



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Developing new bioinformatic methods to supercharge genome-centric metagenomics using machine learning

Heidelbach, Søren; Lamurias, Andre; Sereika, Mantas; Nielsen, Thomas Dyhre; Hose, Katja; Albertsen, Mads

Publication date:
2021

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Heidelbach, S., Lamurias, A., Sereika, M., Nielsen, T. D., Hose, K., & Albertsen, M. (2021). *Developing new bioinformatic methods to supercharge genome-centric metagenomics using machine learning*. Abstract from Danish Microbiological Society congress 2021, Copenhagen, Denmark.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Developing new bioinformatic methods to supercharge genome-centric metagenomics using machine learning

Søren Heidelberg¹, Andre Lamurias², Mantas Sereika¹, Thomas Dyhre Nielsen², Katja Hose², Mads Albertsen¹

¹Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Aalborg, Denmark

²Department of Computer Science, Aalborg University, Aalborg, Denmark

Microbes are everywhere and play important roles in most aspects of life and an important part of complex microbial community investigation is the extraction of single organism genomes. The maturation of metagenomic binning techniques has greatly increased the quality of metagenomic assembled genomes, by utilizing features such as sequence coverage and K-mer frequencies. However, challenges still remain with these approaches. K-mer frequencies depend on long contigs for stabilisation and sequence coverage information can be biased by high copy number sequences. The nanopore sequencing platform, which is already an often integrated step in the metagenomic analysis, produces information rich data containing information on the possible methylation of DNA bases. Methylation represents a powerful feature, as the DNA modification depends on the state of the methylome of the organism. Here we explore incorporation of methylation modification as a feature into metagenomic binning using machine learning to complement challenges inherent in sequence centric binning features.