NANOPORE TAKES OVER 16S rRNA GENE AMPLICON SEQUENCING

Yan Hui[1], Grzegorz Nowicki[2], Josue L. Castro[1], Dennis S. Nielsen[1] ,Lukasz Krych[1,2]

[1]Food Microbiology & Fermentation, Department of Food Science, University of Copenhagen, 1958 Frederiksberg C, Denmark; [22] GenXone S.A., 62-002 Złotniki, Poland

INTRODUCTION

16S rRNA gene amplicon sequencing remains a popular and widely used method to study bacterial community in multiple environments, although its resolution in bacterial classification is somewhere between **genus and species level**.

For over a decade 16S rRNA gene amplicon sequencing was a domain of next generation sequencing (NGS) platforms. **NGS** was first to allow for high throughput analysis of multiple samples despite their main **limitation** that was and still remains: **a short read**. To surpass this limitation only a short fragment of the full 16S rRNA gene was amplified and subjected for sequencing. This resulted in development of multiple protocols targeting different variable regions of 16S rDNA causing problem with data comparability across many studies.

Nanopore sequencing is the only technology that allows **high throughput,** real-time sequencing of near-full-length 16S rRNA gene amplicons at **low cost**, yet still suffers from a **relatively high error rate on a single molecule level**. The utilisation of unique molecular identifiers (UMI) brought a hope for resolving this setback, but in rich bacterial communities would require ultra-deep sequencing what would have jeopardised its cost-effectiveness. It is clear that by fixing the problem related to primers universality (targeting more bacteria) and **reducing the error** on a single molecule level combined with low running costs, would make nanopore based sequencing technology the most obvious approach in 16S rRNA gene amplicon sequencing strategy.

NEW WET-LAB SOLUTION

We have developed a novel wet-lab protocol with multiplex primers targeting the most universal regions of the 16S rRNA gene that according to the in-silico analysis captures **nearly 7,000 bacterial species more** (Silva database) than the best primer combination alone. Also, our wet-lab solution allows for multiplexing of up to **192 samples** that in combination with 96 barcodes from the Native Barcoding Expansion gives the theoretical multiplexing possibility reaching **18,432** samples per single run/flow cell (**PromethION** application).

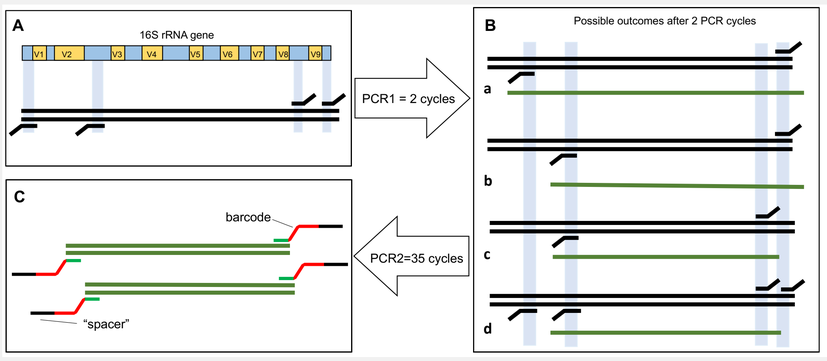

Figure 1 Wet-lab protocol overview for the novel 16S rRNA gene amplicon sequencing with Oxford nanopore Technologies platforms. A) PCR1 with multiple forward and reverse primers to ensure highest universality of 16S rRNA gene targets. B) possible template outcomes after 2 cycles of PCR with: (a) most distant primers alone, (b) proximal forward and distal reverse (c) proximal forward and proximal reverse (d) both forwards and/or both reverse attach at the same time. C) PCR2 for barcoding with 1-192 custom designed barcodes.

NEW PIPELINE: "KAMP"

We have developed a novel pipeline called "**Kamp**" that allows for the **error corrections** and recovery of long reads with the average quality ranging from **97% to 100%** allowing for more accurate **species level identification**. We adopted the binning theory from metagenomics and present an **end-to-end amplicon denoise** workflow based on kmer profiles (Kamp). With snakemake as a scheduler, Kamp is constituted by independent modules of demultiplexing, quality control, kmer binning, consensus polishing, taxonomy assignment, and **phylogenetic reconstruction**. This allows for construction of **phylogenetical trees** and utilization of **UniFrac** distances. Kamp is versatile and flexible for reference-free microbiome analysis of metabarcoding libraries. Incorporated with our wet-lab protocol. Kamp can process **192 samples** in either pooling or single (sample-by-sample) mode for the consideration of sensitivity and computing bottlenecks. Unlike the read-by-read classification tools like EPI2ME, Kamp generates consensus from clusters with similar kmer profiles, which avoids the massive database quires afterward.

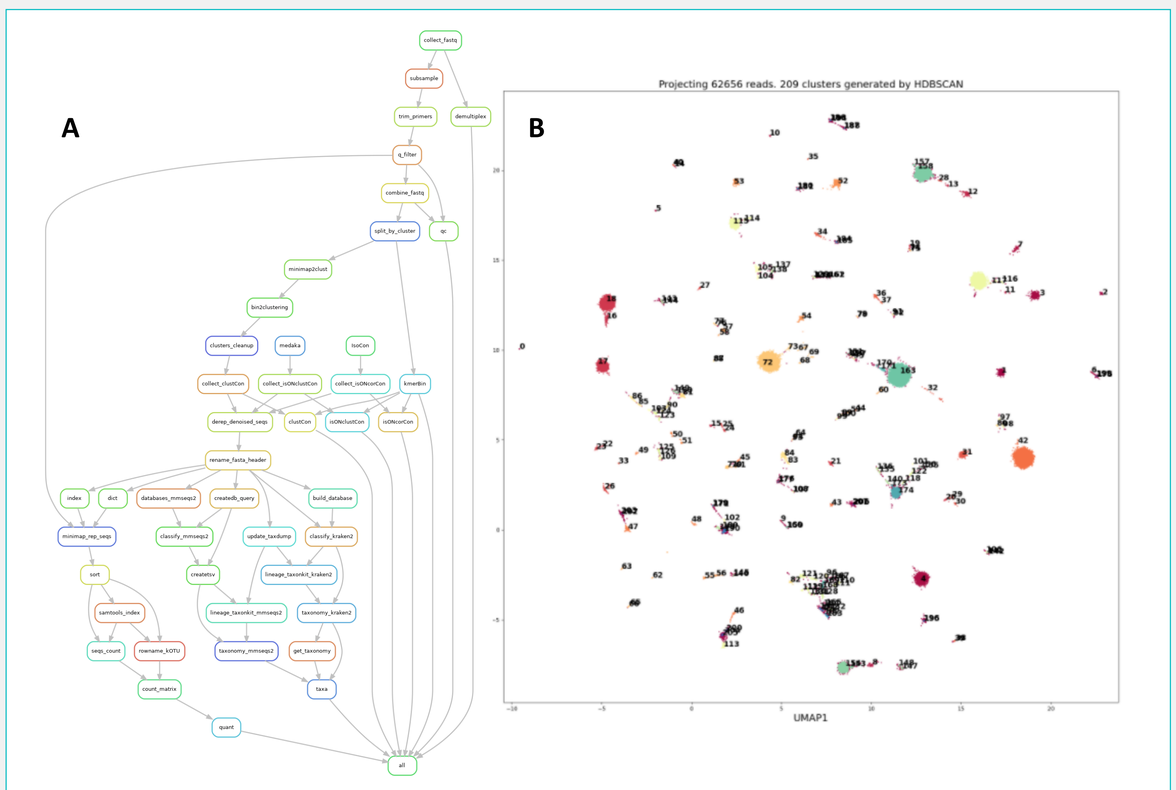

Figure 2 Novel kmer based pipeline for 16S rRNA gene amplicon sequencing with Oxford Nanopore technologies platforms. A) The Kamp workflow overview: Supported by the recent findings, the binning process is aided by Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). The consensus of kmer bins may be heterozygous of similar sequences, especially in a complex community. To address this, we implemented up-to-date alignment (minimap2) and clustering (isONclust) tools for long noisy reads. The representative sequences can be generated from direct kmer bins, or sub-clusters based on alignment or *de novo* clustering as prior requests and polished with racon and medaka. The highly similar gene copies can also be deciphered with IsoCon after error correction with isONcorrect. **B) The PCA depicting the example of kmer based binning**

RESULTS

- In 12 multi-targeted priming libraries for a commercial ten-strain mock community (ZymoBIOMICS), Kamp, at a subsampling depth of 5000 reads, recovered **nearly all copies of 16S rRNA**. Using NCBI RefSeq as a query database, the consensus sequences showed an **average identity of 99.6% (SD = 1.08%).**
- The new wet-lab for sequencing of 96 samples takes **~ 6h** what includes PCR1, clean-up, PCR2 and ligation protocol (SQK-LSK110)
- The analysis with **Kamp** to reach **high quality**, near-full length 16S rRNA sequences is relatively computing demanding and takes **~1 to 1,5 day** for ~12M reads from single R9.4 flowcell on a 30-core computer (64GB of RAM).
- The novel wet-lab protoclol in combination witht the Kamp based analysis **outperformed** the analysis of mock community using ONT read by read classification (Figure 3) but also illumina based sequeing using V3 region (data not shown).

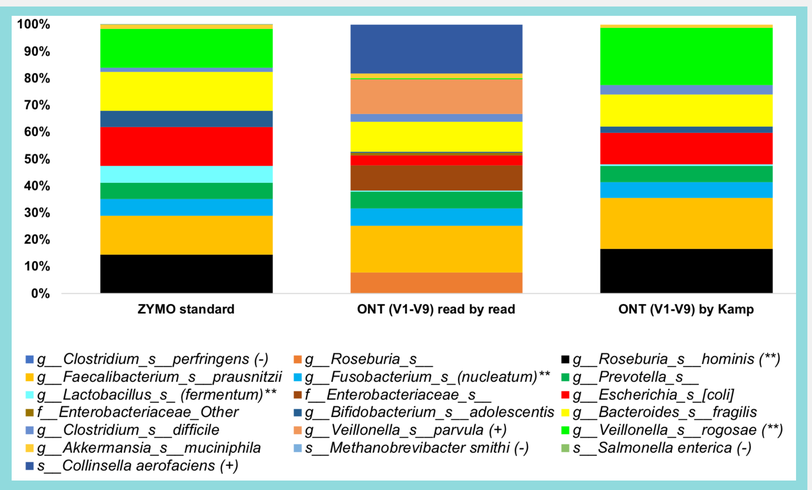

Figure 3. Barchart presenting bacterial relative distribution of ZYMO mock studied with 16S rRNA gene amplicon sequencing analyzed using read by read approach (minimap2) and Kamp pipeline. (-) not detected ; ** species level identification only by Kamp; [] not assigned to the species level

SUMMARY

Taken together, our wet-lab and dry-lab (**Kamp**) solutions dedicated to Oxford Nanopore Technologies offer the **most universal**, **accurate** and **cost-effective** strategy for 16S rRNA gene amplicon sequencing currently available on the market.

UNIVERSITY OF COPENHAGEN

GENXONE SEQUENCING

LC 2022 LONDON CALLING