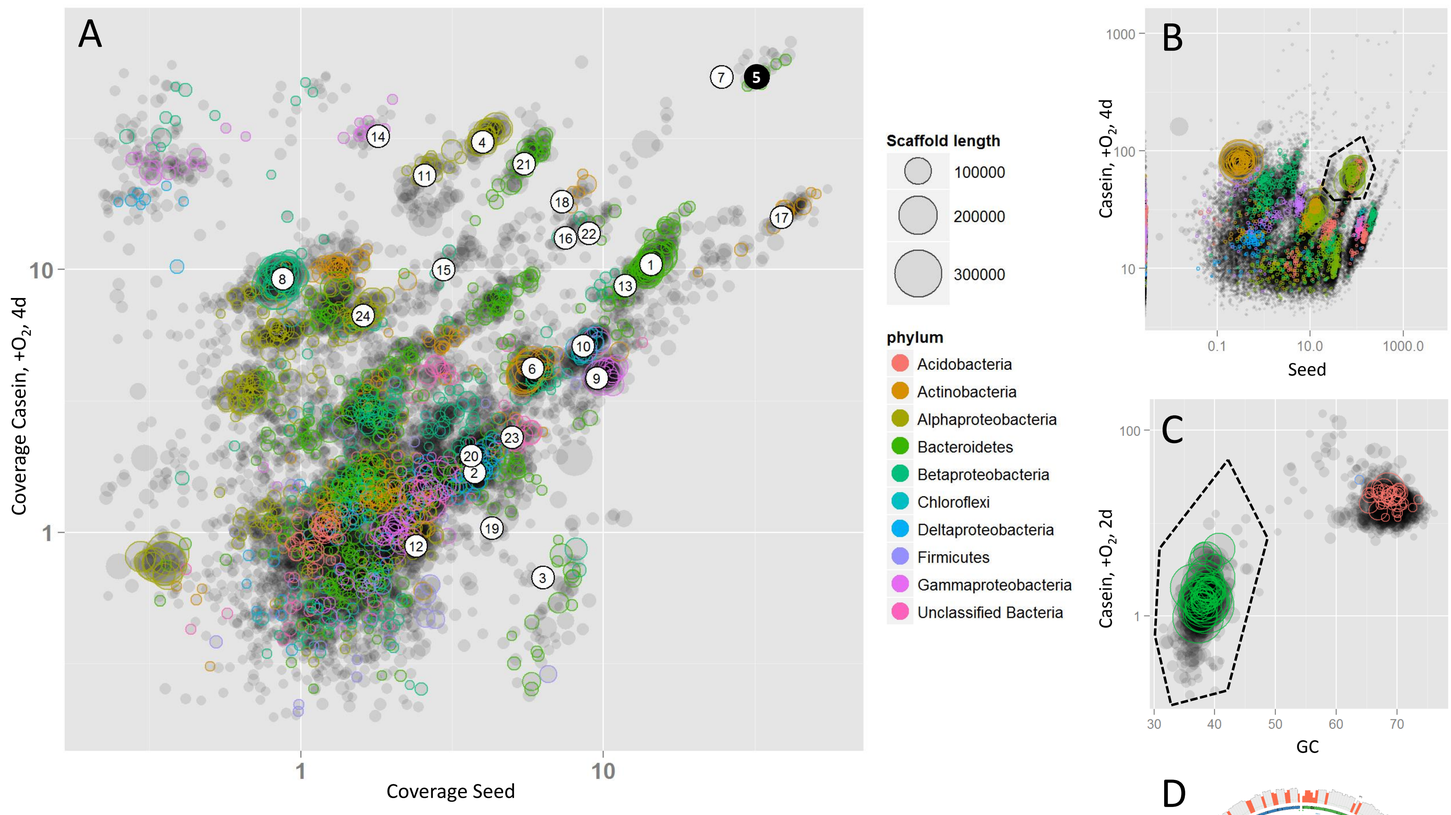
Conclusions

- Enrichments decreased overall complexity, changed relative abundance and facilitated better coverage binning and allowed for numerous high quality genome bins to be extracted.
- Despite the short-term enrichments microdiversity was still present, which compromised recovery of complete genomes from many species.
- The simulation underlines the devastating impact of micro-diversity on assembly and binning.

Results



Screening of short-term enrichments with 16S V4 Amplicon sequencing. Data shown for 3 of 6 incubations. The 16S screening was used to select samples for further metagenome sequencing.



Differential coverage binning and Re-assembly Points represents scaffolds, scaled by their length and colored based on the presence of essential single copy genes. (A) Is a coverage plot of a metagenome assembly of all data. Circled numbers indicate extracted high quality genome bins. In this example bin 5 (A) was targeted for extraction. Bin 5 was found to be enriched in the casein incubation and a sample from this was therefore used for targeted metagenome re-assembly. Bin 5 was binned again from the casein sample metagenome (B & C), and subsequently refined with re-assembly and scaffolding (D). The extracted genome belongs to phylum ***Bacteriodetes*** and genus ***Ferruginibacter***.

	All data		Casein, +O2, 4d		Re-assembly
	Metagenome	Bin 5	Metagenome	Bin 5	Bin 5
Size (Mbp)	2048.3	2.7	125.9	3	3.1
Scaffolds	915 040	702	62422	357	21
Scaffold Max (kbp)	340	34	94	67	1382
Ess. Genes	43522	93/105	2996	104/105	105/105
Species	> 430	1	>29	1	1
Seq. data (Gbp)	300	-	15.5	-	-

Binning Statistics. The statistics show the genome extraction progression from the metagenome assemblies, through binning to the final refinement. It is clear that the assembly of the enriched casein sample with less microdiversity provides a better basis for binning, despite far less sequence data being used.

Simulation of different degrees of microdiversity. Strain identity refers to the percent bp in the two artificial genomes that are identical.

smk@bio.aau.dk

www.cmc.aau.dk

CENTER FOR MICROBIAL COMMUNITIES | AALBORG UNIVERSITY