Performance and Practicality of 16S Nanopore Sequencing for Routine Bacterial Identification in Clinical Samples

Authors: A.U. Geers[1], C. Bütikofer[1], M. A. Terrazos Miani[1], S. Droz[1], A. Zihler Berner[1], I. Lendenmann[1], C. Hirzel[2,3], P.M. Keller[4,5], F. Suter-Riniker[1,2], S. Neuenschwander[1], C. Casanova[1,2], A. Ramette[1,2*]

[1]Institute for Infectious Diseases, University of Bern, Bern, Switzerland

[2]Multidisciplinary Center for Infectious Diseases, University of Bern, Bern, Switzerland

[3]Department of Infectious Diseases, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland.

[4]Division of Clinical Bacteriology and Mycology, University Hospital Basel, Basel, Switzerland

[5]Center for Infectious Disease Diagnostics, University Hospital Basel, Basel, Switzerland

*Correspondence: Alban Ramette. Institute for Infectious Diseases, University of Bern, Friedbühlstrasse 25, Bern 3001, Switzerland. Tel +41 31 684 19 08, Email alban.ramette@unibe.ch

Keywords: low-biomass samples – clinical Microbiology – NGS – Nanopore – 16S rRNA gene – bacterial infections

Abstract

The accurate identification of bacterial pathogens is a critical step in clinical diagnostics, directly impacting effective treatment and patient outcomes. This study addresses the need for standardized, clinically validated protocols for routine infectious disease diagnostics, particularly within the context of low-diversity clinical samples. We validated the performance of a 16S nanopore-based (Oxford Nanopore Technologies [ONT]) sequencing pipeline against an established Illumina Next-Generation Sequencing (NGS) workflow. We first investigated the detection limits and capabilities of ONT and Illumina using dilution series experiments of pure and mixed bacterial cultures. Subsequently, the validated pipeline was prospectively applied to 101 clinical samples, and the results were evaluated in parallel with Illumina data and by clinical review to determine infection plausibility. Species detection patterns and relative abundances demonstrated high consistency between ONT and Illumina across the clinical cohort, achieving an overall read overlap of 93.5% ± 7.6% and a high level of agreement upon clinical review. While Illumina sequencing exhibited slightly greater sensitivity at very low microbial loads in dilution experiments, ONT generated sequences with higher average accuracy (99.8% ± 0.39 vs. 99.8% ± 0.42%). Crucially, the Nanopore workflow demonstrated superior efficiency, requiring approximately 50 hours less total time and proving to be significantly more cost-effective than the Illumina workflow. These findings indicate that 16S nanopore sequencing provides comparable performance to Illumina sequencing for clinical bacterial identification in low-diversity samples. We conclude that 16S nanopore sequencing is a feasible, time-effective, and cost-effective method suitable for direct implementation in routine clinical diagnostic workflows.

1 Introduction

The identification of bacterial pathogens is a critical step in clinical diagnostics, guiding effective treatment and improving patient outcomes. Traditionally, this process has relied on culture-based methods, which can be time-consuming and may fail, due to factors such as prior antibiotic exposure (1) or fastidious growing organisms (2). In recent years, culture-independent approaches, particularly sequencing of the 16S rRNA gene, have become more widely adopted in clinical settings due to their ability to provide rapid taxonomic identification directly from patient samples (3–6). This approach is especially advantageous for complex and urgent cases where timely results are crucial or for samples that yield inconclusive or negative results with conventional culture. Particularly challenging to diagnose are samples with low-bacterial diversity, such as specimens from sterile body sites, since reagent contamination can lead to significant over estimation of potential pathogens (7). Thus, using DNA-free reagents is crucial to limit the detection of false positives. Several sequencing technologies are available for 16S rRNA gene amplicon sequencing. While traditional Sanger sequencing remains widely used due to its low cost and speed, its inability to resolve complex mixtures of bacteria limit its effectiveness for polymicrobial infections (3, 4). Next-generation sequencing (NGS) platforms, such as those from Illumina, offer high precision and throughput. However, the short read lengths produced by Illumina can sometimes limit the ability to achieve species-level taxonomic resolution, particularly for closely related bacteria (2). Additionally, their high hardware costs and the requirement for large sample batches to be cost-effective are significant barriers to implementation in a rapid, on-demand clinical diagnostic workflow.

Third-generation sequencing technologies, most notably Oxford Nanopore Technologies (ONT), present a promising alternative. Nanopore sequencing offers advantages that are particularly well-suited for clinical microbiology, including rapid sample-to-result turnaround times, low initial hardware expenditure, and the ability to process single or small batches of samples efficiently (8, 9). Its portability allows for decentralized testing, further reducing time to diagnosis (10). While ONT sequencing has been associated with a higher rate of sequencing errors compared to Illumina, its capacity to generate long reads spanning the full-length 16S rRNA gene can offer superior taxonomic resolution (8). Despite its potential, the clinical adoption of 16S nanopore sequencing is hindered by the lack of standardized and clinically validated protocols. A comprehensive and direct comparison to established NGS methods like Illumina, especially using a large cohort of clinical samples, has also been lacking. Our study addresses these critical gaps by presenting a clinically tested and used pipeline for 16S nanopore sequencing in routine infectious disease diagnostics. We implemented a cost- and time-effective pipeline utilizing an in vitro diagnostic (IVD) certified amplification protocol that targets the

76 V3-V4 region of the 16S rRNA gene using DNA-free reagents. The pipeline's accuracy was validated on
77 over 100 patient samples, with results confirmed by comparative Illumina sequencing and subjected
78 to rigorous clinical evaluation against the corresponding patient medical history.

79 2 Methods

80 **Clinical samples and pure bacterial cultures**

81 Clinical samples originated from the Insel Gruppe, the largest Swiss hospital group, and were analyzed
82 at the Institute for Infectious Diseases (IFIK) in Bern. The diagnostic division of IFIK (clinical
83 microbiology) is accredited under EN ISO/IEC 17025 to conduct routine bacterial diagnostics on clinical
84 samples. Clinical samples were routinely analyzed with eubacterial 16S PCR and ONT sequencing as
85 described below. The samples were retroactively also sequenced with Illumina. The Ethics Committee
86 of the Canton of Bern (BASEC-Nr 2024-00640) approved this study. For dilution series experiments, *S.*
87 *aureus* (ATCC 25923) and *E. coli* (ATCC 25922) were grown in 5 ml Luria Broth (LB) liquid culture for 2
88 h, and subsequently diluted to 10^{-5}, 10^{-6}, and 10^{-7} in LB. Additionally, undiluted *E. coli* and *S. aureus*
89 liquid cultures were mixed in a ratio of 1:2. The resulting mixture was diluted to 10^{-5}, 10^{-6}, and 10^{-7} in
90 LB (**Figure 1**). 1 ml of each of the diluted liquid cultures was used for further DNA extraction and
91 sequencing (see below), and the liquid cultures were plated on MAC (MacConkey II Agar, Becton
92 Dickinson) and CNA (Columbia CNA Agar, Becton Dickinson) plates to determine bacterial counts
93 (colony forming units; CFU).

94 **DNA extraction and PCR amplification**

95 The commercial Micro-Dx kit (Molzym, Bremen, Germany) was used for automated microbial DNA
96 enrichment, host depletion, microbial DNA extraction and real-time 16S PCR (**Figure 1**). The
97 manufacturer's protocol was followed. In the case of positive PCR result, the PCR product was purified
98 and used for sequencing. PCR mixtures (28 µl) were assembled with 10 µl master mix containing 7.5 µl
99 water, 2.5 µl DNA staining solution and 0.8 µl MolTaq 16S (DNA polymerase; Molzym). Real-time 16S
100 PCR was performed in a LightCycler z480 (Roche) with the following parameters: 95°C for 1 min; 40
101 cycles of 95°C for 5 s, 55°C for 10 s, 72°C for 25 s; and 40°C for 10 s. PCR products were purified with
102 AMPure XP beads (Beckman coulter, Brea, California) following the manufacturer's protocol with
103 minor modifications: An additional 3s centrifugation step was added after the washing step, and the
104 purified DNA was eluted in 80 µl of Tris-HCl (0.01 M, pH 8.0). Amplicon fragment sizes were assessed
105 using the TapeStation D1000 assay (Agilent, Santa Clara, CA, USA), and DNA concentrations were
106 determined with the Qubit double-stranded DNA (dsDNA) broad-range (BR) assay (Thermo Fisher
107 Scientific, Reinach, Switzerland).
108

109 To control for contaminations, we used a total of 14 negative controls (i.e. sterile double-distilled
110 water), including five controls from the DNA extraction step, five controls from the eubacterial PCR
111 step, four from library prep Illumina steps. We were only able to successfully identify species in three
112 negative controls (one extraction negative, two eubacterial PCR negative controls). The species
113 identified in the extraction or PCR negative controls could not be identified in samples from the same
114 extraction and PCR batch, with the exception of one species of *Actinomycetales* bacterium
115 (**Supplementary Figure 1**).

116 **Library preparation and sequencing**

117 Library preparation for ONT sequencing was performed with the kit SQK NBD114.24 (ONT, Oxford, UK)
118 using the supplementary reagents NEBNext Ultra II End repair/dA tailing Module (New England Biolabs
119 [NEB], ON, CA), NEB Blunt/TA Ligase Master Mix (M0367; NEB) and NEBNext Quick Ligation Module
120 (NEB). The manufacturer's protocol (REF NBA_9168_v114_revO_15Sep2022) was modified as follows:
121 Diluted DNA Control Sample (DCS) was not used for End Prep (EP) step, DCS volume was substituted
122 with nuclease free water (NFW). No clean-up and no normalization were performed after EP.
123 Therefore, 3 µl of End Prep Product with 4.5 µl NFW was used for barcoding step instead of 7.5
124 normalized EP product. Ethanol (80%) was used for the post-barcoding washes instead of Short
125 Fragment Buffer (SFB). SFB was used for clean-up step after adapter ligation. A typical library consisted
126 of the pooling of PCR amplicons from 1 to 20 clinical samples, two positive controls (*Streptococcus*
127 *agalactiae* amplicons), one extraction negative and one PCR negative samples. ONT libraries were
128 sequenced on GridION X5 sequencers with the standard flowcells (FLO-MIN114) with real-time
129 basecalling under high accuracy mode (HAC), for up to 1 h or until the positive controls reached at least
130 10,000 reads.

131 Library preparation for Illumina sequencing was performed with the KAPA Unique Dual-Indexed
132 Adapter Kit KR1736 – v2.19 and the KAPA HyperPrep Kit KR0961 – v7.19. The protocol was followed
133 according to the manufacturer's instructions, with the exception of using CleanNGS DNA Beads instead
134 of the KAPA cleanup beads, and two additional clean-up steps with beads after amplification. Each
135 library consisted of 68 to 76 pooled clinical samples and negative controls. The library was sequenced
136 on an Illumina MiSeq sequencer with v2 reagents in 2x250bp paired-end mode. Both ONT and Illumina
137 sequencing were performed at the Institute for Infectious Diseases, University Bern.

138

139 **Bioinformatic analyses**

140 ONT reads were demultiplexed and analyzed with the LORCAN pipeline (11). In brief, the adapters are
141 trimmed, filtered by length (± 100 bp around modal sequence length; i.e. 320-520 bp), and subsampled
142 to a maximum of 9,000 reads (**Figure 1**). Then the reads are grouped into species by mapping to a

143 custom nonredundant reference database. Detailed instructions for database creation can be found
144 online: https://github.com/aramette/ LORCAN/. In brief the reference database, LeBiBi SSU-rDNA-
145 mk45 (12) was trimmed to the region of interest, dereplicated (mothur; version v.1.48.0 (13)), and
146 used to make a custom database. Only species groups with at least 100 reads are further processed.
147 For each species group, the reads are remapped to the reference sequence with the highest number
148 of mapped reads, and the consensus sequence is generated using a 50% majority rule consensus. In a
149 last step, the consensus sequences are mapped again to the custom database using BLAST (v2.12.0) to
150 identify the closest taxonomic neighbors (**Figure 1**).
151
152 The basecalled and demultiplexed Illumina reads were trimmed, quality filtered (fastp v 0.23.2,
153 enabled base correction in overlapped regions, min Phred quality of 20), end trimmed (24 bp) and
154 filtered for minimum length of 240 bp (DADA2, v1.34 (14)). Subsequently the reads were denoised,
155 merged and amplicon sequencing variants (ASVs) generated using the established DADA2 (v1.3.4)
156 pipeline (14). The resulting ASVs were again filtered for a minimum length of 390 bp and maximum of
157 450 bp and their taxonomy was determined by mapping to the reference database using BLAST
158 (v2.12.0). In a last step read counts were combined at species level and all species with less than 100
159 reads were filtered out (**Figure 1**). Read counts were processed with R version (4.4.1), using the
160 packages tidyverse (v2.2.0), data.table (v.17.2), RColorBrewer (v1.1-3), ggvenn (v0.1.10), ggpubr
161 (v0.6.0), ggtext (v0.1.2), svglite (v2.21), to generate the figures and perform statistical analyses.
162
163 **Data availability**
164 Raw sequencing data is available under the BioProject accession number: PRJEB102560.
165 Supplementary tables can be accessed under: 10.5281/zenodo.17811713.
166

167 # 3 Results

168 **Dilution series.** We assessed the detection capabilities of ONT and Illumina sequencing on pure
169 cultures of *E. coli* and *S. aureus*, and on a mixture of the two at a ratio of 2:1. Samples were prepared
170 in a dilution series of 10^{-5}, 10^{-6}, and 10^{-7} in triplicates, corresponding to average CFU counts of 3,040,
171 310, 30 for *E. coli*, and 2,590, 80, 0 for *S. aureus*, respectively. In the mixture there were an average of
172 34,270, 8,330, 2,110 *E. coli* CFU and 14,600, 6,150, 910 *S. aureus* CFUs in the respective dilutions. All
173 samples were sequenced with both technologies. Additionally, we analysed a subsample of 9,000
174 Illumina reads per sample, which represents the maximum number of reads considered in the ONT
175 analysis, to assess the impact of sequencing depth on the results. On average we obtained 1040 raw
176 reads with ONT sequencing and 186,040 raw reads with Illumina sequencing for our dilution series

177 (**Supplementary table 1**). During the bioinformatic analysis, sequences identified as *Escherichia* sp.
178 and *Shigella* sp. were grouped together due to the high genetic similarity of their 16S rRNA gene and
179 the resulting inability of V3V4 16S sequencing to distinguish them (15). Similarly, we made the decision
180 to regroup sequences for *Corynebacterium* spp. together. At the highest concentrations, both
181 sequencing methods, with either the full or subsampled read sets, successfully detected the correct
182 species, confirming their ability to identify species present at high concentrations (**Figure 2**). However,
183 at lower concentrations (<10^{-6} CFU/ml), neither method could detect *E. coli*, suggesting that the true
184 concentration of this species fell below the limit of detection for both sequencing platforms under
185 these experimental conditions. For *S. aureus*, detection failed to be reliable at concentrations below
186 80 CFU/ml; only Illumina sequencing (including the subsampled data) was able to identify the species
187 at the 10^{-6} dilution in 2 out of 3 replicates, indicating that this method possessed a slightly greater
188 sensitivity at low microbial loads. In the polymicrobial sample, both methods detected both species
189 even at the lowest concentration. The observed ratio between the two species was in the same order
190 of magnitude as for the CFU-based counts. For the dilutions 10^{-5}, 10^{-6}, and 10^{-7} using ONT sequencing,
191 we observed ratios between *E. coli* and *S. aureus* of 8, 3 and 7, respectively. With Illumina sequencing
192 these ratios were 4, 2 and 5, respectively (**Figure 2**). Subsampled Illumina reads had no effect on these
193 ratios. Ratios determined by CFU counts were 2.3, 1.3 and 2.3, respectively. This indicates that
194 although both methods can accurately identify species within a mixed community, minor differences
195 in the observed abundances may occur. A notable finding was the more frequent detection of false
196 positives using Illumina sequencing with the DADA2 pipeline. These false positives included *Paracoccus*
197 *hibiscisoli*, *Rhizobium daejeonense*, *Corynebacterium* sp., *Streptococcus parasanguinis*, and
198 *Cutibacterium acnes*, among others. These false positives occurred in both the full or subsampled
199 datasets, indicating that the issue was not related to sequencing depth. We further investigated the
200 ONT raw reads for the presence of these contaminants, and found them to be present as well, albeit
201 below the threshold required for generating consensus sequences (**Supplementary Table 2**).
202
203 **Clinical samples.** We analyzed 101 clinical samples from routine diagnostics collected over a 1.5-year
204 period. These samples represented a diverse range of clinical materials, encompassing 43 distinct
205 sample types, and were predominantly derived from joint and spine biopsies (**Supplementary Table**
206 **3**). All samples underwent sequencing using both ONT and Illumina on the same DNA extracts. Across
207 both sequencing methods, a total of 167 species were detected with at least 100 assigned reads, and
208 115 of these species were shared between the two methods (**Figure 3A, Supplementary Table 4**), while
209 23 species were uniquely detected by ONT and 29 species were exclusively found by Illumina.
210 Subsampling the Illumina reads reduced the number of detected species, resulting in only 109 shared

211 species and 5 species unique to Illumina. The prevalence of species across samples was comparable
212 between methods, with ONT detecting species on average in 2.23 ± 2.97 samples, and Illumina on
213 average in 2.08 ± 3.11 samples (subsampled Illumina data in 2.18 ± 3.10 samples) (**Figure 3B**).
214 *Cutibacterium acnes* was the most common species (in 20 and 30% of the samples, based on ONT and
215 Illumina sequencing), with average relative abundances of 51% ± 40% with ONT, and 39% ± 41% with
216 Illumina sequencing (across samples it was present in). Other common species included those from
217 the *Staphylococcus* and *Streptococcus* genera. We found that the overall species detection pattern was
218 consistent between the two methods when considering the relative abundances of species within each
219 sample (**Figure 3B**). Species uniquely detected by either Illumina or ONT generally had lower average
220 relative abundances (3% on average). Aggregating the abundances of all species across all samples
221 further showed that the detection patterns for Illumina and ONT were highly similar (**Figure 3C**), and
222 subsampling the Illumina reads did not significantly alter these trends. Consistent with the prevalence
223 data, *C. acnes* and several *Staphylococcus* and *Streptococcus* species were highly represented in the
224 total dataset (**Figure 3C**). As an estimate of bacterial diversity per sample, the average number of
225 species per sample was 3.0 ± 2.7 for ONT, and 3.0 ± 3.4 for Illumina (2.5 ± 2.3 for the subsampled
226 Illumina data). Subsampling the Illumina reads significantly reduced the number of species per sample
227 (paired *t* test, p-value = 0.0007), possibly reducing the number of detected contaminants. To precisely
228 evaluate the differences between the two methods, we calculated the overlap of species for each
229 sample. On average, 97.1% ± 5.9%, of species detected by ONT (weighted by their relative abundance)
230 were also found by Illumina, while 98.7% ± 2.8% of Illumina species (weighted by their relative
231 abundance) were also detected by ONT. The overall read overlap with the same species between
232 methods was 93.5% ± 7.6%. Lastly, a comparison of consensus sequence accuracy using BLAST identity
233 showed that sequences generated by ONT had an average sequence identity to the reference of 99.88
234 % ± 0.39 (weighted by reads), while Illumina sequences had an average identity of 99.83 % ± 0.42 % to
235 the reference. This difference was found to be statistically significant (paired *t* test p-value = 0.0001).
236
237 **Clinician review.** An experienced infectiologist reviewed all sequencing results obtained for the clinical
238 samples in order to evaluate the clinical plausibility of the NGS findings. The review found a high level
239 of agreement between the two sequencing methods: For 77 (76%) samples, a likely pathogen was
240 observed, whereas for the remaining 24 samples, the sequencing data suggested a likely
241 contamination. Importantly, there were no cases where one sequencing method identified a pathogen
242 and the other identified a contamination in the same sample. We also found that the number of
243 species detected per sample was significantly higher in samples identified as contaminations compared
244 to those with a likely pathogen, for both Illumina (p=0.02) and ONT (p=0.03) sequencing (**Figure 3D)**.

245 Furthermore, we assessed the potential impact of patient antibiotic treatment prior to sampling and
246 sequencing, which was documented in 94% of the samples (69 samples with treatment, 26 samples
247 without). There was no significant correlation between antibiotic pretreatment and the number of
248 species detected for either sequencing method. Interestingly, in samples with antibiotic pretreatment
249 significantly more pathogens than contaminations were detected (Fisher exact test, p-value = 0.02)
250 (**Supplementary Tables 3 and 4**).
251
252 **Cost and turnaround time estimation.** We estimated that the consumables for a batch size of 24
253 samples processed on a single flow cell resulted in approximate per-sample costs of 35 CHF for ONT
254 and 155 CHF for Illumina. The effective price difference is likely to be even greater in routine diagnostic
255 settings, as batch sizes are often smaller than 24 samples. This is primarily because ONT flow cells can
256 be reused if only a small number of samples is sequenced, while Illumina sequencing cartridges are
257 single-use, regardless of the number of sequenced samples. To estimate the turnaround time of the
258 two sequencing approaches, we considered the time required for library preparation, sequencing, and
259 bioinformatic analysis (including data transfer, quality control, and report generation) for a 24-sample
260 batch. The ONT workflow required approximately 5, 2, and 0.5 h for each of the steps, respectively. In
261 contrast, the Illumina workflow required 7, 43, and 0.5 h, respectively. These time estimates for the
262 bioinformatic analysis assume that pre-established and automated pipelines are in place. These
263 findings underscore the ONT-based workflow's advantage in speed, demonstrating a much faster
264 turnaround than the Illumina-based process.
265

266 # 4 Discussion

267 The study successfully demonstrates that ONT sequencing provides a time- and cost-efficient
268 alternative to established methods of 16S rRNA sequencing in routine clinical diagnostics, especially
269 for low-diversity samples. Using a controlled dilution series and a large panel of 101 routine clinical
270 samples, we showed that ONT 16S rRNA gene sequencing achieves sensitivity and specificity
271 comparable to the gold standard Illumina sequencing. In 76% of clinical samples, it successfully
272 identified a likely pathogen, as confirmed by clinical review of the cases. Furthermore, for batch sizes
273 of around 24 samples, ONT is much more time and cost efficient than Illumina-based workflow.

274 The limit of detection achieved using the ONT platform in our controlled experiments was in the range
275 of 100 to 1,000 CFUs/ml (**Figure 2**). In a recent study standardizing 16S rRNA gene sequencing for
276 culture-negative infections (16), the authors evaluated ONT sensitivity using DNA genome copies per
277 µl, noting that species detection varied, and observed better sensitivity with the MinION flow cell than

278 with Flongle, potentially due to differences in sequencing depth. Another investigation that focused
279 on direct bacterial identification in sterile body fluids estimated a slightly lower detection limit of 90
280 CFUs/ml (17). This difference is likely attributable to their substantially higher sequencing depth,
281 reporting approximately ten times more reads per sample than generated in our study. In comparison,
282 Illumina sequencing demonstrated a slightly lower detection limit in our controlled dilution series
283 (**Figure 2**), irrespective of whether the full or a subsampled dataset was analyzed. This aligns with
284 findings from other 16S studies, where Illumina sequencing was observed to capture greater species
285 richness in respiratory microbial communities (18).

286 A limitation of our current method is the use of the V3-V4 hypervariable region of the 16S rRNA gene,
287 instead of the full-length 16S gene. The V3-V4 region is the established universal target in the IVD-R kit
288 we used, which offers a broad-range detection of >345 bacterial species of clinical relevance (19).
289 Shorter amplicons are generally assumed to lend themselves to better amplification efficiency in PCR.
290 For instance, one study found that shorter 16S amplicon had higher sensitivity in their metagenomic
291 analysis of culture-negative samples (20). However, utilizing the V3-V4 region inherently limits the
292 taxonomic resolution achievable, as full-length 16S sequencing often provides greater species-level
293 differentiation, enabling for example the differentiation of *E. coli* from *Shigella* (17), or between
294 *Staphylococcus* species (21).

295 An interesting finding from the controlled dilution experiments was the detection of more
296 contaminating reads with Illumina than with ONT sequencing, when a minimum threshold of 100 reads
297 per species for identification was set **(Figure 2)**. However, when examining routine clinical samples,
298 both sequencing methods appeared equally likely to detect non-pathogenic contaminants **(Figure 3D)**.
299 Interpreting such low-abundance reads is critical for clinical reporting. One study on mock
300 communities found that ONT displayed remarkably high accuracy compared to Illumina sequencing
301 (22), though the use of full-length 16S for ONT and shorter amplicons for Illumina in that study
302 represents a confounding factor. To address the "tail" of low-abundant species (often contaminants)
303 observed in sequencing data, prior studies have implemented relative abundance filters, such as
304 filtering species below 0.1% (21) or 0.06% (17). While filtering by relative abundance worked well for
305 monomicrobial samples, it was shown to fail for polymicrobial samples (17), underscoring the
306 complexity of robustly differentiating true pathogens from contaminants in complex clinical scenarios.
307 We found *Cutibacterium acnes* to be one of the most common species in the investigated clinical
308 samples **(Figure 3B)**. Because *C. acnes* is a skin commensal and can also be an opportunistic pathogen,
309 differentiating a true infection from a contamination can be challenging (23). In eleven (=11%) of our
310 clinical samples *C. acnes* was the likely pathogen as evaluated by the clinical review, while in the

311 remaining 10-20 samples, the species was detected in, a potential implication as a pathogen could not
312 be unambiguously determined (**Supplementary Tables 3-4**). In those latter cases, *C. acnes* presence
313 might thus represent a contamination occurring during the sampling or laboratory processes. The
314 elevated prevalence of *C. acnes* may also be attributable to the significant proportion of joint biopsies
315 within our clinical samples. *C. acnes* has gained notoriety for its role in implant-associated infections
316 and has also been recognized as a significant pathogen in cerebrovascular infections, breast implant
317 fibrosis, and cardiovascular device infections (23).

318 Overall, we observed an average species overlap of 93.5% (weighted by abundance) between the two
319 platforms. This direct comparison, conducted on a large set of routine clinical specimens prospectively
320 collected over 18 months, confirms that ONT sequencing is equally effective and accurate as the
321 current gold NGS standard, Illumina sequencing, for clinical pathogen diagnostics. Furthermore, the
322 ONT platform in combination with the LORCAN bioinformatic pipeline demonstrated a slightly higher
323 consensus sequences accuracy (99.88% ± 0.39% sequence identity to reference) compared to
324 sequences generated by Illumina and DADA2 (14) (99.83% ± 0.42%). While there are multiple studies
325 exploring the feasibility of ONT 16S rRNA gene sequencing in diagnostics, primarily comparing results
326 to culture (17, 24–27), MALDI-TOF (28, 29), or Sanger sequencing (10, 16, 20, 29), there is a notable
327 absence of studies directly comparing ONT versus Illumina sequencing using a large cohort of clinical
328 samples. A small scale study found that the microbial composition determined with Illumina V3-V4
329 sequencing closely matched to the one determined with ONT full-length sequencing in tracheal
330 aspirates samples (21), as well as in a mock community. Similarly, an older study in 2019 showed both
331 technologies could accurately identify bacterial genera in a well-characterized reference sample (22).

332 Overall, ONT sequencing holds a very positive outlook in clinical diagnostics of low-diversity samples,
333 primarily due to its proven advantages in shorter turnaround time and cost efficiency, especially with
334 small batch sizes. The successful demonstration that ONT sequencing achieves sensitivity and
335 specificity comparable to the gold standard Illumina sequencing, confirms its viability as a mainstream
336 clinical tool. Future efforts should prioritize maximizing the platform's utility by addressing current
337 limitations. Specifically, using longer sequences or complementary genetic targets could be crucial to
338 enhance taxonomic resolution (30), especially in combination with full-length curated databases of 16S
339 rRNA genes (31–33). Finally, continued development is necessary for robust bioinformatic strategies
340 to accurately interpret low-abundance reads, which represent contaminants, especially in complex
341 clinical scenarios where filtering methods successful for monomicrobial samples may fail for
342 polymicrobial samples.

343 # 5 References

344 1. Sigakis MJG, Jewell E, Maile MD, Cinti SK, Bateman BT, Engoren M. 2019. Culture-Negative and
345 Culture-Positive Sepsis: A Comparison of Characteristics and Outcomes. Anesthesia & Analgesia
346 129:1300.

347 2. Church DL, Cerutti L, Gürtler A, Griener T, Zelazny A, Emler S. 2020. Performance and Application
348 of 16S rRNA Gene Cycle Sequencing for Routine Identification of Bacteria in the Clinical
349 Microbiology Laboratory. Clinical Microbiology Reviews 33:10.1128/cmr.00053-19.

350 3. Youssef N, Boutros CF, Dakroub F, Akl F, Reslan L, Finianos M, Moumneh MBM, Dargham TB, Zein
351 ZE, Haddara A, Korman R, Khafaja S, Matar G, Araj GF, Dbaibo GS. 2025. The clinical impact of
352 16S ribosomal RNA PCR and sequencing in the identification of bacterial infections: a 7-year
353 report from a Lebanese tertiary care center. Front Cell Infect Microbiol 15.

354 4. Flurin L, Wolf MJ, Mutchler MM, Daniels ML, Wengenack NL, Patel R. 2022. Targeted
355 Metagenomic Sequencing-based Approach Applied to 2146 Tissue and Body Fluid Samples in
356 Routine Clinical Practice. Clinical Infectious Diseases 75:1800–1808.

357 5. Botan A, Campisciano G, Zerbato V, Di Bella S, Simonetti O, Busetti M, Toc DA, Luzzati R, Comar
358 M. 2024. Performance of 16S rRNA Gene Next-Generation Sequencing and the Culture Method
359 in the Detection of Bacteria in Clinical Specimens. 13. Diagnostics 14:1318.

360 6. Poulsen SH, Søgaard KK, Fuursted K, Nielsen HL. 2023. Evaluating the diagnostic accuracy and
361 clinical utility of 16S and 18S rRNA gene targeted next-generation sequencing based on five years
362 of clinical experience. Infectious Diseases.

363 7. Egli K, Risch M, Risch L, Bodmer T. 2022. Comparison of an automated DNA extraction and 16S
364 rDNA real time PCR/sequencing diagnostic method using optimized reagents with culture during
365 a 15-month study using specimens from sterile body sites. BMC Microbiology 22:119.

366 8. Ivens MJR, Kamminga S, Benchamach K, Akile C, Wessels E, Claas ECJ, Boers SA. 2025. Rapid and
367 reliable species-level identification from clinical samples using 16 S rRNA gene nanopore
368 sequencing analysis. Sci Rep 15:28606.

369 9. Brunet S, Grankvist A, Jaen-Luchoro D, Bergdahl M, Tison J-L, Wester A, Elfving K, Brandenburg J,
370 Gullsby K, Lindsten C, Arvidsson L-O, Larsson H, Eilers H, Strand AS, Lannefors M, Keskitalo J,
371 Rylander F, Welander J, Jungestrom MB, Geörg M, Kaden R, Karlsson I, Linde A-M, Mernelius S,
372 Berglind L, Feuk L, Kerje S, Karlsson L, Sjödin A, Guerra-Blomqvist L, Wallin F, Fagerström A,
373 Vondracek M, Mölling P, Hallbäck ET. 2025. Nationwide multicentre study of Nanopore long-read
374 sequencing for 16S rRNA-species identification. Eur J Clin Microbiol Infect Dis 44:1907–1916.

375 10. Carlisle D, Cunningham-Oakes E, Booth J, Frankland A, McDowell M, Pilgrim J, Rzeszutek A, Evans
376 C, Larkin S, Li A, Loftus C, Samuel M, Scott N, Swithenbank L, Owen V, Darby AC, Smielewska A.

377 2024. Implementing portable, real-time 16S rRNA sequencing in the healthcare sector enhances
378 antimicrobial stewardship. medRxiv https://doi.org/10.1101/2024.09.23.24314079.

379 11. Neuenschwander SM, Terrazos Miani MA, Amlang H, Perroulaz C, Bittel P, Casanova C, Droz S,
380 Flandrois J-P, Leib SL, Suter-Riniker F, Ramette A. 2020. A Sample-to-Report Solution for
381 Taxonomic Identification of Cultured Bacteria in the Clinical Setting Based on Nanopore
382 Sequencing. Journal of Clinical Microbiology 58:10.1128/jcm.00060-20.

383 12. Flandrois J-P, Perrière G, Gouy M. 2015. leBIBIQBPP: a set of databases and a webtool for
384 automatic phylogenetic analysis of prokaryotic sequences. BMC Bioinformatics 16:251.

385 13. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB,
386 Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing
387 mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing
388 and Comparing Microbial Communities. Applied and Environmental Microbiology 75:7537–7541.

389 14. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-
390 resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.

391 15. Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. 2018.
392 Accurate differentiation of *Escherichia coli* and *Shigella* serogroups: challenges and strategies.
393 New Microbes and New Infections 21:58–62.

394 16. Butler I, Turner O, Mohammed K, Akhtar M, Evans D, Lambourne J, Harris K, O'Sullivan DM,
395 Sergaki C. 2025. Standardization of 16S rRNA gene sequencing using nanopore long read
396 sequencing technology for clinical diagnosis of culture negative infections. Front Cell Infect
397 Microbiol 15:1517208.

398 17. Lao H-Y, Wong LL-Y, Hui Y, Ng TT-L, Chan CT-M, Lo HW-H, Yau MC-Y, Leung EC-M, Wong RC-W,
399 Ho AY-M, Yip K-T, Lam JY-W, Chow VC-Y, Luk KS, Que T-L, Chow FWN, Siu GK-H. 2024. The clinical
400 utility of Nanopore 16S rRNA gene sequencing for direct bacterial identification in normally
401 sterile body fluids. Front Microbiol 14.

402 18. Macip G, Soler-Comas A, Palomeque A, Motos A, Llonch B, Canseco-Ribas J, Bueno-Freire L,
403 Calabretta D, Kiarostami K, Cabrera R, Ferrer M, Torres A, Fernandez-Barat L. 2025. Comparative
404 analysis of illumina and oxford nanopore sequencing platforms for 16S rRNA profiling of
405 respiratory microbial communities. Sci Rep 15:33688.

406 19. molzym. Automated Sample Pre-treatment, Bacterial / Fungal DNA Isolation and PCR Analysis.

407 20. Bouchiat C, Ginevra C, Benito Y, Gaillard T, Salord H, Dauwalder O, Laurent F, Vandenesch F. 2022.
408 Improving the Diagnosis of Bacterial Infections: Evaluation of 16S rRNA Nanopore Metagenomics
409 in Culture-Negative Samples. Front Microbiol 13.

410 21. Gupta A, Cooper VS, Zemke AC. 2025. Evaluation of V3–V4 and FL-16S rRNA amplicon sequencing
411 approach for microbiota community analysis of tracheostomy aspirates. mSphere 0:e00388-25.

412 22. Winand R, Bogaerts B, Hoffman S, Lefevre L, Delvoye M, Van Braekel J, Fu Q, Roosens NH, De
413 Keersmaecker SC, Vanneste K. 2020. Targeting the 16S rRNA Gene for Bacterial Identification in
414 Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford
415 Nanopore Technologies) Generation Sequencing Technologies. 1. International Journal of
416 Molecular Sciences 21:298.

417 23. Achermann Y, Goldstein EJC, Coenye T, Shirtliff ME. 2014. Propionibacterium acnes: from
418 Commensal to Opportunistic Biofilm-Associated Implant Pathogen. Clinical Microbiology
419 Reviews 27:419–440.

420 24. Han H-S, Ro DH, Chung J, Kim N, Moon J. 2022. Nanopore 16S amplicon sequencing enables rapid
421 detection of pathogen in knee periprosthetic joint infection. International Journal of Medical
422 Microbiology 312:151570.

423 25. Moon J, Kim N, Kim T-J, Jun J-S, Lee HS, Shin H-R, Lee S-T, Jung K-H, Park K-I, Jung K-Y, Kim M, Lee
424 SK, Chu K. 2019. Rapid diagnosis of bacterial meningitis by nanopore 16S amplicon sequencing:
425 A pilot study. International Journal of Medical Microbiology 309:151338.

426 26. Jang Y, Kim S, Kim N, Son H, Ha EJ, Koh EJ, Phi JH, Park C-K, Kim JE, Kim S-K, Lee SK, Cho W-S,
427 Moon J, Chu K. 2022. Nanopore 16S sequencing enhances the detection of bacterial meningitis
428 after neurosurgery. Annals of Clinical and Translational Neurology 9:312–325.

429 27. Wang M, Fu A, Hu B, Shen G, Liu R, Zhao W, Jiang S, Cai X, Li C, Li J, Wu Q, Feng K, Gu J, Chen J,
430 Shu M, Zhang B, Deng Z, Yu L, Li Y, Liu T. 2020. Same-Day Simultaneous Diagnosis of Bacterial and
431 Fungal Infections in Clinical Practice by Nanopore Targeted Sequencing. medRxiv
432 https://doi.org/10.1101/2020.04.08.20057604.

433 28. Vanhee M, Floré K, Vanthourenhout S, Hellemans J, Muyldermans A, Reynders M. 2024.
434 Implementation of full-length 16S nanopore sequencing for bacterial identification in a clinical
435 diagnostic setting. Diagnostic Microbiology and Infectious Disease 108:116156.

436 29. Baldan R, Cliff PR, Burns S, Medina A, Smith GC, Batra R, Cerda A, Wilson R, Merrill T, Lewis SJ,
437 Patel A, Jeyaratnam D, Wyncoll DL, Barrett N, Chand MA, Edgeworth JD. 2021. Development and
438 evaluation of a nanopore 16S rRNA gene sequencing service for same day targeted treatment of
439 bacterial respiratory infection in the intensive care unit. Journal of Infection 83:167–174.

440 30. Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, Wu Q, Graeber E, Finzer P,
441 Mendling W, Savidge T, Villapol S, Dilthey A, Treangen TJ. 2022. Emu: species-level microbial
442 community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. Nat Methods
443 19:845–853.

444 31. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA
445 ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic
446 Acids Res 41:D590–D596.

32. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res 42:D643–D648.

33. McDonald D, Jiang Y, Balaban M, Cantrell K, Zhu Q, Gonzalez A, Morton JT, Nicolaou G, Parks DH, Karst SM, Albertsen M, Hugenholtz P, DeSantis T, Song SJ, Bartko A, Havulinna AS, Jousilahti P, Cheng S, Inouye M, Niiranen T, Jain M, Salomaa V, Lahti L, Mirarab S, Knight R. 2024. Greengenes2 unifies microbial data in a single reference tree. Nat Biotechnol 42:715–718.

454 # 6 Figure legends

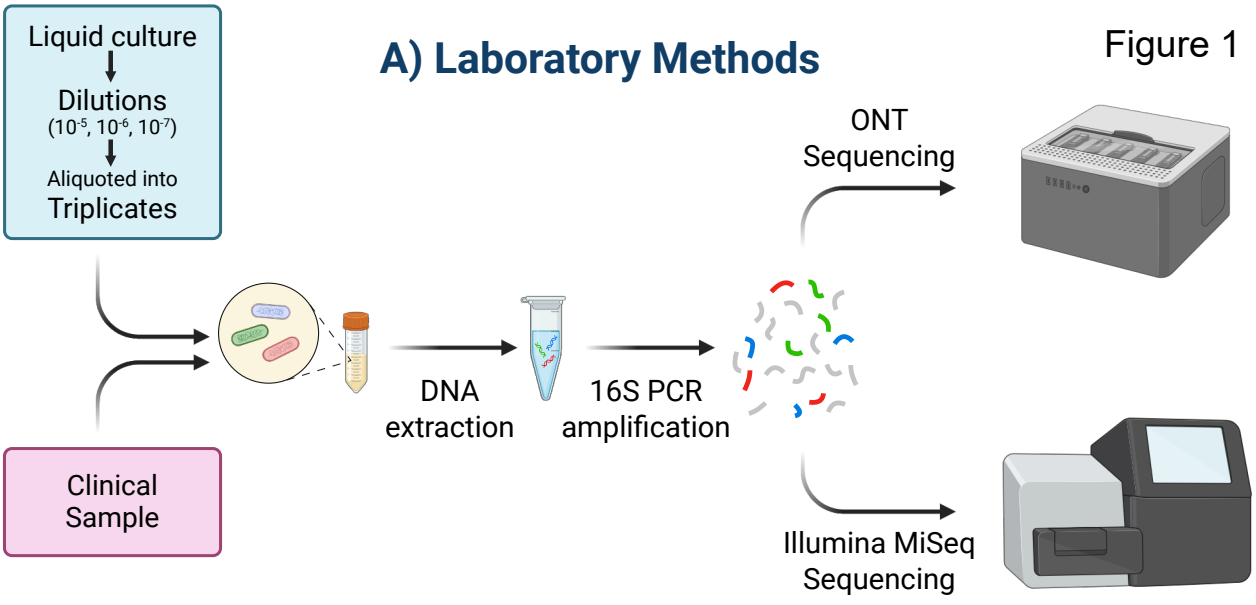
455 **Figure 1. Overview of the laboratory and bioinformatic workflows. A**) Investigated were dilutions
456 series of pure liquid cultures of *E. coli* and *S. aureus*, as well as over 100 clinical samples from routine
457 diagnostics. The DNA was extracted and the 16S rRNA amplified with the Molzym Micro-Dx kit. The
458 resulting amplicons were sequenced with both ONT and Illumina MiSeq. **B**) All raw sequencing data
459 was quality filtered and length filtered. Illumina reads were additionally subsampled to 9,000 reads
460 before further processing. The DADA2 tool was used to generate amplicon sequence variants (ASVs)
461 from the Illumina reads and the LORCAN pipeline was used for the ONT reads. The taxonomy of the
462 consensus sequences was determined using BLAST. For further analysis only species associated with
463 more than 100 reads were kept. The Figure was created with BioRender https://BioRender.com.

464 **Figure 2. Sensitivity and specificity based on dilutions series of pure bacterial liquid cultures.** Serial
465 dilutions for *E. coli*, *S. aureus* and a mixture of both were prepared and subjected to our laboratory
466 and bioinformatic workflows (**Fig. 1**). Reads assigned to *Escherichia* sp. and *Shigella* sp. were grouped
467 into *Escherichia* sp. due to their high similarity of the 16S rRNA gene across said species. Similarly,
468 *Corynebacterium* species were grouped together. Species with less than 100 reads in one sample are
469 not displayed. At the bottom of the figure the observed number of CFUs (colony forming units) per ml
470 of the diluted liquid cultures are shown.

471 **Figure 3. Detected bacterial species in routine clinical samples. A)** Venn-diagram of the number of
472 species shared between sequencing methods in clinical samples (n=101). In parentheses the
473 percentages of all the species across samples are given for each case. The total number of detected
474 species is 167. **B)** Species prevalence (i.e. percent of samples in which a species is present) across
475 sequencing technologies. The average relative abundance across the samples in which a species was
476 found present is indicated by a color range (red = 100%, blue = 0%). Only species that are present in at
477 least 3 samples in at least one method are depicted. **C)** Barplots of the summed relative abundances
478 of detected species across all samples for each of the sequencing methods. Shown are the 12 most
479 abundant species, the remaining are grouped into the category "Other". **D)** Boxplot of the number of
480 species identified by sequencing method, grouped by antibiotic pre-treatment of the samples and by
481 the results of the clinical assessment. Each dot represents an individual sample, with the shape
482 denoting the availability of culture-based identification results. No significant difference between
483 antibiotic treatments in either of the methods could be detected. The number of species was
484 significantly higher in samples where a likely contamination was detected, in contrast to a reasonable
485 pathogen for both Illumina (p=0.02) and ONT (p=0.03).

Figure 1

A) Laboratory Methods



B) Bioinformatic Methods

Illumina MiSeq reads

Trim, quality filter and min length filter

Subsample to 9'000 reads

Denoise and merge (F&R) with DADA2 to find Amplicon sequence variants (ASVs)

Min length filter

BLAST to determine taxonomy

Filter out species with < 100 reads

ONT reads

LORCAN

Quality & length filter

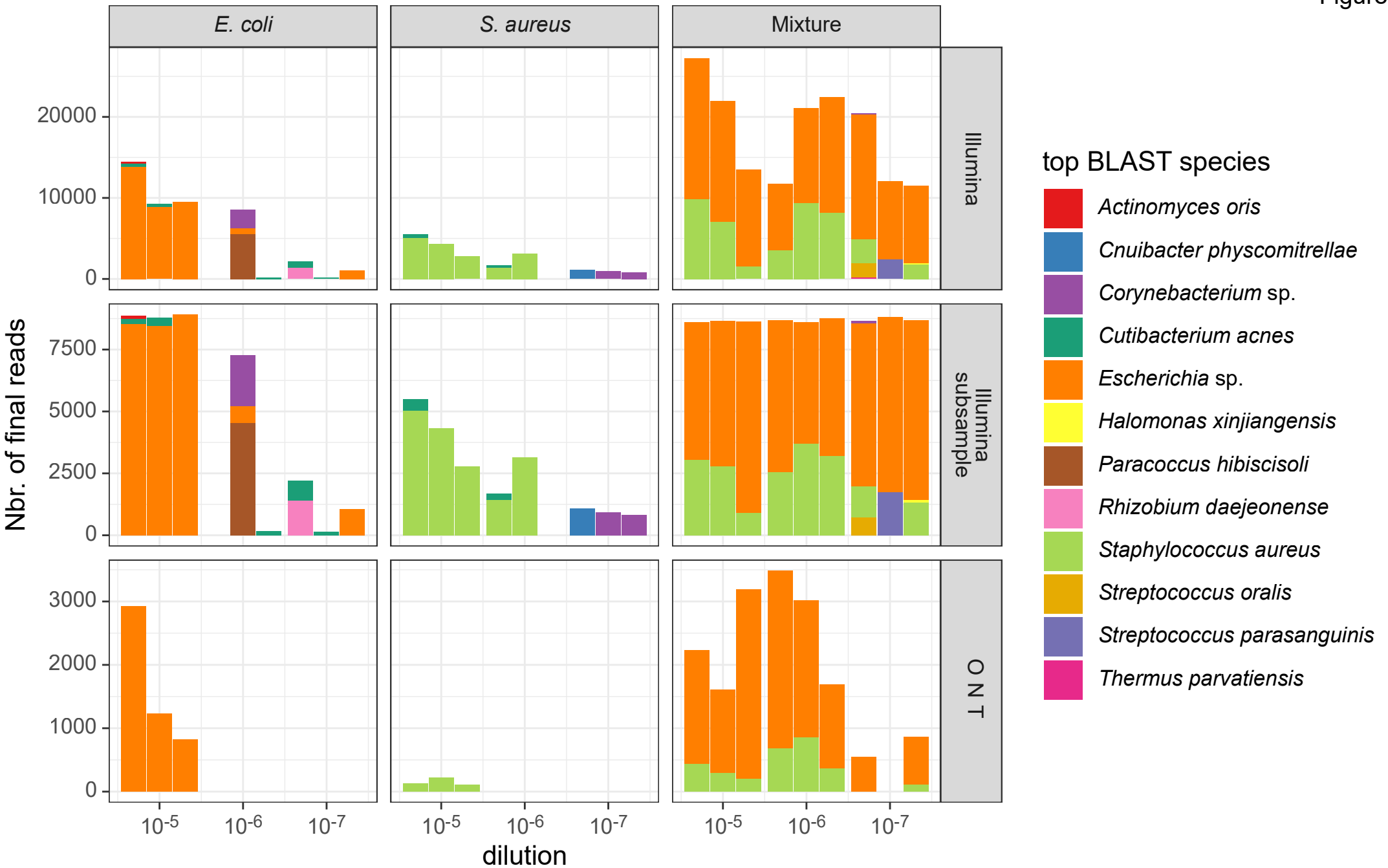
Subsample to 9'000 reads

Group reads into species by mapping to reference sequences

Consensus sequence generation for each group (>100 reads)

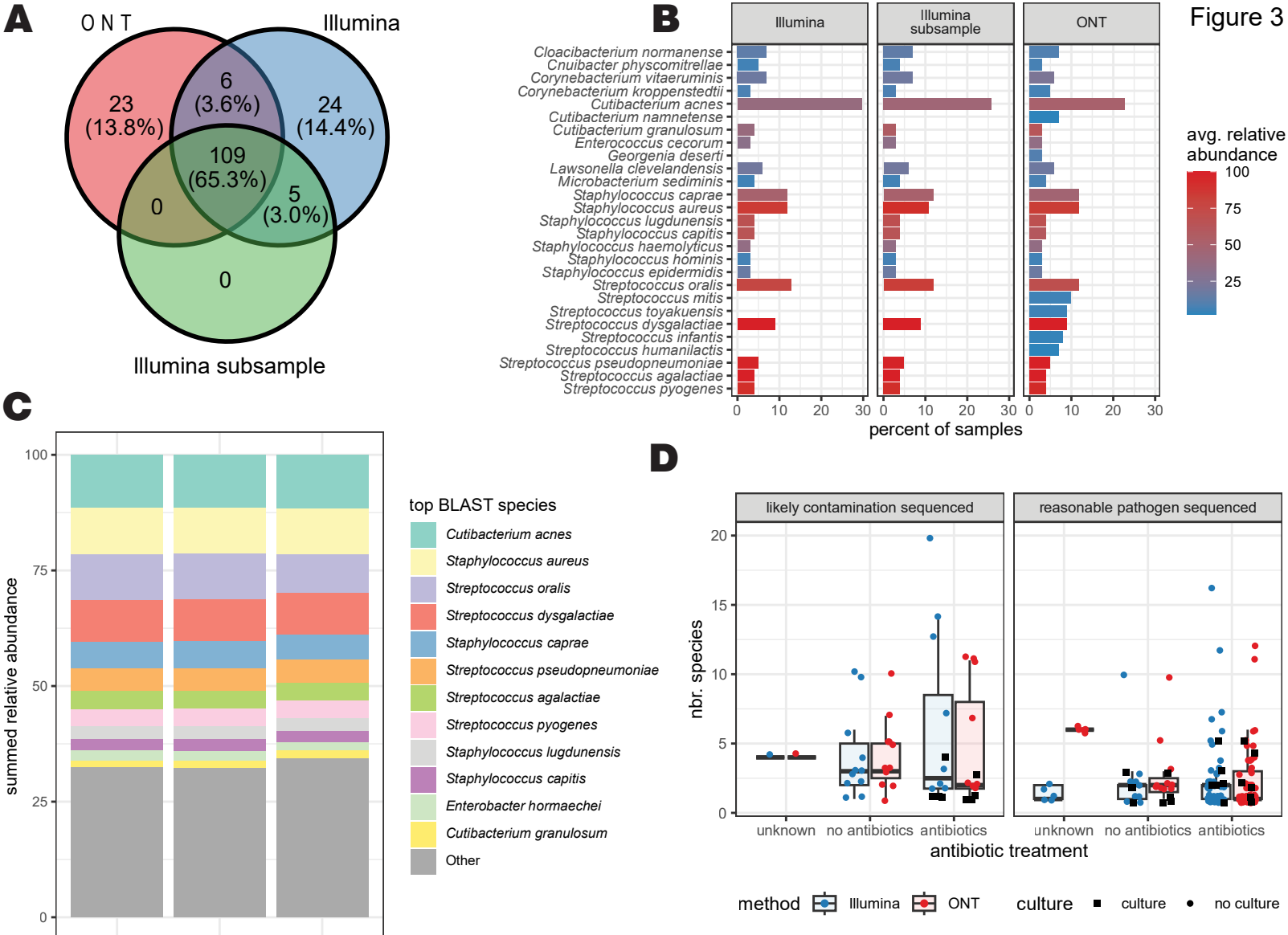
BLAST to determine taxonomy

Figure 2



3040	**310**	**30**				**34270**	**8330**	**2110**	***E. coli* [CFU/ml]**
			2590	**80**	**0**	**14600**	**6150**	**910**	***S. aureus* [CFU/ml]**

Figure 3


A
ONT
Illumina
Illumina subsample
23 (13.8%)
6 (3.6%)
24 (14.4%)
109 (65.3%)
0
5 (3.0%)
0
B
Illumina
Illumina subsample
ONT
Cloacibacterium normanense
Cnuibacter physcomitrellae
Corynebacterium vitaeruminis
Corynebacterium kroppenstedtii
Cutibacterium acnes
Cutibacterium namnetense
Cutibacterium granulosum
Enterococcus cecorum
Georgenia deserti
Lawsonella clevelandensis
Microbacterium sediminis
Staphylococcus caprae
Staphylococcus aureus
Staphylococcus lugdunensis
Staphylococcus capitis
Staphylococcus haemolyticus
Staphylococcus hominis
Staphylococcus epidermidis
Streptococcus oralis
Streptococcus mitis
Streptococcus toyakuensis
Streptococcus dysgalactiae
Streptococcus infantis
Streptococcus humanilactis
Streptococcus pseudopneumoniae
Streptococcus agalactiae
Streptococcus pyogenes
0 10 20 30
percent of samples
avg. relative abundance
100
75
50
25
C
summed relative abundance
100
75
50
25
0
Illumina
Illumina subsample
ONT
top BLAST species
Cutibacterium acnes
Staphylococcus aureus
Streptococcus oralis
Streptococcus dysgalactiae
Staphylococcus caprae
Streptococcus pseudopneumoniae
Streptococcus agalactiae
Streptococcus pyogenes
Staphylococcus lugdunensis
Staphylococcus capitis
Enterobacter hormaechei
Cutibacterium granulosum
Other
D
likely contamination sequenced
reasonable pathogen sequenced
nbr. species
20
15
10
5
0
unknown
no antibiotics
antibiotics
antibiotic treatment
method
Illumina
ONT
culture
culture
no culture