Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Assessing Illumina, Nanopore and PacBio sequencing platforms at recovering high quality genomes from complex microbial communities

Sereika, Mantas; Kirkegaard, Rasmus Hansen; Sørensen, Emil Aarre; Karst, Søren Michael; Yssing Michaelsen, Thomas; Albertsen, Mads

Creative Commons License
Unspecified

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Sereika, M., Kirkegaard, R. H., Sørensen, E. A., Karst, S. M., Yssing Michaelsen, T., & Albertsen, M. (2021). *Assessing Illumina, Nanopore and PacBio sequencing platforms at recovering high quality genomes from complex microbial communities*. Poster presented at 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.

Downloaded from vbn.aau.dk on: November 19, 2022

Assessing Illumina, Nanopore and PacBio sequencing at recovering high quality genomes from complex microbial communities

Mantas Sereika, Rasmus Hansen Kirkegaard , Søren Michael Karst, Thomas Yssing Michaelsen,

Emil Aare Sørensen, Mads Albertsen

Center for Microbial Communities, Aalborg University, Aalborg, Denmark

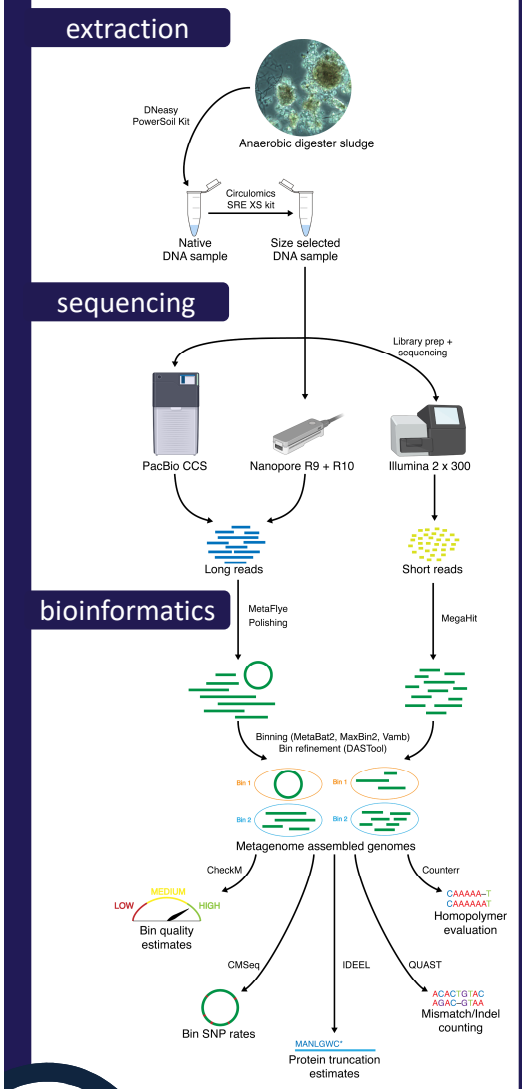
Introduction

Cheap short-read DNA sequencing has led to a massive increase of genome reference databases, although many genome drafts from environmental systems are highly fragmented due to strain heterogeneity, micro-diversity, and repeats that cannot be resolved with short reads. This poses an issue, as the absence of genes in incomplete and fragmented genomes fails to provide important genomic information about microbial species. However, one way to circumvent this problem is to use bigger pieces of DNA to assemble the genomes.

Aim

Investigate the potential of acquiring high quality genomes from complex samples with PacBio CCS and Nanopore sequencing.

Methods

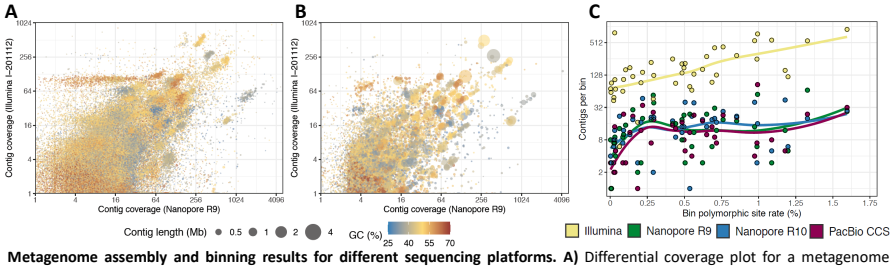


Conclusions

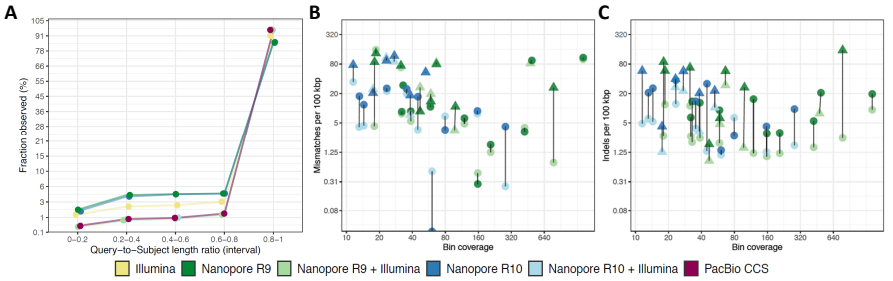
- Long-read sequencing vastly outperforms short reads at genome-centric metagenomics
- Hybrid Nanopore + Illumina method produces genomes in comparable quality as PacBio CCS
- A Nanopore-only approach can be used to get high quality drafts, but some errors still persist
- Nanopore R9 + Illumina was found to be the most cost-effective approach at obtaining high quality genome drafts from complex microbial communities

Results

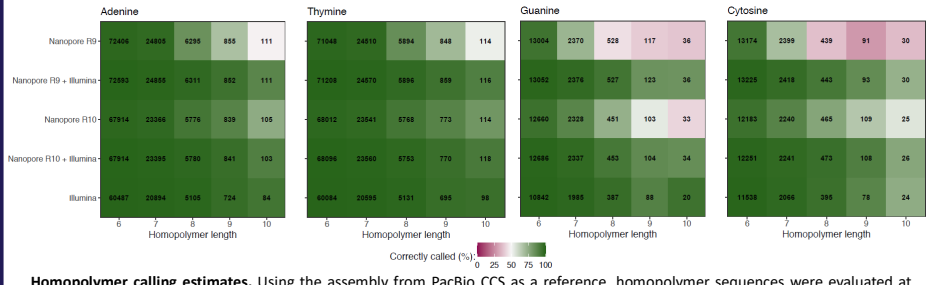
Feature	Illumina	Nanopore R9	Nanopore R9 + Illumina	Nanopore R10	Nanopore R10 + Illumina	PacBio CCS
HQ MAGs	8	64	**86**	33	45	74
MQ MAGs	83	114	95	64	63	72
LQ MAGs	3	28	26	18	12	22
Contaminated MAGs	10	6	13	6	6	14
Reads binned (%)	76	86	85	86	86	83
Reads binned in HQ MAGs (%)	16	46	**49**	39	41	48
Contigs binned (%)	55	68	71	71	74	73
Contigs binned in HQ MAGs (%)	4	22	30	22	29	**34**
Sequencing costs ($)	1,200	811	2,011	811	2,011	4,420
Costs per HQ MAG ($)	150	**13**	23	25	45	60



Metagenome assembly and binning results for different sequencing platforms. A) Differential coverage plot for a metagenome acquired by Illumina read assembly. A total of 145,876 contigs above 1 kb length were acquired with an N50 of 3.5 kb. In contrast, assembling a metagenome with **B)** Nanopore R9 reads produced an assembly of 24,680 contigs (7 circular and above 0.5 Mb length) with an N50 of 79.9 kb. **C)** Genome bins fragmentation levels at varying bin SNP rates. Bins, recovered from Illumina-only assembly, featured significantly lesser bin contiguity (median 183.5 contigs per HQ MAG) than long read platforms (9, 14.5 and 15 median contigs per HQ MAG with PacBio CCS, Nanopore R9 and R10 assemblies, respectively).



Quality evaluation of genome bins from Nanopore sequencing. A) IDEEL test, showcasing that 14.99 and 14.44 % of proteins recovered from Nanopore R9 and R10, respectively, bins exhibited greater than 20 % estimated protein sequence truncation. For the hybrid Nanopore + Illumina approach and PacBio CCS, the estimates protein truncations were similar (mean 4.19 %). Using bins from PacBio CCS as a reference, Illumina read polishing of Nanopore assemblies was observed to reduce the amount of **B)** mismatches by at least 20 % in 6/18 and 8/15 bins for R9 and R10 data, respectively. Also, the reduction of **C)** indel rates from short read polishing was detected for 18 R9 and 14 R10 Nanopore bins. In plots B) and C) high quality bins are marked by circles, while medium quality genome drafts are represented by triangles.



Homopolymer calling estimates. Using the assembly from PacBio CCS as a reference, homopolymer sequences were evaluated at different lengths. Increases in homopolymer errors at longer lengths for all nucleotides were observed in metagenomes from Nanopore R9 and R10, although homopolymer miscalling was more pronounced in R9 data, while Illumina-polished metagenomes exhibited improved homopolymer calling.

mase@bio.aau.dk

https://github.com/Serka-M

POUL DUE JENSEN / GRUNDFOS FOUNDATION

VILLUM FONDEN

MICRO FLORA DANICA