Rapid resistome mapping using nanopore sequencing

Eric van der Helm[1], Lejla Imamovic[1], Mostafa M. Hashim Ellabaan[1], Willem van Schaik[2], Anna Koza[1] and Morten O.A. Sommer[1,*]

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark and [2]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands

ABSTRACT

The emergence of antibiotic resistance in human pathogens has become a major threat to modern medicine. The outcome of antibiotic treatment can be affected by the composition of the gut. Accordingly, knowledge of the gut resistome composition could enable more effective and individualized treatment of bacterial infections. Yet, rapid workflows for resistome characterization are lacking. To address this challenge we developed the poreFUME workflow that deploys functional metagenomic selections and nanopore sequencing to resistome mapping. We demonstrate the approach by functionally characterizing the gut resistome of an ICU (intensive care unit) patient. The accuracy of the poreFUME pipeline is with >97% sufficient for the annotation of antibiotic resistance genes. The poreFUME pipeline provides a promising approach for efficient resistome profiling that could inform antibiotic treatment decisions in the future.

INTRODUCTION

It is estimated that every year 700,000 people die of resistant infections (1). Antibiotic resistance by human pathogens has become a major threat, in particular for hospitalized patients (1,2). Bacterial infections by resistant pathogens are also coupled with an increase in healthcare costs (3). The gut microbiome comprises a diverse set of antibiotic resistance genes which may impact antibiotic treatment outcomes in at least two ways (4–6). First, the gut microbiome can act as a reservoir of resistance genes that can be acquired by infecting human pathogens leading to evolution of resistance during infection. Indeed, a close evolutionary relationship between resistance genes in pathogens and commensals has been found (7). Second, the gut resistome impacts the extent to which the commensal microbiota is affected by antibiotic treatment. Studies of preterm infants and their response to antibiotic treatment suggest that the collateral damage to the commensal microbiota can be predicted from the resistome status at the start of treatment

(8). Accordingly, there is an increasing interest in the development of clinically applicable workflows that enable expedited and comprehensive characterization of the gut resistome. Unfortunately, given the diversity of antibiotic resistance genes in the gut microbiota, sequencing based methods alone provide an incomplete way to representatively characterize the gut resistome. Instead, functional metagenomic selections, which circumvent the culturing step of individual gut microbes, allow less biased interrogation of the gut resistome (9).

The future diagnostics of the whole microbiome are likely going to be important in order to minimize collateral damage of antibiotic treatment on the gut microbiome and, in addition, to limit resistance gene exchange. The general principles for how to guide antibiotic therapy based on a resistome characterization have not been established, largely due to the lack of bioinformatics methods for characterizing the resistome.

Consequently, rapid resistome profiling using a functional-metagenomic approach would be a viable approach to guide personalized antibiotic treatment.

A functional metagenomic workflow consists of several steps, of which the final step is the analysis of metagenomic sequencing data (Figure 1) (10). Traditionally, sequencing data was obtained using Sanger sequencing, (11,12), yet, high-throughput sequencing technologies such as 454 pyrosequencing (13) and Illumina sequencing (6,14) have been applied to analyze functional metagenomic assays as well. The PARFuMS pipeline based on Illumina data was used to profile the antibiotic resistome of soil and of the human microbiome (4,6,14), and PacBio SMRT data has been used to sequence large metagenomic insert libraries (∼40 kb) from fosmids (15). There are several challenges related to data processing and annotation in functional metagenomic selections. For Sanger sequencing, the data annotation is usually done on fragmented assemblies. In each Sanger set, contigs can be closed by using primer walking. However, primer walking is hardly feasible for high-throughputs datasets and requires weeks to complete. Short-read sequencing based on the Illumina platform offers a high-throughput method, yet, contig assembly can be hampered by repetitive sequences in the original insert. A workflow based on PacBio
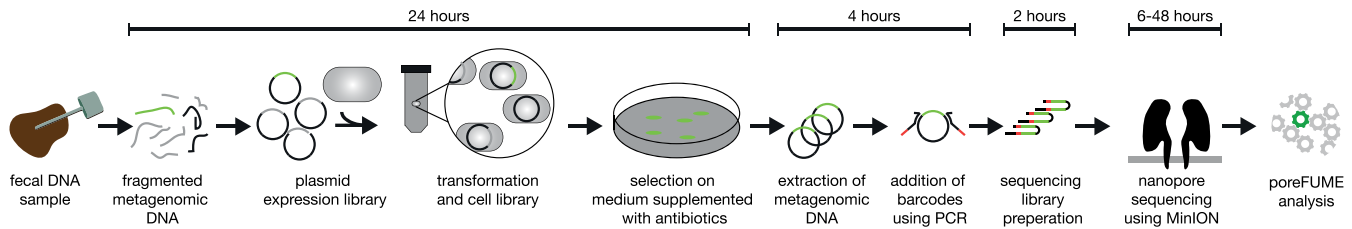
Figure 1. Overview of the poreFUME workflow consisting of the construction and sequencing of a metagenomic library. Fecal DNA is extracted, purified, fragmented and cloned into a shuttle vector. The library of plasmids is transformed in an *Escherichia coli* expression host and grown on selective media supplemented with antibiotics, this process takes ∼24 h. The next day, DNA of the resulting colonies is extracted and barcodes are ligated using a PCR reaction. PCR-amplified DNA is used as input for the sequencing library. This process takes ∼2 h. The sequencing library is loaded into the MinION nanopore sequencer and run between 6 and 48 h. Finally, the obtained sequencing data is demultiplexed, error corrected, polished (20) and annotated using the CARD database (21) in the poreFUME pipeline.

SMRT data circumvents such assembly challenges; however, this technology has a significant capital cost requirement, a large laboratory footprint and is technically demanding, limiting point of care applications (16). In contrast, nanopore sequencing may be able to address these challenges enabling on site monitoring of resistomes in both clinical and environmental settings.

Nanopore sequencing works by threading a DNA molecule through a nanopore embedded in a membrane. When a voltage is applied over the membrane, an ion current is established and this current is modulated when DNA bases pass through the pore. This current-signal depends on the identity of the base that resides in the pore and is converted into a nucleotide sequence by downstream software. Using the MinION, read lengths larger than 100 kb have been reported (http://biorxiv.org/content/early/2015/05/13/019281). The MinION has already been applied to study various aspects related to antibiotic resistance and infection microbiology, including determining the structure and chromosomal insertion site of a bacterial antibiotic resistance island in *Salmonella* Typhi (16), detection of carbapenemases and ESBL genes and their genomic location in Gram-negative pathogenic isolates (17,18), and identification of species and resistance profiles of *Staphylococcus aureus* and *Mycobacterium tuberculosis* isolates (19). Yet, nanopore sequencing has so far not been applied to profile the resistome of a complex microbial community.

In this study, we developed the poreFUME workflow to characterize the resistome of a clinical samples (Figure 1) using nanopore sequencing. Metagenomic expression libraries were constructed using fecal samples from a hospitalized patient as input. The libraries were selected on solid media containing various antibiotics and DNA was extracted from the surviving clones expressing metagenomic inserts conferring antibiotic resistance. The extracted DNA was sequenced using nanopore sequencing. Finally, the sequence data was processed using the poreFUME computational pipeline, which demultiplexes the barcodes, increases the data quality and annotates antibiotic resistance genes.

MATERIALS AND METHODS

Experimental work

Ethics statement. The protocol for this study was reviewed and approved by the institutional review board of the University Medical Center Utrecht (Utrecht, The Netherlands) under number 10/0225. Informed consent for fecal sampling during hospitalization was waived. Written consent was obtained for the collection of fecal samples after hospitalization.

Sample collection. Longitudinal fecal samples were collected from a human adult who received ICU care at the University Medical Center Utrecht in the Netherlands. The subject was a patient who, after a surgical intervention, was admitted to ICU. Five longitudinal feces samples were collected upon admission, during ICU stay and eight months after hospitalization and stored at −80°C. During ICU stay, the patient was treated with cephalosporins and flucloxacillin intravenously and a suspension containing tobramycin (an aminoglycoside antibiotic) and colistin (a polymyxin antibiotic) was delivered to the gut throughout ICU stay as a prophylactic treatment aimed against gut colonization by nosocomial pathogens (22). DNA from feces was extracted as described previously (23).

Construction of metagenomic libraries for functional selections. Construction of metagenomic libraries was done following the protocol (24) with slight modifications. DNA was fragmented using a Covaris shearer to an average size of 2 kb. DNA was blunt-ended and phosphorylated using an End-Repair kit (Epicentre, USA). Sheared DNA (138 μl) was mixed with 20 μl 10× End Repair buffer, 20 μl 2.5 mM dNTP, 20 μl 10 mM ATP and 2 μl of End-It enzyme. After incubation at 22°C for 55 min, the end-repair reaction was heat inactivated at 70°C for 20 min. End-repaired DNA was size selected by electrophoresis. Agarose gel slices selected from the size range 0.5–5 kb were purified using the Gel Purification Kit (Fermentas, USA). Gel purified DNA was ligated into vector pZE21-MCS-1 (25) using Fast-link DNA ligase (Epicentre, USA). For this purpose, the concentration of gel-purified and end-repaired DNA insert was adjusted to 200 ng/μl. Ligation reaction was set up with 2.5 μl of DNA inserts (200 ng/μl), 0.5 μl 10× Ligation buffer, 0.25 μl 10 mM ATP, 0.25 μl dH₂O, 0.5 μl HincII cut pZE21-MCS-1 vector (100 ng/μl). The ligation mixture was incubated at 22°C for 16 h and finally heat inactivated at 70°C for 20 min. Three μl of ligation reaction was electroporated into 50 μl electrocompetent *Escherichia coli* Top10 cells (Invitrogen, USA). After electroporation in a 2 mm cuvette (2000 V, 25 μF, 200R), cells were recovered

in 1 ml SOC medium for 1 h at 37°C. Determination of the transformation efficiency was done by plating out 1 μl and 0.01 μl of recovered cells onto LB agar plates containing 50 μg/ml kanamycin (pZE21-MCS-1 vector contains a selectable marker for kanamycin resistance) (25). Colony forming units (CFUs) were counted after overnight incubation at 37°C. For each library, the insert size distribution was estimated by gel electrophoresis of the PCR products obtained by amplifying the insert using primers flanking the HincII site of the multiple cloning site of the pZE21-MCS-1 vector (pZE21_81_104_57C and pZE21_151_174rc_58C; Supplementary Table S4) (24).

The size of each of the metagenomic libraries for functional selection was determined by multiplying the average PCR based insert size with the number of CFUs. The size of the five metagenomic libraries for functional selection is listed in Supplementary Table S1. The rest of the recovered cells after transformation was inoculated into 10 ml of LB broth supplemented with 50 μg/ml kanamycin and grown overnight at 37°C, 180 rpm. The overnight cultures were stored with 15% (v/v) glycerol at −80°C.

Functional selection of antibiotic resistance clones. The overnight cultures grown and stored at -80°C allowed each clone after transformation to amplify (e.g. total cell count would increase from 5×10^5 CFU containing the plasmid per ml to 5×10^8 CFU/ml after overnight incubation and storage at −80°C). Resulting amplification of the particular clone in the library was resolved by plating each library ∼100× coverage. That is, each unique clone in the library was screened by plating out ∼100 copies. For each library, clones carrying antibiotic resistance determinants were selected by plating onto solid LB agar supplemented with one of the seven antibiotics: tobramycin, spectinomycin, ampicillin, cefotaxime, azithromycin, tetracycline or fosfomycin at concentrations that were inhibiting the wild-type strain (Supplementary Table S2). The CFU was determined after overnight incubation at 37°C (Supplementary Table S2).

Sequencing of antibiotic resistance clones. From each antibiotic plate, a representative number of clones was selected for sequencing and further analysis of antibiotic resistant genes. Single colonies selected on antibiotic resistance plates were picked up into 96-well plates (each of the 96 wells contained 200 μl LB broth supplemented with 50 μg/ml kanamycin). The selected clones were grown overnight at 37°C. The clones were transferred using a 96-pin replicator into new 96-well plates and onto squared LB agar plates supplemented with 50 μg/ml kanamycin. Single clones from 96-well plates were Sanger sequenced by Beckman Genomics, UK with primers listed in table Supplementary Table S4.

The clones from the individual square agar plates were collected by adding 5 ml H₂O, after which the colonies were scraped off the plate with an L-shaped cell scraper. The washing step was repeated twice to remove all the cells from the plate. The bacterial cells were then pelleted by centrifugation at 5000 g for 10 min. The supernatant was discarded and the pellet was dissolved in 10 ml of H₂O. Two ml of the collected bacterial cells was used for plasmid extractions with the Plasmid Mini Kit (Invitrogen, USA). The rest of

the cells were heat inactivated at 95°C for 10 min and stored as a bacterial cell pellet. For nanopore and PacBio sequencing, primers were synthesized that amplify the common region on pZE21-MCS- together with the specific barcodes from PacBio (Supplementary Table S5). One ng of DNA or 1 μl of raw bacterial cell pellet was amplified by PCR. Amplified and barcoded DNA was size selected by electrophoresis. Agarose gel slices selected from the size range 1–5 kb were purified using Gel Purification Kit (Fermentas, USA). In total 39 barcodes (1 and 11–48) were multiplexed according to the scheme in Supplementary Table S7.

Nanopore sequencing library preparation. The nanopore sequencing library B was prepared as briefly described below. DNA QC was performed using Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, USA) and 2200 TapeStation (Agilent, USA). Sequencing library preparation was carried out with Nanopore Genomic Sequencing Kit SQK-MAP006 (Oxford Nanopore, UK) and a PCR-free 'native barcoding' kit according to the manufacturer's protocol. The NEBNext Ultra II End Repair/dA Tailing module (E7546S, NEB, USA) was used to prepare 1000 ng of the functionally selected DNA. End-prepared DNA was ligated with native barcode adapters NB04 from Oxford Nanopore using Blunt/TA Ligase Master Mix (M0367S, NEB, USA). The resulting DNA was pooled with 8 other unrelated barcoded libraries by equivalent weight. The pooled sample was mixed with the 'Native Barcoding Adapter Mix (BAM)' and 'Native Barcoding Hairpin Adapter (BHP)' together with Blunt/TA Ligase Master Mix (M0367S, NEB, USA), and after incubation HP tether was added. The reaction mixture was cleaned up using prewashed Dynabeads MyOne Streptavidin C1 (Thermo Fisher Scientific, USA).

The sequencing library A was prepared using the same protocol as library B, but the native barcoding and pooling steps were omitted.

Nanopore sequencing. The MinION was primed twice for 10 min with 500 μl priming solution (250 μl nuclease free water, 237 μl 2× Running Buffer, 13 μl Fuel Mix). For sequencing, 6 μl library was mixed with 65 μl nuclease free water, 75 μl 2× Running Buffer and 4 μl Fuel Mix (SQK-MAP006, Oxford Nanopore, UK) and immediately loaded to a MinION. The 'SQK-MAP006 Scripts for Yield Monitoring Switch, Bias-Voltage Re-mux Tuning & Pore Shepherding' by John Tyson (personal communication) were used in the MinKNOW software to sequence the library.

Data analysis

Nanopore data processing. The sequencing data was base-called using Metrichor. The Metrichor workflow for sequencing library A included additional native barcode demultiplexing. Poretools (26) was used to extract 2D reads using the *poretools fasta –type 2D* command. Next the 2D reads were analyzed using poreFUME.

poreFUME nanopore sequence analysis. The poreFUME pipeline consists of four steps. First, the reads are demultiplexed on barcode using the Smith-Waterman algorithm
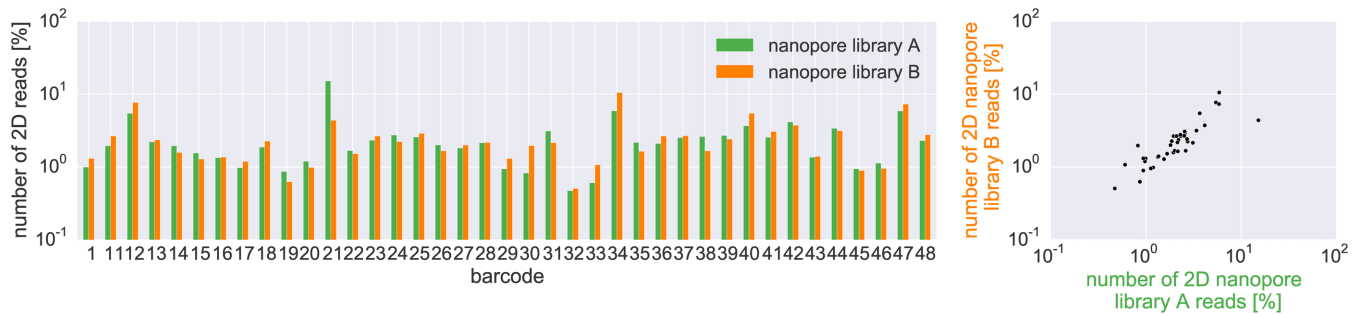
Figure 2. Barcode distribution of the 24 126 2D-reads nanopore library A (green) and control library B (3361 2D-reads, orange). The Pearson correlation test ($R^2 = 0.75$, $P < 10^{-12}$) shows the significance of the log transformed abundance relationship between the two nanopore libraries.

(27). User defined barcodes are detected within 60 (Library A) or 120 (Library B) basepairs of the read ends. Barcode alignment was scored using +2.7 for match, −4.5 for mismatch, −4.7 gap opening and −1.6 for gap extension. A score threshold of >58 was used for the combined score of the asymmetric barcodes. Second, the demultiplexed reads were error corrected using the original nanocorrect protocol (20). The original nanocorrect protocol implements a minimum read coverage of 3×, to ensure that only high-quality data will be outputted. Since we were also interested maximizing sequence diversity, we adjusted the minimum coverage to 1× in the second round of nanocorrect by modifying the *min_coverage* parameter from 3 to 1. In the third step the nanocorrected sequences were polished using nanopolish (20). Briefly, the raw event data is mapped to the nanocorrected sequences via BWA (28) and only nanocorrected reads with a mean raw read coverage of >30× are used in the polishing step. Finally, only nanopolish proposed mutations are accepted if >85% of the reads support this base. In the last step of poreFUME the error corrected and polished reads were mapped against the CARD database (21,29) using blastn (30) (version 2.4.0) with the parameters *max_hsps* 1 and *max_target_seqs* 1000. Closely related genes such as *bla_{TEM}*, *bla_{CTX}*, *mir* and *srt* (ie. *srt-1* and *srt-2*) were masked and only reported as the prototype (i.e. *srt*). For each individual read, the BLAST hits were sorted by bitscore and the highest scoring CARD hit on each segment was kept. For successful CARD gene calling two thresholds were set: a sequence identity of >80% and a >50% coverage of the original gene in the CARD database.

It should be noted that we did not implement the Resistance Gene Identifier (RGI) module of CARD (29). Since RGI relies on open reading frame detection, we noticed that ORF calling is not yet optimized for nanopore data (data not shown), thus instead we made use of the nucleotide data in a blastn query.

Sanger sequence data analysis. The Sanger sequencing resulted in 770 'forward' and 779 'reverse' sequencing reads. Sanger DNA sequences were imported to CLC Genomic Workbench (version 7.6.4). Sequences were vector and quality trimmed (Q 0.01) and assembled using the 'Assemble Sequences' module. Contigs with a length of <500 basepairs were omitted from further analysis.

PacBio data analysis. PacBio sequences were obtained from the Norwegian Sequencing Centre at the University of Oslo in two flowcells on the Pacific Biosciences RS II instrument using P6-C4 chemistry. The metrics for the total set are listed in Supplementary Table S6. Raw PacBio data from the flowcells was analyzed with PacBio SMRT® Portal version 2.3.0 and reads were extracted using the RS_ReadsOfInsert protocol (version 2.3.0). The RS_ReadsOfInsert protocol was run with a minimum predicted accuracy of 99, and minimum read length of insert length of 100 bp. Additionally, paired-end barcode detection was performed using the pacbio_barcodes_paired scheme containing 48 unique barcode pairs. A minimum barcode score of 15 was used in both cells. The extracted reads of insert were grouped by individual barcode and exported in the fastq. The final yield is reported in Supplementary Table S5.

PacBio reads were collapsed with Pbdagcon (https://github.com/PacificBiosciences/pbdagcon, version f19aed1) using dazcon with the flags *–only-proper-overlaps* and *–coverage-sort* and parameter *–min-coverage* 0.

The integron structure was annotated using blastn and IntegronFinder (31).

RESULTS

We constructed metagenomic expression libraries from fecal samples obtained from an Intensive Care Unit (ICU) patient as described in (24) (Materials and methods). The library size ranged between 2.9 and 8.8×10^8 bp of DNA (Supplementary Table S1). The metagenomic libraries were plated on solid LB agar media containing inhibitory concentrations of the antibiotics: tobramycin, spectinomycin, ampicillin, cefotaxime, azithromycin, tetracycline or fosfomycin (Supplementary Table S2). Clones from the metagenomic libraries able to tolerate each of these seven different antibiotics were detected in all libraries (Supplementary Table S2). From each antibiotic plate a representative number of clones were selected (in total 864), pooled, barcoded with custom adapters using PCR and prepared for sequencing using the MinION nanopore sequencer (Materials and methods).

Nanopore sequencing yielded 95.1 Mbase in 62 890 high-quality 'passing filter' two-direction (2D) reads with a mean length of 1513 bp (library A) (Supplementary Figure S1). As an internal control, we multiplexed the sequencing library with eight other unrelated samples (library B). Due
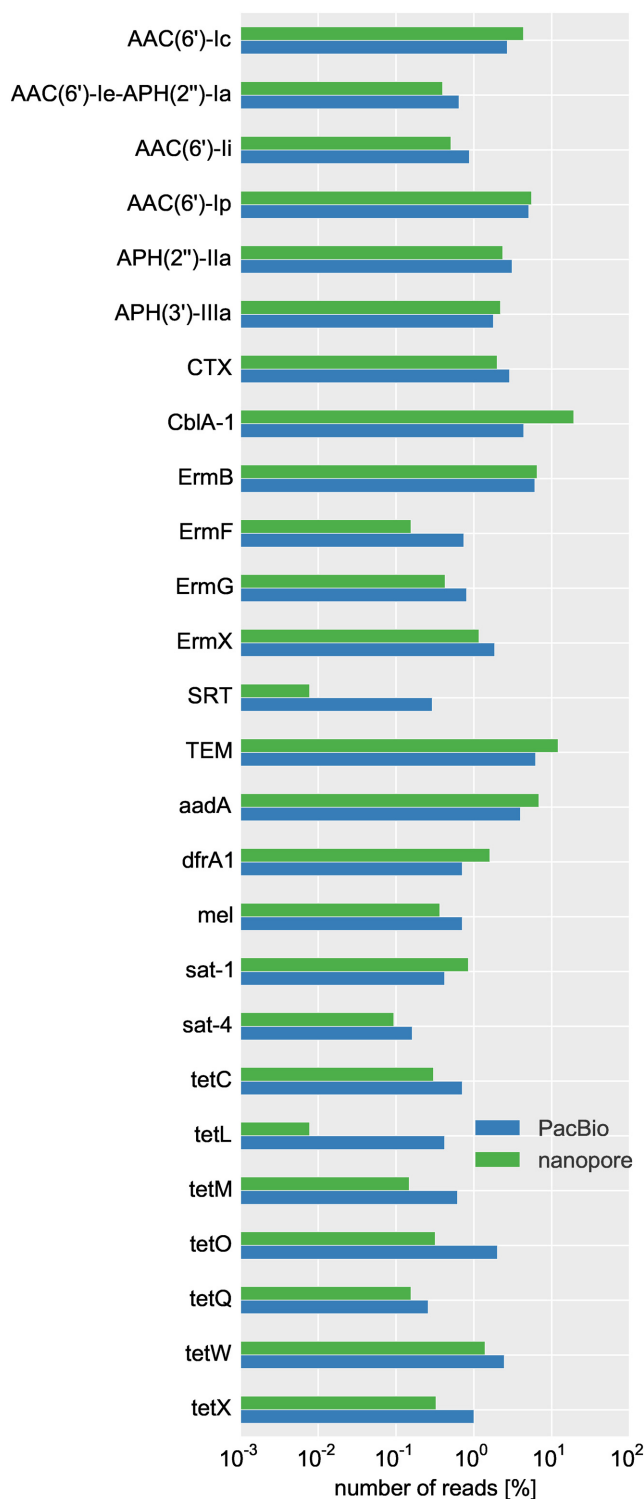
Figure 3. Distribution of the occurrence of the 26 different CARD genes found in the PacBio and nanopore dataset. The Pearson correlation test ($R^2 = 0.71$, $P < 10^{-7}$) showed a significant relationship between the nanopore and PacBio dataset as assessed by the log-transformed proportion of CARD hits found. Threshold for CARD identification are gene coverage of >50% and >80% sequence identity.
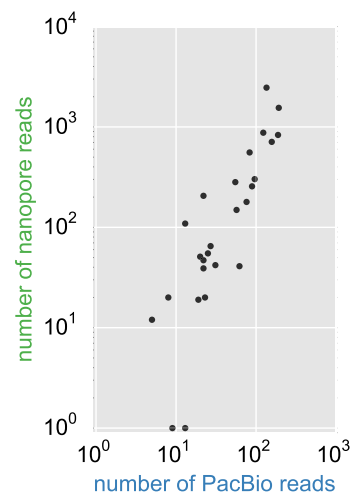

Figure 4. Abundance of 26 CARD genes in the PacBio (blue) and nanopore (green) dataset as show as fraction of the total reads contained in each dataset. The threshold for CARD identification are a gene coverage of >50% and >80% sequence identity.

to multiplexing with unrelated samples, library B generated only 4959 sample specific 2D reads (Supplementary Table S3). The subsequent part of this study focuses exclusively on the use of high-quality 'passing filter' 2D reads.

The first step of poreFUME is to demultiplex the barcodes. We identified all the 39 experimentally attached barcodes in both the nanopore sequencing libraries (Figure 2). The abundance showed a significant correlation of log transformed abundance with the Pearson correlation test ($R^2 = 0.75$, $P < 10^{-12}$) between the two libraries, highlighting the reproducibility of the sequencing and barcode demultiplexing step. Due to the smaller library size of nanopore library B the remainder of this study focuses on nanopore library A. A Pearson correction test of the barcode abundance between nanopore library A and the PacBio library resulted in a R^2 of 0.42 ($P < 10^{-5}$) as shown in Supplementary Figure S4. This lower correlation might be due to a PacBio sequencing bias, as the PacBio instrument preferably loads shorter DNA fragments (32), this is supported by a mean read length of nanopore library A of 1513 bp versus 1370 bp for the PacBio library (Supplementary Table S3).

The sequencing data obtained with MinION nanopore R7 chemistry has an 2D read accuracy of ~85% (20). This relative high error rate can be mitigated using error correction, with tools such as nanocorrect (20). Nanocorrect is implemented as second step in the poreFUME pipeline. Nanocorrect has been applied previously to increase the nanopore read accuracy from 80.5% to 95.9% (20). The algorithm identifies overlapping reads using DALIGNER (33) and calculates a consensus sequence, using partial-order alignment (POA) software (34). Two rounds of error correction where conducted by the poreFUME pipeline. Finally, the error corrected reads were polished using nanopolish (20) which uses the raw nanopore event data to increase the accuracy.

We annotated the error corrected and polished sequencing data for the presence of antibiotic resistance genes us-
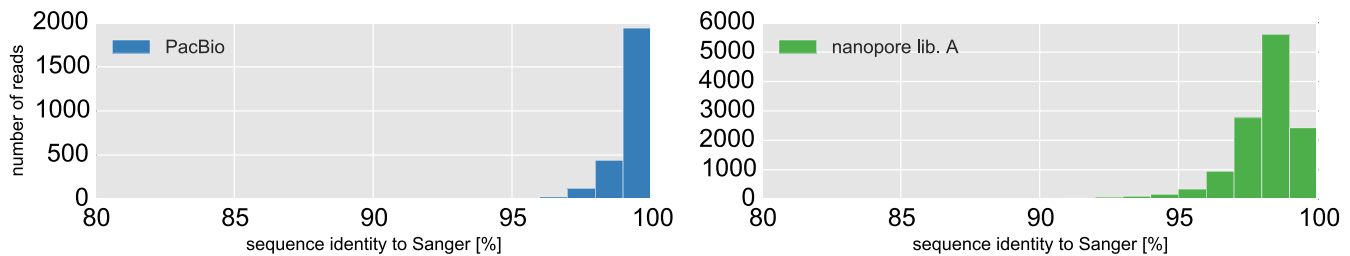
Figure 5. Histogram of the sequence identity of the 2D nanopore sequence reads of library A (12 820 reads) after two rounds of error correction and polishing, showing a mean sequence identity to the Sanger data set of 98.0%. The PacBio dataset after a single round of consensus calling using pbdagcon (3086 reads) shows a mean sequence identity to the Sanger dataset of 99.3%.
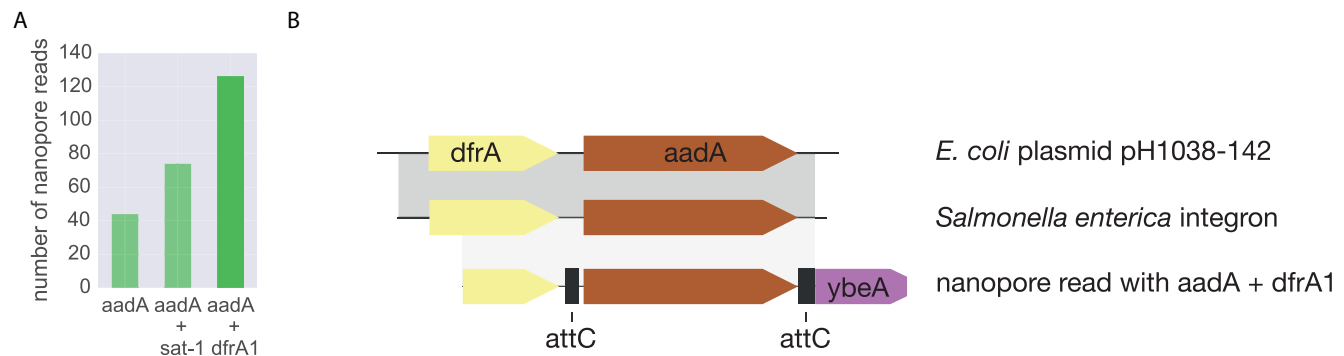
Figure 6. (**A**) Example of CARD genes found in the nanopore dataset that were plated on spectinomycin. Of the 244 nanopore reads that were recovered on spectinomycin, all contain the *aadA,* which encodes an aminoglycoside nucleotidyltransferase known to confer resistance to spectinomycin. In 126 reads *aadA* is flanked by *dfrA1,* which confers resistance to trimethoprim and in 74 reads *aadA* is flanked by *sat-1* which confers resistance to streptothricin. The genes *sat-1* and *dfrA1* do not confer known resistance to spectinomycin and are thus co-selected with *aadA*. (**B**) Alignment against the NT database of the longest corrected nanopore read of the 126 reads containing *aadA* and *dfrA* obtained from functional selection on spectinomycin. The corrected nanopore read shares 99% sequence identify with the two top scoring hits in the NT database: (light gray) integron from *Salmonella enterica* subsp. enterica serovar (GenBank: HQ132378.1) and the *E. coli* plasmid pH1038-142 (genbank: KJ484634) described by Wang *et al*. (38). The *attC* region is a repeat sequence which typically flanks a resistance cassette, *dfrA* encodes a dihydrofolate reductase known to confer resistance against trimethoprim, aadA is an aminoglycoside nucleotidyltransferase known to confer resistance to spectinomycin and ybeA (also known as orfX) which is a pseudocassette and present as conserved 3′ fragment in class 2 integrons (36).

ing Comprehensive Antibiotic Research Database (CARD) (21). Using the CARD database, 26 different antibiotic resistance genes were identified in the nanopore data set (Figure 4). A variety of antibiotic resistance genes were detected with a mean sequence identity of 97.3%, including beta lactamase genes (*bla*$_{CTX-M,}$ *bla*$_{TEM}$ and *cblA*), genes coding aminoglycoside modifying proteins (from different subclasses of AAC, ANT and APH enzymes) and diverse genes encoding ribosomal and efflux mediated resistance towards tetracycline antibiotics, among others (Figure 3 and Supplementary Figure S3). Antibiotic resistance databases such as CARD are commonly used to identify antibiotic resistance genes in metagenomic sequencing data (35). Yet, based on our observation 33.8% inserts had no identifiably match in CARD, indicating the further potential for discoveries of novel antibiotic resistance genes from the human gut (Figure 5).

To validate the nanopore sequencing results, we sequenced the same barcoded DNA using PacBio SMRT technology yielding 93.5 Mbase of DNA in 68 144 reads (with >99% accuracy) from two sequencing cells. After annotation with the CARD database, we observed that the exact same set of 26 antibiotic resistance genes detected in the nanopore dataset were also present in the PacBio dataset

(Figure 4). The mean sequence identity of the genes identified in the CARD database is for the PacBio dataset with 97.8% slightly better than that of the error-corrected and polished nanopore dataset with 97.3% (Supplementary Figure S2). The abundance of reads between the nanopore and PacBio dataset was in good agreement as calculated using the Pearson correlation test ($R^2 = 0.71$, $P < 10^{-7}$) (Figure 3).

To further test the accuracy of our nanopore data set, we sequenced the selected libraries using Sanger sequencing. The sequence identity between the Sanger reads and the non-corrected 2D reads of nanopore library A was 85.8% (Supplementary Figure S2), this confirms the higher error rate of the used MinION reads (16,20). However, by using two rounds of error correction and polishing implemented by poreFUME the sequence identity of the nanopore reads was improved from 86.8% to 98.0%, which enables accurate resistome mapping using CARD.

Nanopore reads are longer and typically capture the entire metagenomic insert. In this way, analysis of the context of the antibiotic resistance gene is simplified. As an example to link genotype and phenotype we investigated the 244 nanopore reads that were selected on plates containing spectinomycin and resulted in a CARD annotation. All the

244 reads contain the *aadA* gene which encodes an aminoglycoside nucleotidyltransferase known to confer resistance to spectinomycin (Figure 6A). In 74 reads *aadA* was flanked by *sat-1* which encodes a streptothricin acetyltransferase and confers resistance to streptothricin. The *sat-1* gene is likely co-selected with *aadA*. In 44 reads the *aadA* was the only gene detected, however in 126 reads *aadA* was flanked by *dfrA1*. The *dfrA1* gene confers resistance to trimethoprim and is not known to confer resistance to spectinomycin. Alignment of the longest nanopore read containing both *dfrA1* and *aadA* against the NT database showed that the two highest scoring hits are part of an integron (Figure 6B). Furthermore, two *attC* sites are detected (31) and an *ybeA* gene cassette (part of the 3′ conserved fragment of class 2 integrons (36)) is found. Nanopore sequencing is thus able to identify such differing contexts of antibiotic resistance genes, which can impact the probability of a pathogen to acquire specific antibiotic resistance genes (37).

DISCUSSION

Hospital settings, including ICUs, are hotspots for the emergence and selection for antibiotic resistant organisms. In this study, we successfully applied nanopore sequencing to characterize the gut resistome of an ICU patient using metagenomic functional selections. Nanopore sequencing currently has a higher error rate compared to short-read Illumina sequencing; however, implementing a double error correction and polishing scheme in poreFUME we are able to achieve accuracies >98%, which enables reliable resistance gene annotation and comparable results to that of PacBio SMRT sequencing at considerable lower costs. In this study, both sequencing platforms enabled the reliable identification of 26 unique antibiotic resistance genes. This, along with the rapid turnaround time of the poreFUME workflow, suggests that it could be applied as a possible resistome monitoring tool.

A benefit of nanopore sequencing is the resulting long reads. Since the sequence analysis is often hampered by assembly of short reads, for example in previous efforts only 50% of the functionally selected clones could be fully assembled (15). Other methods, such as ParFUMS, require several assembly steps and the use of custom scripts, which hamper the uptake of this method by the scientific community (4,6,14).

Only 39% of the nanopore 2D reads had their barcodes successfully demultiplexed in this study. Improvements to the protocol can be made by making use of barcodes with a larger editing distance. In accordance, the currently used barcodes are deprecated and replaced by a set with a larger editing distance, which allows for better separation and a higher barcode identification rate (https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/Barcoding#barcode-sequences). However, even the 61% loss due to barcode demultiplexing does not hamper downstream resistome analysis using poreFUME.

The current data is generated using the R7.3 sequencing pore in kit SQK-MAP006. However, recent developments in nanopore sequencing technologies (R9 pore) resulted in 2D accuracies of >90%. This increase in R9 read accuracy (compared with 86% of 2D R7.3) would therefore contribute to the accuracy of the poreFUME annotation.

CONCLUSION

In conclusion, the poreFUME pipeline provides a promising alternative to other next-generation sequencing alternatives (6,15) and can be used to rapidly profile the resistome of both environmental and gut microbial communities (4–6). We foresee that rapid resistome profiling tools such as poreFUME could aid the implementation of personalized antibiotic treatment in high risk patients.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. O'Neill,J. (2016) Tackling drug-resistant infections globally: final report and recommendations.
2. WHO. (2014) Antimicrobial resistance. *Bull. World Health Organ.*, **61**, 383–394.
3. de Kraker,M.E.A., Davey,P.G. and Grundmann,H. (2011) Mortality and hospital stay associated with resistant Staphylococcus aureus and Escherichia coli bacteremia: estimating the burden of antibiotic resistance in Europe. *PLoS Med.*, **8**, 10.1371/journal.pmed.1001104.
4. Pehrsson,E.C., Tsukayama,P., Patel,S., Mejía-Bautista,M., Sosa-Soto,G., Navarrete,K.M., Calderon,M., Cabrera,L., Hoyos-Arango,W., Bertoli,M.T. *et al.* (2016) Interconnected microbiomes and resistomes in low-income human habitats. *Nature*, **533**, 212–216.
5. Dantas,G., Sommer,M.O.A., Oluwasegun,R.D. and Church,G.M. (2008) Bacteria subsisting on antibiotics. *Science*, **320**, 100–103.
6. Forsberg,K.J., Reyes,A., Wang,B., Selleck,E.M., Sommer,M.O.A. and Dantas,G. (2012) The shared antibiotic resistome of soil bacteria and human pathogens. *Science* , **337**, 1107–1111.
7. Sommer,M.O.A. and Dantas,G. (2011) Antibiotics and the resistant microbiome. *Curr. Opin. Microbiol.*, **14**, 556–563.
8. Gibson,M.K., Wang,B., Ahmadi,S., Burnham,C.-A.D., Tarr,P.I., Warner,B.B. and Dantas,G. (2016) Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *Nat. Microbiol.*, 10.1038/nmicrobiol.2016.24.
9. Forslund,K., Sunagawa,S., Coelho,L.P. and Bork,P. (2014) Metagenomic insights into the human gut resistome and the forces that shape it. *BioEssays*, **36**, 316–329.
10. Handelsman,J., Rondon,M.R., Brady,S.F., Clardy,J. and Goodman,R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.*, **5**, R245–R249.
11. Rondon,M.R., August,P.R., Bettermann,A.D., Brady,S.F., Grossman,T.H., Liles,M.R., Kara,A., Lynch,B.A., Macneil,I.A., Minor,C. *et al.* (2000) Cloning the soil metagenome: a Strategy for accessing the genetic and functional diversity of uncultured microorganisms cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.*, **66**, 2541–2547.
12. MacNeil,I.A., Tiong,C.L., Minor,C., August,P.R., Grossman,T.H., Loiacono,K.A., Lynch,B.A., Phillips,T., Narula,S., Sundaramoorthi,R. *et al.* (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. *J. Mol. Microbiol. Biotechnol.*, **3**, 301–308.
13. Iqbal,H.A., Craig,J.W. and Brady,S.F. (2014) Antibacterial enzymes from the functional screening of metagenomic libraries hosted in Ralstonia metallidurans. *FEMS Microbiol. Lett.*, **354**, 19–26.
14. Forsberg,K.J., Patel,S., Gibson,M.K., Lauber,C.L., Knight,R., Fierer,N. and Dantas,G. (2014) Bacterial phylogeny structures soil resistomes across habitats. *Nature*, **509**, 612–616.
15. Wichmann,F., Udikovic-Kolic,N., Andrew,S. and Handelsman,J. (2014) Diverse antibiotic resistance genes in dairy cow manure. *MBio*, **5**, 10.1128/mBio.01017-13.
16. Ashton,P., Nair,S., Dallman,T., Rubino,S., Rabsch,W., Mwaigwisya,S., Wain,J. and O'Grady,J. (2014) MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat. Biotechnol.*, **33**, 296–300.
17. Judge,K., Harris,S.R., Reuter,S., Parkhill,J. and Peacock,S.J. (2015) Early insights into the potential of the Oxford Nanopore MinION for the detection of antimicrobial resistance genes. *J. Antimicrob. Chemother.*, **70**, 2775–2778.
18. Turton,J.F., Doumith,M., Hopkins,K.L., Perry,C., Meunier,D. and Woodford,N. (2016) Clonal expansion of Escherichia coli ST38 carrying chromosomally-integrated OXA-48 carbapenemase gene. *J. Med. Microbiol.*, **38**, 538–546.
19. Bradley,P., Gordon,N.C., Walker,T.M., Dunn,L., Heys,S., Huang,B., Earle,S., Pankhurst,L.J., Anson,L., de Cesare,M. *et al.* (2015) Rapid antibiotic resistance predictions from genome sequence data for S. aureus and M. tuberculosis. *Nat. Commun.*, **6**, 018564.

20. Loman,N.J., Quick,J. and Simpson,J.T. (2015) A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods*, **12**, 733–735.
21. McArthur,A.G., Waglechner,N., Nizam,F., Yan,A., Azad,M.A., Baylay,A.J., Bhullar,K., Canova,M.J., De Pascale,G., Ejim,L. *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrob. Agents Chemother.*, **57**, 3348–3357.
22. de Smet,A.M.G.A., Kluytmans,J.A.J.W., Cooper,B.S., Mascini,E.M., Benus,R.F.J., van der Werf,T.S., van der Hoeven,J.G., Pickkers,P., Bogaers-Hofman,D., van der Meer,N.J.M. *et al.* (2009) Decontamination of the digestive Tract and oropharynx in ICU patients. *N. Engl. J. Med.*, **360**, 20–31.
23. Buelow,E., Gonzalez,T.B., Versluis,D., Oostdijk,E.A.N., Ogilvie,L.A., van Mourik,M.S.M., Oosterink,E., van Passel,M.W.J., Smidt,H., D'Andrea,M.M. *et al.* (2014) Effects of selective digestive decontamination (SDD) on the gut resistome. *J. Antimicrob. Chemother.*, **69**, 2215–2223.
24. Sommer,M.O.A., Dantas,G. and Church,G.M. (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*, **325**, 1128–1131.
25. Lutz,R. and Bujard,H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.*, **25**, 1203–1210.
26. Loman,N.J. and Quinlan,A.R. (2014) Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics*, **30**, 3399–3401.
27. Smith,T.F. and Waterman,M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol.*, **147**, 195–197.
28. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
29. Jia,B., Raphenya,A.R., Alcock,B., Waglechner,N., Guo,P., Tsang,K.K., Lago,B.A., Dave,B.M., Pereira,S., Sharma,A.N. *et al.* (2016) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.*, 10.1093/nar/gkw1004.
30. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421 10.1186/1471-2105-10-421.
31. Cury,J., Jové,T., Touchon,M., Néron,B. and Rocha,E.P.C. (2016) Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Res.*, **44**, 4539–4550.
32. Myer,P.R., Kim,M.S., Freetly,H.C. and Smith,T.P.L. (2016) Evaluation of 16S rRNA amplicon sequencing using two next-generation sequencing technologies for phylogenetic analysis of the rumen bacterial community in steers. *J. Microbiol. Methods*, **127**, 132–140.
33. Myers,G. (2014) *14th International Workshop, WABI 2014, Wroclaw, Poland, September 8–10, 2014*. Proceedings Brown,D and Morgenstern,B (eds), Springer-Verlag, Berlin, Heidelberg.
34. Lee,C., Grasso,C. and Sharlow,M.F. (2002) Multiple sequence alignment using partial order graphs. *Bioinformatics*, **18**, 452–464.
35. McCall,C.A., Bent,E., Jorgensen,T.S., Dunfield,K.E. and Habash,M.B. (2015) Metagenomic comparison of antibiotic resistance genes associated with liquid and dewatered biosolids. *J. Environ. Qual.*, **470**, doi:10.2134/jeq2015.05.0255.
36. Ramírez,M.S., Piñeiro,S., Centrón,D., Kaufman,S., Kovensky,J., Vay,C., Tokumoto,M., Fernández,A., Famiglietti,A., Couto,E. *et al.* (2010) Novel insights about class 2 integrons from experimental and genomic epidemiology. *Antimicrob. Agents Chemother.*, **54**, 699–706.
37. Martínez,J.L., Coque,T.M. and Baquero,F. (2015) What is a resistance gene? Ranking risk in resistomes. *Nat. Rev. Microbiol.*, **13**, 116–123.
38. Wang,J., Stephan,R., Power,K., Yan,Q., Hachler,H. and Fanning,S. (2014) Nucleotide sequences of 16 transmissible plasmids identified in nine multidrug-resistant Escherichia coli isolates expressing an ESBL phenotype isolated from food-producing animals and healthy humans. *J. Antimicrob. Chemother.*, **69**, 2658–2668.