Technical Note No. 1/2025V02:1-8

Targeted 16S Amplicon ONT MinION Workflow

Keywords: culture-independent, bacteria, polymicrobial infections, 16S rDNA PCR, Micro-Dx™, SelectNA™*plus*, automated DNA extraction, Sanger sequencing, Next Generation Sequencing, Nanopore Sequencing, MinION

Björn Abendroth – Molzym GmbH & Co. KG, Bremen, Germany

Jakub Kantor MD - Department of Medical Microbiology, Second Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

Abstract

Nanopore sequencing is a powerful tool for fast and comprehensive microbiome analysis and pathogen identification.

In this technical note, we describe a 16S rRNA amplicon workflow using the CE IVD-marked Micro-Dx™ kit (Molzym) for isolation of microbial DNA and amplification of bacterial DNA by 16S broad-range PCR in combination with Oxford Nanopore Technology (ONT) sequencing. For the bioinformatic analyses, different Bio-IT approaches were performed and evaluated regarding the taxonomic classification outcome in comparison to known species or Sanger sequencing results.

Introduction

Nanopore sequencing as an alternative to Sanger sequencing

The 16S rRNA targeted Nanopore sequencing with ONT is a fast and comprehensive method for a microbial species identification. It is a good alternative to Sanger sequencing, with the advantage that microbial communities or mixed infections can also be analyzed.

Using culture-independent molecular analysis of bacteria in clinical samples often faces challenges due to the low abundance of pathogens. A significant issue is the overwhelming presence of host DNA, which can be several thousand times more abundant than microbial target sequences, thereby negatively affecting the sensitivity and specificity of amplification assays targeting 16S rRNA and other genes. This effect can be counteracted using the MolYsis™ technology (Molzym) which efficiently depletes host DNA through a differential lysis approach.

For our here described workflow, the CE IVD-marked Micro-Dx™ kit and SelectNA™*plus* device (Molzym) were used. The 16S broad-range PCR generated amplicons underwent Sanger and targeted Nanopore sequencing (ONT). NGS data analysis was conducted with EPI2ME wf-16S [1] and the EMU data evaluation [2].

The aim was the development of a targeted 16S amplicon sequencing workflow with ONT MinION using artificial and clinical samples. Questions to be answered are the needed time for achieving appr. 10.000 reads per sample, the concordance of Sanger and ONT sequencing results and the potential influence of the analysis pipeline.

Material and Methods

Specimens

Artificial samples and clinical samples were used:

i. a positive DNA control containing a defined concentration of *Bacillus subtilis* was used in several dilutions.
ii. bovine tissue specimens spiked with defined concentrations of *Streptococcus agalactia* and *Escherichia coli*.
iii. clinical samples are various specimens with confirmed microbial infections, e.g. from Motol University Hospital (Prague, Czech Republic).

Sample preparation and DNA isolation

DNA extraction was done using Micro-Dx™ kit (Molzym), including the pre-treatment for tissue specimens prior to automated human DNA depletion and microbial DNA extraction on the SelectNA™*plus* device.

PCR amplification

DNA eluates underwent broad-range 16S real-time PCR analysis according to the Micro-Dx™ manual with assay MA Bac amplifying the V3/V4 region of the 16S rRNA gene, enabling the sensitive detection down to femtograms of bacterial DNA using 40 PCR cycles. Protocols for amplification are validated for various PCR cycler, among them the LightCycler® 96 (Roche) or CFX96™ (BioRad). PCR was performed twice per sample, one for each of the following identification methods.

Sanger sequencing

For Sanger sequencing, 25 µl of amplicons were purified with QIAquick® PCR Purification Kit (Qiagen) and DNA was eluted in 30 µl. Sanger sequencing was conducted by a third party sequencing provider. Bacteria were identified by BLAST analyses, in which the sequences obtained were compared to known sequences in databases, e.g. NCBI or SepsiTest™-BLAST.

Targeted Nanopore sequencing

a. *Purification of amplicons for library preparation*

For the library preparation the Rapid Barcoding Kit SQK-RBK114.96 (ONT) was used. Pre library preparation, the amplicons were purified using AMPure XP Beads (AXP) (incl. in Rapid Barcoding Kit) by adding a 1.8-time volume of AXP per sample (i.e. 25 µL PCR product + 45 µl AXP) into a 1.5 mL DNA LoBind® Tube (Eppendorf) and incubation for 5 minutes at room temperature (RT) on a rotator mixer. Afterwards, the mixture was placed on a magnetic rack and the pellet incl. AXP and DNA was washed twice with 80% ethanol. DNA was eluted in 11 µL of DNA-free water (Molzym).

b. Determination of DNA concentration

DNA concentration was determined using 1 µl of the AXP purified amplicon with the Qubit™ dsDNA HS Assay Kit (Invitrogen) in combination with the Qubit™ 4 Fluorometer (Invitrogen).

c. *Library preparation and ONT sequencing*

For the library preparation, 10 µl of amplicon DNA per sample were diluted in DNA-free water (Molzym) up to the recommended amount of 50 ng per required 9 µL input volume, if possible. For the remaining workflow from amplicon barcoding to loading of the flow cell the manufacturer's instructions "rapid-sequencing-v14-amplicon-sequencing-sqk-rbk114-24" were followed.
Sequencing was performed using the MinION Mk1B sequencing device, MinION Flow Cell (R.10.4) and MinKNOW 24.06.10 software from ONT. In order to reach the target of 10,000 reads per sample, an optimal sequencing duration of up to 5 hours was determined, covering also samples with low microbial DNA content (see results section).

d. *Bioinformatic analysis of Nanopore sequencing data*

Raw data processing was done using the ONT provided software Guppy (version 6.5.7.1) for basecalling, barcoding and trimming.
To exclude large differences within the NGS data analysis, two different analysis pipelines for species identification were used and compared. The first pipeline for species identification was based on the wf-16S (v0.0.4.) of the open-analysis platform EPI2ME (v5.2.3) including minimap2 (v2.1.3) and the NCBI-db (ncbi_16s_18s_28s_ITS) were used. The second pipeline was based on the open source software package EMU (Expectation-Maximization for microbial community profiling; v3.5.0) including minimap2 (v2.28), python (v3.7) and reference data bases (rrnDB v5.6 and NCBI 16S RefSeq from 17 Sept. 2020, available at https://github.com/treangenlab/EMU) [2].
The abundance of identified species with both data analysis pipelines was determined and for clinical samples compared to Sanger sequencing results.

Results

The time needed to achieve appr. 10.000 reads per sample for artificial and clinical samples was determined and is summarized in Table 1.

Table 1: Time needed for 10.000 total reads and 10.000 Top Hit reads with ONT sequencing in hours for artificial samples and clinical samples, respectively.

PCR Input	10.000 total Reads	10.000 Reads Top Hit
10 pg	< 2 h	> 2 h
2.5 pg	> 2 h	> 3 h
0.25 pg	~7250 reads after 24 h	~3000 reads after 24 h
Clinical Samples	> 2-6 h	> 2,5–7 h

Consequently, the run time for all samples, artificial and clinical, was set to 5 hours with generated reads up to 100.000. Relevant results were found from 1.000 reads.

For the seven artificial samples, an overview of the known species in comparison to the top-hit species with the highest abundance of the targeted ONT sequencing is shown in Table 2.

Using EMU, all known species were correctly identified. With wf-16S, a large proportion of the samples were accurately identified except tissue 1 and 2. Here, only the spiked *Streptococcus*

agalactiae was found whereas *Escherichia coli* was not correctly identified or could not be detected at all by wf-16S. Using EMU, both spiked in species were found.

Table 2: Top-hit species ID and abundance per pipeline for spiked samples.

Sample	Library Input DNA [ng]	Spike species and PCR Input DNA	wf-16S		EMU	
			Abundance	Species ID	Abundance	Species ID
A	50	*B. subtilis, 10 pg*	18,3%	*B. subtilis*	72,8%	*B. subtilis*
B	50,2	*B. subtilis, 10 pg*	20,1%	*B. subtilis*	79,0%	*B. subtilis*
C	49,2	*B. subtilis, 2.5 pg*	19,7%	*B. subtilis*	78,0%	*B. subtilis*
D	4,8	*B. subtilis, 0.25 pg*	18,9%	*B. subtilis*	76,0%	*B. subtilis*
E	49,7	*B. subtilis, 1.25 pg*	21,9%	*B. subtilis*	78,4%	*B. subtilis*
Tissue 1	51	*S. agalactiae 10^4 CFU/ml* *E. coli 10^6 CFU/ml*	83,3% 0,5%	*S. agalactiae* *E. fergusonii*	71,5% 2,3%	*S. agalactiae* *E. coli*
Tissue 2	3,8	*S. agalactiae, 10^1 CFU/ml* *E. coli 10^2 CFU/ml*	68,7%	*S. agalactiae*	57,1% 1,6%	*S. agalactiae* *E. coli*

Compared to the wf-16S, using EMU resulted in a higher species-related abundance. It is shown that the correct species of the artificial samples were detected in a range of 3.8 to 51 ng library input DNA (50 ng DNA input is suggested in the library preparation protocol). Even when using samples with low bacterial load, e.g. 0.25 pg DNA, and resulting in lower input DNA amounts for the library preparation, a species identification was possible with similar abundance as for samples with higher bacterial load.

For the 14 clinical samples, the identified species by Sanger sequencing in comparison to the top-hit species with the highest abundance of the targeted ONT sequencing per analysis pipeline are displayed in Table 3.

Pathogens initially identified using PCR and Sanger sequencing were largely confirmed in all samples using targeted NGS.

Clinically relevant pathogens (printed in bold) were identified in 13 of the 14 samples. Three samples show a partial match with Sanger sequencing results. In four samples, ONT identified more clinical relevant pathogens compared to Sanger sequencing.

For the abscess sample, *P. aeruginosa* and *S. aureus* were identified as relevant pathogens using both Sanger sequencing and EMU. However, with wf-16S, only *P. aeruginosa* was found as a relevant pathogen.

In one sample (heart valve D) only irrelevant bacteria were found and the species identified are likely contaminants from sample handling, as they were detected equally in both Sanger sequencing and NGS.

Table 3: Top-hit species ID and abundance per pipeline for clinical samples in correlation to CT value of 16S PCR, library input DNA and Sanger sequencing results; clinically relevant findings in bold.

Specimen	Ct MA Bac	Library Input DNA [ng]	Sanger sequencing	wf-16S		EMU	
			Species ID	Abundance	Species ID	Abundance	Species ID
Heart valve A	23,3	50	***S. mutans***	13,50%	***S. mutans***	98,20%	***S. mutans***
Heart valve B	25,5	50	***S. mutans***	13,10%	***S. mutans***	98,20%	***S. mutans***
Debridement soft tissue lumbal l. sin	14,0	50	***S. pyogenes*** ***C. diphteriae***	6,6% 2,6%	***S. pyogenes*** ***C. diphtheriae***	55,1% 44,9%	***S. pyogenes*** ***C. diphtheriae***
Debridement soft tissue lumbal l. sin	28,6	50	***S. pyogenes*** ***C. diphteriae***	10,70%	***S. pyogenes***	97,50%	***S. pyogenes***
Tissue - metatarsus II l.sin	n.d.	50	***S. aureus*** ***P. harei***	0,6% 0,3% 0,2%	***S. aureus*** ***P. harei*** ***F. magna***	63,6% 17,3% 7,9%	***S. aureus*** ***P. harei*** ***F. magna***
Tissue - metatarsus II l.sin	29,8	50	***S. aureus***	3,4% 1,6%	***S. aureus*** ***F. magna***	78,4% 10,6% 5,4% 3,3%	***S. aureus*** ***F. magna*** ***C. simulans*** ***P. harei***
Heart valve C	30,5	50	***S. oralis/mitis/infantis*** *M. osloensis* *E. aerosaccus*	44,10%	***S. oralis/mitis/ infantis***	92,50%	***S. mitis/oralis***
Heart valve D	33,2	17,2	*A. ebreus* *C. composti* *D. nitroreducens* */polyhydroxybutyrativorans* *T. arfidensis/ thermarum* */ignava* *C. leidyia*	30,5% 13,0% 4,0%	*C. denitrificans* *C. normanense* *S. epidermidis*	52,0% 21,4% 16,0%	*C. denitrificans* *C. normanense* *S. epidermidis*
Mitral valve	23,6	50	***S. agalactiae***	87,30%	***S. agalactiae***	81,00%	***S. agalactiae***
Medial knee prosthesis puncture	31,3	50	***B. thetaiotaomicron*** ***/congonensis/faecis*** ***/oleiciplenus/fragilis***	27,9	***B. thetaiotaomicron/*** ***faecis/luhongzhouii***	88,00%	***B. thetaiotaomicron***
Pancreas cyst fluid	32,3	15,1	***Anaerococcus sp.*** ***Dialister sp.***	22,9% 8,1% 4,5%	***C. gracilis*** ***A. prevotii*** *P. chinchillae*	30% 27% 8,3%	***C. gracilis*** ***A. prevotii*** ***P. goldsteinii***
Tissue (not specified)	19,9	50	***A. actino-*** ***mycetemcomitans***	80,50%	***A. actino-*** ***mycetemcomitans***	100%	***A. actino-*** ***mycetemcomitans***
Abscess	30,3	50	***P. aeruginosa*** ***S. aureus***	33,4% 1,2%	***P. aeruginosa*** *S. hominis*	94% 3,3% 1,7%	***P. aeruginosa*** *S. hominis* ***S. aureus***
Pre-sternal tissue	20,5	50	***G. morbillorum*** ***A. aphrophilus*** ***S. intermedius*** ***P. micra*** ***Peptostreptococcus***	25,4% 18,7% 17,0% 11,5% 3,6% 1,8%	***G. morbillorum*** ***A. aphrophilus*** ***S. intermedius*** ***Ca. peptostreptococcus*** ***massiliensis*** ***S. anginosus*** ***P. micra***	34% 26% 23% 15% 2,4%	***G morbillorum*** ***S. intermedius*** ***A. aphrophilus*** ***P. micra*** ***H. parainfluenzae***

As demonstrated with the spiked samples, EMU successfully identified the pathogens down to the species level. In eight clinical samples, the abundance exceeded 80%, in four samples it was above 50%, and in two samples it reached 30%. In contrast, with wf-16S, only two samples showed an abundance of 80%, while the remaining samples ranged between 0.6% and 44.1%. The higher abundance observed with EMU enables significantly better species differentiation, leading to a more accurate taxonomic classification.

In the pie charts in Figure 1, a high concordance between Sanger and both used targeted NGS data evaluation methods is shown. For the artificial samples, using EMU data analysis all species (n=7) were identified correctly. With wf-16S, the results for the artificial samples were concordant in 71,4% (n=5) and partially concordant in 28,6% (n=2).
Analyzing clinical samples with EMU 64,3% (n=9) of samples show full concordance and 35,7% (n=5) a partial concordance was reached in comparison to Sanger sequencing, whereby EMU identified in most cases additional clinically relevant pathogens. Using wf-16S a 57,1% (n=8) concordance and 42,9% (n=6) partial concordance was found. No discordance was detected.

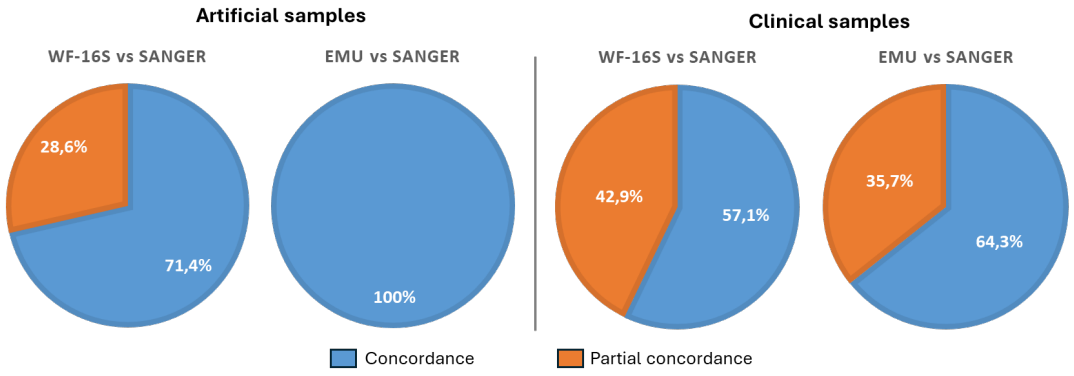

Figure 1: Concordance of clinical relevant species ID results derived by Sanger sequencing and 16s-wf, and Sanger sequencing and EMU are displayed in this pie chart for artificial and clinical samples, respectively.

Discussion

Our findings demonstrate the feasibility and efficiency of 16S targeted nanopore sequencing in combination with Micro-Dx™ for microbial identification in comparison to Sanger sequencing.

The sequencing time required to obtain a sufficient number of reads per sample varied depending on the sample type and DNA concentration. While samples with higher concentrations reach 10.000 reads in under two hours, lower input can result in fewer reads. However, an adequate read count could be achieved within a few hours for all samples demonstrating that even shorter sequencing runs can yield meaningful pathogen identifications. For reliable detection using our workflow, approximately 10,000 reads are to be recommended although, relevant results were already obtained from as few as 1,000 reads using an adequate analysis method.

Using the described targeted NGS approach in combination with the EMU pipeline, all bacteria in artificial and spiked samples were correctly identified in concordance to Sanger sequencing,

even at very low bacterial load. For the 14 clinical samples, pathogens were largely confirmed in most samples using targeted NGS with additional findings related to polymicrobial infections.

However, clinical relevance needs to be carefully evaluated. For example, heart valve D likely contained contaminants introduced during sample handling, as these species were found consistently across both sequencing methods. The bacteria were accurately identified but are not clinically relevant.

The workflow combined with the wf-16S pipeline exhibited limitations in two artificial samples (Tissue 1 and Tissue 2). In Tissue 1, *E. fergusonii* was incorrectly identified instead of *E.coli*. In Tissue 2, wf-16S failed to detect *E. coli*, whereas EMU correctly identified both spiked-in species. These discrepancies highlight the improved species differentiation and detection sensitivity of EMU compared to wf-16S, particularly for samples with multiple bacterial species.

For the abscess, *P. aeruginosa* and *S. aureus* were confirmed as relevant pathogens by both Sanger sequencing and EMU, while wf-16S detected only *P. aeruginosa*. This suggests a reduced sensitivity of wf-16S in identifying co-infections or species present at lower abundance levels.

A key advantage of EMU over wf-16S was its ability to provide higher species-level abundance, which significantly improved taxonomic classification. EMU achieved an abundance of over 80% in 57% of samples, whereas wf-16S reached this threshold in only two cases (14%). In samples with lower bacterial input concentrations, EMU still successfully identified species with comparable abundance levels to samples with higher concentrations. This suggests that EMU provides robust species identification even when sequencing input is limited, making it particularly useful for low-biomass samples.

In dependence from the data evaluation method only a low background was found, with no influence on the determination of pathogens.

As shown in Figure 1, a high degree of concordance was observed between Sanger sequencing and both data evaluation methods. Importantly, no cases of discordance were detected, confirming the reliability of both methods in microbial identification. However, the improved species differentiation and higher read abundance provided by EMU suggest that it is a more suitable approach for clinical diagnostics, particularly in polymicrobial cases where precise taxonomic resolution is required.

The reference databases used can be a limiting factor. The accuracy of results depends on how up-to-date and well-curated the database is. Therefore, regular benchmarking is essential.

All in all, the results show that targeted nanopore sequencing with EMU provides a reliable, rapid, and accurate alternative to Sanger sequencing for microbial identification. The ability to detect pathogens at low input concentrations and the differentiation between closely related species makes it a valuable tool for clinical microbiology.

Conclusion

The Micro-Dx™ amplicon workflow in combination with ONT sequencing and EMU pipeline represents a promising method for targeted bacterial diagnostics. The combination of rapid sequencing and accurate taxonomic classification enables efficient analysis of bacterial pathogens and communities in clinical applications and is therefore a great alternative to Sanger sequencing and BLAST analysis.

References

1. EPI2ME Labs. (n.d.). wf-16S: Taxonomic classification of single reads from amplicon-targeted sequencing. GitHub repository as of 09.12.2024-28.03.2025], https://github.com/epi2me-labs/wf-16s
2. Curry et al. (2022) Nat Methods. 2022 Jul; 19(7):845-853. https://doi.org/10.1038/s41592-022-01520-4