UNIVERSITÄT DUISBURG ESSEN

LOW-BIOMASS NANOPORE METAGENOMICS

Sophie Anja Simon, M.Sc.
Group for Environmental Metagenomics (GEM)
University of Duisburg-Essen
Department of Chemistry
Research Center One Health Ruhr

sophie.simon@uni-due.de

Planetary Protection: Metagenomics in Spaceflight Workshop | November 20, 2024

GEYSERS & SPRINGS

Gates into the deep biosphere

Cold-water geyser, Andernach, Germany

Sulfuric Spring “Islinger Mühlbach” (MSI), Regensburg, Germany

- Geysers and springs to sample groundwater originating from the terrestrial subsurface
- Average DNA yields for given sampling sites:
  - Andernach: 6.0 ng/µl
  - MSI: 5.86 ng/µl
- Recommended DNA quantity for Nanopore sequencing often exceeds the recoverable amount of DNA of environmental aquatic samples

GENOME-RESOLVED NANOPORE METAGENOMICS

With little to (almost) no DNA?

Ligation sequencing DNA V14 (SQK-LSK114)

Version: GDE_9161_v114_revL_29Jun2022
Last update: 06/01/2023

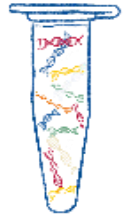
Oxford NANOPORE Technologies

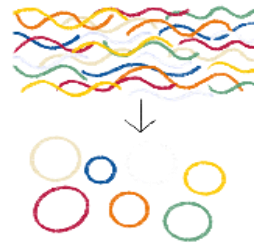
Before start checklist

Materials

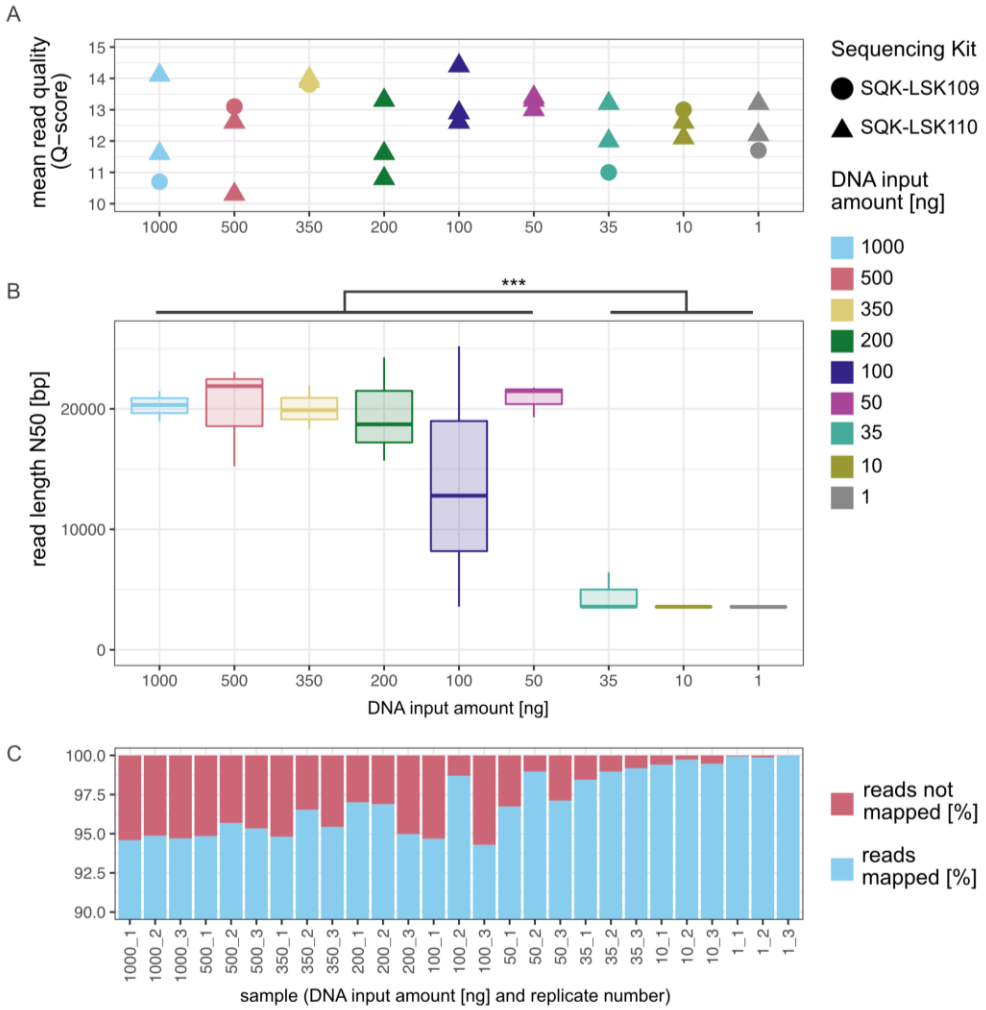
- [ ] 1 µg (or 100-200 fmol) high molecular weight genomic/amplicon DNA

- Microbial DNA Standard (HMW ZymoBIOMICS)
  - DNA from seven bacterial and one fungal strain
  - High molecular weight: > 50 kb in size
- 27 metagenomes
  - Nine DNA input level in triplicates
- Sequencing depth: 1 Gb (if possible)
- Evaluating the results at multiple levels of a genome-resolved metagenomics pipeline
  - Sequencing quality
  - Community composition
  - Assembly quality
  - Recovery of MAGs
- Impact of Nanopore reads generated with lower-than-recommended DNA amounts on hybrid assemblies in combination with Illumina reads



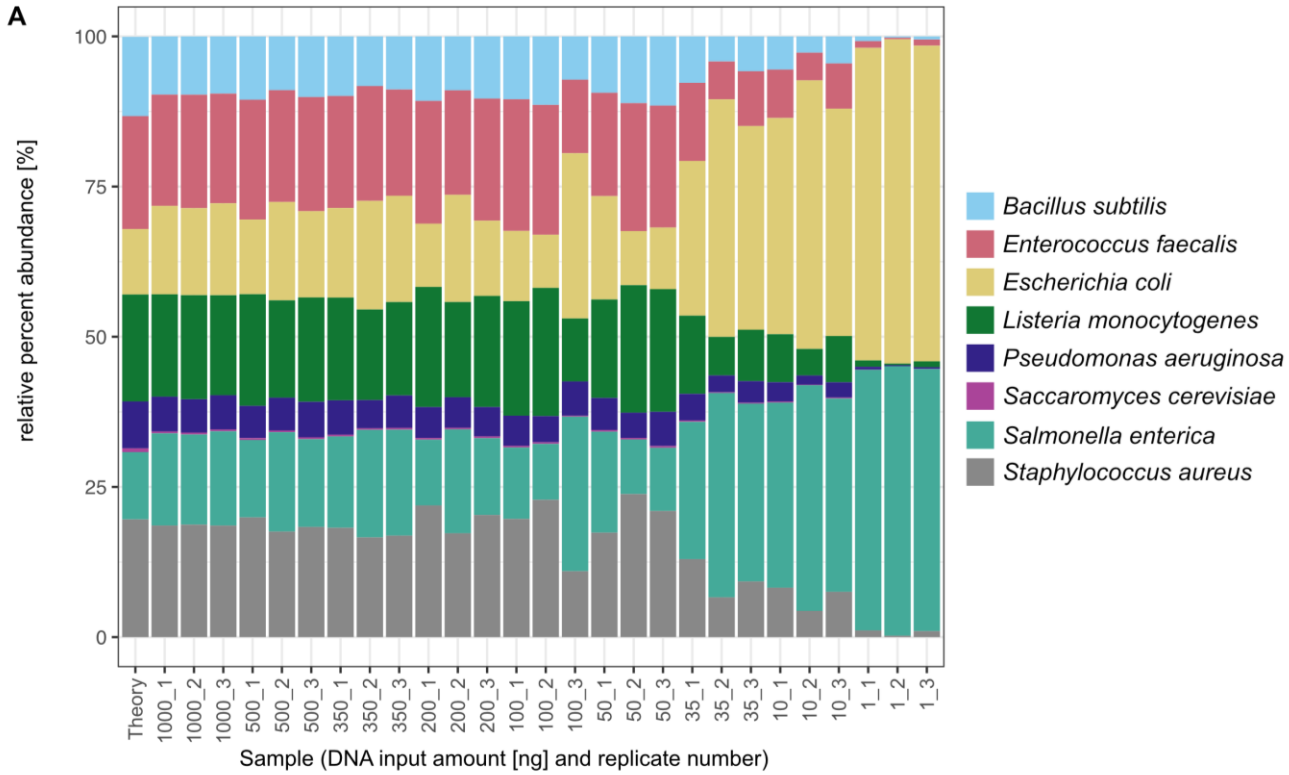


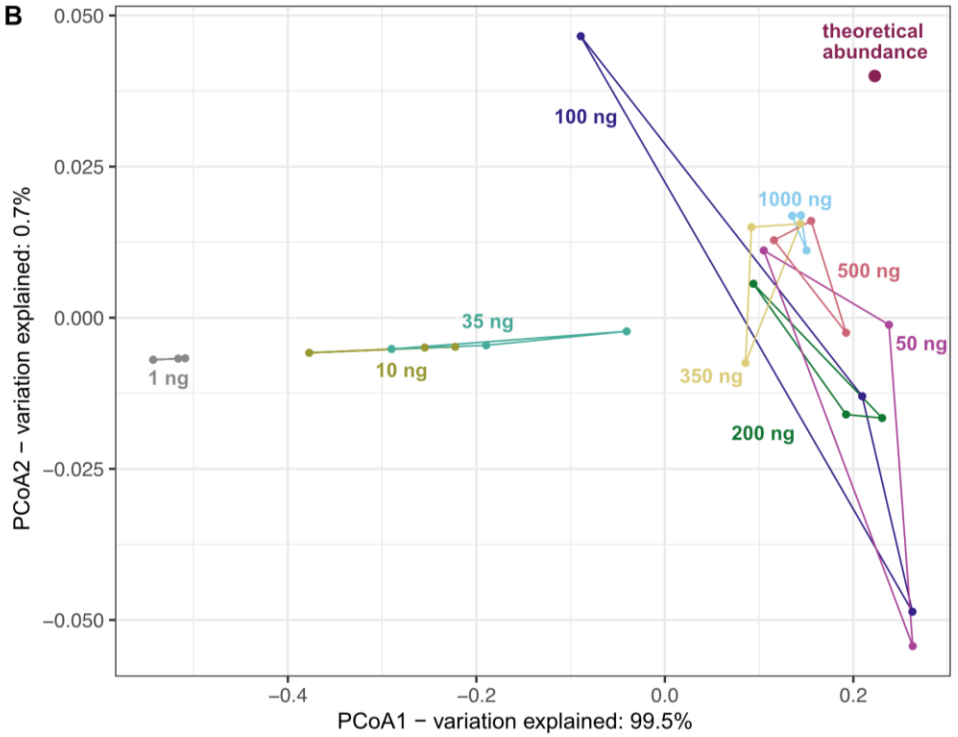
SEQUENCING QUALITY



- Q-Score remains stable across all sequenced input quantities
  - Phred quality scores Q are logarithmically related to the base-calling error probabilities;
  - Q-Score 10 = 90% base call accuracy
- Read length N50 is > 20 kb down to 50 ng DNA input
- Read-to-reference alignments highlights the consistent sequencing quality
  - Interesting: mapping rate improves for smaller input
    - Shorter reads might map better?

COMMUNITY COMPOSITION

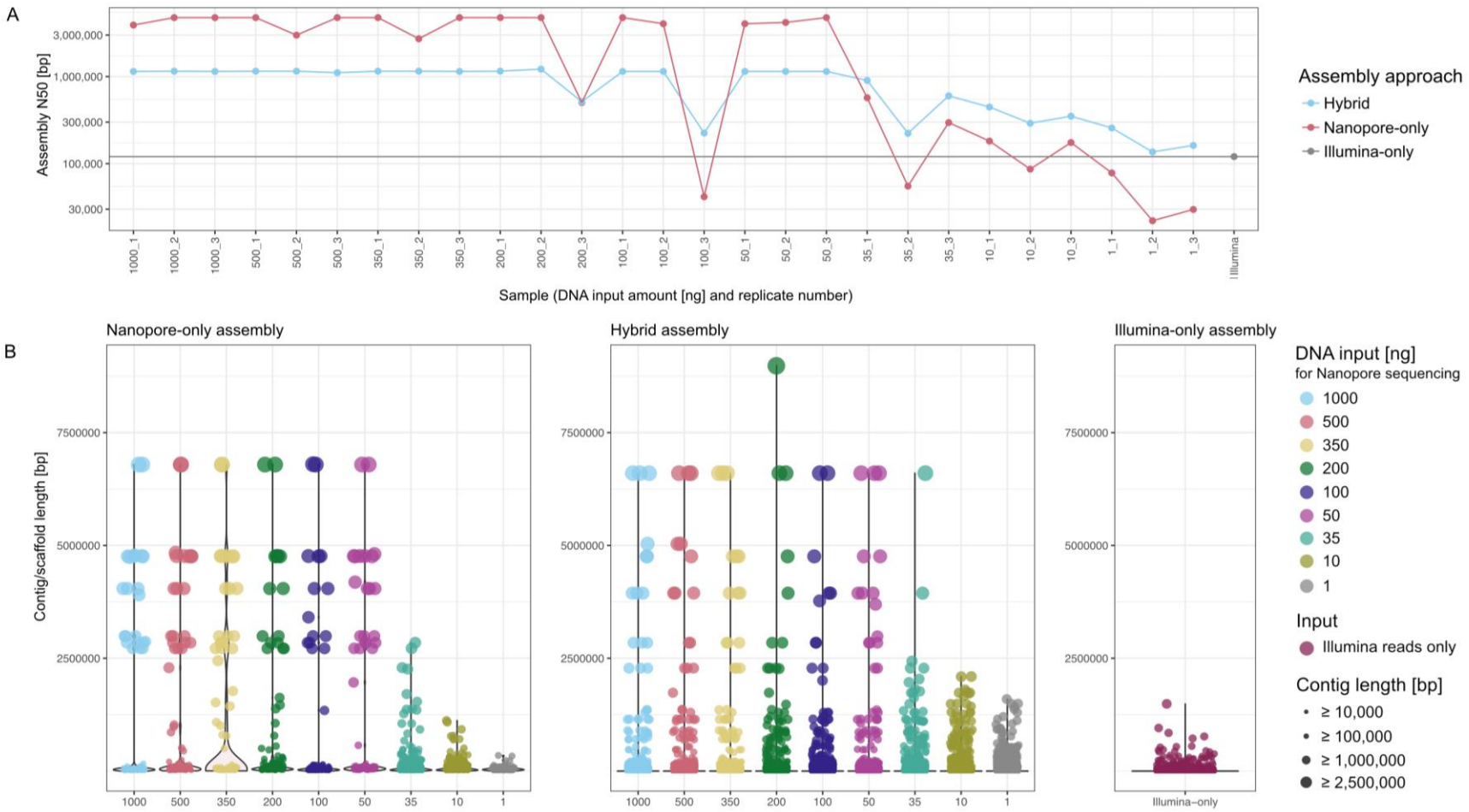




⇒ Each species is qualitatively detectable down to 1 ng

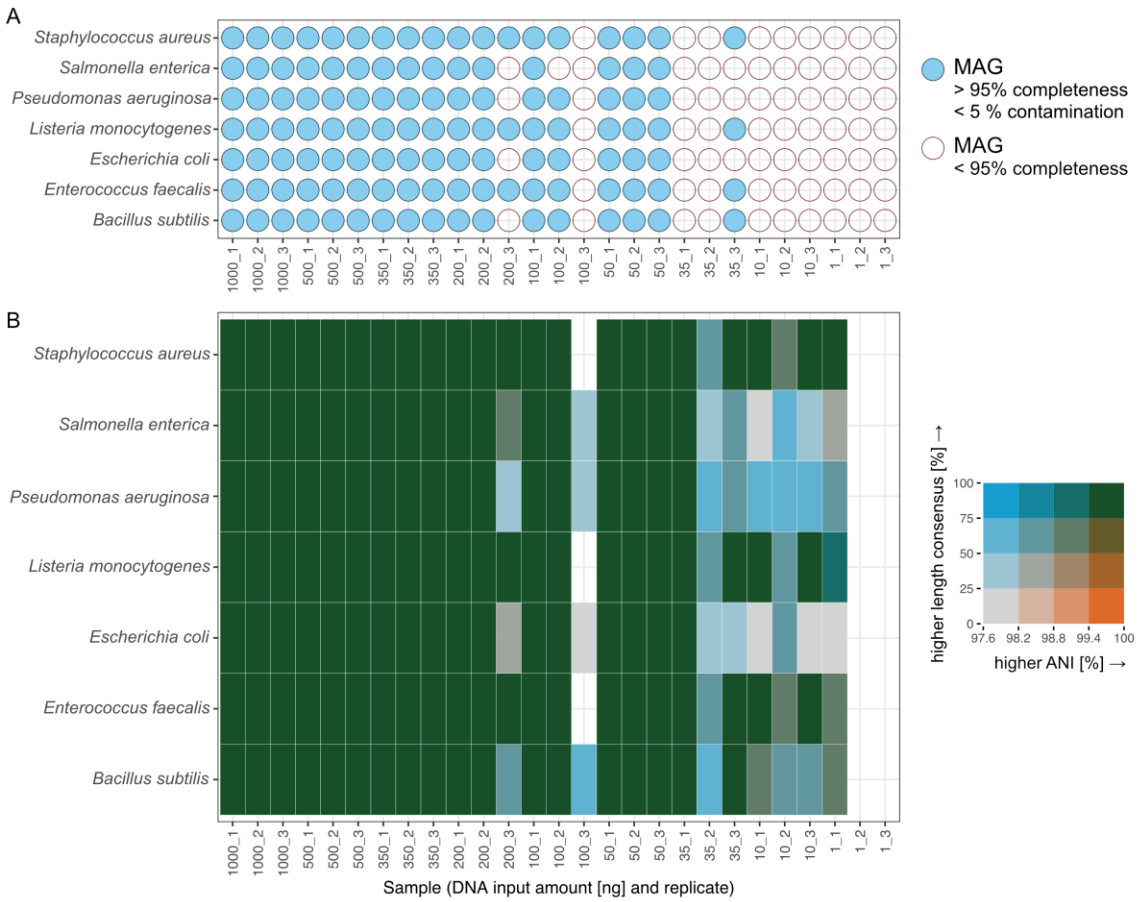
⇒ Significant correlation of relative abundance and theoretical abundance down to 50 ng

A READ OF LENGTH IS A JOY FOREVER: ASSEMBLY STATISTICS

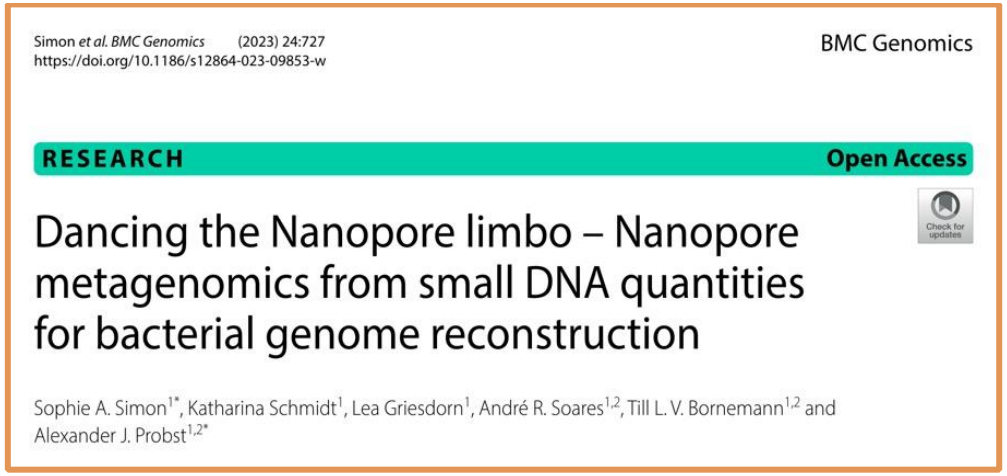


- A “handful“ of Nanopore-long reads improves hybrid assemblies compared to short-read only
- Recovery of cMAGS (closed MAGs) only possible when using long-reads

RECONSTRUCTION OF GENOMES



- Manual binning of Nanopore-only metagenomes
- Contamination/completeness based on 51 universal bacterial single-copy genes
- ANI = average nucleotide identity
- Length consensus = metric comparing how well reference and obtained MAG align to each other

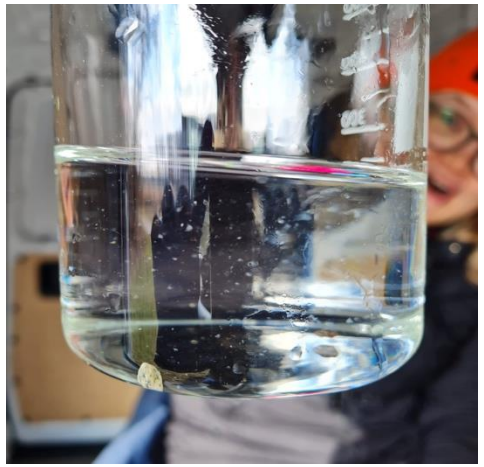

Simon et al. BMC Genomics (2023) 24:727
https://doi.org/10.1186/s12864-023-09853-w
BMC Genomics
RESEARCH
Open Access

Dancing the Nanopore limbo – Nanopore metagenomics from small DNA quantities for bacterial genome reconstruction

Sophie A. Simon[1*], Katharina Schmidt[1], Lea Griesdorn[1], André R. Soares[1,2], Till L. V. Bornemann[1,2] and Alexander J. Probst[1,2*]

APPLIED LOW-INPUT NANOPORE METAGENOMICS

Example: Ca. Altiarchaeum hamiconexum (MSI)

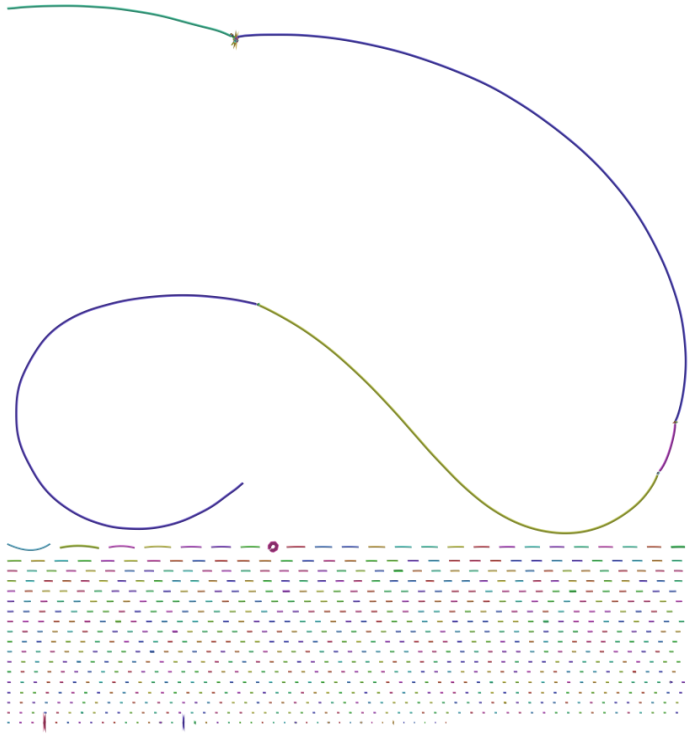


Nearly pure biofilm flocks formed by *Candidatus* Altiarchaeum hamiconexum (MSI)

- Uncultivated archaeal genus
  - Globally distributed
  - Primary producers
- Several times sequenced (short- and long-reads)

  ⇒ highly fragmented genomes
- Here: Sequencing of one single biofilm flock to reduce strain heterogeneity
- DNA extraction from ten single Alti flocks:
  - Highest concentration: 0.30 ng/µL DNA
- Sequencing of ~15 ng of DNA using Nanopore Sequencing via MinION resulted in 2.1 Gb

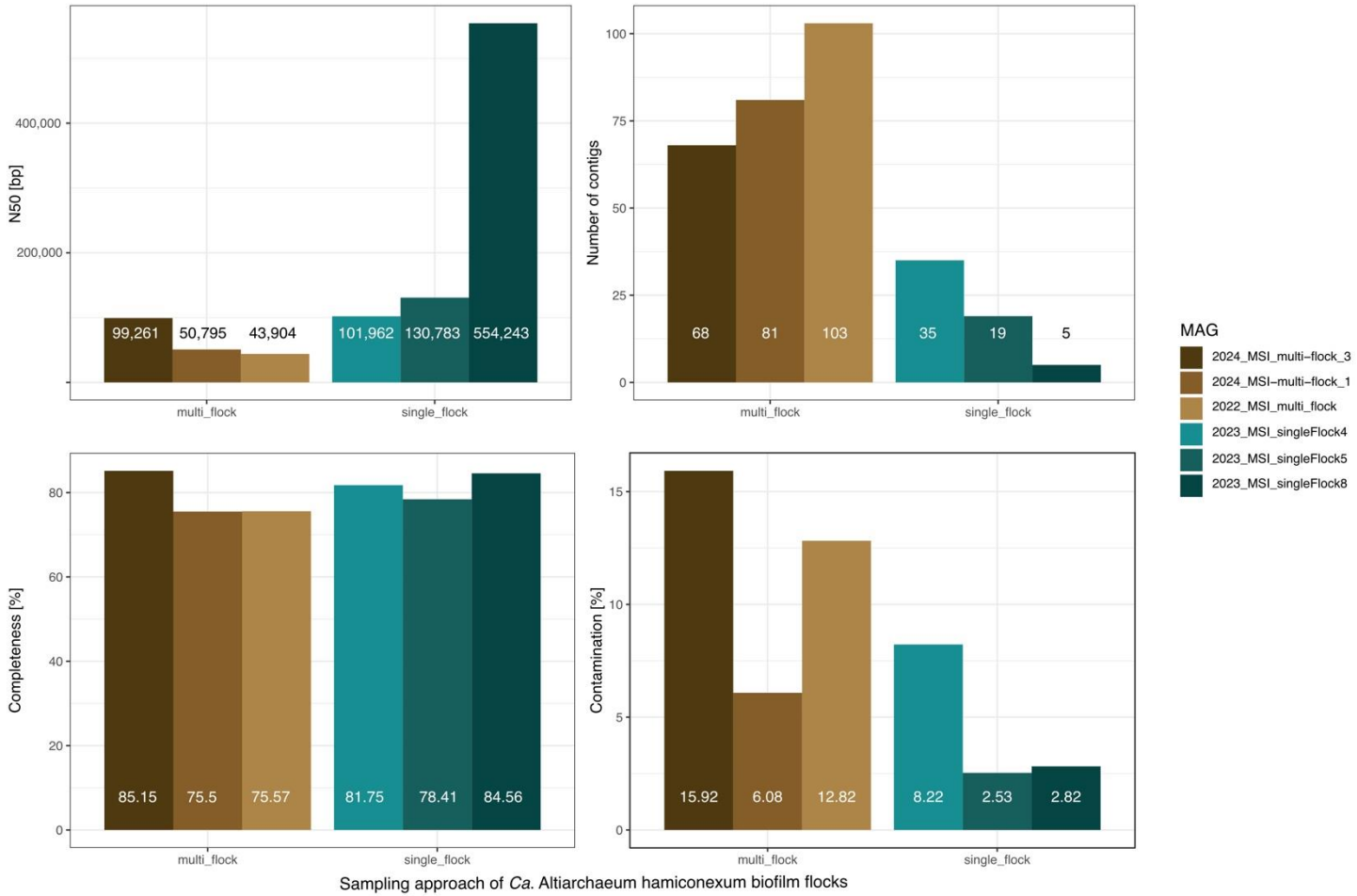
  ⇒ best reconstruction of the genome (so far)



Assembly Graph

IMPROVED QUALITY OF MAGS

Applying low-biomass Nanopore metagenomics



- Sequencing of two additional single biofilm flocks
  - 0.188 ng/µL
  - 0.252 ng/µL
- Reduced strain-heterogeneity improved assemblies

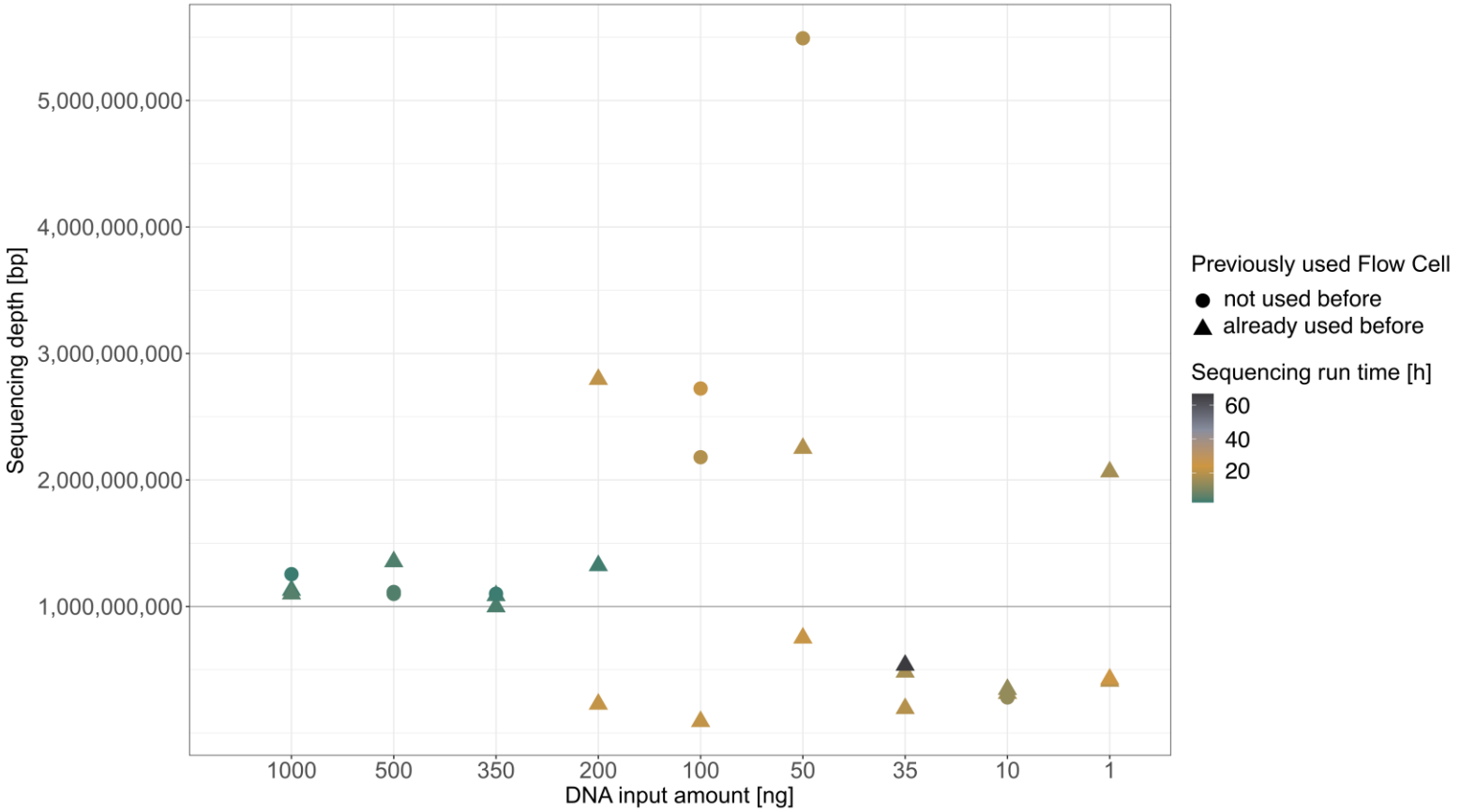
⇒ **Use of long-reads in every assembly-based metagenomic study including those from low-biomass environments is highly recommended**

ACKNOWLEDGEMENTS

SPONSORED BY THE

Federal Ministry of Education and Research

BACK UP: SEQUENCING DEPTH



- The smaller the DNA input, the lower the sequencing depth
- Sequencing runs were manually stopped